



**HAL**  
open science

# Histogramme, régressogramme et polygone de fréquences en temps continu

François-Xavier Lejeune

► **To cite this version:**

François-Xavier Lejeune. Histogramme, régressogramme et polygone de fréquences en temps continu. Statistiques [math.ST]. Université Pierre et Marie Curie - Paris VI, 2007. Français. NNT : 2007PA066350 . tel-00809777

**HAL Id: tel-00809777**

**<https://theses.hal.science/tel-00809777>**

Submitted on 9 Apr 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT DE L'UNIVERSITÉ PARIS 6

*Spécialité* : Mathématiques

*Option* : Statistique

*présentée par*

**François-Xavier LEJEUNE**

*pour obtenir le grade de*

DOCTEUR DE L'UNIVERSITÉ PARIS 6

Sujet de la thèse :

**Histogramme, régressogramme et polygone de fréquences  
en temps continu**

Soutenue le 7 décembre 2007 devant le jury composé de :

<i>Président</i>	M. Paul DEHEUELS	Université Paris 6
<i>Directrice de thèse</i>	M <sup>lle</sup> Delphine BLANKE	Université Paris 6
<i>Co-directeur</i>	M. Denis BOSQ	Université Paris 6
<i>Rapporteurs</i>	M. Michel CARBON	Université Rennes 2
	M. Gérard GRÉGOIRE	Université Grenoble 2
	M. Lanh Tat TRAN	Université de l'Indiana
<i>Examineur</i>	M. Alois KNEIP	Université de Bonn



# Remerciements

J'ai eu la chance de pouvoir compter sur de nombreuses personnes durant ces dernières années au Laboratoire de Statistique Théorique et Appliquée ; sans elles, ce travail aurait été plus difficile sinon impossible. Je tiens ici à les saluer et les remercier.

En tout premier lieu, mes plus vifs remerciements s'adressent à Monsieur Paul Deheuvels pour m'avoir ouvert les portes du L.S.T.A. et pour m'avoir permis de travailler dans les meilleures conditions qui soient jusqu'à la fin de la rédaction de ce mémoire. Je lui suis très reconnaissant d'avoir accepté de présider le jury de cette thèse.

Ce fut pour moi un immense privilège de travailler sous la co-direction de Mademoiselle Delphine Blanke et de Monsieur Denis Bosq. Je leur adresse mes remerciements les plus chaleureux pour tout ce qu'ils m'ont appris, leurs encouragements et leur grande disponibilité. En particulier, je suis très fier de partager avec Boris le statut enviable de "premier doctorant" de Mademoiselle Blanke ; sa générosité, sa modestie et ses hautes compétences scientifiques ont été des qualités inestimables pour mener ce travail à terme. J'espère de tout cœur que notre collaboration se prolongera bien au-delà de cette thèse.

Je tiens à remercier Messieurs Michel Carbon, Gérard Grégoire et Lanh Tat Tran de l'intérêt et du temps précieux qu'ils ont bien voulu accorder à l'expertise de cette thèse en acceptant d'en être les rapporteurs. Je suis de plus particulièrement honoré que Monsieur Alois Kneip ait accepté de faire partie de mon jury.

Aux professeurs et maîtres de conférences du L.S.T.A. ; je leur serai éternellement redevable de l'excellente formation qu'ils m'ont donnée et d'avoir été des interlocuteurs concernés : Gérard Biau, Michel Broniatowski, Jérôme Dedecker, Stéphane Gaïffas, Emmanuel Guerre, Armelle Guillou, Agathe Guilloux, Djamel Louani, Mounir Mesbah, Giovanni Peccati, Daniel Pierre-Loti-Viaud, Philippe Saint-Pierre, Annick Valibouze.

À Louise Lamart et à Anne Durrande pour cette dose de bonne humeur quotidienne qui fait chaud au cœur et pour tout ce qu'elles font pour le bien-être des doctorants ; et à Pascal Epron, parce qu'il n'y pas de recherche possible sans bibliothécaire compétent et impliqué.

À tous mes amis doctorants pour les bons moments passés ensemble mais aussi pour avoir été présents dans les moments plus difficiles. Aux "anciens" : Jean-Baptiste & Samuela Aubin, Noureddine Berrahou, Annamaria Bianchi, Salim Bouzebda, Mohammed

Debbarh, Ségolen Geffray, Rosalba Ignaccolo, Amor Keziou, Esterina Masiello, Jean-Renaud Pycke, Fatiha Rachedi, Pierre Ribéreau . . . ; aux “prochains” : Aboubacar Amiri, Véronique Boisson, Olivier Bouaziz, Julien Cornebise, David Degras, Omar El-Dakkak, Kaouthar El Fassi, Mamadou Koné, Nabil Nessigha, Nafissa Rekik, Ravan Roumefort, Camille Sabbah, Rawane Samb, Gwladys Toulemonde, Clara Zelli . . . ; aux vieux complices de Paris 6 : Emmanuel Roy, Bahaeddine Taoufik et Jacques Touboul ; et puis enfin à la dream team des “non paramétriciens” : Lahcen Douge, Olivier Faugeras, Boris Labrador et Mory Souaré.

J’ai une pensée reconnaissante pour toutes les personnes que j’ai eu le plaisir de côtoyer dans le cadre de mes activités d’enseignement à l’I.S.U.P.. Je pense en particulier à Michel Delecroix, à Josiane Confais, à Agathe et à Jérôme auprès de qui j’ai beaucoup appris du métier d’enseignant. Un grand merci également à Séverine Elisabeth, à Patrice Dubois et aux étudiants Isupiens.

Enfin, les années de thèse sont aussi l’occasion de belles rencontres. Je pense à Sophie Dabo-Niang et à Nadia Bensaïd pour l’intérêt qu’elles ont manifesté vis-à-vis de mon travail ; et à Monique Le Guen pour m’avoir éclairé sur des questions de SAS et de statistique exploratoire.

Je tiens à remercier aussi Philippe Bastien qui a été un responsable de stage exceptionnel pendant mon année de D.E.A. ; il m’a énormément appris sur l’utilisation du logiciel SAS et c’est un plaisir pour moi que de pouvoir compter sur son soutien indéfectible.

Enfin, je remercie du fond du cœur mes parents, mes frères, ma famille et mes amis : ils ont été, sont et resteront mes plus fervents supporteurs ; il va de soi que sans eux rien n’aurait été possible. J’ai aussi une tendre pensée pour Tonton François qui est parti trop tôt pour partager ce moment avec moi.

À tous, merci !

# Table des matières

Introduction générale	1
<b>1 Histogramme et polygone de fréquences en temps continu</b>	<b>7</b>
1.1 Cadre d'étude	7
1.1.1 Mesure de dépendance	8
1.1.2 Critères de convergence	8
1.2 Estimation et convergence en moyenne quadratique	9
1.2.1 L'histogramme	10
1.2.2 Le polygone de fréquences ( $d = 1$ )	14
1.2.3 Exemples de processus	19
1.3 Normalité asymptotique ( $d = 1$ )	20
1.3.1 L'histogramme	20
1.3.2 Le polygone de fréquences	21
1.4 Convergence presque sûre	22
1.4.1 Convergence ponctuelle presque sûre	23
1.4.2 Convergence uniforme presque sûre	24
1.4.3 Le polygone de fréquences ( $d = 1$ )	26
1.5 Discussion	27
1.6 Preuves	27
1.6.1 Convergence en moyenne quadratique	27
1.6.2 Normalité asymptotique	38
1.6.3 Convergence presque sûre	42
<b>2 Régressogramme en temps continu</b>	<b>57</b>
2.1 Cadre d'étude	57
2.2 Convergence ponctuelle en moyenne quadratique	61
2.3 Discussion	62
2.4 Preuves	62
<b>3 Estimation pour des trajectoires discrétisées et simulations</b>	<b>67</b>
3.1 Preliminaries and notations	67
3.1.1 Sampling schemes	67
3.1.2 Mean integrated squared error	69

3.2	Histogram . . . . .	70
3.2.1	Definitions and assumptions . . . . .	70
3.2.2	Rates of convergence . . . . .	72
3.3	Frequency polygon . . . . .	75
3.3.1	Definition and assumptions . . . . .	75
3.3.2	Rates of convergence . . . . .	76
3.4	Examples . . . . .	78
3.5	Simulations . . . . .	79
3.6	Discussion . . . . .	82
3.7	Proofs . . . . .	83
3.7.1	Histogram . . . . .	83
3.7.2	Frequency polygon . . . . .	92
	<b>Bibliographie</b>	<b>94</b>

# Introduction générale

Dans ce mémoire de thèse, nous nous intéressons à des méthodes non paramétriques d'estimation de paramètres fonctionnels pour des processus à temps continu. Parmi ces paramètres, les propriétés d'estimateurs de la densité marginale et de la régression basés sur l'observation d'une trajectoire de processus sur  $[0, T]$ , avec  $T \rightarrow \infty$ , ont fait l'objet d'une attention particulière. Les premiers résultats en temps continu ont ainsi été obtenus par Banon (1978) dans le cas d'une diffusion stationnaire, puis par Banon et Nguyen (1978, 1981), Nguyen (1979) et Nguyen et Pham (1980). Plus généralement, le cas des processus stationnaires est étudié par Delecroix (1980) pour une classe d'estimateurs regroupant le noyau, les fonctions orthogonales et pour certains résultats les estimateurs splines. En fait, ce cadre va trouver un intérêt majeur avec l'apparition des vitesses de convergence et la possibilité, pour des processus à trajectoires irrégulières, de construire des estimateurs avec une erreur quadratique qui converge avec une vitesse "suroptimale" (ou "paramétrique") en  $T^{-1}$ . Nous devons ce résultat surprenant à Castellana et Leadbetter (1986) (CL) pour des estimateurs de la densité du type delta-séquences, sous une condition "locale" garantissant l'intégrabilité de la densité jointe du couple  $(X_0, X_u)$  pour les valeurs de  $|u|$  proches de zéro. Cet article historique a par la suite inspiré de nombreux travaux en temps continu. On pourra se référer par exemple aux ouvrages de Bosq (1998), de Bosq et Blanke (2007) pour une étude complète des estimateurs à noyau et par projection, et de Kutoyants (2004) pour l'inférence des processus de diffusion ergodiques, ainsi qu'à leurs bibliographies respectives.

## Chapitre 1 : Histogramme et polygone de fréquences en temps continu

Le premier chapitre porte sur l'estimation de la fonction de densité du processus par des versions "temps continu" de l'histogramme et du polygone de fréquences que l'on construit en reliant les points centraux des sommets du précédent histogramme. L'histogramme est notamment le plus ancien des estimateurs non paramétriques dont l'idée pourrait être attribuée à John Graunt pour un recensement statistique de la population londonienne publié en 1662 sous le titre "Natural and Political Observations on the Bills of Mortality". À partir des variables aléatoires  $X_1, \dots, X_n$  et pour une partition  $\Pi_n$  de  $\mathbb{R}^d$  ( $d \geq 1$ ) en hypercubes de volume  $h_n^d : \Pi_n = \{\pi_{nj}, j \in \mathbb{Z}^d\}$ , rappelons en temps discret la forme de l'histogramme :

$$\hat{f}_j := \hat{f}_n^H(x) = \frac{1}{nh_n^d} \sum_{i=1}^n \mathbb{1}_{\pi_{nj}}(X_i), \quad \forall x \in \pi_{nj},$$

et celle du polygone de fréquences univarié ( $d = 1$ ) :

$$\hat{f}_n^{PF}(x) = \left(\frac{x - c_j}{h_T}\right) \hat{f}_{j+1} + \left(\frac{c_{j+1} - x}{h_T}\right) \hat{f}_j, \quad \forall x \in [c_j, c_{j+1}[,$$

où  $\mathbb{1}_A$  est la fonction indicatrice de l'ensemble  $A$  et où  $c_j$  désigne le milieu de l'intervalle  $\pi_{nj}$ . Pour des variables indépendantes et identiquement distribuées (i.i.d.), les principales propriétés de ces estimateurs élémentaires ont été étudiées par Scott (1992) (chapitres 3 et 4); celles-ci ont ensuite été étendues au cas fortement mélangeant par Tran (1994) et par Carbon *et al.* (1997). L'étude de ces estimateurs en temps continu peut sembler, au premier abord, anachronique étant donné l'existence d'estimateurs plus sophistiqués tels que les fonctions orthogonales, les histo-splines (ou "histogrammes lissés"), les ondelettes ou encore les estimateurs à noyau (ou "histogrammes généralisés") introduits par Rosenblatt (1956b) et Parzen (1962) :

$$\hat{f}_n^{PR}(x) = \frac{1}{nh_n^d} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right), \quad \forall x \in \mathbb{R}^d,$$

où le noyau  $K$  est une application de  $\mathbb{R}^d$  dans  $\mathbb{R}$ , bornée, intégrable par rapport à la mesure de Lebesgue et d'intégrale 1. On trouvera, par exemple, une vaste liste des noyaux usuels dans l'article de Deheuvels (1977). Le choix particulier du noyau unité (ou noyau "naïf") *i.e.*  $K(u) = \mathbb{1}_{[-1/2, 1/2]^d}(u)$  définit alors un histogramme à fenêtre mobile centrée au point  $x = (x_1, \dots, x_d)'$  :

$$\hat{f}_n^{FM}(x) = \frac{1}{nh_n^d} \sum_{i=1}^n \mathbb{1}(X_i \in \prod_{j=1}^d [x_j - h_n/2; x_j + h_n/2]), \quad \forall x \in \mathbb{R}^d.$$

Néanmoins, du fait de leur simplicité de mise en œuvre, histogrammes et polygones de fréquences restent des outils populaires en pratique et particulièrement bien adaptés pour traiter de grandes quantités de données en temps réel; ce qui arrive, par exemple, en traitement de l'image ou dans les problèmes de réseaux de communication (voir Berlinet *et al.* (2002)). Ils peuvent en outre s'avérer performants d'un point de vue théorique : pour l'erreur en moyenne quadratique intégrée, Scott (1985) a ainsi montré que la vitesse de convergence de l'estimateur à noyau en  $n^{-4/5}$  était également atteinte par le polygone de fréquences sous des conditions de régularité équivalentes sur  $f$ .

La section 1.2 est une étude de l'erreur en moyenne quadratique (ponctuelle et intégrée) de ces graphiques de fréquences. En temps continu, Delecroix (1980) a établi une condition suffisante pour la convergence de l'erreur ponctuelle en moyenne quadratique d'une classe générale d'estimateurs – comprenant en particulier les estimateurs à noyau et par projection – pour des processus fortement mélangeants. Les vitesses de convergence de l'erreur quadratique ont été étudiées dans le cas ponctuel pour différents estimateurs : estimateur à noyau (Bosq (1998), section 4.2), estimateur basé sur le temps local (Bosq et Davydov (1999)), estimateur par projection dans le cas particulier des bases de fonctions orthogonales uniformément bornées (Frenay (2001)); et dans le cas intégré : estimateur à ondelettes (Leblanc (1993, 1997)), estimateurs par projection basés sur des procédures

adaptatives et sur l'estimateur du temps local (Comte et Merlevède (2002, 2005)). Nous proposons ici de compléter ces résultats avec l'histogramme et le polygone de fréquences. Sous certaines hypothèses de régularité et de mélange fort, on montre que l'erreur quadratique (ponctuelle et intégrée) converge avec des vitesses comparables à celles (optimales) du cas i.i.d. qui sont de l'ordre  $T^{-2/(d+2)}$  pour l'histogramme et  $T^{-4/5}$  ( $d = 1$ ) pour le polygone de fréquences. Pour le polygone de fréquences, la vitesse  $T^{-4/5}$  est de plus minimale pour le temps continu sur une large classe de processus que l'on explicitera. Sous une condition locale de type 'CL', nous améliorons ensuite ces vitesses au sens où ces erreurs quadratiques s'écrivent comme un  $\mathcal{O}(T^{-1})$ . Pour des observations i.i.d., rappelons que Boyd et Steele (1978) ont prouvé qu'il n'existait pas d'estimateur non paramétrique de la densité avec une erreur en moyenne quadratique intégrée meilleure que  $\mathcal{O}(n^{-1})$ . En temps continu, les vitesses optimale et suroptimale obtenues pour le polygone de fréquences sont donc comparables à celles de l'estimateur à noyau.

La section 1.3 consiste en une application directe d'un théorème central-limite pour les tableaux triangulaires démontré et utilisé par Bosq *et al.* (1999) pour donner la normalité asymptotique de l'estimateur à noyau sous les conditions 'CL'. Nous déduisons de ce théorème, la normalité asymptotique de l'histogramme et du polygone de fréquences ainsi que la possibilité, pour ce dernier, d'obtenir des régions de confiance.

Dans la section 1.4, nous étudions le critère de convergence presque sûre (ponctuelle et uniforme) qui nous donne asymptotiquement une information plus précise de la forme générale de la vraie densité  $f$ . Dans le cas i.i.d., nous devons à Bertrand-Retali (1974) une condition nécessaire et suffisante de convergence uniforme en probabilité :  $h_n \rightarrow 0$ ,  $(nh_n^d)/\ln n \rightarrow \infty$ ,  $n \rightarrow \infty$ , qui implique la convergence presque complète (et *a fortiori* presque sûre) de l'histogramme. Hall et Hannan (1988) ont donné le choix optimal de la fenêtre qui minimise à fois les critères de complexité stochastique et de description à longueur minimale (minimum description length) pour l'histogramme et qui est de l'ordre  $(n^{-1} \ln n)^{1/(d+2)}$ . Avec ce choix de fenêtre, Tran (1994) a établi que la vitesse de convergence en  $(n^{-1} \ln n)^{1/3}$  pouvait être atteinte dans le cadre univarié et fortement mélangeant. Il s'agit notamment de la vitesse optimale des estimateurs non paramétriques de la densité obtenue dans le cas i.i.d. (voir Stone (1983)). Malgré leur grande simplicité de mise en œuvre, les histogrammes sont donc des estimateurs très performants du point de vue de la convergence uniforme. Ces résultats ont été étendus par Carbon et Tran (1996) aux processus linéaires qui ne sont pas en général fortement mélangeants. La démonstration de Tran a ensuite été reprise par Carbon *et al.* (1997) pour établir la convergence uniforme du polygone de fréquences et par Carbon (2006) pour l'extension aux champs aléatoires. En temps continu, les résultats obtenus par Delecroix (1980) impliquent aussi des conditions suffisantes de convergence ponctuelle presque sûre d'une classe générale d'estimateurs comprenant les estimateurs à noyau et par projection. La convergence uniforme (avec des vitesses) a ensuite été étudiée pour différents estimateurs : estimateur à ondelettes (Leblanc (1993)), estimateur à noyau (Bosq (1998), section 4.3), estimateur du temps local (Bosq et Davydov (1999)), estimateur par projection pour les fonctions orthogonales uni-

formément bornées (Frenay (2001)), noyau adaptatif (Blanke (2004)), estimateur par projection adaptatif (Bosq et Blanke (2004)), estimateur du  $k_T$ -temps d'occupation (Labrador (2006)) – version “temps continu” de l'estimateur des  $k$  plus proches voisins. La méthode utilisée dans nos preuves repose essentiellement sur l'approximation de variables aléatoires dépendantes par des variables indépendantes; soit une technique similaire à celle déjà utilisée par de nombreux auteurs dont Tran (1994) (en temps discret), Wu (1997) (pour des variables échantillonnées) et Blanke (2004) (en temps continu). En particulier, l'erreur associée à l'estimateur à noyau converge uniformément vers zéro à la vitesse optimale de l'ordre  $\ln_m T (T^{-1} \ln T)^{r/(2r+d)}$ , avec  $m \geq 1$  et  $r$  le nombre de dérivées de la vraie densité  $f$ . Si  $r = d = 1$ , nous retrouvons alors la vitesse du cas i.i.d. à un facteur logarithmique près. Pour des classes de processus à temps continu qui satisfont les conditions ‘CL’, Bosq (1997) a également établi une vitesse suroptimale de convergence de l'estimateur à noyau de l'ordre  $\ln_m T (T^{-1} \ln T)^{1/2}$ ,  $m \geq 1$ .

Enfin, nous proposons des perspectives de recherche dans la continuité de nos travaux (section 1.5) et les preuves des principaux résultats sont données dans la section 1.6.

## Chapitre 2 : Régressogramme en temps continu

Dans le deuxième chapitre, pour un processus bivarié  $\{(X_t, Y_t), t \in \mathbb{R}\}$  à temps continu tel que chaque paire  $(X_t, Y_t)$  ait la même loi de probabilité que  $(X_0, Y_0)$ , nous nous intéressons à l'estimation non paramétrique d'une version  $r$  de la régression non linéaire de  $m(Y_0)$  sur  $X_0$  définie par

$$r : x \mapsto \mathbb{E}(m(Y_0) | X_0 = x),$$

où  $m(\cdot)$  est une fonction borélienne à valeurs réelles (choisie par le statisticien). La connaissance du paramètre de régression  $r$  est notamment utile pour estimer les valeurs futures de  $m(Y)$  sachant que  $X = x$  et pour comprendre le lien entre les variables  $X$  et  $Y$ . En particulier, l'étude porte sur l'estimation de la fonction  $r$  par une version “temps continu” du régressogramme dont la construction est analogue à celle de l'histogramme à partir de l'observation complète d'une trajectoire du processus sur  $[0, T]$ . Cet estimateur introduit par Tukey (1961) conserve donc l'avantage d'être à la fois simple à mettre en œuvre et économique en termes de calculs. On peut trouver dans la littérature plusieurs estimateurs non paramétriques de la fonction de régression  $r$  parmi lesquels ceux bien connus de type noyau introduits par Nadaraya (1964) et Watson (1964) et par Gasser et Müller (1979) (voir aussi Collomb (1981) pour une revue bibliographique). Pour comprendre les idées qui ont amené à l'introduction de ces estimateurs dans le cadre des variables aléatoires réelles et i.i.d., il faut alors remonter au régressogramme qui est constant par morceaux sur les éléments  $\{\pi_{nj}\}$  d'une partition  $\Pi_n$  d'un intervalle compact  $[a, b]$  :

$$\hat{r}_n^{reg}(x) = \frac{\sum_{i=1}^n Y_i \mathbb{1}_{\pi_{nj}}(X_i)}{\sum_{i=1}^n \mathbb{1}_{\pi_{nj}}(X_i)}, \quad \forall x \in \pi_{nj},$$

soit la moyenne arithmétique des  $Y_i$  tels que  $X_i$  appartienne à  $\pi_{nj}$ . Pour éviter le problème classique lié au choix du positionnement des bords des éléments de la partition, un nouvel

estimateur dit de la “fenêtre mobile” peut être construit en remplaçant la partition *a priori*  $\Pi_n$  par un seul intervalle qui varie de manière continue :

$$\hat{r}_n^{FM}(x) = \frac{\sum_{i=1}^n Y_i \mathbb{1}(X_i \in [x - h_n; x + h_n])}{\sum_{i=1}^n \mathbb{1}(X_i \in [x - h_n; x + h_n])}, \quad \forall x \in \mathbb{R}.$$

L’estimateur précédent présente encore le désavantage d’être discontinu par nature. Ainsi, sa généralisation naturelle est l’estimateur à noyau ou estimateur de Nadaraya-Watson défini de la manière suivante :

$$\hat{r}_n^{NW}(x) = \frac{\sum_{i=1}^n Y_i K\left(\frac{x-X_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)}, \quad \forall x \in \mathbb{R}.$$

Nous retrouvons alors l’estimateur  $\hat{r}_n^{FM}$  lorsque  $K$  est le noyau uniforme :

$$K(u) = \mathbb{1}_{[-1,1]}(u).$$

Il est à noter que dans toutes ces définitions la convention “0/0 = 0” est implicite. Pour un aperçu complet des principales propriétés de l’estimateur de Nadaraya-Watson, on pourra se référer aux livres de Györfi *et al.* (1989) (chapitre 3) et de Bosq (1998) (chapitre 3 en temps discret et chapitre 5 en temps continu), ainsi qu’à leurs propres bibliographies. En temps continu, Chèze-Payaud (1994a,b) et Bosq (1998) ont obtenu pour l’erreur ponctuelle en moyenne quadratique les vitesses optimale et suroptimale qui sont du même ordre que celles des estimateurs à noyau de la densité. Il existe encore une littérature dense pour d’autres estimateurs non paramétriques de la régression basés sur les splines (voir Wahba (1990) et Eubank (1999)) et les ondelettes (voir Nason et Silverman (2000)). De manière générale, l’estimation non paramétrique de la régression est également traitée dans les ouvrages de Härdle (1990) et de Györfi *et al.* (2002), ainsi que dans un cours disponible en ligne par Ferraty et Vieu (2003).

Étant donné un échantillon  $(X_1, Y_1), \dots, (X_n, Y_n)$ , les propriétés statistiques du régressogramme ont d’abord été étudiées par Bosq (1969) avec une condition suffisante de convergence uniforme presque sûre vers zéro sur  $[a, b]$  qui est donnée par  $h_n = \mathcal{O}(n^{-\beta})$ ,  $0 < \beta < 1$ . Partant d’une méthodologie communiquée par Geffroy en 1975 et publiée *a posteriori* (Geffroy (1980)) avec une condition nécessaire et suffisante de convergence uniforme presque complète sur  $[0, 1]^d$  donnée par  $h_n \rightarrow 0$  et  $h_n^d \gg n^{-1} \log n$ , Sabry (1978) a obtenu pour  $d = 1$  une condition suffisante de convergence uniforme presque sûre sur des intervalles de la forme  $[0, \sqrt{\log n / \log \log n}]$  où  $\sqrt{\log n / \log \log n} \rightarrow \infty$  quand  $n \rightarrow \infty$  pourvu que  $h_n \gg n^{-1+(1/\sqrt{\log \log n})} \log n$  et  $h_n \ll n^{-1/\sqrt{\log \log n}}$ . Ces résultats ont ensuite été améliorés par Lecoutre (1982) pour estimer la régression sur  $\mathbb{R}$  tout entier. Des propriétés asymptotiques pour le biais et la variance du régressogramme au point  $x$  ont aussi été énoncées par Collomb (1978) sans démonstration. Enfin, Major (1973) a donné la loi limite et la loi du logarithme associées à la déviation maximale  $\sup_{x \in [0,1]} |\hat{r}_n^{reg}(x) - r(x)|$ . Toutes ces propriétés sont notamment regroupées dans le livre de Bosq et Lecoutre (1987) (chapitre 6, partie B) dans le cas i.i.d. et pour un seul régresseur. Dans la section 2.1, nous définissons

une version “temps continu” du régressogramme dont la construction est analogue à celle de l’histogramme étudié dans le chapitre 1 et, dans la section 2.2, nous établissons les vitesses optimale et suroptimale de l’erreur ponctuelle en moyenne quadratique.

Enfin, nous discutons des travaux en perspective en vue de compléter l’étude du régressogramme en temps continu (section 2.3) et les preuves sont données dans la section 2.4.

### Chapitre 3 : Estimation pour des trajectoires discrétisées et simulations

Dans le troisième chapitre, nous complétons notre étude de l’estimation de la densité dans le cas des données échantillonnées. Le temps continu est en effet un contexte d’étude théorique dont tout l’intérêt est d’éclairer le statisticien sur les vitesses idéales d’un estimateur. Cependant, dans certains cas, il n’est pas possible pour des raisons techniques – dues aux imprécisions des appareils de mesure ou encore lorsque les données ne sont pas disponibles à tout instant  $t$  – d’enregistrer une trajectoire dans la continuité ; et la plupart des phénomènes physiques “observés en continu” ne produisent en fait que des observations discrètes, éventuellement très rapprochées dans le temps, puis interpolées. Pour toutes ces raisons, nous avons recherché un cadre plus réaliste pour tenter d’approcher les vitesses obtenues pour nos estimateurs en temps continu. Nous proposons alors trois schémas de discrétisation afin de laisser au statisticien un plus grand choix pour modéliser les données réelles suivant leurs caractéristiques et leur fréquence d’apparition. Les deux premiers schémas aléatoires – tels que définis par Masry (1983) – sont ainsi particulièrement bien adaptés pour des mesures disponibles à basse fréquence et irrégulièrement espacées dans le temps, ce qui est courant dans le cadre de transactions boursières (voir un exemple dans Aït-Sahalia et Mykland (2003)). Un troisième schéma utilisé par Bosq (1997) s’applique ensuite au cas des mesures recueillies à haute fréquence ou bien sur une longue période de temps. Ce modèle est en particulier très intéressant puisqu’il permet de retrouver une vitesse “suroptimale” pourvu que la trajectoire du processus soit suffisamment irrégulière. À titre d’exemple pour ce type de données, il est fait mention dans un article de Bibby *et al.* (2005) de 5415 mesures de la vitesse du vent effectuées en 1985 à Ferring (Danemark) avec une fréquence de 10 Hertz.

Dans la section 3.1, nous présentons le cadre d’étude avec les différentes procédures d’échantillonnages utilisées pour estimer  $f$ .

La section 3.2 introduit les conditions de régularité imposées à  $f$  et les hypothèses sur le processus qui sont utilisées pour établir les vitesses optimale et “suroptimale” de l’histogramme. Nous déduisons ensuite de cette étude les vitesses du polygone de fréquences dans la section 3.3. Des exemples de processus qui satisfont nos hypothèses sont donnés dans la section 3.4 : notamment, cette liste vient compléter les exemples du chapitre 1 et inclut des processus de diffusion ergodiques multivariés. Dans la section 3.5, des simulations tendent à valider les résultats obtenus ; et nous comparons l’efficacité de l’histogramme, du polygone de fréquences et de l’estimateur à noyau en termes d’erreur quadratique.

La section 3.6 contient une brève discussion de nos résultats et les preuves sont données dans la section 3.7.

# Chapitre 1

## Histogramme et polygone de fréquences en temps continu

**Résumé.** – Dans ce chapitre, nous étudions les propriétés de graphiques de fréquences pour estimer la fonction de densité d'un processus à temps continu, faiblement stationnaire, à valeurs dans  $\mathbb{R}^d$  et observé sur  $[0, T]$ . Plus précisément, sous certaines hypothèses de régularité et d'indépendance asymptotique, les erreurs en moyenne quadratique (ponctuelles et intégrées) de l'histogramme et du polygone de fréquences convergent avec les mêmes vitesses optimales du cas i.i.d; et sous une condition locale plus faible que celle proposée par Castellana et Leadbetter (1986), la vitesse suroptimale en  $T^{-1}$  est obtenue. La normalité asymptotique de ces estimateurs se déduit ensuite d'un résultat de Bosq *et al.* (1999). Enfin, nous établissons les vitesses optimales ainsi que les vitesses suroptimales en  $(T^{-1}(\ln T)^3)^{1/2}$  (histogramme) et en  $(T^{-1} \ln T)^{1/2}$  (polygone de fréquences) pour les convergences ponctuelle et uniforme presque sûre.

### 1.1 Cadre d'étude

Soit  $X = \{X_t, t \in \mathbb{R}\}$  un processus à temps continu, mesurable sur l'espace de probabilité  $(\Omega, \mathcal{F}, P)$  et à valeurs dans  $\mathbb{R}^d$ ,  $d \geq 1$ ; chaque  $X_t$  ayant la même loi de probabilité  $\mu$  de densité  $f$  relativement à la mesure de Lebesgue  $\lambda_d$  sur  $\mathbb{R}^d$ , la fonction  $f$  étant inconnue. Dans tout le chapitre, on suppose en outre l'existence de la densité jointe  $f_{(X_s, X_t)}$  du couple  $(X_s, X_t)$  pour  $s \neq t$  et telle que  $f_{(X_s, X_t)} = f_{(X_0, X_{|t-s|})} =: f_{|t-s|}$ , ce qui est une condition faible de stationnarité sur le processus. On note aussi  $f_{u|0}(\cdot|\cdot)$  la densité conditionnelle de  $X_u$  sachant  $X_0$  et  $g_u$  la fonction de dépendance définie pour tout  $u > 0$  par  $g_u := f_u - f \otimes f$  où  $(f \otimes f)(y, z) = f(y)f(z)$ . Certaines conditions complémentaires d'indépendance asymptotique (incluant une condition de mélange fort) seront données plus tard avec nos hypothèses.

### 1.1.1 Mesure de dépendance

Nous donnons ici quelques rappels sur le coefficient  $\alpha$  de mélange fort introduit par Rosenblatt (1956a).

**Définition 1.1.** Soit  $(\Omega, \mathcal{A}, P)$  un espace probabilisé et soient  $\mathcal{B}$  et  $\mathcal{C}$  deux sous-tribus de  $\mathcal{A}$ . Afin de mesurer la dépendance de  $\mathcal{B}$  et  $\mathcal{C}$ , on définit le coefficient  $\alpha$  de mélange fort par

$$\alpha := \alpha(\mathcal{B}, \mathcal{C}) = \sup_{B \in \mathcal{B}, C \in \mathcal{C}} |P(B \cap C) - P(B)P(C)|.$$

Ce coefficient est tel que  $0 \leq \alpha \leq \frac{1}{4}$  et  $\mathcal{B} \perp \mathcal{C} \Leftrightarrow \alpha = 0$ .

**Définition 1.2.** Le processus  $X$  est dit  $\alpha$ -mélangeant si

$$\alpha_X(u) = \sup_{t \in \mathbb{R}} \alpha(\mathcal{F}_{-\infty}^t, \mathcal{F}_{t+u}^\infty) \xrightarrow{u \rightarrow \infty} 0,$$

où  $\mathcal{F}_m^t$  désigne la  $\sigma$ -algèbre des événements engendrés par  $\{X_s, m \leq s \leq t\}$  pour  $-\infty \leq m \leq t \leq \infty$ .

Notons  $\sigma(X)$  la  $\sigma$ -algèbre des événements engendrés par la variable  $X$  et  $\|X\|_q = (\mathbb{E}|X|^q)^{1/q}$  pour  $1 \leq q < \infty$ .  $X \in L^q(P)$  signifie alors que  $\|X\|_q < \infty$ . On rappelle l'inégalité de Davydov (cf. Bosq (1998), p. 21) :

**Lemme 1.1.** Soient  $X \in L^q(P)$  et  $Y \in L^r(P)$ , si  $q > 1$ ,  $r > 1$  et  $\frac{1}{r} + \frac{1}{q} = 1 - \frac{1}{p}$ , alors

$$|\text{Cov}(X, Y)| \leq 2p \left[ 2\alpha(\sigma(X), \sigma(Y)) \right]^{1/p} \|X\|_q \|Y\|_r.$$

**Remarque 1.1.** Précisons que le coefficient de mélange  $\alpha$  est notamment plus faible que les coefficients de mélange  $\phi$ ,  $\beta$  ou  $\rho$  qui sont aussi utilisés usuellement (cf. Doukhan (1994)). En ce sens, les résultats obtenus concernent une classe plus large de processus.

### 1.1.2 Critères de convergence

Soit  $\hat{g}_T$  un estimateur d'une fonction réelle  $g$  construit à partir d'une trajectoire de  $X$  observée sur  $[0, T]$ . Pour évaluer la précision de nos estimateurs non paramétriques, nous utiliserons dans notre étude les deux critères de convergence suivants :

- **Convergence dans  $L^2$**

L'erreur en moyenne quadratique est un critère très répandu dans la littérature pour évaluer la précision d'une valeur estimée en un point  $x$ . En intégrant sur  $\mathbb{R}^d$ , nous obtenons une mesure globale de cette erreur quadratique.

**Définition 1.3 (Erreur en moyenne quadratique).** L'erreur en moyenne quadratique (EMQ) de  $\hat{g}_T$  au point  $x \in \mathbb{R}^d$  est définie par

$$\text{EMQ}(\hat{g}_T(x)) = \mathbb{E} \left( \hat{g}_T(x) - g(x) \right)^2,$$

et l'estimateur  $\hat{g}_T$  est dit convergent en moyenne quadratique au point  $x \in \mathbb{R}^d$  si

$$\text{EMQ}(\hat{g}_T(x)) \xrightarrow{T \rightarrow \infty} 0.$$

**Définition 1.4 (Erreur en moyenne quadratique intégrée).** L'erreur en moyenne quadratique intégrée (EMQI) de  $\hat{g}_T$  est définie par

$$\text{EMQI}(\hat{g}_T) = \mathbb{E} \int_{\mathbb{R}^d} (\hat{g}_T(x) - g(x))^2 dx = \int_{\mathbb{R}^d} \mathbb{E} (\hat{g}_T(x) - g(x))^2 dx,$$

et l'estimateur  $\hat{g}_T$  est dit convergent en moyenne quadratique intégrée si

$$\text{EMQI}(\hat{g}_T) \xrightarrow{T \rightarrow \infty} 0.$$

**Remarque 1.2.** Notons que l'EMQ de  $\hat{g}_T(x)$  peut s'écrire simplement comme la somme d'un terme de biais au carré et d'un terme de variance :

$$\text{EMQ}(\hat{g}_T(x)) = \left( \mathbb{E}(\hat{g}_T(x)) - f(x) \right)^2 + \text{Var}(\hat{g}_T(x)).$$

De même, l'EMQI de  $\hat{g}_T$  se décompose en une somme d'un terme de biais au carré intégré (BCI) et d'un terme de variance intégrée (VI) :

$$\text{BCI}(\hat{g}_T) = \int_{\mathbb{R}^d} \left( \mathbb{E}(\hat{g}_T(x)) - f(x) \right)^2 dx \quad \text{et} \quad \text{VI}(\hat{g}_T) = \int_{\mathbb{R}^d} \text{Var}(\hat{g}_T(x)) dx.$$

- **Convergence presque sûre**

Nous définissons deux critères de convergence ponctuelle et uniforme comme suit :

**Définition 1.5 (Convergence ponctuelle presque sûre).** On dit que  $\hat{g}_T$  converge presque sûrement vers  $g$  au point  $x \in \mathbb{R}^d$  et on note

$$\hat{g}_T(x) \xrightarrow[T \rightarrow \infty]{p.s.} g(x),$$

si

$$P \left( \omega \in \Omega; \lim_{T \rightarrow \infty} \hat{g}_T(x) = g(x) \right) = 1.$$

**Définition 1.6 (Convergence uniforme presque sûre).** On dit que  $\hat{g}_T$  converge uniformément presque sûrement vers  $g$  si

$$\sup_{x \in \mathbb{R}^d} |\hat{g}_T(x) - g(x)| \xrightarrow[T \rightarrow \infty]{p.s.} 0.$$

**Remarque 1.3.** Rappelons que les convergences en moyenne quadratique et presque sûre entraînent la convergence en probabilité.

## 1.2 Estimation et convergence en moyenne quadratique

Dans cette section, nous définissons une version “temps continu” de l'histogramme et du polygone de fréquences et nous étudions la convergence en moyenne quadratique (ponctuelle et intégrée) de ces estimateurs.

### 1.2.1 L'histogramme

L'histogramme est le plus ancien et le plus connu des estimateurs non paramétriques. Simple et rapide à mettre en œuvre, il est encore fréquemment utilisé en pratique. Dans le cas i.i.d., les principales propriétés de cet estimateur sont développées dans les ouvrages de Bosq et Lecoutre (1987) (chapitre 6, partie A) et de Scott (1992) (chapitre 3). L'étude qui suit propose des extensions de ces résultats au cadre continu et fortement mélangeant sous des hypothèses faibles de stationnarité.

**Construction de l'estimateur.** On obtient un histogramme en réalisant dans un premier temps une partition équilibrée  $\Pi_T$  de  $\mathbb{R}^d$  en hypercubes de volume  $h_T^d$  tel que  $T \rightarrow \infty$ ,  $h_T \rightarrow 0^+$ ,  $Th_T^d \rightarrow \infty$  :

$$\Pi_T = \{\pi_{Tj}, j \in \mathbb{Z}^d\},$$

et

$$\pi_{Tj} = \prod_{k=1}^d [b_{j_k}, b_{j_k+1}[ = \prod_{k=1}^d \left[ c_{j_k} - \frac{h_T}{2}, c_{j_k} + \frac{h_T}{2} \right], \quad j = (j_1, \dots, j_d) \in \mathbb{Z}^d,$$

où  $b_j = (b_{j_1}, \dots, b_{j_d})' \in \mathbb{R}^d$ ,  $b_{j_k+1} - b_{j_k} = h_T$  et  $c_{j_k} = \frac{b_{j_k} + b_{j_k+1}}{2}$ . Le paramètre  $h_T$  est usuellement désigné comme le *paramètre de lissage* ou la *fenêtre* de nos estimateurs. Remarquons que nous pouvons tout aussi bien partitionner  $\mathbb{R}^d$  par des hyperrectangles de volume  $h_{1,T} \times \dots \times h_{d,T}$ , les preuves sont alors similaires avec des notations plus lourdes.

**Définition 1.7.** Au point  $x \in \mathbb{R}^d$  et relativement à la partition  $\Pi_T$ , l'histogramme est défini par

$$\hat{f}_T^H(x) = \sum_j \left[ \frac{1}{Th_T^d} \int_0^T \mathbf{1}_{\pi_{Tj}}(X_t) dt \right] \mathbf{1}_{\pi_{Tj}}(x) =: \sum_j \hat{f}_j \mathbf{1}_{\pi_{Tj}}(x),$$

qui s'interprète comme le temps moyen que passe le processus dans l'hypercube  $\pi_{Tj}$  sur la période  $[0, T]$  normalisé par la mesure  $h_T^d$  de  $\pi_{Tj}$ .

Désignons par  $\|\cdot\|$  une norme de  $\mathbb{R}^d$ . On rappelle la définition de l'intégrale généralisée de Riemann (IGR) qui sera utile pour démontrer nos principaux résultats dans le cadre de la convergence intégrée (cf. Bertrand-Retali (1978)) :

**Définition 1.8.** Soit  $g$  une fonction réelle définie sur  $\mathbb{R}^d$ . Si  $g_n^-$  et  $g_n^+$  sont deux suites de fonctions  $\Pi_n$ -simples, Riemann-intégrables sur  $\mathbb{R}^d$  et telles que

$$g_n^- \leq g_{n+1}^- \leq g \leq g_{n+1}^+ \leq g_n^+ \quad \text{et} \quad \lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} g_n^-(x) dx = \lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} g_n^+(x) dx = I,$$

alors  $I$  est l'intégrale généralisée de Riemann de la fonction  $g$ . Il en résulte que  $g$  est Lebesgue-intégrable sur  $\mathbb{R}^d$  et que  $\int_{\mathbb{R}^d} g(x) dx = I$ . De plus,

$$\forall \varepsilon > 0, \quad g_\varepsilon(x) = \sup_{t \in \mathbb{R}^d} \{g(t); \|t - x\| \leq \varepsilon\}$$

est Lebesgue-intégrable sur  $\mathbb{R}^d$  et  $\int_{\mathbb{R}^d} g_\varepsilon(x) dx$  reste bornée quand  $\varepsilon \rightarrow 0$ .

Nous admettrons dans la suite que toute fonction intégrable l'est au sens généralisé de Riemann. Désignons par  $\nu_x$  un voisinage ouvert fixé du point  $x$  dans  $\mathbb{R}^d$  et notons  $\mathcal{C}^k(E)$  l'ensemble des fonctions  $k$  fois continûment dérivables sur le sous-ensemble  $E$  de  $\mathbb{R}^d$  et  $L^k(\mathbb{R}^d)$  l'ensemble des fonctions de puissance  $k$  intégrable sur  $\mathbb{R}^d$ . En prenant  $\Lambda = \nu_x$  dans le cas ponctuel puis  $\Lambda = \mathbb{R}^d$  dans le cas intégré, les principales hypothèses portant sur la vraie densité et le processus sont les suivantes :

### Hypothèses R

- (i)  $\exists K_f^x > 0, |f(y) - f(z)| \leq K_f^x \|y - z\|$  pour  $(y, z) \in \nu_x^2$ ;
- (i)'  $f \in \mathcal{C}^2(\mathbb{R}^d)$  et les dérivées partielles et leurs carrés admettent une IGR;
- (ii)  $f$  est continue sur  $\Lambda$  et  $\sup_{x \in \Lambda} f(x) \leq M_1 < \infty$  (où  $M_1 =: M_1^x$  si  $\Lambda = \nu_x$ ).

Hypothèses H<sub>1</sub> Il existe un borélien de  $\mathbb{R}^2$  :  $\Gamma = \{(s, t) \in \mathbb{R}^2, |t - s| \leq u_0, u_0 > 0\}$  tel que

- (i)  $g_{s,t}$  existe pour  $(s, t) \notin \Gamma$ ;
- (ii)  $\sup_{(y,z) \in \Lambda^2} f_{u|0}(z|y) \leq M_2 < \infty$  pour  $u \geq u_0$  (où  $M_2 =: M_2^x$  si  $\Lambda = \nu_x$ );
- (iii)  $X$  est arithmétiquement fortement mélangeant de coefficient  $\alpha_X$  :  $\alpha_X(u) \leq a_0 u^{-\varrho}$  pour  $u \geq u_0, a_0 > 0$  et  $\varrho > 2$ .

### Hypothèses H<sub>2</sub> (d = 1)

- (i)  $g_{s,t}$  existe pour  $s \neq t$ ;
- (ii)  $\int_0^\infty \|g_u\|_\infty du < \infty$  et  $g_u(\cdot, \cdot)$  est continue en  $(x, x)$  pour tout  $u > 0$ ;
- (ii)' a)  $\forall y \in \Lambda, \sup_{z \in \Lambda} \int_0^\infty |g_u(y, z)| du \leq \varphi(y)$  où  $\varphi$  est une fonction définie sur  $\Lambda$ , positive, continue et bornée,
- b)  $\varphi \in L^1(\mathbb{R})$ .

Les différentes hypothèses sont classiques en estimation non paramétrique. Les conditions  $R$  portent sur la régularité de la densité  $f$ . En particulier, la condition lipschitzienne  $R(i)$  que l'on retrouve dans Tran (1994) est faible et permet d'obtenir l'ordre de la convergence ponctuelle du terme de biais (sans constante asymptotique). Dans le cas intégré, nous la remplaçons par la condition plus forte  $R(i)'$  donnée par Lecoutre (1985) faisant intervenir la dérivée seconde de  $f$  pour donner une expression asymptotique exacte du biais au carré intégré. Concernant le processus, l'existence de  $\Gamma$  (hypothèses  $H_1$ ),  $H_2(ii)$  et  $H_2(ii)'$  sont des conditions locales spécifiques au temps continu. L'introduction du borélien  $\Gamma$  est ainsi utilisée par Bosq (1998) dans le contexte de l'estimateur à noyau pour une classe plus générale de boréliens qui satisfont la condition  $\limsup_{T \rightarrow \infty} \frac{1}{T} \iint_{[0,T]^2 \cap \Gamma} ds dt < \infty$ ; cette technique qui exclut seulement l'information apportée par le couple  $(X_s, X_t)$  pour les proches valeurs de  $s$  et  $t$  est peu contraignante. Plus restrictive, la condition  $H_2(ii)$  de Castellana et Leadbetter (1986) est liée aux propriétés locales des trajectoires et tend à contrôler le comportement explosif de l'intégrale en  $u$  de la densité jointe de  $(X_0, X_u)$  au voisinage de zéro. Celle-ci peut-être affaiblie par la condition  $H_2(ii)'$  introduite par

Leblanc (1997) et qui est notamment vérifiée par une large classe de processus de diffusion à valeurs réelles (voir aussi Veretennikov (1999)).  $H_1(iii)$  est une condition sur l'ordre de la décroissance du coefficient de mélange  $\alpha$  (ici plus faible que la décroissance géométrique) qui assure l'indépendance asymptotique des variables suffisamment éloignées dans le temps. De même, les conditions locales  $H_2(ii)$  et  $H_2(ii)'$  font aussi office de conditions d'indépendance asymptotique pour les grandes valeurs de  $u$ . Nous utiliserons alternativement  $H_1$  et  $H_2$  avec  $R$  pour obtenir les deux vitesses extrêmes de convergence dites optimale et suroptimale. On supposera ensuite, pour tout  $x \in \mathbb{R}^d$ , l'existence d'un indice  $j(x, T) \in \mathbb{Z}^d$  tel que  $x \in \pi_{Tj(x, T)}$  ( $=: \pi_{Tj}$ ). Pour l'étude spécifique du cas ponctuel, on notera l'existence d'un rang  $T_0 > 0$  tel que pour tout  $T \geq T_0 : \pi_{Tj} \subset \nu_x$ ; les constantes, notées  $K_f^x$ ,  $M_1^x$  et  $M_2^x$ , seront alors dépendantes du point  $x$ .

- **Convergence ponctuelle en moyenne quadratique ( $\Lambda = \nu_x$ )**

Le lemme suivant donne l'ordre de la convergence ponctuelle du terme de biais.

**Lemme 1.2.** *Si la condition  $R(i)$  est satisfaite, alors on a*

$$\limsup_{T \rightarrow \infty} h_T^{-1} \cdot \left| \mathbb{E} \hat{f}_T^H(x) - f(x) \right| \leq C_f(x) < \infty, \quad x \in \mathbb{R}^d,$$

où  $C_f(x)$  est une constante positive qui dépend à la fois de  $x$  et de  $f$ .

Les résultats suivants apportent des majorations de la variance de l'histogramme et des vitesses de convergence de l'erreur en moyenne quadratique au point  $x \in \mathbb{R}^d$  :

**Théorème 1.1.** *Si les conditions  $R(ii)$  et  $H_1$  sont vérifiées, alors il existe  $T_0 > 0$  tel que pour tout  $T \geq T_0$  et pour tout  $1 < p \leq \varrho - 1$  :*

$$\begin{aligned} Th_T^d \cdot \text{Var}(\hat{f}_T^H(x)) &\leq f(\xi_j) (1 - h_T^d f(\xi_j)) \cdot 2u_0 \left(1 - \frac{u_0}{2T}\right) + 2M^x f(\xi_j) \cdot h_T^\varepsilon \\ &\quad + \frac{4p^2 (2a_0)^{1/p}}{\varrho - p} f(\xi_j)^{1 - \frac{1}{p}} \cdot h_T^{\frac{1}{p} \{(d-\varepsilon)(\varrho-p) - d\}}, \end{aligned}$$

avec  $M^x = \max(M_1^x, M_2^x)$  (dépendante de  $x$ ),  $0 \leq \varepsilon \leq d \left(1 - \frac{1}{\varrho-p}\right)$ ,  $x \in \pi_{Tj}$  et  $\xi_j \in \pi_{Tj}$ .

**Remarque 1.4.** Sous  $R$  et  $H_1$ , on retrouve un résultat semblable à celui obtenu, en particulier, en temps discret par Rio (2000) (chapitre 1) dans le cas d'un processus strictement stationnaire et fortement mélangeant. En changeant  $H_1(iii)$  pour la condition plus forte de mélange géométrique :  $\alpha_X(u) \leq a_0 \varrho^u$ ,  $u > 0$ ,  $a_0 > 0$ ,  $0 < \varrho < 1$ , la condition  $1 < p \leq \varrho - 1$  devient  $p > 1$ .

**Théorème 1.2.** (i) *Si les conditions  $H_2(i)(ii)$  sont vérifiées, alors*

$$T \cdot \text{Var}(\hat{f}_T^H(x)) \xrightarrow{T \rightarrow \infty} 2 \int_0^\infty g_u(x, x) dx, \quad x \in \mathbb{R};$$

(ii) *Si les conditions  $H_2(i)(ii)'_a$  sont vérifiées, alors*

$$T \cdot \text{Var}(\hat{f}_T^H(x)) \leq 2\varphi(\eta_j), \quad x \in \pi_{Tj}, \quad \eta_j \in \pi_{Tj}.$$

**Remarque 1.5.** Sous les conditions  $H_2(ii)$  de Castellana-Leadbetter, on obtient le comportement exact de la variance ponctuelle de l'histogramme avec la même constante asymptotique que pour le noyau. Cette constante est aussi minimax pour des processus de diffusion ergodiques (voir Kutoyants (1998)).

De là, pour des choix *ad hoc* du paramètre  $h_T$ , on obtient les vitesses optimale et suroptimale de la convergence ponctuelle en moyenne quadratique pour l'histogramme :

**Corollaire 1.1.** (i) Cas "optimal" : sous les conditions  $R(i)(ii)$  et  $H_1$ , le choix  $h_T = c_T \cdot T^{-\frac{1}{d+2}}$  tel que  $c_T \rightarrow c$ ,  $0 < c < \infty$ , entraîne

$$\text{EMQ}(\hat{f}_T^H(x)) = \mathcal{O}\left(T^{-\frac{2}{d+2}}\right), \quad x \in \mathbb{R}^d;$$

(ii) Cas "suroptimal" ( $d = 1$ ) : si les conditions  $R(i)$  et  $H_2(i)(ii)$  sont vérifiées, alors pour  $h_T = o\left(T^{-\frac{1}{2}}\right)$  :

$$\lim_{T \rightarrow \infty} T \cdot \text{EMQ}(\hat{f}_T^H(x)) = 2 \int_0^\infty g_u(x, x) du, \quad x \in \mathbb{R},$$

et si on remplace la condition  $H_2(ii)$  par  $H_2(ii)'_a$ , on a

$$\limsup_{T \rightarrow \infty} T \cdot \text{EMQ}(\hat{f}_T^H(x)) \leq 2\varphi(x), \quad x \in \mathbb{R}.$$

• **Convergence en moyenne quadratique intégrée ( $\Lambda = \mathbb{R}^d$ )**

Notons  $f'_i := \partial f / \partial x_i$  et définissons le *coefficient de rigidité* ( $\mathcal{R}$ ) de  $f'_i$  par sa norme  $L^2$  i.e.  $\mathcal{R}(f'_i) := \int_{\mathbb{R}^d} f'_i(x)^2 dx$ . Pour une classe plus restreinte de densités, ici deux fois continûment différentiables, le lemme suivant dû à Lecoutre (1985) nous donne une expression du BCI de l'histogramme multivarié :

**Lemme 1.3.** Si la condition  $R(i)'$  est vérifiée, alors

$$\text{BCI}(\hat{f}_T^H) = \frac{h_T^2}{12} \mathcal{R}_d(f') \times (1 + o(1)),$$

avec  $\mathcal{R}_d(f') := \sum_{i=1}^d \mathcal{R}(f'_i)$ .

Les majorations de la variance intégrée se déduisent ensuite directement des théorèmes 1.1 et 1.2.

**Corollaire 1.2.** (i) Cas "optimal" : si les conditions  $R(ii)$  et  $H_1$  sont vérifiées et si  $f^{1-\frac{1}{p}} \in L^1(\mathbb{R}^d)$  avec  $1 < p \leq \varrho - 1$ , alors pour tout  $0 \leq \varepsilon \leq d \left(1 - \frac{1}{\varrho-p}\right)$  :

$$Th_T^d \cdot \text{VI}(\hat{f}_T^H) \leq \left\{ 2u_0 + 2M \cdot h_T^\varepsilon + \frac{4p^2(2a_0)^{1/p}}{\varrho - p} \|f^{1-\frac{1}{p}}\|_1 \cdot h_T^{\frac{1}{p}\{(d-\varepsilon)(\varrho-p)-d\}} \right\} (1 + o(1));$$

(ii) Cas "suroptimal" ( $d = 1$ ) : si les conditions  $H_2(i)(ii)'_{ab}$  sont vérifiées, alors

$$\limsup_{T \rightarrow \infty} T \cdot \text{VI}(\hat{f}_T^H) \leq 2\|\varphi\|_1.$$

Par passage à la limite et pour les choix de  $h_T$  du cas ponctuel, on retrouve les mêmes vitesses optimale et suroptimale de convergence pour l'histogramme :

**Corollaire 1.3.** (i) Cas “optimal” : sous les conditions  $R(i)'(ii)$  et  $H_1$  et si  $f^{1-\frac{1}{p}} \in L^1(\mathbb{R}^d)$  pour  $1 < p \leq \varrho - 1$ , le choix  $h_T = c_T \cdot T^{-\frac{1}{d+2}}$  tel que  $c_T \rightarrow c$ ,  $0 < c < \infty$ , entraîne

$$\limsup_{T \rightarrow \infty} T^{\frac{2}{d+2}} \cdot \text{EMQI}(\hat{f}_T^H) \leq \begin{cases} \frac{c^2}{12} \mathcal{R}_d(f') + \frac{1}{c^d} \left( 2u_0 + 2M + \frac{4p^2(2a_0)^{1/p}}{\varrho-p} \|f^{1-\frac{1}{p}}\|_1 \right), \\ \text{si } p = \varrho - 1; \\ \frac{c^2}{12} \mathcal{R}_d(f') + \frac{2u_0}{c^d}, \text{ si } p < \varrho - 1; \end{cases}$$

(ii) Cas “suroptimal” ( $d = 1$ ) : si les conditions  $R(i)'$  et  $H_2(i)(ii)'_{ab}$  sont vérifiées, alors pour  $h_T = o\left(T^{-\frac{1}{2}}\right)$  :

$$\limsup_{T \rightarrow \infty} T \cdot \text{EMQI}(\hat{f}_T^H) \leq 2\|\varphi\|_1.$$

**Remarque 1.6.** Sous  $R$  et  $H_1$ , on retrouve dans la partie (i) du corollaire 1.3 des résultats similaires à ceux obtenus dans le cas i.i.d. par Scott (1979), Freedman et Diaconis (1981) puis Lecoutre (1985) en multivarié. Dans la partie (ii), sous les conditions  $R$  et  $H_2$ , on obtient aussi la vitesse suroptimale, propre au temps continu, avec une variance asymptotique ne dépendant plus de  $h_T$  (contrairement au cas discret). Rappelons que la condition sur  $p$  se réduit à  $p > 1$  dans le cas géométriquement mélangeant.

## 1.2.2 Le polygone de fréquences ( $d = 1$ )

Il est possible d'améliorer la vitesse de convergence du biais de l'histogramme en introduisant le polygone de fréquences. Les principales propriétés de cet estimateur ont notamment été explorées par Scott (1992) (chapitre 4) en i.i.d., par Carbon *et al.* (1997) dans le cas mélangeant et, plus récemment, par Carbon (2005, 2006) et par Bensaïd et Dabo-Niang (2007) (discrétisation) pour l'extension aux champs aléatoires. En particulier, Scott (1985) a démontré le premier ce résultat remarquable, à savoir que la vitesse de convergence du noyau en  $n^{-4/5}$  était également atteinte par le polygone de fréquences pour des classes de densités bornées et deux fois continûment différentiables. L'étude qui suit est une extension au temps continu et, ce afin de simplifier la présentation, elle sera considérée dans le cadre *unidimensionnel* ( $d = 1$ ).

Notations. On désigne par  $f'$  et  $f''$  les deux premières dérivées de  $f$  et on rappelle que  $\mathcal{R}(f'') := \int_{\mathbb{R}} f''(x)^2 dx$ .

**Construction de l'estimateur.** On construit le polygone de fréquences en traçant une droite segmentée reliant les points centraux des sommets de chaque rectangle de l'histogramme. L'estimateur ainsi obtenu est continu sur  $\mathbb{R}$ .

**Définition 1.9.** Au point  $x \in \mathbb{R}$ , on définit l'estimateur  $\hat{f}_T^{PF}$  de  $f$  par

$$\hat{f}_T^{PF}(x) = \sum_j \left[ \left( \frac{\hat{f}_{j+1} - \hat{f}_j}{h_T} \right) x + \left( \frac{c_{j+1} \hat{f}_j - c_j \hat{f}_{j+1}}{h_T} \right) \right] \mathbb{1}_{[c_j, c_{j+1}[}(x),$$

qui peut s'écrire également sous la forme symétrique de Lagrange

$$\hat{f}_T^{PF}(x) = \sum_j \left[ \left( \frac{x - c_j}{h_T} \right) \hat{f}_{j+1} + \left( \frac{c_{j+1} - x}{h_T} \right) \hat{f}_j \right] \mathbb{1}_{[c_j, c_{j+1}]}(x),$$

ou encore sous la forme des différences divisées de Newton

$$\hat{f}_T^{PF}(x) = \sum_j \left[ \hat{f}_j + (x - c_j) \frac{\hat{f}_{j+1} - \hat{f}_j}{h_T} \right] \mathbb{1}_{[c_j, c_{j+1}]}(x).$$

**Remarque 1.7.** On trouve dans la littérature des variantes du polygone de fréquences (“Frequency Polygon”) en modifiant la méthode d’interpolation entre les rectangles de l’histogramme. Parmi elles, le “Edge Frequency Polygon”, introduit par Jones *et al.* (1998), s’obtient en calculant les valeurs moyennes  $\{(\hat{f}_{j-1} + \hat{f}_j)/2\}$  aux bords  $\{b_j\}$  des intervalles de la partition  $\Pi_T$  puis en reliant ces valeurs par des segments de droite. Le “Generalized Edge Frequency Polygon”, proposé ensuite par Dong et Zheng (2001), généralise le précédent estimateur en remplaçant la valeur calculée à chacun des bords  $\{b_j\}$  par une moyenne pondérée des  $2k$ ,  $k \geq 1$ , valeurs  $\hat{f}_{j-k}, \dots, \hat{f}_j, \dots, \hat{f}_{j+k-1}$ . En i.i.d. et en univarié, notons que l’erreur en moyenne quadratique intégrée asymptotique de ces différentes versions du polygone de fréquences garde la même forme à des constantes de biais et de variance près :

$$\text{EMQI}(\hat{f}_n^{FP}) \approx \frac{49}{2880} \mathcal{R}(f'') h_n^4 + \frac{2}{3} \frac{1}{nh_n};$$

$$\text{EMQI}(\hat{f}_n^{EFP}) \approx \frac{23}{360} \mathcal{R}(f'') h_n^4 + \frac{5}{12} \frac{1}{nh_n};$$

$$\begin{aligned} \text{EMQI}(\hat{f}_n^{GEFP}) &\approx \frac{23 + 90\tau + 90\tau^2}{360} \mathcal{R}(f'') h_n^4 \\ &+ \left( \frac{1}{4} \omega_1^2 + \frac{1}{6} \omega_k^2 + \frac{1}{6} \sum_{i=1}^{k-1} (\omega_i^2 + \omega_i \omega_{i+1} + \omega_{i+1}^2) \right) \frac{1}{nh_n}, \end{aligned}$$

où  $k \geq 2$  est fixé,  $\sum_{i=1}^k \omega_i = 1$ ,  $\omega_i \geq 0$  et  $\tau = \sum_{i=1}^k \omega_i (-i + i^2)$ . Si  $k = 1$  ( $\Leftrightarrow \omega_1 = 1$ ), le “GEFP” coïncide avec le “EFP”. Par ailleurs, des choix optimaux du paramètre  $h_n$  minimisant ces erreurs sont également proposés dans les références mentionnées ci-dessus.

- **Convergence ponctuelle en moyenne quadratique ( $\Lambda = \nu_x$ )**

Dans le cas du polygone de fréquences, nous modifions légèrement les hypothèses de régularité sur  $f$  données par  $R$  :

**Hypothèses  $\bar{R}$**

- (i)  $f \in \mathcal{C}^2(\nu_x)$  et  $f''$  est bornée sur  $\nu_x$  ;
- (ii)  $f$  est continue sur  $\nu_x$  et  $\sup_{u \in \nu_x} f(u) \leq M_1^x < \infty$ .

Le lemme suivant donne l’ordre de la convergence ponctuelle du terme de biais.

**Lemme 1.4.** *Si la condition  $\bar{R}(i)$  est satisfaite, alors on a*

$$\limsup_{T \rightarrow \infty} h_T^{-2} \cdot \left| \mathbb{E} \hat{f}_T^{PF}(x) - f(x) \right| \leq \frac{5}{16} \sup_{u \in \nu_x} |f''(u)| < \infty, \quad x \in \mathbb{R}.$$

Sous les mêmes hypothèses  $\bar{R}(ii)$ ,  $H_1$  ou  $H_2$  (avec  $H_2(ii)$  ou  $H_2(ii)'_a$ ), le comportement asymptotique de la variance ponctuelle du polygone de fréquences s'obtient à partir des résultats donnés par les théorèmes 1.1 et 1.2 pour l'histogramme. En particulier, on retrouve avec les conditions  $H_2(ii)$  de Castellana-Leadbetter le comportement limite exact de la variance ponctuelle.

**Corollaire 1.4.** (i) *Cas "optimal" : si les conditions  $\bar{R}(i)$  et  $H_1$  sont vérifiées, alors*

$$\limsup_{T \rightarrow \infty} T h_T \cdot \text{Var}(\hat{f}_T^{PF}(x)) \leq 2u_0 f(x), \quad x \in \mathbb{R};$$

(ii) *Cas "suroptimal" : si les conditions  $H_2(i)(ii)$  sont vérifiées, alors*

$$T \cdot \text{Var}(\hat{f}_T^{PF}(x)) \xrightarrow{T \rightarrow \infty} 2 \int_0^\infty g_u(x, x) dx, \quad x \in \mathbb{R},$$

*et si on remplace la condition  $H_2(ii)$  par  $H_2(ii)'_a$ , alors*

$$\limsup_{T \rightarrow \infty} T \cdot \text{Var}(\hat{f}_T^{PF}(x)) \leq 2\varphi(x), \quad x \in \mathbb{R}.$$

Pour de nouveaux choix du paramètre  $h_T$ , on obtient les vitesses optimale et suroptimale de la convergence ponctuelle en moyenne quadratique du polygone de fréquences :

**Corollaire 1.5.** (i) *Cas "optimal" : sous les conditions  $\bar{R}(i)(ii)$  et  $H_1$ , le choix  $h_T = c_T \cdot T^{-\frac{1}{5}}$  tel que  $c_T \rightarrow c$ ,  $0 < c < \infty$ , entraîne*

$$\text{EMQ}(\hat{f}_T^{PF}(x)) = \mathcal{O}\left(T^{-\frac{4}{5}}\right), \quad x \in \mathbb{R};$$

(ii) *Cas "suroptimal" : si les conditions  $\bar{R}(i)$  et  $H_2(i)(ii)$  sont vérifiées, alors pour  $h_T = o\left(T^{-\frac{1}{4}}\right)$  :*

$$\lim_{T \rightarrow \infty} T \cdot \text{EMQ}(\hat{f}_T^{PF}(x)) = 2 \int_0^\infty g_u(x, x) du, \quad x \in \mathbb{R},$$

*et si on remplace la condition  $H_2(ii)$  par  $H_2(ii)'_a$ , on a*

$$\limsup_{T \rightarrow \infty} T \cdot \text{EMQ}(\hat{f}_T^{PF}(x)) \leq 2\varphi(x), \quad x \in \mathbb{R}.$$

**Remarque 1.8.** Notons  $r$  le nombre de dérivées associé à la fonction  $f$ ,  $f^{(k)}$  la dérivée d'ordre  $k$  de  $f$  et introduisons l'ensemble  $\mathcal{C}_r^d(\ell)$  ( $r = k + \lambda$ ,  $0 < \lambda \leq 1$ ,  $k \in \mathbb{N}$ ) des fonctions réelles définies sur  $\mathbb{R}^d$  et  $k$  fois continûment différentiables telles que

$$\left| \frac{\partial f^{(k)}}{\partial x_1^{j_1} \dots \partial x_d^{j_d}}(y) - \frac{\partial f^{(k)}}{\partial x_1^{j_1} \dots \partial x_d^{j_d}}(z) \right| \leq \ell \|y - z\|^\lambda,$$

avec  $(y, z) \in \mathbb{R}^{2d}$ ,  $j_1 + \dots + j_d = k$ . La vitesse en  $T^{-4/5}$  obtenue dans la partie (i) du corollaire 1.5 est optimale au sens où elle est effectivement atteinte par un processus

construit de manière analogue à Bosq (1998) (p. 96) et appartenant à une certaine classe  $\mathcal{X}$  de processus  $X = \{X_t, t \in \mathbb{R}\}$  (non nécessairement stationnaires toutefois) telle que, pour chaque  $X \in \mathcal{X}$  admettant une densité  $f_X \in \mathcal{C}_r^d(\ell)$  (dans notre cas, pour  $d = 1$  et  $r = 2$ ), on ait

$$\left\{ \begin{array}{l} \sup_{x \in \mathbb{R}^d} f_X(x) < \infty; \\ \frac{1}{T} \iint_{[0, T]^2 \cap \Gamma_X} ds dt \leq L_0 \left(1 + \frac{L_0}{T}\right), \quad L_0 > 0; \\ \exists p_0 \in ]2, \infty], \quad \sup_{(s, t) \notin \Gamma_X} \left( \iint_{\mathbb{R}^{2d}} |g_{s, t}(y, z)|^{p_0} dy dz \right)^{1/p_0} < \infty; \\ \alpha_X^{(2)}(|t - s|) \leq a_0 |t - s|^{-\varrho}, \quad a_0 < \infty, \quad \varrho \geq \frac{2p_0 - 1}{p_0 - 2}, \end{array} \right.$$

où le coefficient  $\alpha_X^{(2)}$  est défini par

$$\alpha_X^{(2)}(u) := \sup_{t > 0} \alpha\left(\sigma(X_t), \sigma(X_{t+u})\right), \quad u > 0.$$

Cette vitesse est de plus minimax d'après un résultat de Bosq (1998) (théorème 4.3, p. 97) pour cette même classe de processus. Si  $\mathcal{F}_T$  est la classe des estimateurs mesurables basés sur l'observation d'une trajectoire de  $X$  sur  $[0, T]$ , on a alors

$$\liminf_{T \rightarrow \infty} \inf_{\hat{f}_T \in \mathcal{F}_T} \sup_{X \in \mathcal{X}} T^{\frac{2r}{2r+d}} \cdot \mathbb{E}_X \left( \hat{f}_T(x) - f_X(x) \right)^2 > 0, \quad x \in \mathbb{R}^d.$$

### • Convergence en moyenne quadratique intégrée ( $\Lambda = \mathbb{R}$ )

Pour donner une expression asymptotique exacte du BCI de  $\hat{f}_T^{PF}$ , nous modifions encore les hypothèses de régularité sur  $f$  données par  $R$  :

#### Hypothèses $\tilde{R}$

- (i)  $f \in \mathcal{C}^2(\mathbb{R})$ ,  $f(x) > 0$   $\mu - p.p.$ ,  $f'' \in L^1(\mathbb{R})$  et  $f, f'' \in L^2(\mathbb{R})$ ;
- (ii)  $|f''(y) - f''(z)| \leq l_0 |y - z|^\nu$  pour  $l_0 > 0$ ,  $\nu \in ]0, 1]$  et  $(y, z) \in \mathbb{R}^2$ ;
- (iii)  $f$  est continue sur  $\mathbb{R}$  et  $\|f\|_\infty = M_1 < \infty$ .

Le lemme suivant nous donne le BCI du polygone de fréquences. Il s'agit d'une version analogue au résultat de Scott (1985) obtenue sous des conditions légèrement différentes. Notons qu'avec la condition höldérienne  $\tilde{R}(ii)$ , nous n'avons pas besoin de supposer l'existence de la dérivée troisième de  $f$  tout en obtenant la même constante asymptotique.

**Lemme 1.5.** *Si les conditions  $\tilde{R}(i)(ii)$  sont vérifiées, alors*

$$\text{BCI}(\hat{f}_T^{PF}) = \frac{49}{2880} \mathcal{R}(f'') h_T^4 \times (1 + o(1)).$$

Les majorations de la variance intégrée du polygone de fréquences se déduisent directement du corollaire 1.4.

**Corollaire 1.6.** (i) *Cas "optimal" : si les conditions  $H_1$  sont vérifiées et si  $f^{1-\frac{1}{p}} \in L^1(\mathbb{R})$  avec  $1 < p \leq \varrho - 1$ , alors pour  $0 \leq \varepsilon \leq 1 - \frac{1}{\varrho - p}$  :*

$$Th_T \cdot \text{VI}(\hat{f}_T^{PF}) \leq \left\{ 2u_0 + 2M \cdot h_T^\varepsilon + \frac{4p^2(2a_0)^{1/p}}{\varrho - p} \|f^{1-\frac{1}{p}}\|_1 \cdot h_T^{\frac{1}{p}\{(1-\varepsilon)(\varrho-p)-1\}} \right\} (1 + o(1));$$

(ii) Cas “suroptimal” : si les conditions  $H_2(i)(ii)'_{ab}$  sont vérifiées, alors

$$\limsup_{T \rightarrow \infty} T \cdot \text{VI}(\hat{f}_T^{PF}) \leq 2\|\varphi\|_1.$$

Finalement, pour les choix de  $h_T$  du cas ponctuel, on obtient les vitesses optimale et suroptimale du polygone de fréquences :

**Corollaire 1.7.** (i) Cas “optimal” : si les conditions  $\tilde{R}$  et  $H_1$  sont vérifiées et si  $f^{1-\frac{1}{p}} \in L^1(\mathbb{R})$  pour  $1 < p \leq \varrho - 1$ , le choix  $h_T = k_T \cdot T^{-\frac{1}{5}}$  tel que  $k_T \rightarrow k$ ,  $0 < k < \infty$ , entraîne

$$\limsup_{T \rightarrow \infty} T^{\frac{4}{5}} \cdot \text{EMQI}(\hat{f}_T^{PF}) \leq \begin{cases} \frac{49}{2880} k^4 \mathcal{R}(f'') + \frac{1}{k} \left( 2u_0 + 2M + \frac{4p^2(2a_0)^{1/p}}{\varrho-p} \|f^{1-\frac{1}{p}}\|_1 \right), \\ \text{si } p = \varrho - 1; \\ \frac{49}{2880} k^4 \mathcal{R}(f'') + \frac{2u_0}{k}, \text{ si } p < \varrho - 1; \end{cases}$$

(ii) Cas “suroptimal” : si les conditions  $\tilde{R}(i)(ii)$  et  $H_2(i)(ii)'_{ab}$  sont vérifiées et si  $h_T = o\left(T^{-\frac{1}{4}}\right)$ , alors

$$\limsup_{T \rightarrow \infty} T \cdot \text{EMQI}(\hat{f}_T^{PF}) \leq 2\|\varphi\|_1.$$

**Remarque 1.9.** Dans la partie (i) du corollaire 1.7, on retrouve la même vitesse de l’erreur intégrée que celle en  $n^{-4/5}$  obtenue par Scott (1985) dans le cas i.i.d. puis par Carbon *et al.* (1997) dans le cas mélangeant.

**Remarque 1.10 (Taille des échantillons).** Des simulations sont proposées par Scott (1985) dans le contexte i.i.d.; celles-ci mettent en évidence l’intérêt de l’estimateur par polygone de fréquences en ce qui concerne la taille des échantillons. Notamment, lorsque la fonction  $f$  théorique est la densité d’une Gaussienne  $\mathcal{N}(0, 1)$ , pour obtenir une erreur en moyenne quadratique intégrée inférieure à  $1/400$  : 2254 observations sont nécessaires avec un histogramme, 431 avec un noyau d’Epanechnikov et 546 avec un polygone de fréquences. Simonoff et Hurvich (1993) présentent des cas de figure où le polygone de fréquence est meilleur que l’estimateur à noyau et soulignent les bonnes performances du polygone pour estimer les queues lourdes de distribution.

**Remarque 1.11 (Positionnement des bords).** Il est clair que la construction des graphiques de fréquences dépend directement de la partition  $\Pi_T$ , c’est-à-dire des choix de la position des bords  $\{b_j\}$  et du paramètre de lissage  $h_T$ . Lorsque  $h_T =: h > 0$  est fixé, la forme des graphiques peut ainsi varier sensiblement avec les bords et induire des interprétations erronées de la densité théorique (cf. Silverman (1986) en temps discret, p. 7-11). Néanmoins, l’influence des bords sur les caractéristiques asymptotiques de ces graphiques apparaît négligeable d’après des simulations de Scott (1992) (p. 65) en i.i.d.. Pour l’histogramme univarié, deux partitions sont notamment considérées où le point  $x = 0$  coïncide tantôt avec le milieu d’un intervalle tantôt avec un bord. Le critère EMQI exact est minimal dans le premier cas de figure et si la fonction  $f$  est la densité d’une loi Gaussienne  $\mathcal{N}(0, 1)$ , l’écart entre les deux erreurs associées est de 1,29% avec 25 observations

et inférieur à  $10^{-5}$  à partir de 100 observations. Pour une loi log-normale, cet écart est aussi inférieur à  $10^{-5}$  à partir de 400 observations. En multivarié, le problème se complique encore avec un choix supplémentaire, celui de l'orientation des axes (ou des hypercubes). Dans un cadre non asymptotique et pour des observations i.i.d. positives, le problème du choix de la partition pour l'histogramme est traité par Baraud et Birgé (2006) au sens du critère de distance de Hellinger ; la procédure proposée est notamment universelle pour une classe importante de fonctions positives qui inclut la fonction de densité, la moyenne d'un vecteur de variables aléatoires positives, l'intensité d'un processus de Poisson non homogène, la fonction de risque (hazard function) et la densité de transition d'un processus Markovien.

### 1.2.3 Exemples de processus

Nous donnons ici, dans le cas réel ( $d = 1$ ), des exemples de processus de diffusion homogènes qui satisfont nos hypothèses. En particulier, ce type de processus est aussi important du point de vue des applications pour décrire la dynamique de phénomènes physiques sujets à des fluctuations aléatoires (par exemple, dans des domaines aussi variés que la médecine, la météorologie ou l'économie). Rappelons que ces processus sont définis comme les solutions de l'équation différentielle stochastique (EDS) :

$$dX_t = S(X_t)dt + \sigma(X_t)dW_t, \quad X_0, \quad 0 \leq t \leq T, \quad (1.1)$$

où  $\{W_t, 0 \leq t \leq T\}$  désigne un processus de Wiener standard et  $X_0$  une valeur initiale, éventuellement aléatoire, indépendante du Wiener. Les fonctions  $S(\cdot)$  et  $\sigma^2(\cdot)$  sont usuellement appelées *dérive* et *coefficient de diffusion*. En mathématiques financières, elles s'interprètent comme la tendance et la volatilité de l'évolution du cours d'une option financière, ce cours étant modélisé par le processus  $X$ . On suppose en outre que  $X$  est un processus de diffusion Markovien admettant une densité de transition  $p_u(z|y)$  telle que  $f_u(y, z) = f(y)p_u(z|y)$ . On trouve alors dans Leblanc (1997) (proposition 11) des conditions de régularité sur  $S$  et  $\sigma$  de telle sorte à avoir des processus strictement stationnaires et géométriquement  $\beta$ -mélangeants qui soient solutions de l'équation (1.1) et qui satisfont  $p_u(z|y) \leq Cu^{-1/2}$ ,  $0 < u \leq 1$ . On retrouve encore cette inégalité dans Veretennikov (1999) si  $\sigma$  est non dégérée, et si  $S$  et  $\sigma$  sont bornées et  $C^\alpha$ -höldériennes. En particulier, ces processus satisfont les conditions locales  $H_2(ii)$  de Castellana-Leadbetter, et celles plus faibles données par  $H_2(ii)'$ , ainsi que l'hypothèse  $H_1(ii)$  portant sur la densité conditionnelle  $f_{u|0}(\cdot|\cdot)$  pour  $u_0 \leq 1$ . Parmi ces solutions, citons le processus d'Ornstein-Uhlenbeck vérifiant l'équation de Langevin :  $dX_t = \theta(\mu - X_t)dt + \sigma dW_t$ ,  $X_0$ ,  $0 \leq t \leq T$ , où  $(\theta, \sigma) > 0$ ,  $\mu \in \mathbb{R}$ , de loi de probabilité invariante  $\mathcal{N}(\mu, \sigma^2/(2\theta))$  ; et le processus "Double Exponentiel" solution de l'EDS :  $dX_t = -\theta \operatorname{sgn}(X_t)dt + dW_t$ ,  $\theta > 0$ ,  $X_0$ ,  $0 \leq t \leq T$ , de densité invariante  $f(x) = \theta e^{-2\theta|x|}$ .

### 1.3 Normalité asymptotique ( $d = 1$ )

Dans cette section, nous supposons le processus  $X$  réel ( $d = 1$ ) et strictement stationnaire. “ $\xrightarrow{d}$ ” désignera la convergence en distribution.

#### 1.3.1 L’histogramme

Lorsque les observations sont i.i.d., la convergence en distribution de l’histogramme multivarié vers une Gaussienne est donnée par Bosq et Lecoutre (1987) (p. 153) avec une vitesse en  $(nh_n^d)^{-1/2}$  sous la condition  $nh_n^{d+2} \rightarrow 0$  quand  $n \rightarrow \infty$ . Pour l’extension au temps continu, on rappelle un théorème central-limite proposé par Bosq *et al.* (1999) pour les tableaux triangulaires multidimensionnels et appliqué pour prouver la normalité asymptotique de l’estimateur à noyau dans les cas discret et continu.

Soit  $Y = \{Y_{n,i}, i \in \mathbb{Z}\}$ ,  $n \geq 1$ , un processus à temps discret, strictement stationnaire, à valeurs dans  $\mathbb{R}^m$ ,  $m \in \mathbb{N}^*$ , et tel que  $Y_{n,i} = \left(Y_{n,i}^{(1)}, \dots, Y_{n,i}^{(m)}\right)'$ . Le coefficient de mélange fort associé à  $Y$  est donné par

$$\alpha_Y(k) = \sup_{n \geq 1} \alpha\left(\sigma(Y_{n,i}, i \leq 0), \sigma(Y_{n,i}, i \geq k)\right).$$

#### Hypothèses $H_3$

- (i)  $E\left(Y_{n,i}^{(r)}\right) = 0$  et  $\left|Y_{n,i}^{(r)}\right| \leq M_n$  pour tout  $n \geq 1$ ,  $i \geq 1$  et  $1 \leq r \leq m$ , où  $M_n$  est une constante qui ne dépend que de  $n$ ;
- (ii)  $\sup_{i \geq 1, 1 \leq r \leq m} E\left(Y_{n,i}^{(r)}\right)^2 < \infty$ ;
- (iii) pour tout  $1 \leq r \leq m$ ,  $1 \leq s \leq m$  et pour toute suite  $(b_n)_{n \geq 1}$  telle que  $b_n \rightarrow \infty$  quand  $n \rightarrow \infty$  et  $b_n \leq n$ , on a

$$\lim_{n \rightarrow \infty} \frac{1}{b_n} E\left(\sum_{i=1}^{b_n} Y_{n,i}^{(r)} \sum_{j=1}^{b_n} Y_{n,j}^{(s)}\right) = \sigma_{r,s};$$

- (iv) il existe un réel  $1 < a < \infty$  tel que  $\sum_{k \geq 1} k \alpha_Y^{\frac{a-1}{a}}(k) < \infty$ ;
- (v)  $M_n \leq C \cdot n^{\frac{a^2}{(3a-1)(2a-1)}}$  pour une constante  $C > 0$  et pour tout  $n \geq 1$ .

**Théorème 1.3.** *Si les conditions  $H_3$  sont vérifiées, alors*

$$\frac{\sum_{i=1}^n Y_{n,i}}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{d} N^{(m)} \sim \mathcal{N}\left(0, \Sigma^{(m)}\right),$$

avec  $\Sigma^{(m)} := (\sigma_{r,s})_{1 \leq r, s \leq m}$ .

La normalité asymptotique de l’histogramme en temps continu se déduit ensuite du théorème 1.3 et des conditions  $H_2(ii)$  de Castellana-Leadbetter. Pour ce faire, nous considérons un  $m$ -uplet  $(x_1, \dots, x_m)$  de valeurs réelles distinctes. Au point  $x_r$ ,  $1 \leq r \leq m$ , on suppose l’existence d’un indice  $j(x_r, T)$  tel que  $x_r \in \pi_{Tj(x_r, T)}$  et l’estimateur par histogramme vaut

$\hat{f}_T^H(x_r) = \frac{1}{Th_T} \int_0^T \mathbb{1}_{\pi_{Tj}(x_r, T)}(X_t) dt$ . Dans cette section, le coefficient  $\alpha_X$  de mélange fort associé au processus  $X$  est donné par

$$\alpha_X(k) = \alpha(\mathcal{F}_{-\infty}^0, \mathcal{F}_k^\infty) \downarrow_{k \rightarrow \infty} 0.$$

**Théorème 1.4.** *Supposons les conditions  $H_2(i)(ii)$  satisfaites sur  $\mathbb{R}$  et que, pour tout  $m$ -uplet  $(x_1, \dots, x_m)$  de valeurs réelles distinctes tel que  $\Sigma^{(m)} = (2 \int_0^\infty g_u(x_r, x_s) du)_{1 \leq r, s \leq m}$  soit définie positive, la fonction  $g_u(\cdot, \cdot)$  est continue au point  $(x_r, x_s)$  pour  $u > 0$  et pour tout  $1 \leq r, s \leq m$ . Si de plus, il existe un réel  $1 < a < \infty$  tel que*

$$\sum_{k \geq 1} k \alpha_X^{(a-1)/a}(k) < \infty,$$

alors pour  $h_T \geq c \cdot T^{-a^2/((3a-1)(2a-1))}$ ,  $0 < c < \infty$  :

$$\sqrt{T} \left( \hat{f}_T^H(x_r) - \mathbb{E} \hat{f}_T^H(x_r), 1 \leq r \leq m \right) \xrightarrow[T \rightarrow \infty]{d} N^{(m)} \sim \mathcal{N}^{(m)}(0, \Sigma^{(m)}).$$

**Remarque 1.12.** La principale hypothèse du théorème 1.4 repose sur l'existence d'un réel  $1 < a < \infty$  tel que le coefficient de mélange fort associé à  $X$  satisfasse la vitesse de convergence  $\sum_{k \geq 1} k \alpha_X^{(a-1)/a}(k) < \infty$ . Notons que lorsque  $a$  est proche de 1, l'hypothèse  $H_3(iv)$  est encore vérifiée si le processus  $X$  est géométriquement fortement mélangeant. Enfin, le cas  $a = \infty$  est également possible (cf. Bosq *et al.* (1999)).

### 1.3.2 Le polygone de fréquences

De façon analogue à l'histogramme, nous établissons dans le contexte suroptimal la normalité asymptotique du polygone de fréquences en temps continu.

**Théorème 1.5.** *Si les conditions du théorème 1.4 sont satisfaites, alors*

$$\sqrt{T} \left( \hat{f}_T^{PF}(x_r) - \mathbb{E} \hat{f}_T^{PF}(x_r), 1 \leq r \leq m \right) \xrightarrow[T \rightarrow \infty]{d} N^{(m)} \sim \mathcal{N}^{(m)}(0, \Sigma^{(m)}).$$

Le corollaire suivant est une version utile du précédent théorème pour construire des régions de confiance du vecteur  $(f(x_1), \dots, f(x_m))$ .

**Corollaire 1.8.** *Sous les conditions du théorème 1.5 et si de plus  $f \in \mathcal{C}^2(\mathbb{R})$  avec  $f''$  bornée sur  $\mathbb{R}$ , alors pour  $h_T = o(T^{-\frac{1}{4}})$  :*

$$\sqrt{T} \left( \hat{f}_T^{PF}(x_r) - f(x_r), 1 \leq r \leq m \right) \xrightarrow[T \rightarrow \infty]{d} N^{(m)} \sim \mathcal{N}^{(m)}(0, \Sigma^{(m)}).$$

**Remarque 1.13.** Les conditions  $h_T = \mathcal{O}(T^{-a^2/((3a-1)(2a-1))})$  et  $h_T = o(T^{-1/4})$  sont compatibles pour des valeurs de  $a \in ]1, (5 + \sqrt{17})/4[$  (cf. Bosq *et al.* (1999)). Dans le cas d'un processus géométriquement mélangeant, *i.e.* si  $\alpha_X(u) \leq cu^{-\beta}$ ,  $c > 0$ , la série  $\sum_{k \geq 1} k \alpha_X^{(a-1)/a}(k)$  est alors convergente pour les valeurs  $\beta > 2a/(a-1)$ .

Finalement, pour déduire des régions de confiance du corollaire 1.8, nous devons encore chercher à estimer la matrice  $\Sigma^{(m)}$  car la valeur de ses éléments est, la plupart du temps, inconnue. Ce problème qui n'est pas abordé ici ouvre une perspective de recherche intéressante pour prolonger ce travail. Pour la construction d'intervalles de confiance (si  $m = 1$ ), des travaux récents traitent la question de l'estimation de la variance asymptotique  $\text{VA}(x) := 2 \int_0^\infty g_u(x, x) dx$  et apportent des réponses adaptées. Guillou et Merlevède (2001) proposent ainsi deux estimateurs construits de façons différentes à partir de l'estimateur à noyau. Une autre approche qui utilise l'estimateur basé sur le temps local est aussi examinée par Blanke et Merlevède (2000). Enfin, un autre estimateur combinant la fonction de répartition empirique, l'estimateur à noyau et l'estimateur du temps local est également proposé par Dehay et Kutoyants (2004). Dans la plupart des cas (par exemple, pour le processus d'Ornstein-Uhlenbeck), l'estimateur du temps local atteint de meilleures vitesses de convergence pour l'erreur en moyenne quadratique. Cependant, l'usage des estimateurs à noyau est plus simple en pratique. En choisissant l'une de ces méthodes, nous obtenons alors un intervalle de confiance de niveau asymptotique  $1 - \varepsilon$ ,  $\varepsilon \in (0, 1)$ , pour la valeur estimée de  $f(x)$  :

$$\left[ \hat{f}_T^{PF}(x) - \frac{z_\varepsilon \widehat{\text{VA}}_T(x)^{1/2}}{\sqrt{T}}, \hat{f}_T^{PF}(x) + \frac{z_\varepsilon \widehat{\text{VA}}_T(x)^{1/2}}{\sqrt{T}} \right],$$

où  $\widehat{\text{VA}}_T(x)$  désigne une estimation de la variance asymptotique au point  $x$  et  $z_\varepsilon$  le quantile  $(1 - \varepsilon)$  de la loi Gaussienne centrée réduite.

## 1.4 Convergence presque sûre

Dans cette section, nous étudions respectivement la convergence ponctuelle presque sûre et la convergence uniforme presque sûre de l'histogramme sur un ensemble compact  $D$  puis sur l'espace  $\mathbb{R}^d$  tout entier. Notons  $\nu_x$  un voisinage ouvert du point  $x$  dans  $\mathbb{R}^d$  et  $D$  un sous-ensemble compact de  $\mathbb{R}^d$ . Nos principales hypothèses seront alors progressivement renforcées en prenant respectivement  $\Lambda = \nu_x$ ,  $\Lambda = D$  et  $\Lambda = \mathbb{R}^d$ .

Les hypothèses de régularité sur la vraie densité  $f$  sont les suivantes :

### Hypothèses $\mathbf{R}^*$

- (i)  $\exists K_f > 0$ ,  $|f(y) - f(z)| \leq K_f \|y - z\|$  pour  $(y, z) \in \Lambda^2$  (où  $K_f =: K_f^x$  si  $\Lambda = \nu_x$ ) ;
- (ii)  $\sup_{x \in \Lambda} f(x) \leq M_1 < \infty$  (où  $M_1 =: M_1^x$  si  $\Lambda = \nu_x$ ).

Les principales hypothèses portant sur le processus sont les suivantes :

### Hypothèse $\mathbf{H}_0^*$

$X$  est géométriquement fortement mélangeant de coefficient  $\alpha_X$  :  $\alpha_X(u) \leq \alpha_0 \varrho^u$  pour  $u > 0$ ,  $\alpha_0 > 0$  et  $0 < \varrho < 1$ .

**Hypothèses  $H_1^*$**  Il existe un borélien de  $\mathbb{R}^2$  :  $\Gamma = \{(s, t) \in \mathbb{R}^2, |t - s| \leq u_0, u_0 > 0\}$  tel que

- (i)  $g_{s,t}$  existe pour  $(s, t) \notin \Gamma$ ;
- (ii)  $\sup_{(y,z) \in \Lambda^2} f_{u|0}(z|y) \leq M_2 < \infty$  pour  $u \geq u_0$  (où  $M_2 =: M_2^x$  si  $\Lambda = \nu_x$ ).

**Hypothèses  $H_2^*$  ( $d = 1$ )**

- (i)  $g_{s,t}$  existe pour  $s \neq t$ ;
- (ii)  $\forall y \in \Lambda, \sup_{z \in \Lambda} \int_0^\infty |g_u(y, z)| du \leq \varphi(y)$  où  $\varphi$  est une fonction définie sur  $\Lambda$ , positive, continue et bornée.

Notons que les hypothèses  $H_1^*$  et  $H_2^*$  ont déjà été introduites pour étudier la convergence  $L^2$  de nos estimateurs. Elles seront utilisées séparément avec les hypothèses de régularité  $R^*$  pour obtenir les vitesses optimale et suroptimale. En particulier, nous supposons seulement que  $f$  est lipschitzienne et bornée. Pour l'étude spécifique du cas ponctuel, rappelons encore que les constantes, notées  $K_f^x, M_1^x$  et  $M_2^x$ , dépendent du point  $x$ .

### 1.4.1 Convergence ponctuelle presque sûre

On prend  $\Lambda = \nu_x$  et on suppose que le processus est observé sur les intervalles de temps  $[0, T_n]$ , où les suites  $(T_n)$  vérifient  $T_{n+1} - T_n \geq \tau > 0$  et  $T_n \uparrow \infty$ . Autrement dit, les trajectoires sont délivrées à des instants croissants  $T_1 < T_2 < \dots < T_n$ .

**Théorème 1.6.** (i) Cas "optimal" : sous les conditions  $R^*, H_0^*, H_1^*$  et si  $(h_{T_n})$  est telle que

$$\frac{T_n h_{T_n}^d}{(\ln T_n)^3} \longrightarrow \infty, \quad T_n \uparrow \infty,$$

alors

$$\limsup_{T_n \uparrow \infty} \left( \frac{T_n h_{T_n}^d}{\ln T_n} \right)^{\frac{1}{2}} \cdot \left| \hat{f}_{T_n}^H(x) - \mathbb{E} \hat{f}_{T_n}^H(x) \right| \leq 2\sqrt{2u_0 f(x)} \quad p.s., \quad x \in \mathbb{R}^d;$$

(ii) Cas "suroptimal" ( $d = 1$ ) : sous les conditions  $H_0^*, H_2^*$  et si  $(h_{T_n})$  est telle que

$$\frac{T_n h_{T_n}^2}{(\ln T_n)^3} \longrightarrow \infty, \quad T_n \uparrow \infty,$$

alors

$$\limsup_{T_n \uparrow \infty} \left( \frac{T_n}{\ln T_n} \right)^{\frac{1}{2}} \cdot \left| \hat{f}_{T_n}^H(x) - \mathbb{E} \hat{f}_{T_n}^H(x) \right| \leq 2\sqrt{2\varphi(x)} \quad p.s., \quad x \in \mathbb{R}.$$

**Remarque 1.14.** Si nous renforçons l'hypothèse  $H_2^*(ii)$  par l'hypothèse :

$$H_2^*(ii)' : \int_0^\infty \|g_u\|_\infty du < \infty \text{ et } g_u(\cdot, \cdot) \text{ est continue en } (x, x) \text{ pour tout } u > 0,$$

alors la constante asymptotique devient  $2\sqrt{2\int_0^\infty g_u(x, x) du}$  et nous retrouvons la limite exacte de la variance ponctuelle de l'histogramme sous la racine carrée (voir la preuve du théorème).

En utilisant les résultats du théorème 1.6 et du lemme 1.2 avec des choix *ad hoc* du paramètre de lissage  $h_T$ , nous établissons des vitesses optimale et suroptimale de la convergence ponctuelle presque sûre de l'histogramme.

**Corollaire 1.9.** (i) Cas “optimal” : si les conditions  $R^*$ ,  $H_0^*$  et  $H_1^*$  sont vérifiées alors le choix  $h_{T_n} = c_{T_n} \cdot (T_n^{-1} \ln T_n)^{1/(d+2)}$  tel que  $c_{T_n} \rightarrow c$ ,  $0 < c < \infty$ , entraîne

$$\left| \hat{f}_{T_n}^H(x) - f(x) \right| = \mathcal{O} \left( \left( \frac{T_n}{\ln T_n} \right)^{-\frac{1}{d+2}} \right) \quad p.s., \quad x \in \mathbb{R}^d;$$

(ii) Cas “suroptimal” ( $d = 1$ ) : si les conditions  $R^*$ ,  $H_0^*$  et  $H_2^*$  sont vérifiées alors le choix  $h_{T_n} = c'_{T_n} \cdot (T_n^{-1} (\ln T_n)^3)^{1/2}$  tel que  $c'_{T_n} \rightarrow c'$ ,  $0 < c' < \infty$ , entraîne

$$\left| \hat{f}_{T_n}^H(x) - f(x) \right| = \mathcal{O} \left( \left( \frac{T_n}{(\ln T_n)^3} \right)^{-\frac{1}{2}} \right) \quad p.s., \quad x \in \mathbb{R}.$$

#### 1.4.2 Convergence uniforme presque sûre

Concernant l'ordre de convergence du terme de biais, le lemme suivant est analogue au lemme 1.2 pour  $\Lambda = D$  ou  $\Lambda = \mathbb{R}^d$ .

**Lemme 1.6.** Si la condition  $R^*(i)$  est satisfaite alors on a

$$\limsup_{T \rightarrow \infty} h_T^{-1} \cdot \left| \mathbb{E} \hat{f}_T^H(x) - f(x) \right| \leq C_f < \infty, \quad x \in \mathbb{R}^d,$$

où  $C_f$  est une constante positive dépendante de  $f$ .

#### • Convergence uniforme sur un compact ( $\Lambda = D$ )

**Théorème 1.7.** (i) Cas “optimal” : sous les conditions  $R^*$ ,  $H_0^*$ ,  $H_1^*$  et si  $(h_{T_n})$  est telle que

$$\frac{T_n h_{T_n}^d}{(\ln T_n)^3} \longrightarrow \infty, \quad T_n \uparrow \infty,$$

alors

$$\limsup_{T_n \uparrow \infty} \left( \frac{T_n h_{T_n}^d}{\ln T_n} \right)^{\frac{1}{2}} \cdot \sup_{x \in D} \left| \hat{f}_{T_n}^H(x) - \mathbb{E} \hat{f}_{T_n}^H(x) \right| \leq 2\sqrt{4u_0 M_1} \quad p.s.;$$

(ii) Cas “suroptimal” ( $d = 1$ ) : sous les conditions  $H_0^*$ ,  $H_2^*$  et si  $(h_{T_n})$  est telle que

$$\frac{T_n h_{T_n}^2}{(\ln T_n)^3} \longrightarrow \infty, \quad T_n \uparrow \infty,$$

alors

$$\limsup_{T_n \uparrow \infty} \left( \frac{T_n}{\ln T_n} \right)^{\frac{1}{2}} \cdot \sup_{x \in D} \left| \hat{f}_{T_n}^H(x) - \mathbb{E} \hat{f}_{T_n}^H(x) \right| \leq 2\sqrt{3\|\varphi\|_\infty} \quad p.s..$$

En utilisant les résultats du théorème 1.7 et du lemme 1.6 avec les mêmes choix de  $h_T$  que dans le cas de la convergence ponctuelle presque sûre, nous retrouvons les mêmes vitesses pour la convergence uniforme presque sûre de l'histogramme sur un compact.

**Corollaire 1.10.** (i) Cas “optimal” : si les conditions  $R^*$ ,  $H_0^*$  et  $H_1^*$  sont vérifiées alors le choix  $h_{T_n} = c_{T_n} \cdot (T_n^{-1} \ln T_n)^{1/(d+2)}$  tel que  $c_{T_n} \rightarrow c$ ,  $0 < c < \infty$ , entraîne

$$\sup_{x \in D} \left| \hat{f}_{T_n}^H(x) - f(x) \right| = \mathcal{O} \left( \left( \frac{T_n}{\ln T_n} \right)^{-\frac{1}{d+2}} \right) \text{ p.s.};$$

(ii) Cas “suroptimal” ( $d = 1$ ) : si les conditions  $R^*$ ,  $H_0^*$  et  $H_2^*$  sont vérifiées alors le choix  $h_{T_n} = c'_{T_n} \cdot (T_n^{-1} (\ln T_n)^3)^{1/2}$  tel que  $c'_{T_n} \rightarrow c'$ ,  $0 < c' < \infty$ , entraîne

$$\sup_{x \in D} \left| \hat{f}_{T_n}^H(x) - f(x) \right| = \mathcal{O} \left( \left( \frac{T_n}{(\ln T_n)^3} \right)^{-\frac{1}{2}} \right) \text{ p.s..}$$

• **Convergence uniforme sur  $\mathbb{R}^d$  ( $\Lambda = \mathbb{R}^d$ )**

Nous établissons dans un premier temps la convergence uniforme de l’histogramme sur une suite croissante d’hyperboules  $D_T := \{x : \|x\| \leq T_n^a\}$ , sous-ensembles compacts de  $\mathbb{R}^d$  centrés à l’origine.

**Théorème 1.8.** (i) Cas “optimal” : sous les conditions  $R^*$ ,  $H_0^*$ ,  $H_1^*$  et si  $(h_{T_n})$  est telle que

$$\frac{T_n h_{T_n}^d}{(\ln T_n)^3} \longrightarrow \infty, \quad T_n \uparrow \infty,$$

alors

$$\limsup_{T_n \uparrow \infty} \left( \frac{T_n h_{T_n}^d}{\ln T_n} \right)^{\frac{1}{2}} \cdot \sup_{\|x\| \leq T_n^a} \left| \hat{f}_{T_n}^H(x) - \mathbb{E} \hat{f}_{T_n}^H(x) \right| \leq 2\sqrt{(ad+2)2u_0 M_1} \text{ p.s.};$$

(ii) Cas “suroptimal” ( $d = 1$ ) : sous les conditions  $H_0^*$ ,  $H_2^*$  et si  $(h_{T_n})$  est telle que

$$\frac{T_n h_{T_n}^2}{(\ln T_n)^3} \longrightarrow \infty, \quad T_n \uparrow \infty,$$

alors

$$\limsup_{T_n \uparrow \infty} \left( \frac{T_n}{\ln T_n} \right)^{\frac{1}{2}} \cdot \sup_{\|x\| \leq T_n^a} \left| \hat{f}_{T_n}^H(x) - \mathbb{E} \hat{f}_{T_n}^H(x) \right| \leq 2\sqrt{(2a+3)\|\varphi\|_\infty} \text{ p.s..}$$

Pour établir la convergence uniforme presque sûre de l’histogramme sur l’espace  $\mathbb{R}^d$  tout entier, nous introduisons une hypothèse supplémentaire sur le processus.

**Hypothèse  $H_3^*$**

$X$  est un processus strictement stationnaire tel que  $\sup_{0 \leq t \leq T} \|X_t\|$  est mesurable pour tout  $T > 0$  et  $\mathbb{E}(\sup_{0 \leq t \leq 1} \|X_t\|) < \infty$ . De plus,  $\lim_{\|u\| \rightarrow \infty} \|u\|f(u) = 0$ .

En utilisant les résultats du théorème 1.8 et du lemme 1.6 avec  $H_3^*$ , nous retrouvons les mêmes vitesses pour la convergence uniforme presque sûre de l’histogramme sur  $\mathbb{R}^d$ .

**Corollaire 1.11.** (i) Cas “optimal” : si les conditions  $R^*$ ,  $H_0^*$  et  $H_1^*$  sont vérifiées alors le choix  $h_{T_n} = c_{T_n} \cdot (T_n^{-1} \ln T_n)^{1/(d+2)}$  tel que  $c_{T_n} \rightarrow c$ ,  $0 < c < \infty$ , entraîne

$$\sup_{\|x\| \leq T_n^a} \left| \hat{f}_{T_n}^H(x) - f(x) \right| = \mathcal{O} \left( \left( \frac{T_n}{\ln T_n} \right)^{-\frac{1}{d+2}} \right) \text{ p.s.};$$

(ii) Cas “suroptimal” ( $d = 1$ ) : si les conditions  $R^*$ ,  $H_0^*$  et  $H_2^*$  sont vérifiées alors le choix  $h_{T_n} = c'_{T_n} \cdot (T_n^{-1} (\ln T_n)^3)^{1/2}$  tel que  $c'_{T_n} \rightarrow c'$ ,  $0 < c' < \infty$ , entraîne

$$\sup_{\|x\| \leq T_n^a} \left| \hat{f}_{T_n}^H(x) - f(x) \right| = \mathcal{O} \left( \left( \frac{T_n}{(\ln T_n)^3} \right)^{-\frac{1}{2}} \right) \text{ p.s.}$$

**Corollaire 1.12.** Avec l’hypothèse supplémentaire  $H_3^*$ , les résultats du corollaire 1.11 restent valables avec  $\sup_{x \in \mathbb{R}^d} |\hat{f}_{T_n}^H(x) - f(x)|$  au lieu de  $\sup_{\|x\| \leq T_n^a} |\hat{f}_{T_n}^H(x) - f(x)|$ .

### 1.4.3 Le polygone de fréquences ( $d = 1$ )

En introduisant le polygone de fréquences, nous retrouvons sous les mêmes hypothèses nos résultats précédents. Pour une classe plus restreinte de densités qui vérifient

$$R^*(i)' : f \in \mathcal{C}^2(\Lambda) \text{ et } f'' \text{ est bornée sur } \Lambda,$$

nous améliorons l’ordre du biais et la vitesse optimale de convergence.

**Lemme 1.7.** (i) Si la condition  $R^*(i)$  est satisfaite alors on a

$$\limsup_{T \rightarrow \infty} h_T^{-1} \cdot \sup_{x \in \Lambda} \left| \mathbb{E} \hat{f}_T^{PF}(x) - f(x) \right| < \infty, \quad \Lambda = \nu_x, D \text{ et } \mathbb{R};$$

(ii) Si la condition  $R^*(i)'$  est satisfaite alors on a

$$\limsup_{T \rightarrow \infty} h_T^{-2} \cdot \sup_{x \in \Lambda} \left| \mathbb{E} \hat{f}_T^{PF}(x) - f(x) \right| \leq \frac{5}{16} \sup_{x \in \Lambda} |f''(x)| < \infty, \quad \Lambda = \nu_x, D \text{ et } \mathbb{R}.$$

À partir des hypothèses  $R^*(i)(ii)$ ,  $H_0^*$ ,  $H_1^*$  ou  $H_2^*$  et  $H_3^*$ , toute l’étude du polygone de fréquences avec les mêmes vitesses de convergence découle de nos résultats sur l’histogramme.

**Corollaire 1.13.** Sous les mêmes conditions d’application, les résultats des théorèmes 1.6, 1.7 et 1.8 et des corollaires 1.9, 1.10, 1.11 et 1.12 obtenus pour l’histogramme restent valables pour le polygone de fréquences.

À partir des hypothèses  $R^*(i)'(ii)$ ,  $H_0^*$ ,  $H_1^*$  ou  $H_2^*$  et  $H_3^*$ , nous améliorons les vitesses optimale et suroptimale pour de nouveaux choix du paramètre  $h_{T_n}$ .

**Corollaire 1.14.** Sous l’hypothèse renforcée  $R^*(i)'$ , nous obtenons une nouvelle vitesse optimale de l’ordre  $(T_n^{-1} \ln T_n)^{2/5}$  pour le choix  $h_{T_n} = c_{T_n} (T_n^{-1} \ln T_n)^{1/5}$  tel que  $c_{T_n} \rightarrow c$ ,  $0 < c < \infty$ ; ainsi qu’une nouvelle vitesse suroptimale de l’ordre  $(T_n^{-1} \ln T_n)^{1/2}$  pour le choix  $h_{T_n} = o(T_n^{-1} \ln T_n)^{1/4}$ .

**Remarque 1.15.** Les vitesses optimales obtenues dans les corollaires 1.13 et 1.14 sont comparables à celles obtenues par Carbon *et al.* (1997) en temps discret avec des hypothèses similaires.

## 1.5 Discussion

Nous avons présenté des résultats théoriques qui valident le bon potentiel d'estimateurs élémentaires fréquemment choisis parmi les praticiens. Le polygone de fréquences est particulièrement intéressant avec des vitesses de convergence comparables à celles des estimateurs à noyau et la possibilité d'obtenir des régions de fréquences si la matrice des variances et covariances asymptotiques est convenablement estimée. En ce sens, un estimateur de type polygone pour cette matrice pourra être envisagé plus tard. Par ailleurs, rappelons que le contexte de cette étude, où la trajectoire est parfaitement observée sur  $[0, T]$ , est théorique mais il donne déjà une bonne indication sur les vitesses idéales. Nous regarderons dans le chapitre 3 comment approcher ces vitesses dans des situations pratiques où, par exemple, le processus n'est observé qu'à des instants discrétisés suivant des schémas aléatoires ou déterministes ; ce cadre sera également favorable à la mise en œuvre de simulations pour justifier empiriquement du bon comportement de ces estimateurs. Enfin et toujours en temps continu, nous pourrons chercher à généraliser les propriétés de ces estimateurs pour des classes plus vastes de fonctions positives (par exemple, celles considérées dans Baraud et Birgé (2006)). L'estimation de la régression est traitée dans le chapitre suivant.

## 1.6 Preuves

### 1.6.1 Convergence en moyenne quadratique

#### Preuve du lemme 1.2

Notons  $p_k := P(X_0 \in \pi_{T_k})$ ,  $k \in \mathbb{Z}^d$ . Pour  $x \in \pi_{T_j}$ , l'espérance de  $\hat{f}_T^H(x)$  s'écrit

$$\mathbb{E}(\hat{f}_j) = \frac{1}{Th_T^d} \mathbb{E} \int_0^T \mathbf{1}_{\pi_{T_j}}(X_t) dt = \frac{1}{h_T^d} p_j = \frac{1}{h_T^d} \int_{\pi_{T_j}} f(u) du.$$

D'où

$$\mathbb{E}(\hat{f}_j) - f(x) = \frac{1}{h_T^d} \int_{\pi_{T_j}} (f(u) - f(x)) du.$$

De là, pour  $T$  assez grand de telle sorte que  $\pi_{T_j} \subset \nu_x$ , on peut appliquer l'hypothèse lipschitzienne de continuité  $R(i)$  :

$$\left| \mathbb{E}(\hat{f}_j) - f(x) \right| \leq \frac{1}{h_T^d} \int_{\pi_{T_j}} |f(u) - f(x)| du \leq \frac{1}{h_T^d} \int_{\pi_{T_j}} K_f^x \underbrace{\|u - x\|}_{\leq Ch_T} du \leq C_f(x) h_T,$$

où  $C > 0$  et où  $C_f(x)$  désigne une constante positive qui dépend à la fois de  $x$  et de  $f$ , ce qui prouve que

$$\limsup_{T \rightarrow \infty} h_T^{-1} \cdot \left| \mathbb{E} \hat{f}_T^H(x) - f(x) \right| \leq C_f(x) < \infty.$$

■

### Preuve du théorème 1.1

Pour  $x \in \pi_{T_j}$ , on a

$$Th_T^d \cdot \text{Var}(\hat{f}_j) = \frac{1}{Th_T^d} \iint_{[0,T]^2} \text{Cov}\left(\mathbf{1}_{\pi_{T_j}}(X_s), \mathbf{1}_{\pi_{T_j}}(X_t)\right) dsdt = \frac{1}{Th_T^d} (I_1 + I_2),$$

où

$$\begin{aligned} I_1 &:= \iint_{[0,T]^2 \cap \Gamma} \text{Cov}\left(\mathbf{1}_{\pi_{T_j}}(X_s), \mathbf{1}_{\pi_{T_j}}(X_t)\right) dsdt; \\ I_2 &:= \iint_{[0,T]^2 \cap \Gamma^c} \text{Cov}\left(\mathbf{1}_{\pi_{T_j}}(X_s), \mathbf{1}_{\pi_{T_j}}(X_t)\right) dsdt. \end{aligned}$$

Majoration de la première intégrale avec l'inégalité de Cauchy-Schwarz :

$$\begin{aligned} I_1 &\leq \iint_{[0,T]^2 \cap \Gamma} \sqrt{\text{Var}(\mathbf{1}_{\pi_{T_j}}(X_s)) \text{Var}(\mathbf{1}_{\pi_{T_j}}(X_t))} dsdt \\ &\leq \iint_{[0,T]^2 \cap \Gamma} \text{Var}(\mathbf{1}_{\pi_{T_j}}(X_0)) dsdt \leq p_j(1-p_j) \iint_{[0,T]^2 \cap \Gamma} dsdt. \end{aligned}$$

Remarquons maintenant qu'il existe un réel positif  $T_0$  tel que  $\pi_{T_j} \subset \nu_x$  pour tout  $T \geq T_0$ . Compte tenu de l'hypothèse de continuité  $R(ii)$ , nous pouvons utiliser le théorème de la moyenne : il existe un  $\xi_j \in \pi_{T_j}$  tel que  $\int_{\pi_{T_j}} f(x) dx = h_T^d f(\xi_j)$ . D'où

$$I_1 \leq h_T^d f(\xi_j) (1 - h_T^d f(\xi_j)) (2u_0 T - u_0^2).$$

Majoration de la seconde intégrale :

$$\begin{aligned} I_2 &= \iint_{[0,T]^2 \cap \Gamma^c} \left( P(X_s \in \pi_{T_j}, X_t \in \pi_{T_j}) - p_j^2 \right) dsdt \\ &\leq 2T \int_{u_0}^{\infty} \left| P(X_0 \in \pi_{T_j}, X_u \in \pi_{T_j}) - p_j^2 \right| du = 2T(I_{21} + I_{22}), \end{aligned}$$

grâce à l'hypothèse de stationnarité et où on a posé

$$\begin{aligned} I_{21} &:= \int_{u_0}^{h_T^{-d+\varepsilon}} \left| \iint_{\pi_{T_j}^2} (f_u(y, z) - f(y)f(z)) dydz \right| du; \\ I_{22} &:= \int_{h_T^{-d+\varepsilon}}^{\infty} \left| \text{Cov}(\mathbf{1}_{\pi_{T_j}}(X_0), \mathbf{1}_{\pi_{T_j}}(X_u)) \right| du, \end{aligned}$$

avec  $\varepsilon$  réel positif dont on précisera les valeurs ensuite. Comme  $f$  est majorée par  $M_1^x$ , on déduit des hypothèses  $H_1(i)(ii)$  :

$$\begin{aligned} I_{21} &\leq \int_{u_0}^{h_T^{-d+\varepsilon}} \iint_{\pi_{T_j}^2} f(y) |f_{u|0}(z|y) - f(z)| dydz du \\ &\leq h_T^d \cdot \max(M_1^x, M_2^x) \int_{u_0}^{h_T^{-d+\varepsilon}} \int_{\pi_{T_j}} f(y) dy du \\ &\leq (h_T^{-d+\varepsilon} - u_0) h_T^{2d} \cdot \max(M_1^x, M_2^x) f(\xi_j). \end{aligned}$$

D'où, en posant  $M^x = \max(M_1^x, M_2^x)$  :

$$I_{21} \leq h_T^{d+\varepsilon} \cdot M^x f(\xi_j).$$

En utilisant l'inégalité de Davydov (cf. lemme 3.3) pour un couple  $(p, q) \in ]1, \varrho - 1] \times \left[1 + \frac{\varrho}{\varrho-2}, \infty\right]$  tel que  $\frac{2}{q} + \frac{1}{p} = 1$ , on a

$$I_{22} \leq \int_{h_T^{-d+\varepsilon}}^{\infty} 2p \cdot 2^{1/p} \cdot \|\mathbf{1}_{\pi_{T_j}}(X_0)\|_q \|\mathbf{1}_{\pi_{T_j}}(X_u)\|_q \cdot \alpha_X^{1/p}(u) du,$$

soit avec l'hypothèse de mélange arithmétique  $H_1(iii)$  :

$$\begin{aligned} I_{22} &\leq 2p \cdot 2^{1/p} \cdot p_j^{2/q} \cdot \int_{h_T^{-d+\varepsilon}}^{\infty} a_0^{1/p} u^{-\varrho/p} du \\ &\leq h_T^{\frac{1}{p}\{(d-\varepsilon)(\varrho-p)+(p-1)d\}} \cdot \frac{2p^2(2a_0)^{1/p}}{\varrho-p} f(\xi_j)^{1-\frac{1}{p}}. \end{aligned}$$

D'où

$$I_2 \leq Th_T^d \cdot \left( h_T^\varepsilon \cdot 2M^x f(\xi_j) + h_T^{\frac{1}{p}\{(d-\varepsilon)(\varrho-p)-d\}} \cdot \frac{4p^2(2a_0)^{1/p}}{\varrho-p} f(\xi_j)^{1-\frac{1}{p}} \right).$$

Ainsi pour tout  $\varepsilon \in \left[0, d \left(1 - \frac{1}{\varrho-p}\right)\right]$  :

$$\begin{aligned} Th_T^d \cdot \text{Var}(\hat{f}_j) &\leq f(\xi_j) \left(1 - h_T^d f(\xi_j)\right) \cdot 2u_0 \left(1 - \frac{u_0}{2T}\right) + 2M^x f(\xi_j) \cdot h_T^\varepsilon \\ &\quad + \frac{4p^2(2a_0)^{1/p}}{\varrho-p} f(\xi_j)^{1-\frac{1}{p}} \cdot h_T^{\frac{1}{p}\{(d-\varepsilon)(\varrho-p)-d\}} < \infty. \end{aligned}$$

■

### Preuve du théorème 1.2

Partie (i) : avec l'hypothèse de stationnarité et  $H_2(i)$ , on écrit

$$\begin{aligned} T \cdot \text{Var}(\hat{f}_j) &= \frac{1}{Th_T^2} \iint_{[0, T]^2} \text{Cov}\left(\mathbf{1}_{\pi_{T_j}}(X_{\tau_1}), \mathbf{1}_{\pi_{T_j}}(X_{\tau_2})\right) d\tau_1 d\tau_2 \\ &= \frac{2}{Th_T^2} \int_0^T \int_u^T \text{Cov}\left(\mathbf{1}_{\pi_{T_j}}(X_0), \mathbf{1}_{\pi_{T_j}}(X_u)\right) dv du \\ &= \frac{2}{h_T^2} \iint_{\pi_{T_j}^2} \int_0^T \left(1 - \frac{u}{T}\right) g_u(y, z) du dy dz. \end{aligned}$$

On cherche maintenant un équivalent asymptotique de cette expression pour  $x \in \pi_{T_j} = [b_j, b_{j+1}[$  tel que  $b_j \uparrow_{T \rightarrow \infty} x$  et  $b_{j+1} = b_j + h_T \downarrow_{T \rightarrow \infty} x$ .

$$T \cdot \text{Var}(\hat{f}_j) = 2 \iint_{[0, 1]^2} \int_0^\infty \mathbf{1}_{[0, T]}(u) \left(1 - \frac{u}{T}\right) g_u(b_j + h_T s, b_j + h_T t) du ds dt.$$

De là, avec les conditions  $H_2(ii)$  de Castellana-Leadbetter, on a pour tout  $u > 0$  :

$$\mathbf{1}_{[0, T]}(u) \left(1 - \frac{u}{T}\right) g_u(b_j + h_T s, b_j + h_T t) \xrightarrow{T \rightarrow \infty} g_u(x, x);$$

$$\left| \mathbb{1}_{[0,T]}(u) \left(1 - \frac{u}{T}\right) g_u(b_j + h_T s, b_j + h_T t) \right| \leq \|g_u\|_\infty \in L^1([0, \infty[ \times [0, 1]^2),$$

et on déduit du théorème de convergence dominée de Lebesgue que

$$\lim_{T \rightarrow \infty} T \cdot \text{Var}(\hat{f}_j) = 2 \iint_{[0,1]^2} \int_0^\infty g_u(x, x) du \, dydz = 2 \int_0^\infty g_u(x, x) du.$$

Partie (ii) : en substituant l'hypothèse  $H_2(ii)$  par celle plus faible  $H_2(ii)'_a$ , nous avons la majoration suivante :

$$\begin{aligned} T \cdot \text{Var}(\hat{f}_j) &= \frac{2}{h_T^2} \iint_{\pi_{Tj}^2} \int_0^T \left(1 - \frac{u}{T}\right) g_u(y, z) du \, dydz \\ &\leq \frac{2}{h_T^2} \iint_{\pi_{Tj}^2} \underbrace{\sup_{z \in \nu_x} \int_0^\infty |g_u(y, z)| du}_{\leq \varphi(y)} dydz \\ &\leq \frac{2}{h_T^2} h_T \cdot h_T \varphi(\eta_j) \quad \text{où } \eta_j \in \pi_{Tj}. \end{aligned}$$

Finalement,

$$T \cdot \text{Var}(\hat{f}_j) \leq 2\varphi(\eta_j). \quad \blacksquare$$

### Preuve du corollaire 1.1

On obtient la partie (i) du corollaire en combinant les résultats du lemme 1.2 et du théorème 1.1 :

$$\min \left\{ h_T^{-2}, Th_T^d \right\} \cdot \text{EMQ}(\hat{f}_T^H(x)) \leq h_T^{-2} \cdot \left( \mathbb{E} \hat{f}_T^H(x) - f(x) \right)^2 + Th_T^d \cdot \text{Var}(\hat{f}_T^H(x)),$$

et en choisissant  $h_T \sim T^{-1/(d+2)}$  de telle sorte à équilibrer les termes de biais et de variance. De même, on déduit la partie (ii) du corollaire en combinant les résultats du lemme 1.2 et du théorème 1.2 avec le choix  $h_T = o(T^{-1/2})$ .  $\blacksquare$

### Preuve du lemme 1.3

Partant de la définition de l'estimateur en temps continu pour  $x \in \pi_{Tj}$ , on écrit

$$\mathbb{E}(\hat{f}_j) = \frac{1}{Th_T^d} \mathbb{E} \int_0^T \mathbb{1}_{\pi_{Tj}}(X_t) dt = \frac{1}{h_T^d} P(X_0 \in \pi_{Tj}) = \frac{1}{h_T^d} \int_{\pi_{Tj}} f(x) dx.$$

La preuve est alors identique à celle donnée par Lecoutre (1985) dans le cas i.i.d. par un développement de Taylor de  $f$  à l'ordre 2 au voisinage du centre  $c_j$  de l'hypercube  $\pi_{Tj}$ .  $\blacksquare$

**Preuve du corollaire 1.2**

On écrit  $\text{VI}(\hat{f}_T^H) = \sum_j \int_{\pi_{Tj}} \text{Var}(\hat{f}_j) dx$ . Sous les hypothèses du théorème 1.1 pour tout  $0 \leq \varepsilon \leq d \left(1 - \frac{1}{\varrho - p}\right)$ , on a

$$Th_T^d \cdot \int_{\pi_{Tj}} \text{Var}(\hat{f}_j) dx \leq h_T^d \left\{ f(\xi_j) (1 - h_T^d f(\xi_j)) \cdot 2u_0 \left(1 - \frac{u_0}{2T}\right) + 2M f(\xi_j) \cdot h_T^\varepsilon \right. \\ \left. + \frac{4p^2(2a_0)^{1/p}}{\varrho - p} f(\xi_j)^{1 - \frac{1}{p}} \cdot h_T^{\frac{1}{p}\{(d-\varepsilon)(\varrho-p)-d\}} \right\},$$

avec  $M = \max(M_1, M_2)$  ne dépendant plus du point  $x$ . Il reste maintenant à sommer sur l'ensemble des indices  $j$  des éléments de la partition  $\Pi_T$ . Comme les fonctions  $f^{1-\frac{1}{p}}$ ,  $f$  et  $f^2$  sont supposées Riemann-intégrables sur  $\mathbb{R}^d$ , on peut écrire

$$\sum_j h_T^d f^\kappa(\xi_j) = \int_{\mathbb{R}^d} f^\kappa(x) dx + o(1) \quad \text{pour } \kappa = 1 - \frac{1}{p}, 1, 2.$$

D'où,

$$Th_T^d \cdot \text{VI}(\hat{f}_T^H) \leq \left\{ 2u_0 + 2M \cdot h_T^\varepsilon + \frac{4p^2(2a_0)^{1/p}}{\varrho - p} \|f^{1-\frac{1}{p}}\|_1 \cdot h_T^{\frac{1}{p}\{(d-\varepsilon)(\varrho-p)-d\}} \right\} (1 + o(1)).$$

Avec les conditions  $H_2(i)(ii)'_{ab}$ , on a

$$T \cdot \text{Var}(\hat{f}_j) \leq 2\varphi(\eta_j), \quad \eta_j \in \pi_{Tj},$$

et on en déduit que

$$T \cdot \text{VI}(\hat{f}_T^H) \leq 2 \sum_j h_T^d \varphi(\eta_j) \leq 2\|\varphi\|_1 + o(1).$$

■

**Preuve du corollaire 1.3**

En utilisant le résultat du lemme 1.3 avec le choix  $h_T = c_T \cdot T^{-\frac{1}{d+2}}$  tel que  $c_T \rightarrow c$ ,  $0 < c < \infty$ , on a d'une part :

$$T^{\frac{2}{d+2}} \cdot \text{BCI}(\hat{f}_T^H) = \frac{c_T^2}{12} \mathcal{R}_d(f') + o(1) \quad \text{soit} \quad \lim_{T \rightarrow \infty} T^{\frac{2}{d+2}} \cdot \text{BCI}(\hat{f}_T^H) = \frac{c^2}{12} \mathcal{R}_d(f').$$

D'autre part, avec la partie (i) du corollaire 1.2, si  $p = \varrho - 1 (\Leftrightarrow \varepsilon = 0)$  :

$$T^{\frac{2}{d+2}} \cdot \text{VI}(\hat{f}_T^H) \leq \frac{1}{c_T^d} \left\{ 2u_0 + 2M + \frac{4p^2(2a_0)^{1/p}}{\varrho - p} \|f^{1-\frac{1}{p}}\|_1 \right\} (1 + o(1)),$$

soit

$$\limsup_{T \rightarrow \infty} T^{\frac{2}{d+2}} \cdot \text{VI}(\hat{f}_T^H) \leq \frac{1}{c^d} \left\{ 2u_0 + 2M + \frac{4p^2(2a_0)^{1/p}}{\varrho - p} \|f^{1-\frac{1}{p}}\|_1 \right\},$$

et si  $p < \varrho - 1$  ( $\Leftrightarrow \varepsilon > 0$ ) :

$$T^{\frac{2}{d+2}} \cdot \text{VI}(\hat{f}_T^H) \leq \frac{1}{c_T^d} \left\{ 2u_0 + 2M \cdot h_T^\varepsilon + \frac{4p^2(2a_0)^{1/p}}{\varrho - p} \|f^{1-\frac{1}{p}}\|_1 \cdot h_T^{\frac{1}{p}\{(d-\varepsilon)(\varrho-p)-d\}} \right\} (1 + o(1)),$$

soit

$$\limsup_{T \rightarrow \infty} T^{\frac{2}{d+2}} \cdot \text{VI}(\hat{f}_T^H) \leq \frac{2u_0}{c^d}.$$

Finalement, pour  $p = \varrho - 1$ , on obtient

$$\limsup_{T \rightarrow \infty} T^{\frac{2}{d+2}} \cdot \text{EMQI}(\hat{f}_T^H) \leq \frac{c^2}{12} \mathcal{R}_d(f') + \frac{1}{c^d} \left\{ 2u_0 + 2M + \frac{4p^2(2a_0)^{1/p}}{\varrho - p} \|f^{1-\frac{1}{p}}\|_1 \right\},$$

et pour  $p < \varrho - 1$  :

$$\limsup_{T \rightarrow \infty} T^{\frac{2}{d+2}} \cdot \text{EMQI}(\hat{f}_T^H) \leq \frac{c^2}{12} \mathcal{R}_d(f') + \frac{2u_0}{c^d}.$$

Si le paramètre  $h_T$  est un  $o\left(T^{-\frac{1}{2}}\right)$ , le terme de biais est négligeable devant le terme de variance et la vitesse suroptimale de l'histogramme se déduit immédiatement de la partie (ii) du corollaire 1.2 :

$$\limsup_{T \rightarrow \infty} T \cdot \text{EMQI}(\hat{f}_T^H) \leq 2\|\varphi\|_1.$$

■

#### Preuve du lemme 1.4

On suppose sans perte de généralité que  $x \in \pi_{Tj}^+ := \pi_{Tj} \cap [c_j, c_{j+1}[$ . Pour  $T$  assez grand de telle sorte que  $(\pi_{Tj} \cap \pi_{Tj+1}) \subset \nu_x$  et comme  $f$  est deux fois continûment différentiable sur  $\nu_x$ , on peut écrire un développement de Taylor à l'ordre 2 au voisinage de  $b_{j+1}$  ( $=: m_j$ ), milieu de l'intervalle  $[c_j, c_{j+1}[$  :

$$f(x) = f(m_j) + (x - m_j)f'(m_j) + \frac{1}{2}(x - m_j)^2 f''(m_j + \theta(x - m_j)), \quad \theta \in ]0, 1[.$$

On a alors

$$\begin{aligned} \mathbb{E}(\hat{f}_j) &= f(m_j) - \frac{f'(m_j)}{2} h_T + \frac{1}{h_T} \int_{\pi_{Tj}} \frac{1}{2} (u - m_j)^2 f''(m_j + \theta(u - m_j)) du, \\ \mathbb{E}(\hat{f}_{j+1}) &= f(m_j) + \frac{f'(m_j)}{2} h_T + \frac{1}{h_T} \int_{\pi_{Tj+1}} \frac{1}{2} (u - m_j)^2 f''(m_j + \theta(u - m_j)) du. \end{aligned}$$

D'où

$$\begin{aligned} \mathbb{E}\hat{f}_T^{PF}(x) - f(x) &= \left(\frac{x - c_j}{h_T}\right) \frac{1}{h_T} \int_{\pi_{Tj+1}} \frac{1}{2} (u - m_j)^2 f''(m_j + \theta(u - m_j)) du \\ &\quad + \left(\frac{c_{j+1} - x}{h_T}\right) \frac{1}{h_T} \int_{\pi_{Tj}} \frac{1}{2} (u - m_j)^2 f''(m_j + \theta(u - m_j)) du \\ &\quad - \frac{1}{2} (x - m_j)^2 f''(m_j + \theta(x - m_j)), \end{aligned}$$

et comme  $f''$  est bornée sur  $\nu_x$ , on a

$$\left| \mathbb{E} \hat{f}_T^{PF}(x) - f(x) \right| \leq \left[ \frac{1}{2} + 1 + 1 \right] \times \frac{1}{2} \left( \frac{h_T}{2} \right)^2 \sup_{u \in \nu_x} |f''(u)|,$$

soit

$$h_T^{-2} \cdot \left| \mathbb{E} \hat{f}_T^{PF}(x) - f(x) \right| \leq \frac{5}{16} \sup_{u \in \nu_x} |f''(u)|.$$

■

### Preuve du corollaire 1.4

Pour  $x \in \pi_{Tj}^+ := \pi_{Tj} \cap [c_j, c_{j+1}[$ , on a

$$\hat{f}_T^{PF}(x) = \left( \frac{x - c_j}{h_T} \right) \hat{f}_{j+1} + \left( \frac{c_{j+1} - x}{h_T} \right) \hat{f}_j,$$

et la preuve est basée sur la décomposition suivante :

$$\begin{aligned} \text{Var}(\hat{f}_T^{PF}(x)) &= \left( \frac{x - c_j}{h_T} \right)^2 \text{Var}(\hat{f}_{j+1}) + \left( \frac{c_{j+1} - x}{h_T} \right)^2 \text{Var}(\hat{f}_j) \\ &\quad + 2 \left( \frac{x - c_j}{h_T} \right) \left( \frac{c_{j+1} - x}{h_T} \right) \text{Cov}(\hat{f}_j, \hat{f}_{j+1}). \end{aligned} \quad (1.2)$$

Notons  $\overline{E}_j^x$  une expression (dépendante de  $x$ ) qui majore la variance de  $\hat{f}_j$ ,  $j \in \mathbb{Z}$ . En utilisant l'inégalité de Cauchy-Schwarz, on a alors

$$\left| \text{Cov}(\hat{f}_j, \hat{f}_{j+1}) \right| \leq \sqrt{\text{Var}(\hat{f}_j) \text{Var}(\hat{f}_{j+1})} \leq \frac{1}{2} \left\{ \text{Var}(\hat{f}_j) + \text{Var}(\hat{f}_{j+1}) \right\} \leq \frac{1}{2} \left\{ \overline{E}_j^x + \overline{E}_{j+1}^x \right\}.$$

Partie (i) : en utilisant la majoration établie dans le théorème 1.1 sous les conditions  $H_1$  d'“optimalité” pour  $T$  assez grand, on a

$$\begin{aligned} \overline{E}_j^x &= \frac{1}{Th_T} \left\{ f(\xi_j) (1 - h_T f(\xi_j)) 2u_0 \left( 1 - \frac{u_0}{2T} \right) + h_T^\varepsilon \cdot 2M^x f(\xi_j) \right. \\ &\quad \left. + h_T^{\frac{1}{p} \{(1-\varepsilon)(\varrho-p)-1\}} \cdot \frac{4p^2 (2a_0)^{1/p}}{\varrho - p} f(\xi_j)^{1-\frac{1}{p}} \right\}, \end{aligned}$$

que l'on peut écrire grâce à la continuité de  $f$  sur  $\nu_x$  :

$$Th_T \cdot \overline{E}_j^x = 2u_0 f(x) \times (1 + o(1)).$$

De même, comme  $\xi_{j+1} \in \pi_{Tj+1} = [b_j + h_T, b_{j+1} + h_T[$  avec  $b_j \uparrow_{T \rightarrow \infty} x$ ,  $b_{j+1} = b_j + h_T \downarrow_{T \rightarrow \infty} x$  et  $h_T \downarrow_{T \rightarrow \infty} 0^+$ , on peut écrire

$$f(\xi_{j+1}) = f(x) \times (1 + o(1)) \implies Th_T \cdot \overline{E}_{j+1}^x = 2u_0 f(x) \times (1 + o(1)).$$

Partant de la relation (1.2),

$$\text{Var}(\hat{f}_T^{PF}(x)) \leq \left( \frac{x - c_j}{h_T} \right)^2 \overline{E}_{j+1}^x + \left( \frac{c_{j+1} - x}{h_T} \right)^2 \overline{E}_j^x + 2 \left( \frac{x - c_j}{h_T} \right) \left( \frac{c_{j+1} - x}{h_T} \right) \frac{1}{2} \left\{ \overline{E}_j^x + \overline{E}_{j+1}^x \right\},$$

et on en déduit que

$$\begin{aligned} Th_T \cdot \text{Var}(\hat{f}_T^{PF}(x)) &\leq \left\{ \left( \frac{x - c_j}{h_T} \right)^2 + \left( \frac{c_{j+1} - x}{h_T} \right)^2 + 2 \left( \frac{x - c_j}{h_T} \right) \left( \frac{c_{j+1} - x}{h_T} \right) \right\} \times 2u_0 f(x) \times (1 + o(1)) \\ &\leq 2u_0 f(x) \times (1 + o(1)). \end{aligned}$$

Partie (ii) : en utilisant la majoration établie dans le théorème 1.2 sous les conditions  $H_2(ii)'_a$  de “suroptimalité”, on a

$$\bar{E}_j^x = \frac{1}{T} \{2\varphi(\eta_j)\} \quad \text{soit} \quad T \cdot \bar{E}_j = 2\varphi(x) \times (1 + o(1)),$$

et

$$T \cdot \bar{E}_{j+1}^x = 2\varphi(x) \times (1 + o(1)),$$

et on en déduit que

$$\begin{aligned} T \cdot \text{Var}(\hat{f}_T^{PF}(x)) &\leq \left\{ \left( \frac{x - c_j}{h_T} \right)^2 + \left( \frac{c_{j+1} - x}{h_T} \right)^2 + 2 \left( \frac{x - c_j}{h_T} \right) \left( \frac{c_{j+1} - x}{h_T} \right) \right\} \times 2\varphi(x) \times (1 + o(1)) \\ &\leq 2\varphi(x) \times (1 + o(1)). \end{aligned}$$

Finalement, avec les conditions  $H_2(ii)$  de Castellana-Leadbetter, on obtient le comportement asymptotique exact de la covariance de l’histogramme :

$$\begin{aligned} T \cdot \text{Cov}(\hat{f}_j, \hat{f}_{j+1}) &= \frac{1}{Th_T^2} \iint_{[0, T]^2} \text{Cov}(\mathbf{1}_{\pi_{Tj}}(X_{\tau_1}), \mathbf{1}_{\pi_{Tj+1}}(X_{\tau_2})) d\tau_1 d\tau_2 \\ &= \frac{2}{h_T^2} \iint_{\pi_{Tj} \times \pi_{Tj+1}} \int_0^T \left(1 - \frac{u}{T}\right) g_u(y, z) du dy dz \\ &= 2 \iint_{[0, 1]^2} \int_0^\infty \mathbf{1}_{[0, T]}(u) \left(1 - \frac{u}{T}\right) g_u(b_j + h_T s, b_{j+1} + h_T t) du ds dt. \end{aligned}$$

En appliquant le théorème de convergence dominée, on obtient alors

$$\lim_{T \rightarrow \infty} T \cdot \text{Cov}(\hat{f}_j, \hat{f}_{j+1}) = 2 \iint_{[0, 1]^2} \int_0^\infty g_u(x, x) du dy dz = 2 \int_0^\infty g_u(x, x) du,$$

et comme

$$T \cdot \text{Var}(\hat{f}_j) \xrightarrow{T \rightarrow \infty} 2 \int_0^\infty g_u(x, x) du, \quad T \cdot \text{Var}(\hat{f}_{j+1}) \xrightarrow{T \rightarrow \infty} 2 \int_0^\infty g_u(x, x) du,$$

on conclut directement à partir de la relation (1.2) :

$$\lim_{T \rightarrow \infty} T \cdot \text{Var}(\hat{f}_T^{PF}(x)) = 2 \int_0^\infty g_u(x, x) du.$$

■

### Preuve du corollaire 1.5

Les résultats du corollaire découlent du lemme 1.4 et du corollaire 1.4. Pour la partie (i) correspondant au cas “optimal” :

$$\min \left\{ h_T^{-4}, Th_T \right\} \cdot \text{EMQ}(\hat{f}_T^{PF}(x)) \leq h_T^{-4} \cdot \left( \mathbb{E} \hat{f}_T^{PF}(x) - f(x) \right)^2 + Th_T \cdot \text{Var}(\hat{f}_T^{PF}(x)),$$

et on conclut en choisissant  $h_T \sim T^{-1/5}$ . La partie (ii) découle du choix  $h_T = o(T^{-1/4})$ .

■

### Preuve du lemme 1.5

Nous reprenons ici les grandes lignes de la preuve apportée par Scott (1985) en concluant avec la condition höldérienne  $\tilde{R}(ii)$  sur  $f''$ . Soit  $x \in [c_j, c_{j+1}[$ ,  $\hat{f}_T^{PF}(x) = \left(\frac{x-c_j}{h_T}\right)\hat{f}_{j+1} + \left(\frac{c_{j+1}-x}{h_T}\right)\hat{f}_j$ . Dans un premier temps, on écrit un développement limité de  $f$  à l'ordre 2 au voisinage de  $b_{j+1}$  ( $=: m_j$ ), milieu de l'intervalle  $[c_j, c_{j+1}[$  :

$$f(x) = f(m_j) + (x - m_j)f'(m_j) + \frac{1}{2}(x - m_j)^2 f''(m_j + \theta(x - m_j)), \quad \theta \in ]0, 1[.$$

On calcule  $\mathbb{E}(\hat{f}_j)$  et  $\mathbb{E}(\hat{f}_{j+1})$  :

$$\begin{aligned} \mathbb{E}(\hat{f}_j) &= \frac{1}{h_T} \int_{\pi_{Tj}} f(x) dx = f(m_j) - \frac{f'(m_j)}{2} h_T + \frac{f''(\xi_j)}{6} h_T^2, \\ \mathbb{E}(\hat{f}_{j+1}) &= \frac{1}{h_T} \int_{\pi_{Tj+1}} f(x) dx = f(m_j) + \frac{f'(m_j)}{2} h_T + \frac{f''(\xi_{j+1})}{6} h_T^2, \end{aligned}$$

où  $\xi_j \in \pi_{Tj}$  et  $\xi_{j+1} \in \pi_{Tj+1}$ . On en déduit alors une expression du biais de  $\hat{f}_T^{PF}$  au point  $x$  :

$$\begin{aligned} \mathbb{E} \hat{f}_T^{PF}(x) - f(x) &= \frac{1}{6}(x - c_j) f''(\xi_{j+1}) h_T + \frac{1}{6}(c_{j+1} - x) f''(\xi_j) h_T \\ &\quad - \frac{1}{2}(x - m_j)^2 f''(m_j + \theta(x - m_j)). \end{aligned}$$

On élève ensuite le biais au carré :

$$\begin{aligned} \left[ \mathbb{E} \hat{f}_T^{PF}(x) - f(x) \right]^2 &= \left( \frac{x - c_j}{6} \right)^2 f''(\xi_{j+1})^2 h_T^2 + \left( \frac{c_{j+1} - x}{6} \right)^2 f''(\xi_j)^2 h_T^2 \\ &\quad + \frac{1}{4}(x - m_j)^4 f''(m_j + \theta(x - m_j))^2 \\ &\quad + \frac{(x - c_j)(c_{j+1} - x)}{18} f''(\xi_j) f''(\xi_{j+1}) h_T^2 \\ &\quad - \left( \frac{x - c_j}{6} \right) (x - m_j)^2 f''(\xi_{j+1}) f''(m_j + \theta(x - m_j)) h_T \\ &\quad - \left( \frac{c_{j+1} - x}{6} \right) (x - m_j)^2 f''(\xi_j) f''(m_j + \theta(x - m_j)) h_T \\ &=: \mathbf{T}_1(x) + \mathbf{T}_2(x) + \mathbf{T}_3(x) + \mathbf{T}_4(x) - \mathbf{T}_5(x) - \mathbf{T}_6(x), \end{aligned}$$

et on intègre chacun des termes sur  $[c_j, c_{j+1}[$  :

$$\begin{aligned} \int_{c_j}^{c_{j+1}} T_1(x) dx &= \frac{f''(\xi_{j+1})^2}{108} h_T^5, & \int_{c_j}^{c_{j+1}} T_2(x) dx &= \frac{f''(\xi_j)^2}{108} h_T^5, \\ \int_{c_j}^{c_{j+1}} T_3(x) dx &= \frac{f''(\tilde{x}_j)^2}{320} h_T^5, & \int_{c_j}^{c_{j+1}} T_4(x) dx &= \frac{f''(\xi_j) f''(\xi_{j+1})}{108} h_T^5, \\ \int_{c_j}^{c_{j+1}} T_5(x) dx &= \frac{f''(\hat{x}_j) f''(\xi_{j+1})}{144} h_T^5, & \int_{c_j}^{c_{j+1}} T_6(x) dx &= \frac{f''(\bar{x}_j) f''(\xi_j)}{144} h_T^5, \end{aligned}$$

avec  $(\tilde{x}_j, \hat{x}_j, \bar{x}_j) \in [c_j, c_{j+1}[^3$ . D'où

$$\begin{aligned} \int_{c_j}^{c_{j+1}} \left[ E \hat{f}_T^{PF}(x) - f(x) \right]^2 dx &= h_T^5 \left\{ \frac{f''(\xi_{j+1})^2}{108} + \frac{f''(\xi_j)^2}{108} + \frac{f''(\tilde{x}_j)^2}{320} \right. \\ &\quad \left. + \frac{f''(\xi_j) \cdot f''(\xi_{j+1})}{108} - \frac{f''(\hat{x}_j) \cdot f''(\xi_{j+1})}{144} - \frac{f''(\bar{x}_j) \cdot f''(\xi_j)}{144} \right\}. \end{aligned}$$

Avec  $\tilde{R}(ii)$ , on écrit

$$\begin{aligned} f''(\xi_{j+1}) &= f''(\xi_j) + A_j \quad \text{où } A_j = f''(\xi_{j+1}) - f''(\xi_j) \quad \text{et } |A_j| \leq l_0 h_T^\nu, \\ f''(\hat{x}_j) &= f''(\xi_{j+1}) + B_j \quad \text{où } B_j = f''(\hat{x}_j) - f''(\xi_{j+1}) \quad \text{et } |B_j| \leq l_0 h_T^\nu, \\ f''(\bar{x}_j) &= f''(\xi_j) + C_j \quad \text{où } C_j = f''(\bar{x}_j) - f''(\xi_j) \quad \text{et } |C_j| \leq l_0 h_T^\nu, \end{aligned}$$

et en remarquant que

$$\begin{aligned} &\left| h_T^4 \sum_j h_T \left( \frac{f''(\xi_j) \cdot A_j}{108} - \frac{f''(\xi_{j+1}) \cdot B_j}{144} - \frac{f''(\xi_j) \cdot C_j}{144} \right) \right| \\ &\leq h_T^{4+\nu} \cdot l_0 \left| \frac{1}{108} + \frac{1}{144} + \frac{1}{144} \right| \left( \int_{\mathbb{R}} |f''(x)| dx + o(1) \right) \\ &\leq h_T^{4+\nu} \cdot \frac{5l_0 \int_{\mathbb{R}} |f''(x)| dx}{216} \times (1 + o(1)), \end{aligned}$$

on obtient

$$\sum_j h_T^5 \left( \frac{f''(\xi_j) \cdot A_j}{108} - \frac{f''(\xi_{j+1}) \cdot B_j}{144} - \frac{f''(\xi_j) \cdot C_j}{144} \right) = \mathcal{O}(h_T^{4+\nu}).$$

On en déduit l'expression du biais au carré intégré :

$$\text{BCI}(\hat{f}_T^{PF}) = \left\{ \frac{1}{108} + \frac{1}{108} + \frac{1}{320} + \frac{1}{108} - \frac{1}{144} - \frac{1}{144} \right\} \mathcal{R}(f'') h_T^4 + o(h_T^4).$$

Finalement,

$$\text{BCI}(\hat{f}_T^{PF}) = \frac{49}{2880} \mathcal{R}(f'') h_T^4 \times (1 + o(1)).$$

■

### Preuve du corollaire 1.6

La preuve est basée sur la décomposition (1.2). La variance intégrée de  $\hat{f}_T^{PF}$  s'obtient ensuite en intégrant sur  $[c_j, c_{j+1}[$  et en sommant sur les indices  $j$ . Notons  $\bar{E}_j$  une expression (indépendante de  $x$ ) qui majore  $\text{Var}(\hat{f}_j)$ ,  $j \in \mathbb{Z}$ . On a alors

$$|\text{Cov}(\hat{f}_j, \hat{f}_{j+1})| \leq \sqrt{\text{Var}(\hat{f}_j)\text{Var}(\hat{f}_{j+1})} \leq \frac{1}{2} \left\{ \text{Var}(\hat{f}_j) + \text{Var}(\hat{f}_{j+1}) \right\} \leq \frac{1}{2} \{ \bar{E}_j + \bar{E}_{j+1} \}.$$

Ainsi,

$$\int_{c_j}^{c_{j+1}} \text{Var}(\hat{f}_T^{PF}(x)) dx \leq \frac{h_T}{3} \bar{E}_{j+1} + \frac{h_T}{3} \bar{E}_j + \frac{h_T}{6} \{ \bar{E}_j + \bar{E}_{j+1} \} = \frac{h_T}{2} \{ \bar{E}_j + \bar{E}_{j+1} \},$$

et la suite découle des majorations établies dans les théorèmes 1.1 et 1.2. Si les conditions  $\tilde{R}(iii)$  et  $H_1$  sont vérifiées et si  $f^{1-\frac{1}{p}} \in L^1(\mathbb{R})$  avec  $1 < p \leq \varrho - 1$ , alors pour  $0 \leq \varepsilon \leq 1 - \frac{1}{\varrho - p}$  :

$$\begin{aligned} Th_T \cdot \int_{c_j}^{c_{j+1}} \text{Var}(\hat{f}_T^{PF}(x)) dx &\leq \frac{h_T}{2} \left\{ f(\xi_j)(1 - h_T f(\xi_j)) \cdot 2u_0 \left(1 - \frac{u_0}{2T}\right) + h_T^\varepsilon \cdot 2M f(\xi_j) \right. \\ &\quad + h_T^{\frac{1}{p}\{(1-\varepsilon)(\varrho-p)-1\}} \cdot \frac{4p^2(2a_0)^{1/p}}{\varrho-p} f(\xi_j)^{1-\frac{1}{p}} \\ &\quad + f(\xi_{j+1})(1 - h_T f(\xi_{j+1})) \cdot 2u_0 \left(1 - \frac{u_0}{2T}\right) + h_T^\varepsilon \cdot 2M f(\xi_{j+1}) \\ &\quad \left. + h_T^{\frac{1}{p}\{(1-\varepsilon)(\varrho-p)-1\}} \cdot \frac{4p^2(2a_0)^{1/p}}{\varrho-p} f(\xi_{j+1})^{1-\frac{1}{p}} \right\}. \end{aligned}$$

Si les conditions  $H_2(i)(ii)'_a$  sont vérifiées, alors

$$T \cdot \int_{c_j}^{c_{j+1}} \text{Var}(\hat{f}_T^{PF}(x)) dx \leq \frac{h_T}{2} \{2\varphi(\eta_j) + 2\varphi(\eta_{j+1})\} \leq h_T \{\varphi(\eta_j) + \varphi(\eta_{j+1})\}.$$

Il reste maintenant à sommer sur l'ensemble des indices  $j$  des éléments de la partition  $\Pi_T$ . Comme les fonctions  $f^{1-\frac{1}{p}}$ ,  $f$  et  $f^2$  sont supposées Riemann-intégrables sur  $\mathbb{R}$ , on peut écrire

$$\sum_j h_T f^\kappa(\xi_j) = \int_{\mathbb{R}} f^\kappa(x) dx + o(1) \quad \text{pour } \kappa = 1 - \frac{1}{p}, 1, 2.$$

Par conséquent, sous  $\tilde{R}(iii)$  et  $H_1$  pour  $0 \leq \varepsilon \leq 1 - \frac{1}{\varrho - p}$  :

$$Th_T \cdot \text{VI}(\hat{f}_T^{PF}) \leq \left\{ 2u_0 + 2M \cdot h_T^\varepsilon + \frac{4p^2(2a_0)^{1/p}}{\varrho-p} \|f^{1-\frac{1}{p}}\|_1 \cdot h_T^{\frac{1}{p}\{(1-\varepsilon)(\varrho-p)-1\}} \right\} (1 + o(1)),$$

et sous  $H_2(i)(ii)'_{ab}$  :

$$T \cdot \text{VI}(\hat{f}_T^{PF}) \leq 2\|\varphi\|_1 + o(1).$$

■

### Preuve du corollaire 1.7

Par passage à la limite, on obtient les vitesses de convergence de l'erreur en moyenne quadratique intégrée du polygone de fréquences à partir du lemme 1.5 et du corollaire 1.6. Sous  $\tilde{R}$  et  $H_1$  et si  $f^{1-\frac{1}{p}} \in L^1(\mathbb{R})$  pour  $1 < p \leq \varrho - 1$ , avec le choix  $h_T = k_T \cdot T^{-\frac{1}{5}}$  tel que  $k_T \rightarrow k$ ,  $0 < k < \infty$ , on a d'une part :

$$T^{\frac{4}{5}} \cdot \text{BCI}(\hat{f}_T^{PF}) = \frac{49}{2880} k_T^4 \mathcal{R}(f'') + o(1) \quad \text{soit} \quad \lim_{T \rightarrow \infty} T^{\frac{4}{5}} \cdot \text{BCI}(\hat{f}_T^{PF}) = \frac{49}{2880} k^4 \mathcal{R}(f'').$$

D'autre part, si  $p = \varrho - 1$  ( $\Leftrightarrow \varepsilon = 0$ ) :

$$T^{\frac{4}{5}} \cdot \text{VI}(\hat{f}_T^{PF}) \leq \frac{1}{k_T} \left\{ 2u_0 + 2M \cdot h_T^\varepsilon + \frac{4p^2(2a_0)^{1/p}}{\varrho - p} \|f^{1-\frac{1}{p}}\|_1 \cdot h_T^{\frac{1}{p}\{(1-\varepsilon)(\varrho-p)-1\}} \right\} (1 + o(1)),$$

soit

$$\limsup_{T \rightarrow \infty} T^{\frac{4}{5}} \cdot \text{VI}(\hat{f}_T^{PF}) \leq \frac{1}{k} \left\{ 2u_0 + 2M + \frac{4p^2(2a_0)^{1/p}}{\varrho - p} \|f^{1-\frac{1}{p}}\|_1 \right\},$$

et si  $p < \varrho - 1$ , on peut améliorer la constante en choisissant un  $\varepsilon > 0$  tel que

$$\limsup_{T \rightarrow \infty} T^{\frac{4}{5}} \cdot \text{VI}(\hat{f}_T^{PF}) \leq \frac{2u_0}{k}.$$

Par conséquent, pour  $p = \varrho - 1$ , on obtient

$$\limsup_{T \rightarrow \infty} T^{\frac{4}{5}} \cdot \text{EMQI}(\hat{f}_T^{PF}) \leq \frac{49}{2880} k^4 \mathcal{R}(f'') + \frac{1}{k} \left\{ 2u_0 + 2M + \frac{4p^2(2a_0)^{1/p}}{\varrho - p} \|f^{1-\frac{1}{p}}\|_1 \right\},$$

et pour  $p < \varrho - 1$  :

$$\limsup_{T \rightarrow \infty} T^{\frac{4}{5}} \cdot \text{EMQI}(\hat{f}_T^{PF}) \leq \frac{49}{2880} k^4 \mathcal{R}(f'') + \frac{2u_0}{k}.$$

Sous  $\tilde{R}(i)(ii)$  et  $H_2(i)(ii)'_{ab}$  et si le paramètre  $h_T$  est un  $o(T^{-\frac{1}{4}})$ , le biais est négligeable devant la variance et on obtient immédiatement la vitesse suroptimale du polygone de fréquences :

$$\limsup_{T \rightarrow \infty} T \cdot \text{EMQI}(\hat{f}_T^{PF}) \leq 2\|\varphi\|_1.$$

■

## 1.6.2 Normalité asymptotique

### Preuve du théorème 1.4

Notons  $[x]$  la partie entière du réel  $x$ . La preuve du théorème repose sur une application directe du théorème 1.3. Considérons un  $m$ -uplet  $(x_1, \dots, x_m)$  de valeurs réelles distinctes. Au point  $x_r$ ,  $1 \leq r \leq m$ , on suppose l'existence d'un indice  $j(x_r, T) =: j(r)$  tel que  $x_r \in \pi_{Tj(r)}$ , et à partir des valeurs  $\hat{f}_T^H(x_r)$ , pour tout  $n \geq 1$ , on construit une suite strictement stationnaire de variables aléatoires réelles centrées  $\{Y_{n,i}^{(r)}, 1 \leq i \leq n\}$  définies par

$$Y_{n,i}^{(r)} = \frac{1}{\sqrt{\delta} h_T} \left\{ \int_{\delta(i-1)}^{\delta i} \mathbb{1}_{\pi_{Tj(r)}}(X_t) dt - \mathbb{E} \int_{\delta(i-1)}^{\delta i} \mathbb{1}_{\pi_{Tj(r)}}(X_t) dt \right\},$$

où  $n\delta = T_n = T$ ,  $n = \lfloor T \rfloor$  ( $T \geq 1$ ) et ainsi  $1 \leq \delta < 2$ . Notons  $Y_{n,i}$  le vecteur aléatoire à valeurs dans  $\mathbb{R}^m$  défini par

$$Y_{n,i} = \left\{ \left( Y_{n,i}^{(1)}, \dots, Y_{n,i}^{(m)} \right)', 1 \leq i \leq n \right\}.$$

Pour tout  $n \geq 1$  et  $1 \leq r \leq m$ , on a alors

$$\frac{\sum_{i=1}^n Y_{n,i}}{\sqrt{n}} = \left\{ \sqrt{T} \left( \hat{f}_T^H(x_1) - \mathbb{E} \hat{f}_T^H(x_1) \right), \dots, \sqrt{T} \left( \hat{f}_T^H(x_m) - \mathbb{E} \hat{f}_T^H(x_m) \right) \right\}.$$

Pour conclure à la convergence en loi de  $(\sum_{i=1}^n Y_{n,i}/\sqrt{n})_{n \geq 1}$  vers un vecteur Gaussien, nous devons au préalable vérifier que les hypothèses  $H_3(i) - (v)$  du théorème 1.3 sont satisfaites :

(i) Pour tout  $n \geq 1$ ,  $i \geq 1$  et  $1 \leq r \leq m$ ,

$$\begin{aligned} \left| Y_{n,i}^{(r)} \right| &= \frac{1}{\sqrt{\delta} h_T} \left| \int_{\delta(i-1)}^{\delta i} \mathbb{1}_{\pi_{Tj(r)}}(X_t) dt - \mathbb{E} \int_{\delta(i-1)}^{\delta i} \mathbb{1}_{\pi_{Tj(r)}}(X_t) dt \right| \\ &= \frac{1}{\sqrt{\delta} h_T} \left| \int_{\delta(i-1)}^{\delta i} \mathbb{1}_{\pi_{Tj(r)}}(X_t) dt - \delta P(X_0 \in \pi_{Tj(r)}) \right| < \frac{\sqrt{2}}{h_T} < \infty. \end{aligned}$$

(ii) Pour tout  $i \geq 1$  et chaque  $1 \leq r \leq m$ ,

$$\begin{aligned} \mathbb{E} \left( Y_{n,i}^{(r)} \right)^2 &= \mathbb{E} \left( Y_{n,1}^{(r)} \right)^2 = \text{Var} \left( \frac{1}{\sqrt{\delta} h_T} \int_0^\delta \mathbb{1}_{\pi_{Tj(r)}}(X_t) dt \right) \\ &= \frac{2}{h_T^2} \iint_{\pi_{Tj(r)}^2} \int_0^\delta \left( 1 - \frac{u}{\delta} \right) g_u(y, z) du dy dz, \end{aligned}$$

et avec la condition  $H_2(ii)$  de Castellana-Leadbetter, on a

$$\mathbb{E} \left( Y_{n,i}^{(r)} \right)^2 \leq 2 \int_0^\infty \|g_u\|_\infty du < \infty.$$

(iii) Soit  $a_n$  une suite de nombres entiers telle que  $a_n \rightarrow \infty$  et  $a_n \leq n$  pour tout  $n \geq 1$ . Pour chaque  $1 \leq r, s \leq m$ ,

$$\begin{aligned} &\frac{1}{a_n} \mathbb{E} \left( \sum_{i=1}^{a_n} Y_{n,i}^{(r)} \sum_{j=1}^{a_n} Y_{n,j}^{(s)} \right) \\ &= \frac{1}{\delta a_n h_T^2} \iint_{[0, \delta a_n]^2} \text{Cov} \left( \mathbb{1}_{\pi_{Tj(r)}}(X_{\tau_1}), \mathbb{1}_{\pi_{Tj(s)}}(X_{\tau_2}) \right) d\tau_1 d\tau_2 \\ &= \frac{2}{h_T^2} \iint_{\pi_{Tj(r)} \times \pi_{Tj(s)}} \int_0^{\delta a_n} \left( 1 - \frac{u}{\delta a_n} \right) g_u(y, z) du dy dz \\ &= 2 \iint_{[0,1]^2} \int_0^\infty \mathbb{1}_{[0, \delta a_n]}(u) \left( 1 - \frac{u}{\delta a_n} \right) g_u(b_{j(r)} + h_T p, b_{j(s)} + h_T q) du dp dq. \end{aligned}$$

et de manière similaire à la preuve de la partie (i) du théorème 1.2, le théorème de convergence dominée entraîne

$$\lim_{n \rightarrow \infty} \frac{1}{a_n} \mathbb{E} \left( \sum_{i=1}^{a_n} Y_{n,i}^{(r)} \sum_{j=1}^{a_n} Y_{n,j}^{(s)} \right) = 2 \int_0^\infty g_u(x_r, x_s) du < \infty.$$

(iv) Comme  $\alpha_Y(k) \leq \alpha_X(k)$ , la condition  $\sum_{k \geq 1} k \alpha_X^{(a-1)/a}(k) < \infty$  implique la condition  $\sum_{k \geq 1} k \alpha_Y^{(a-1)/a}(k) < \infty$ .

(v) Cette condition est clairement vérifiée pour le choix de  $h_T \geq c \cdot T^{-a^2/((3a-1)(2a-1))}$ ,  $0 < c < \infty$ , ce qui achève la preuve du théorème. ■

### Preuve du théorème 1.5

Considérons un  $m$ -uplet  $(x_1, \dots, x_m)$  de valeurs réelles distinctes. Au point  $x_r$ ,  $1 \leq r \leq m$ , on suppose l'existence d'un indice  $j(x_r, T) =: j(r)$  tel que  $x_r \in \pi_{Tj(r)}$  et l'estimateur par polygone de fréquences vaut

$$\hat{f}_T^{PF}(x_r) = \begin{cases} \left( \frac{x_r - c_{j(r)}}{h_T} \right) \hat{f}_{j(r)+1} + \left( \frac{c_{j(r)+1} - x_r}{h_T} \right) \hat{f}_{j(r)} & \text{si } x_r \in \pi_{Tj(r)}^+, \\ \left( \frac{x_r - c_{j(r)-1}}{h_T} \right) \hat{f}_{j(r)} + \left( \frac{c_{j(r)} - x_r}{h_T} \right) \hat{f}_{j(r)-1} & \text{si } x_r \in \pi_{Tj(r)}^-, \end{cases}$$

où  $\pi_{Tj(r)}^+ := \pi_{Tj(r)} \cap [c_{j(r)}, c_{j(r)+1}[$  et  $\pi_{Tj(r)}^- := \pi_{Tj(r)} \cap [c_{j(r)-1}, c_{j(r)}[$ . Sans perte de généralité, on supposera que  $x_r \in \pi_{Tj(r)}^+$  et que

$$\hat{f}_T^{PF}(x_r) = \left( \frac{x_r - c_{j(r)}}{h_T} \right) \hat{f}_{j(r)+1} + \left( \frac{c_{j(r)+1} - x_r}{h_T} \right) \hat{f}_{j(r)}.$$

À partir de  $\hat{f}_T^{PF}(x_r)$ , pour tout  $n \geq 1$  et chaque  $1 \leq r \leq m$ , on construit une suite strictement stationnaire de variables aléatoires réelles centrées  $\{Y_{n,i}^{(r)}, 1 \leq i \leq n\}$  définies par

$$Y_{n,i}^{(r)} = \frac{(x_r - c_{j(r)})}{\sqrt{\delta} h_T^2} \left\{ \int_{\delta(i-1)}^{\delta i} \mathbb{1}_{\pi_{Tj(r)+1}}(X_t) dt - \mathbb{E} \int_{\delta(i-1)}^{\delta i} \mathbb{1}_{\pi_{Tj(r)+1}}(X_t) dt \right\} \\ + \frac{(c_{j(r)+1} - x_r)}{\sqrt{\delta} h_T^2} \left\{ \int_{\delta(i-1)}^{\delta i} \mathbb{1}_{\pi_{Tj(r)}}(X_t) dt - \mathbb{E} \int_{\delta(i-1)}^{\delta i} \mathbb{1}_{\pi_{Tj(r)}}(X_t) dt \right\},$$

où  $n\delta = T_n = T$ ,  $n = \lfloor T \rfloor$  ( $T \geq 1$ ) et  $1 \leq \delta < 2$ .

Pour tout  $n \geq 1$  et  $1 \leq r \leq m$ , on a alors

$$\frac{\sum_{i=1}^n Y_{n,i}^{(r)}}{\sqrt{n}} = \left\{ \sqrt{T} \left( \hat{f}_T^{PF}(x_1) - \mathbb{E} \hat{f}_T^{PF}(x_1) \right), \dots, \sqrt{T} \left( \hat{f}_T^{PF}(x_m) - \mathbb{E} \hat{f}_T^{PF}(x_m) \right) \right\}.$$

Il nous reste à vérifier que les hypothèses du théorème 1.3 sont satisfaites :

(i) Pour tout  $n \geq 1$ ,  $i \geq 1$  et  $1 \leq r \leq m$ ,

$$\left| Y_{n,i}^{(r)} \right| \leq \left| \frac{x_r - c_{j(r)}}{h_T} \right| \frac{\sqrt{\delta}}{h_T} + \left| \frac{c_{j(r)+1} - x_r}{h_T} \right| \frac{\sqrt{\delta}}{h_T} < \frac{\sqrt{2}}{h_T} < \infty.$$

(ii) Pour tout  $i \geq 1$  et chaque  $1 \leq r \leq m$ ,  $E\left(Y_{n,i}^{(r)}\right)^2 = E\left(Y_{n,1}^{(r)}\right)^2$  et

$$\begin{aligned} E\left(Y_{n,1}^{(r)}\right)^2 &= \frac{2(x_r - c_{j(r)})^2}{h_T^4} \iint_{\pi_{Tj(r)+1}^2} \int_0^\delta \left(1 - \frac{u}{\delta}\right) g_u(y, z) du \, dydz \\ &\quad + \frac{2(c_{j(r)+1} - x_r)^2}{h_T^4} \iint_{\pi_{Tj(r)}^2} \int_0^\delta \left(1 - \frac{u}{\delta}\right) g_u(y, z) du \, dydz \\ &\quad + \frac{4(x_r - c_{j(r)})(c_{j(r)+1} - x_r)}{h_T^4} \iint_{\pi_{Tj(r)} \times \pi_{Tj(r)+1}} \int_0^\delta \left(1 - \frac{u}{\delta}\right) g_u(y, z) du \, dydz, \end{aligned}$$

et les conditions  $H_2(ii)$  de Castellana-Leadbetter entraînent

$$\int_0^\infty \mathbb{1}_{[0,\delta]}(u) \left(1 - \frac{u}{\delta}\right) g_u(y, z) du \leq \int_0^\infty \|g_u\|_\infty du < \infty,$$

soit

$$E\left(Y_{n,i}^{(r)}\right)^2 \leq 2 \int_0^\infty \|g_u\|_\infty du < \infty.$$

(iii) Soit  $a_n$  une suite de nombres entiers telle que  $a_n \rightarrow \infty$  et  $a_n \leq n$  pour tout  $n \geq 1$ . Avec un calcul analogue à celui détaillé dans les preuves des théorèmes 1.2 et 1.4, on a pour chaque  $1 \leq r, s \leq m$ ,

$$\begin{aligned} \frac{1}{a_n} E\left(\sum_{i=1}^{a_n} Y_{n,i}^{(r)} \sum_{j=1}^{a_n} Y_{n,j}^{(s)}\right) &= \frac{2(x_r - c_{j(r)})}{h_T^2} \left[(x_s - c_{j(s)})J_1 + (c_{j(s)+1} - x_s)J_2\right] \\ &\quad + \frac{2(c_{j(r)+1} - x_r)}{h_T^2} \left[(x_s - c_{j(s)})J_3 + (c_{j(s)+1} - x_s)J_4\right], \end{aligned}$$

avec

$$\begin{aligned} J_1 &:= \iint_{[0,1]^2} \int_0^\infty \mathbb{1}_{[0,\delta a_n]}(u) \left(1 - \frac{u}{\delta a_n}\right) g_u(b_{j(r)+1} + h_T p, b_{j(s)+1} + h_T q) du \, dpdq, \\ J_2 &:= \iint_{[0,1]^2} \int_0^\infty \mathbb{1}_{[0,\delta a_n]}(u) \left(1 - \frac{u}{\delta a_n}\right) g_u(b_{j(r)+1} + h_T p, b_{j(s)} + h_T q) du \, dpdq, \\ J_3 &:= \iint_{[0,1]^2} \int_0^\infty \mathbb{1}_{[0,\delta a_n]}(u) \left(1 - \frac{u}{\delta a_n}\right) g_u(b_{j(r)} + h_T p, b_{j(s)+1} + h_T q) du \, dpdq, \\ J_4 &:= \iint_{[0,1]^2} \int_0^\infty \mathbb{1}_{[0,\delta a_n]}(u) \left(1 - \frac{u}{\delta a_n}\right) g_u(b_{j(r)} + h_T p, b_{j(s)} + h_T q) du \, dpdq, \end{aligned}$$

et tel que par convergence dominée :

$$J_k = \int_0^\infty g_u(x_r, x_s) du \times (1 + o(1)), \quad 1 \leq k \leq 4.$$

D'où

$$\lim_{n \rightarrow \infty} \frac{1}{a_n} E\left(\sum_{i=1}^{a_n} Y_{n,i}^{(r)} \sum_{j=1}^{a_n} Y_{n,j}^{(s)}\right) = 2 \int_0^\infty g_u(x_r, x_s) du < \infty.$$

(iv) Comme  $\alpha_Y(k) \leq \alpha_X(k)$ , la condition  $\sum_{k \geq 1} k \alpha_X^{(a-1)/a}(k) < \infty$  implique la condition  $\sum_{k \geq 1} k \alpha_Y^{(a-1)/a}(k) < \infty$ .

(v) Clairement vérifiée pour le même choix de  $h_T$  que dans le théorème 1.4, ce qui achève la preuve.  $\blacksquare$

### Preuve du corollaire 1.8

Comme  $f \in \mathcal{C}^2(\mathbb{R})$ , on a

$$\left| \mathbb{E} \hat{f}_T^{PF}(x_r) - f(x_r) \right| = \mathcal{O}(h_T^2).$$

Par conséquent, en prenant  $h_T = o\left(T^{-\frac{1}{4}}\right)$ , le résultat du corollaire est immédiat. ■

### 1.6.3 Convergence presque sûre

Pour établir la convergence presque sûre de nos estimateurs, nous aurons besoin d'un lemme de couplage dû à Rio (2000) (lemme 5.2, p. 72) :

**Lemme 1.8.** *Soit  $\mathcal{A}$  une sous-tribu de  $(\Omega, \mathcal{F}, P)$  et  $X$  une variable aléatoire réelle, prenant ses valeurs dans un intervalle compact  $[a, b]$ . Soit  $U$  une variable de loi Uniforme sur  $[0, 1]$ , indépendante de la tribu engendrée par  $X$  et  $\mathcal{A}$ . Alors il existe une variable aléatoire  $X^*$ , mesurable pour la tribu  $\mathcal{A} \vee \sigma(X) \vee \sigma(U)$ , indépendante de  $\mathcal{A}$  et de même loi que  $X$ , telle que*

$$\mathbb{E} |X - X^*| \leq 2(b - a)\alpha(\mathcal{A}, X).$$

Pour la suite, nous rappelons aussi l'inégalité de Bernstein (cf. Pollard (1984), p. 193) et le lemme de Borel-Cantelli (cf. Neveu (1970), p. 122).

**Lemme 1.9 (Inégalité de Bernstein).** *Soient  $Y_1, \dots, Y_n$  des variables aléatoires réelles centrées et indépendantes telles que  $|Y_i| \leq M$  (p.s.) pour  $0 \leq i \leq n$  et une constante  $M$ . Supposons de plus que  $V \geq \sum_{i=1}^n \mathbb{E}(Y_i^2)$  alors, pour tout  $\eta > 0$ , on a*

$$P\left(\left|\sum_{i=1}^n Y_i\right| > \eta\right) \leq 2 \exp\left(-\frac{\eta^2}{2V + \frac{2}{3}M\eta}\right).$$

Pour une suite d'événements  $(A_n, n \geq 1)$ , rappelons que

$$\limsup_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k \geq n} A_k = \lim_{n \rightarrow \infty} \downarrow \bigcup_{k \geq n} A_k,$$

qui s'écrit encore

$$\limsup_{n \rightarrow \infty} A_n = \left\{ \omega \in \Omega, \omega \in A_n \text{ pour une infinité de } n \right\} = \left\{ \omega \in \Omega, \sum_{n=1}^{\infty} \mathbf{1}_{A_n}(\omega) = \infty \right\}.$$

**Lemme 1.10 (Borel-Cantelli).** *Soit  $(A_n, n \geq 1)$  une suite d'événements.*

(i) *Si  $\sum_{n=1}^{\infty} P(A_n) < \infty$  alors  $P(\limsup_{n \rightarrow \infty} A_n) = 0$  ;*

(ii) *Si les  $A_n$  sont indépendants et si  $\sum_{n=1}^{\infty} P(A_n) = \infty$  alors  $P(\limsup_{n \rightarrow \infty} A_n) = 1$ .*

### Preuve du théorème 1.6

PREMIÈRE ÉTAPE : La preuve est similaire à celle détaillée par Blanke (2004) pour l'estimateur à noyau. Pour  $x \in \pi_{T_j}$ , on pose

$$Z_t := \frac{1}{h_{T_n}^d} \mathbb{1}_{\pi_{T_j}}(X_t) - \mathbb{E} \frac{1}{h_{T_n}^d} \mathbb{1}_{\pi_{T_j}}(X_t) = \frac{1}{h_{T_n}^d} (\mathbb{1}_{\pi_{T_j}}(X_t) - p_j),$$

où  $p_j := P(X_0 \in \pi_{T_j})$ . Notons  $\lfloor x \rfloor$  la partie entière du réel  $x$ . Pour des fonctions positives  $q_T := q_{T_n}$  et  $r_T := r_{T_n}$  telles que  $T_n = 2q_T r_T$  et  $q_T \uparrow \infty$ , nous définissons alors  $(2\lfloor q_T \rfloor + 1)$  blocs comme suit :

$$S'_{T_n}(x) := \sum_{k=1}^{\lfloor q_T \rfloor + 1} V_{T_n}(2k-1), \quad S''_{T_n}(x) := \sum_{k=1}^{\lfloor q_T \rfloor} V_{T_n}(2k),$$

où  $V_{T_n}(k) := \int_{(k-1)r_T}^{kr_T} Z_t dt$ ,  $k = 1, \dots, 2\lfloor q_T \rfloor$ , et  $V_{T_n}(2\lfloor q_T \rfloor + 1) = \int_{2\lfloor q_T \rfloor r_T}^{2q_T r_T} Z_t dt$ . Nous pouvons noter que  $S'_{T_n}(x)$  et  $S''_{T_n}(x)$  constituent respectivement les sommes des blocs de rangs impairs et pairs. Si  $T_n \neq 2\lfloor q_T \rfloor r_T$  alors le dernier bloc de rang  $(2\lfloor q_T \rfloor + 1)$  est d'amplitude à valeurs dans  $]0, 2r_T]$  mais son influence s'avèrera négligeable dans les calculs. Enfin, il est important de remarquer que

$$\frac{1}{T_n} (S'_{T_n}(x) + S''_{T_n}(x)) = \hat{f}_{T_n}^H(x) - \mathbb{E} \hat{f}_{T_n}^H(x).$$

DEUXIÈME ÉTAPE : Pour tout  $\epsilon > 0$ , on a

$$P \left( \left| \hat{f}_{T_n}^H(x) - \mathbb{E} \hat{f}_{T_n}^H(x) \right| > \epsilon \right) \leq P \left( \left| S'_{T_n}(x) \right| > \frac{T_n \epsilon}{2} \right) + P \left( \left| S''_{T_n}(x) \right| > \frac{T_n \epsilon}{2} \right). \quad (1.3)$$

Puisque chacune des deux sommes  $S'_{T_n}(x)$  et  $S''_{T_n}(x)$  est traitée de la même façon, nous pourrions restreindre notre étude à  $S'_{T_n}(x)$  en utilisant de manière récursive le lemme 1.8 pour échanger les variables  $(V_{T_n}(1), \dots, V_{T_n}(2\lfloor q_T \rfloor + 1))$  avec des variables indépendantes : soit  $\{U_k, k \geq 1\}$  une suite de variables aléatoires uniformément distribuées sur  $[0, 1]$ , mutuellement indépendantes et indépendantes de  $\{V_{T_n}(k), k \geq 1\}$ . Si  $W_{T_n}(1) = V_{T_n}(1)$  alors pour tout  $2 \leq k \leq q_T$ , il existe une variable aléatoire  $W_{T_n}(2k-1)$  telle que

1.  $W_{T_n}(2k-1)$  est une fonction mesurable de  $V_{T_n}(1), \dots, V_{T_n}(2k-1), U_{2k-1}$  ;
2.  $W_{T_n}(2k-1)$  est indépendante de  $V_{T_n}(1), \dots, V_{T_n}(2k-3)$  ;
3.  $W_{T_n}(2k-1)$  a la même loi que  $V_{T_n}(2k-1)$ ,

et

$$\mathbb{E} |W_{T_n}(2k-1) - V_{T_n}(2k-1)| \leq 4 \|V_{T_n}(2k-1)\|_\infty \sup_{A \in \mathcal{A}, B \in \mathcal{B}} |P(A \cap B) - P(A)P(B)|, \quad (1.4)$$

où  $\mathcal{A}$  et  $\mathcal{B}$  désignent les  $\sigma$ -algèbres des événements engendrés par  $(V_{T_n}(1), \dots, V_{T_n}(2k-3))$  et  $V_{T_n}(2k-1)$  respectivement.

D'où, pour tout  $\kappa > 0$ , on obtient

$$P\left(|S'_{T_n}(x)| > \frac{T_n \epsilon}{2}\right) \leq P\left(\left|\sum_{k=1}^{\lfloor q_T \rfloor + 1} V_{T_n}(2k-1) - W_{T_n}(2k-1)\right| > \frac{T_n \epsilon \kappa}{2(1+\kappa)}\right) \\ + P\left(\left|\sum_{k=1}^{\lfloor q_T \rfloor + 1} W_{T_n}(2k-1)\right| > \frac{T_n \epsilon}{2(1+\kappa)}\right) =: \text{I} + \text{II},$$

où la constante  $\kappa$  est introduite pour réduire les constantes asymptotiques.

Majoration du terme I : en utilisant l'inégalité de Markov, nous obtenons

$$\text{I} \leq \frac{2(1+\kappa)}{T_n \epsilon \kappa} \sum_{k=1}^{\lfloor q_T \rfloor + 1} \mathbb{E}|V_{T_n}(2k-1) - W_{T_n}(2k-1)|.$$

Par ailleurs,

$$\max_{1 \leq k \leq 2\lfloor q_T \rfloor} |V_{T_n}(k)| \leq \frac{r_T}{h_{T_n}^d} \quad \text{et} \quad |V_{T_n}(2\lfloor q_T \rfloor + 1)| \leq 2 \frac{r_T}{h_{T_n}^d}, \quad (1.5)$$

et nous déduisons de (1.4) :

$$\text{I} \leq \frac{8(1+\kappa)r_T}{T_n \epsilon \kappa h_{T_n}^d} \sum_{k=1}^{\lfloor q_T \rfloor + 1} \alpha_X(r_T) \leq \frac{4(1+\kappa)}{\epsilon \kappa h_{T_n}^d} \alpha_X(r_T) \times (1 + o(1)).$$

où  $o(1)$  englobe l'ultime bloc négligeable pour  $k = \lfloor q_T \rfloor + 1$ . Comme  $\alpha_X(\cdot)$  décroît géométriquement d'après  $H_0^*$ , nous avons

$$\text{I} \leq \frac{4(1+\kappa)}{\kappa} \alpha_0 \frac{e^{-r_T \ln(1/\varrho)}}{\epsilon h_{T_n}^d} \times (1 + o(1)). \quad (1.6)$$

Majoration du terme II : du fait de l'indépendance des  $\{W_{T_n}(2k-1)\}$  et en utilisant (1.5), nous pouvons appliquer l'inégalité de Bernstein (cf. lemme 1.9) :

$$\text{II} \leq 2 \exp\left(-\frac{T_n^2 \epsilon^2}{8(1+\kappa)^2 \sum_{k=1}^{\lfloor q_T \rfloor + 1} \mathbb{E}(W_{T_n}^2(2k-1)) + \frac{8}{3}(1+\kappa) \frac{r_T}{h_{T_n}^d} T_n \epsilon}\right). \quad (1.7)$$

Nous désignons ensuite par  $R_x(T_n)$  une expression qui majore  $\mathbb{E}(W_n^2(2k-1))$  pour tout  $k = 1, \dots, \lfloor q_T \rfloor$  et qui dépend de nos hypothèses.

*Majoration sous les conditions  $H_1^*$  d'“optimalité” :*

$$\mathbb{E}(W_{T_n}^2(2k-1)) = \mathbb{E}(V_{T_n}^2(2k-1)) = \frac{1}{h_{T_n}^{2d}} \iint_{[0, r_T]^2} \text{Cov}(\mathbf{1}_{\pi_{T_j}}(X_s), \mathbf{1}_{\pi_{T_j}}(X_t)) ds dt,$$

qui s'écrit encore sous l'hypothèse faible de stationnarité :

$$\mathbb{E}(V_{T_n}^2(2k-1)) = \frac{1}{h_{T_n}^{2d}} \int_{[0, r_T]^2 \cap \Gamma} \text{Cov}(\mathbf{1}_{\pi_{T_j}}(X_s), \mathbf{1}_{\pi_{T_j}}(X_t)) ds dt \\ + \frac{2r_T}{h_{T_n}^{2d}} \int_{u_0}^{h_{T_n}^{-d+\epsilon}} \left(1 - \frac{u}{r_T}\right) \iint_{\pi_{T_j}^2} (f_u(y, z) - f(y)f(z)) dy dz du \\ + \frac{2r_T}{h_{T_n}^{2d}} \int_{h_{T_n}^{-d+\epsilon}}^{r_T} \left(1 - \frac{u}{r_T}\right) \text{Cov}(\mathbf{1}_{\pi_{T_j}}(X_0), \mathbf{1}_{\pi_{T_j}}(X_u)) du =: C_{n,1} + C_{n,2} + C_{n,3}.$$

De plus, nous établissons simplement que  $E(V_{T_n}^2(2[q_T] + 1)) \leq 2(C_{n,1} + C_{n,2} + C_{n,3})$ . Remarquons maintenant qu'il existe un réel positif  $T_N$  tel que  $\pi_{T_n j} \subset \nu_x$  pour tout  $T_n \geq T_N$ . Compte tenu de l'hypothèse de continuité  $R^*(i)$ , nous pourrons alors utiliser le théorème de la moyenne :  $\int_{\pi_{T_j}} f(x)dx = h_{T_n}^d f(\xi_j)$ ,  $\xi_j \in \pi_{T_j}$ . D'après l'inégalité de Cauchy-Schwarz et le théorème de la moyenne, on a

$$\begin{aligned} |C_{n,1}| &\leq \frac{1}{h_{T_n}^{2d}} \int_{[0,r_T]^2 \cap \Gamma} \sqrt{\text{Var}(\mathbb{1}_{\pi_{T_j}}(X_s)) \text{Var}(\mathbb{1}_{\pi_{T_j}}(X_t))} ds dt \\ &\leq \frac{1}{h_{T_n}^{2d}} p_j (1 - p_j) \int_{[0,r_T]^2 \cap \Gamma} ds dt \\ &\leq \frac{1}{h_{T_n}^d} f(\xi_j) (1 - h_{T_n}^d f(\xi_j)) (2u_0 r_T - u_0^2). \end{aligned} \quad (1.8)$$

En utilisant le fait que la densité conditionnelle soit bornée,

$$\begin{aligned} |C_{n,2}| &\leq \frac{2r_T}{h_{T_n}^{2d}} \int_{u_0}^{h_{T_n}^{-d+\varepsilon}} \iint_{\pi_{T_j}^2} f(y) |f_{s|0}(z|y) - f(z)| dy dz ds \\ &\leq \frac{r_T}{h_{T_n}^d} 2f(\xi_j) \max(M_1^x, M_2^x) \cdot h_{T_n}^\varepsilon. \end{aligned} \quad (1.9)$$

En utilisant l'inégalité de Davydov pour un couple  $(p, q) \in ]1, \rho - 1] \times \left[1 + \frac{\rho}{\rho - 2}, \infty\right]$ ,  $\rho > 2$ , tel que  $\frac{2}{q} + \frac{1}{p} = 1$ , on a

$$|C_{n,3}| \leq \frac{2r_T}{h_{T_n}^{2d}} \int_{h_{T_n}^{-d+\varepsilon}}^{r_T} 2p \cdot 2^{1/p} \cdot \|\mathbb{1}_{\pi_{T_j}}(X_0)\|_q \|\mathbb{1}_{\pi_{T_j}}(X_u)\|_q \cdot \alpha_X^{1/p}(s) ds,$$

soit avec l'hypothèse  $H_0^*$  qui implique la décroissance arithmétique de  $\alpha_X$  i.e.  $\alpha_X(u) \leq a_0 u^{-\rho}$  pour  $u \geq u_0$ ,  $a_0 > 0$  et  $\rho > 2$  :

$$\begin{aligned} |C_{n,3}| &\leq \frac{2r_T}{h_{T_n}^{2d}} \cdot 2p \cdot 2^{1/p} \cdot p_j^{2/q} \cdot \int_{h_{T_n}^{-d+\varepsilon}}^{\infty} a_0^{1/p} s^{-\rho/p} ds \\ &\leq \frac{r_T}{h_{T_n}^d} \cdot \frac{4p^2 (2a_0)^{1/p}}{\rho - p} f(\xi_j)^{1 - \frac{1}{p}} \cdot h_{T_n}^{\frac{1}{p} \{(d-\varepsilon)(\rho-p) - d\}}. \end{aligned} \quad (1.10)$$

D'où, en combinant les expressions (1.8), (1.9) et (1.10) :

$$\begin{aligned} r_T^{-1} h_{T_n}^d \cdot E(V_{T_n}^2(2k - 1)) &\leq f(\xi_j) (1 - h_{T_n}^d f(\xi_j)) 2u_0 \left(1 - \frac{u_0}{2r_T}\right) + 2f(\xi_j) \max(M_1^x, M_2^x) \cdot h_{T_n}^\varepsilon \\ &\quad + \frac{4p^2 (2a_0)^{1/p}}{\rho - p} f(\xi_j)^{1 - \frac{1}{p}} \cdot h_{T_n}^{\frac{1}{p} \{(d-\varepsilon)(\rho-p) - d\}}. \end{aligned}$$

Ainsi pour  $\varepsilon \in ]0, d \left(1 - \frac{1}{\rho-p}\right)[$  et par continuité de  $f$  au point  $x$ , on obtient

$$r_T^{-1} h_{T_n}^d \cdot E(V_{T_n}^2(2k - 1)) \leq 2u_0 f(x) \times (1 + o(1)). \quad (1.11)$$

On posera pour la suite :  $d_1(x) := 2u_0 f(x)$ .

*Majoration sous les conditions  $H_2^*$  de "suroptimalité" :*

(1) Avec l'hypothèse  $H_2^*(ii)$  sur la densité jointe du couple et le théorème de Fubini, on a

$$\begin{aligned} \mathbb{E}(V_{T_n}^2(2k-1)) &\leq \frac{2r_T}{h_{T_n}^2} \int_{[0, r_T]} \left(1 - \frac{u}{r_T}\right) \iint_{\pi_{T_j}^2} |g_u(y, z)| dy dz du \\ &\leq \frac{2r_T}{h_{T_n}} \int_{\pi_{T_j}} \underbrace{\sup_{z \in \nu_x} \int_0^\infty |g_u(y, z)| du}_{\leq \varphi(y)} dy, \end{aligned}$$

et comme  $\pi_{T_j} = [b_j, b_{j+1}[$  de telle sorte que  $b_j \uparrow_{T \rightarrow \infty} x$  et  $b_{j+1} = b_j + h_{T_n} \downarrow_{T \rightarrow \infty} x$ , on peut écrire aussi

$$\mathbb{E}(V_{T_n}^2(2k-1)) \leq 2r_T \int_{[0, 1[} \varphi(b_j + h_{T_n} s) ds.$$

Par le théorème de convergence dominée de Lebesgue, on a alors

$$\mathbb{E}(V_{T_n}^2(2k-1)) \leq 2r_T \varphi(x) \times (1 + o(1)),$$

soit

$$r_T^{-1} \mathbb{E}(V_{T_n}^2(2k-1)) \leq 2\varphi(x) \times (1 + o(1)). \quad (1.12)$$

On posera :  $d_2(x) := 2\varphi(x)$ .

(2) Avec l'hypothèse renforcée  $H_2^*(ii)'$ , on a

$$\begin{aligned} \mathbb{E}(V_{T_n}^2(2k-1)) &= \frac{2r_T}{h_{T_n}^2} \iint_{\pi_{T_j}^2} \int_0^{r_T} \left(1 - \frac{u}{r_T}\right) g_u(y, z) du dy dz \\ &= 2r_T \iint_{[0, 1]^2} \int_0^\infty \mathbb{1}_{[0, r_T]}(u) \left(1 - \frac{u}{r_T}\right) g_u(b_j + h_{T_n} s, b_j + h_{T_n} t) du ds dt. \end{aligned}$$

D'où par convergence dominée :

$$r_T^{-1} \mathbb{E}(V_{T_n}^2(2k-1)) = 2 \int_0^\infty g_u(x, x) du \times (1 + o(1)). \quad (1.13)$$

On posera :  $d_3(x) := 2 \int_0^\infty g_u(x, x) du$ .

TROISIÈME ÉTAPE : En combinant les expressions (1.3), (1.6), (1.7), (1.11), (1.12) et (1.13), on obtient

$$\begin{aligned} P\left(\left|\hat{f}_{T_n}^H(x) - \mathbb{E}\hat{f}_{T_n}^H(x)\right| > \epsilon_{T_n}\right) &\leq \frac{8(1+\kappa)}{\kappa} \alpha_0 \frac{e^{-r_T \ln(1/\varrho)}}{\epsilon_{T_n} h_{T_n}^d} \times (1 + o(1)) \\ &\quad + 4 \exp\left(-\frac{T_n \epsilon_{T_n}^2}{8(1+\kappa)^2 \left(\frac{q_T+1}{T_n}\right) R_x(T_n) + \frac{8}{3}(1+\kappa) \frac{r_T}{h_{T_n}^d} \epsilon_{T_n}}\right), \end{aligned}$$

où

$$R_x(T_n) = \begin{cases} \frac{r_T}{h_{T_n}^d} \cdot d_1(x) \times (1 + o(1)), & \text{cas "optimal"}; \\ r_T \cdot d_2(x) \times (1 + o(1)), & \text{cas "suroptimal" avec } H_2^*(ii); \\ r_T \cdot d_3(x) \times (1 + o(1)), & \text{cas "suroptimal" avec } H_2^*(ii)'. \end{cases}$$

Sous les conditions  $H_1^*$  d'“optimalité” : pour les choix  $\epsilon_{T_n} = \eta \left( \frac{\ln T_n}{T_n h_{T_n}^d} \right)^{1/2}$  et  $r_T = M \ln T_n$ , on a

$$P \left( \left| \hat{f}_{T_n}^H(x) - E \hat{f}_{T_n}^H(x) \right| > \epsilon_{T_n} \right) \leq \frac{8(1+\kappa)}{\eta\kappa} \alpha_0 \frac{T_n^{1/2} e^{-M \ln T_n \ln(1/\varrho)}}{h_{T_n}^{d/2} (\ln T_n)^{1/2}} \times (1 + o(1)) \\ + 4 \exp \left( - \frac{\eta^2 \ln T_n}{4(1+\kappa)^2 d_1(x) (1 + o(1)) + \frac{8}{3}(1+\kappa) M (\ln T_n)^{3/2} T_n^{-1/2} h_{T_n}^{-d/2}} \right). \quad (1.14)$$

De là, dans le terme issu de l'inégalité de Bernstein de l'expression (1.14), on remarque que le premier terme au dénominateur est dominant par rapport au second lorsque  $h_{T_n}$  satisfait la condition :  $(\ln T_n)^3 / (T_n h_{T_n}^d) \rightarrow 0$  quand  $T_n \uparrow \infty$ . Cela implique que

$$\exists \varsigma_1 > 0, \exists T_N, \forall T_n \geq T_N, \quad h_{T_n}^d \geq \frac{(\ln T_n)^3}{\varsigma_1 T_n},$$

tel que

$$\forall \eta > 2(1+\kappa) \sqrt{d_1(x)}, \quad \forall M > \frac{2}{\ln(1/\varrho)}, \quad \exists \lambda_1 > 0, \quad \exists \lambda_2 > 1,$$

où

$$\lambda_2 = \min \left( \frac{\eta^2}{4(1+\kappa)^2 d_1(x)}; M \ln(1/\varrho) - 1 \right) \quad \text{et} \quad d_1(x) = 2u_0 f(x),$$

et tel que

$$P \left( \left| \hat{f}_{T_n}^H(x) - E \hat{f}_{T_n}^H(x) \right| > \eta \left\{ \frac{\ln T_n}{T_n h_{T_n}^d} \right\}^{1/2} \right) \leq \lambda_1 T_n^{-\lambda_2}.$$

Sous les conditions  $H_2^*$  de “suroptimalité” : avec  $H_2^*(ii)$  ou  $H_2^*(ii)'$ , les choix  $\epsilon_{T_n} = \eta \left( \frac{\ln T_n}{T_n} \right)^{1/2}$  et  $r_T = M \ln T_n$  entraînent

$$P \left( \left| \hat{f}_{T_n}^H(x) - E \hat{f}_{T_n}^H(x) \right| > \epsilon_{T_n} \right) \leq \frac{8(1+\kappa)}{\eta\kappa} \alpha_0 \frac{T_n^{1/2} e^{-M \ln T_n \ln(1/\varrho)}}{h_{T_n} (\ln T_n)^{1/2}} \times (1 + o(1)) \\ + 4 \exp \left( - \frac{\eta^2 \ln T_n}{4(1+\kappa)^2 d_i(x) (1 + o(1)) + \frac{8}{3}(1+\kappa) M (\ln T_n)^{3/2} T_n^{-1/2} h_{T_n}^{-1}} \right), \quad i = 2, 3.$$

La condition sur  $h_{T_n}$  devient :  $(\ln T_n)^{3/2} / (T_n^{1/2} h_{T_n}) \rightarrow 0$  quand  $T_n \uparrow \infty$ . Cela implique que

$$\exists \varsigma_2 > 0, \exists T_N, \forall T_n \geq T_N, \quad h_{T_n} \geq \frac{(\ln T_n)^{3/2}}{\varsigma_2 T_n^{1/2}},$$

tel que

$$\forall \eta > 2(1+\kappa) \sqrt{d_i(x)}, \quad i = 2, 3, \quad \forall M > \frac{2}{\ln(1/\varrho)}, \quad \exists \lambda_1 > 0, \quad \exists \lambda_{2,i} > 1, \quad i = 2, 3,$$

où

$$\lambda_{2,i} = \min \left( \frac{\eta^2}{4(1+\kappa)^2 d_i(x)}; M \ln(1/\varrho) - 1 \right), \quad i = 2, 3,$$

avec

$$d_2(x) = 2 \int_0^\infty g_u(x, x) du \quad \text{et} \quad d_3(x) = 2\varphi(x),$$

et tel que

$$P \left( \left| \hat{f}_{T_n}^H(x) - \mathbb{E} \hat{f}_{T_n}^H(x) \right| > \eta \left\{ \frac{\ln T_n}{T_n} \right\}^{1/2} \right) \leq \lambda_1 T_n^{-\lambda_2}.$$

QUATRIÈME ÉTAPE : Le choix initial de la suite  $(T_n)$  telle que  $T_{n+1} - T_n \geq \tau > 0$  implique que  $\sum_{n=1}^\infty T_n^{-\lambda_2} < \infty$  et nos résultats découlent du lemme de Borel-Cantelli (cf. lemme 1.10). ■

On posera ensuite  $\Psi_1(T_n) := \left( \frac{T_n h_{T_n}^d}{\ln T_n} \right)^{1/2}$  et  $\Psi_2(T_n) := \left( \frac{T_n}{\ln T_n} \right)^{1/2}$ .

### Preuve du corollaire 1.9

La preuve est une conséquence directe du théorème 1.6, du lemme 1.2 et de l'inégalité triangulaire :

$$\min \left\{ \Psi_j(T_n), h_{T_n}^{-1} \right\} \left| \hat{f}_{T_n}^H(x) - f(x) \right| \leq \Psi_j(T_n) \left| \hat{f}_{T_n}^H(x) - \mathbb{E} \hat{f}_{T_n}^H(x) \right| + h_{T_n}^{-1} \left| \mathbb{E} \hat{f}_{T_n}^H(x) - f(x) \right|.$$

De manière générale, on a donc

$$\left| \hat{f}_{T_n}^H(x) - f(x) \right| = \mathcal{O} \left( \max \left\{ \Psi_j^{-1}(T_n), h_{T_n} \right\} \right) \quad p.s..$$

La partie (i) du corollaire découle ensuite du choix  $h_{T_n} \sim (T_n^{-1} \ln T_n)^{1/(d+2)}$  pour  $j = 1$  (cas optimal). Relaxons la condition portant sur la suite  $(h_{T_n})$  telle que  $T_n h_{T_n}^2 / (\ln T_n)^3 \rightarrow c$ ,  $0 < c < \infty$ ; celle-ci conserve encore la vitesse de convergence du théorème 1.6 avec une constante asymptotique dégradée. La partie (ii) découle du choix  $h_{T_n} \sim (T_n^{-1} (\ln T_n)^3)^{1/2}$  pour  $j = 2$  (cas suroptimal). Dans ce cas, la vitesse de convergence de l'histogramme est alors donnée par le terme de biais. ■

### Preuve du lemme 1.6

Similaire à celle du lemme 1.2. ■

### Preuve du théorème 1.7

Chaque sous-ensemble compact  $D$  de  $\mathbb{R}^d$  peut être couvert par  $M_T^d$  hypercubes  $\pi_{T_1}^*, \dots, \pi_{TM_T^d}^*$  de la partition  $\Pi_T$  qui sert à définir  $\hat{f}_T^H$  et on notera  $x_j$  un point de  $\pi_{T_j}^*$ . Pour tout  $\epsilon > 0$ , on a

$$\begin{aligned} P \left( \sup_{x \in D} \left| \hat{f}_{T_n}^H(x) - \mathbb{E} \hat{f}_{T_n}^H(x) \right| > \epsilon \right) &\leq P \left( \max_{1 \leq j \leq M_T^d} \left| \hat{f}_{T_n}^H(x_j) - \mathbb{E} \hat{f}_{T_n}^H(x_j) \right| > \epsilon \right) \\ &\leq \sum_{j=1}^{M_T^d} P \left( \left| \hat{f}_{T_n}^H(x_j) - \mathbb{E} \hat{f}_{T_n}^H(x_j) \right| > \epsilon \right), \end{aligned}$$

où

$$P\left(\left|\hat{f}_{T_n}^H(x_j) - \mathbb{E}\hat{f}_{T_n}^H(x_j)\right| > \epsilon\right) \leq P\left(\left|S'_{T_n}(x_j)\right| > \frac{T_n\epsilon}{2}\right) + P\left(\left|S''_{T_n}(x_j)\right| > \frac{T_n\epsilon}{2}\right).$$

Comme précédemment, les deux termes de la majoration se traitent de la même façon à partir de la décomposition

$$P\left(\left|S'_{T_n}(x_j)\right| > \frac{T_n\epsilon}{2}\right) \leq P\left(\left|\sum_{k=1}^{\lfloor q_T \rfloor + 1} V_{2k-1}(x_j) - W_{2k-1}(x_j)\right| > \frac{T_n\epsilon\kappa}{2(1+\kappa)}\right) \\ + P\left(\left|\sum_{k=1}^{\lfloor q_T \rfloor + 1} W_{2k-1}(x_j)\right| > \frac{T_n\epsilon}{2(1+\kappa)}\right),$$

avec  $\kappa > 0$  et  $V_k(x_j) := \frac{1}{h_{T_n}^d} \int_{(k-1)r_T}^{kr_T} [\mathbb{1}_{\pi_{T_j}^*}(X_t) - \mathbb{E}\mathbb{1}_{\pi_{T_j}^*}(X_t)] dt$  et, sous nos hypothèses renforcées  $R^*(i)$ ,  $H_1^*$  et  $H_2^*$ , nous établissons que

$$P\left(\left|\hat{f}_{T_n}^H(x_j) - \mathbb{E}\hat{f}_{T_n}^H(x_j)\right| > \epsilon_{T_n}\right) \leq \frac{8(1+\kappa)}{\kappa} \alpha_0 \frac{e^{-r_T \ln(1/\varrho)}}{\epsilon_{T_n} h_{T_n}^d} \times (1 + o(1)) \\ + 4 \exp\left(-\frac{T_n\epsilon^2}{8(1+\kappa)^2 \left(\frac{q_T+1}{T_n}\right) R(T_n) + \frac{8}{3}(1+\kappa) \frac{r_T}{h_{T_n}^d} \epsilon_{T_n}}\right),$$

où  $R(T_n)$  est l'analogue uniforme de  $R_x(T_n)$  :

$$R(T_n) = \begin{cases} \frac{r_T}{h_{T_n}^d} \cdot d_1 \times (1 + o(1)), & \text{cas "optimal"}; \\ r_T \cdot d_2 \times (1 + o(1)), & \text{cas "suroptimal" avec } H_2^*(ii), \end{cases}$$

où  $o(1)$  est uniforme en  $x$  et avec  $d_1 := 2u_0M_1$  et  $d_2 := 2\|\varphi\|_\infty$ .

Sous les conditions  $H_1^*$  d'"optimalité" : les choix  $\epsilon_{T_n} = \eta\Psi_1^{-1}(T_n)$  et  $r_T = M \ln T_n$  entraînent

$$P\left(\Psi_1(T_n) \left|\hat{f}_{T_n}^H(x_j) - \mathbb{E}\hat{f}_{T_n}^H(x_j)\right| > \eta\right) \leq \frac{8(1+\kappa)}{\eta\kappa} \alpha_0 \frac{T_n^{1/2} e^{-M \ln T_n \ln(1/\varrho)}}{h_{T_n}^{d/2} (\ln T_n)^{1/2}} \times (1 + o(1)) \\ + 4 \exp\left(-\frac{\eta^2 \ln T_n}{4(1+\kappa)^2 d_1 (1 + o(1)) + \frac{8}{3}(1+\kappa) M (\ln T_n)^{3/2} T_n^{-1/2} h_{T_n}^{-d/2}}\right). \quad (1.15)$$

Clairement, on a

$$M_T^d \geq \left\lceil \frac{\lambda_d(D)}{h_{T_n}^d} \right\rceil,$$

où  $\lceil x \rceil$  désigne le plus petit entier qui est strictement plus grand que le réel  $x$ , et on écrira ensuite

$$M_T^d = c_d h_{T_n}^{-d}, \quad 0 < c_d < \infty. \quad (1.16)$$

La condition sur  $h_{T_n}$  est  $(\ln T_n)^3 / (T_n h_{T_n}^d) \rightarrow 0$  quand  $T_n \uparrow \infty$ , ce qui implique

$$\exists \varsigma_1 > 0, \exists T_N, \forall T_n \geq T_N, \quad h_{T_n}^{-d} \leq \varsigma_1 \frac{T_n}{(\ln T_n)^3}. \quad (1.17)$$

Dans la suite,  $c$  et  $c'$  désignent deux constantes positives universelles dont on ne précisera pas les valeurs afin de simplifier les écritures. À partir de (1.15) et pour la valeur de  $M_T^d$  donnée par (1.16), on obtient

$$\begin{aligned} P\left(\sup_{x \in D} \Psi_1(T_n) \left| \hat{f}_{T_n}^H(x) - \mathbb{E} \hat{f}_{T_n}^H(x) \right| > \eta\right) &\leq \sum_{j=1}^{M_T^d} P\left(\Psi_1(T_n) \left| \hat{f}_{T_n}^H(x_j) - \mathbb{E} \hat{f}_{T_n}^H(x_j) \right| > \eta\right) \\ &\leq c \cdot (h_{T_n}^{3d} \ln T_n)^{-1/2} \cdot T_n^{-(M \ln(1/\varrho) - \frac{1}{2})} + c' \cdot h_{T_n}^{-d} \cdot T_n^{-\left(\frac{\eta^2}{4(1+\kappa)^2 d_1}\right)}. \end{aligned}$$

D'après (1.17) :

$$(h_{T_n}^{3d} \ln T_n)^{-1/2} \leq \varsigma_1^{3/2} \frac{T_n^{3/2}}{(\ln T_n)^5} \quad \text{et} \quad h_{T_n}^{-d} \leq \varsigma_1 \frac{T_n}{(\ln T_n)^3},$$

et on a

$$\begin{aligned} P\left(\sup_{x \in D} \Psi_1(T_n) \left| \hat{f}_{T_n}^H(x) - \mathbb{E} \hat{f}_{T_n}^H(x) \right| > \eta\right) \\ \leq c \cdot \frac{1}{(\ln T_n)^5} \cdot T_n^{-(M \ln(1/\varrho) - 2)} + c' \cdot \frac{1}{(\ln T_n)^3} \cdot T_n^{-\left(\frac{\eta^2}{4(1+\kappa)^2 d_1} - 1\right)}. \end{aligned}$$

Ainsi pour  $\eta > 2\sqrt{2}(1+\kappa)d_1^{1/2}$  et  $M > 3/\ln(1/\varrho)$ , la série de l'expression majorante converge et on peut conclure avec le lemme de Borel-Cantelli.

*Sous les conditions  $H_2^*$  de "suroptimalité" : les choix  $\epsilon_{T_n} = \eta \Psi_2^{-1}(T_n)$  et  $r_T = M \ln T_n$  entraînent*

$$\begin{aligned} P\left(\Psi_2(T_n) \left| \hat{f}_{T_n}^H(x_j) - \mathbb{E} \hat{f}_{T_n}^H(x_j) \right| > \eta\right) &\leq \frac{8(1+\kappa)}{\eta \kappa} \alpha_0 \frac{T_n^{1/2} e^{-M \ln T_n \ln(1/\varrho)}}{h_{T_n} (\ln T_n)^{1/2}} \times (1 + o(1)) \\ &+ 4 \exp\left(-\frac{\eta^2 \ln T_n}{4(1+\kappa)^2 d_2 (1 + o(1)) + \frac{8}{3}(1+\kappa) M (\ln T_n)^{3/2} T_n^{-1/2} h_{T_n}^{-1}}\right). \quad (1.18) \end{aligned}$$

La condition sur  $h_{T_n}$  est  $(\ln T_n)^{3/2} / (T_n^{1/2} h_{T_n}) \rightarrow 0$  quand  $T_n \uparrow \infty$ , ce qui implique

$$\exists \varsigma_2 > 0, \exists T_N, \forall T_n \geq T_N, \quad h_{T_n}^{-1} \leq \varsigma_2 \frac{T_n^{1/2}}{(\ln T_n)^{3/2}}. \quad (1.19)$$

À partir de (1.18) et pour la valeur de  $M_T$  donnée par (1.16), on obtient

$$\begin{aligned} P\left(\sup_{x \in D} \Psi_2(T_n) \left| \hat{f}_{T_n}^H(x) - \mathbb{E} \hat{f}_{T_n}^H(x) \right| > \eta\right) &\leq \sum_{j=1}^{M_T} P\left(\Psi_2(T_n) \left| \hat{f}_{T_n}^H(x_j) - \mathbb{E} \hat{f}_{T_n}^H(x_j) \right| > \eta\right) \\ &\leq c \cdot (h_{T_n}^4 \ln T_n)^{-1/2} \cdot T_n^{-(M \ln(1/\varrho) - \frac{1}{2})} + c' \cdot h_{T_n}^{-1} \cdot T_n^{-\left(\frac{\eta^2}{4(1+\kappa)^2 d_2}\right)}. \end{aligned}$$

D'après (1.19) :

$$(h_{T_n}^4 \ln T_n)^{-1/2} \leq \varsigma_2^2 \frac{T_n}{(\ln T_n)^{7/2}} \quad \text{et} \quad h_{T_n}^{-1} \leq \varsigma_2 \frac{T_n^{1/2}}{(\ln T_n)^{3/2}},$$

et on a

$$\begin{aligned} P \left( \sup_{x \in D} \Psi_2(T_n) \left| \hat{f}_{T_n}^H(x) - \mathbb{E} \hat{f}_{T_n}^H(x) \right| > \eta \right) \\ \leq c \cdot \frac{1}{(\ln T_n)^{7/2}} \cdot T_n^{-(M \ln(1/\varrho) - \frac{3}{2})} + c' \cdot \frac{1}{(\ln T_n)^{3/2}} \cdot T_n^{-\left(\frac{\eta^2}{4(1+\kappa)^2 d_2} - \frac{1}{2}\right)}. \end{aligned}$$

Ainsi pour  $\eta > \sqrt{6}(1 + \kappa)d_2^{1/2}$  et  $M > 5/(2 \ln(1/\varrho))$ , la série de l'expression majorante converge et on peut conclure avec le lemme de Borel-Cantelli. ■

### Preuve du corollaire 1.10

La preuve est une conséquence directe du théorème 1.7, du lemme 1.6 et de la relation :

$$\sup_{x \in D} \Psi_j(T_n) \left| \hat{f}_{T_n}^H(x) - f(x) \right| \leq \sup_{x \in D} \Psi_j(T_n) \left| \hat{f}_{T_n}^H(x) - \mathbb{E} \hat{f}_{T_n}^H(x) \right| + \sup_{x \in D} h_{T_n}^{-1} \left| \mathbb{E} \hat{f}_{T_n}^H(x) - f(x) \right|.$$

Les choix du paramètre  $h_{T_n}$  sont donnés dans la preuve du corollaire 1.9. ■

### Preuve du théorème 1.8

La preuve est similaire à celle donnée pour le théorème 1.7. Chaque compact  $D_T$  peut être lui-même couvert par  $M_T^d$  hypercubes  $\pi_{T_1}^*, \dots, \pi_{TM_T^d}^*$  de la partition  $\Pi_T$ . Comme la mesure de Lebesgue de l'hyperboule  $D_T$  de rayon  $T_n^a$  est de l'ordre  $T_n^{ad}$ , on pourra écrire

$$M_T^d = c_d \frac{T_n^{ad}}{h_{T_n}^d}, \quad 0 < c_d < \infty. \quad (1.20)$$

**Remarque 1.16.** Par exemple, dans le cas de la norme euclidienne, si  $\|x\| = (x_1^2 + \dots + x_d^2)^{1/2}$  pour  $x = (x_1, \dots, x_d)' \in \mathbb{R}^d$ , la mesure de Lebesgue de l'hyperboule  $D_T$  de rayon  $T_n^a$  vaut  $\frac{\pi^{d/2}}{\Gamma(\frac{d}{2}+1)} T_n^{ad}$  avec  $\Gamma(n) = (n-1)!$  et  $\Gamma(t) = \int_0^\infty e^{-t} t^{x-1} dt$ . Pour un autre choix de norme, par exemple si  $\|x\| = \max_{1 \leq i \leq d} |x_i|$ ,  $D_T$  est alors un hypercube centré à l'origine de volume  $2^d T_n^{ad}$ .

*Sous les conditions  $H_1^*$  d'“optimalité”* : en reprenant l'expression (1.15) avec la valeur de  $M_T^d$  donnée par (1.20), on obtient

$$\begin{aligned} P \left( \sup_{x \in D_T} \Psi_1(T_n) \left| \hat{f}_{T_n}^H(x) - \mathbb{E} \hat{f}_{T_n}^H(x) \right| > \eta \right) \leq \sum_{j=1}^{M_T^d} P \left( \Psi_1(T_n) \left| \hat{f}_{T_n}^H(x_j) - \mathbb{E} \hat{f}_{T_n}^H(x_j) \right| > \eta \right) \\ \leq c \cdot (h_{T_n}^{3d} \ln T_n)^{-1/2} \cdot T_n^{-(M \ln(1/\varrho) - ad - \frac{1}{2})} + c' \cdot h_{T_n}^{-d} \cdot T_n^{-\left(\frac{\eta^2}{4(1+\kappa)^2 d_1} - ad\right)}. \end{aligned}$$

La condition (1.17) implique alors

$$\begin{aligned} P \left( \sup_{x \in D_T} \Psi_1(T_n) \left| \hat{f}_{T_n}^H(x) - \mathbb{E} \hat{f}_{T_n}^H(x) \right| > \eta \right) \\ \leq c \cdot \frac{1}{(\ln T_n)^5} \cdot T_n^{-(M \ln(1/\varrho) - ad - 2)} + c' \cdot \frac{1}{(\ln T_n)^3} \cdot T_n^{-\left(\frac{\eta^2}{4(1+\kappa)^2 d_1} - ad - 1\right)}, \end{aligned}$$

et pour  $\eta > 2(1+\kappa)\sqrt{(ad+2)d_1}$  et  $M > (ad+3)/\ln(1/\varrho)$ , on peut conclure avec le lemme de Borel-Cantelli.

Sous les conditions  $H_2^*$  de "suroptimalité" : en reprenant l'expression (1.18) avec la valeur de  $M_T$  donnée par (1.20), on obtient

$$\begin{aligned} P\left(\sup_{x \in D_T} \Psi_2(T_n) \left| \hat{f}_{T_n}^H(x) - \mathbb{E} \hat{f}_{T_n}^H(x) \right| > \eta\right) &\leq \sum_{j=1}^{M_T} P\left(\Psi_2(T_n) \left| \hat{f}_{T_n}^H(x_j) - \mathbb{E} \hat{f}_{T_n}^H(x_j) \right| > \eta\right) \\ &\leq c \cdot (h_{T_n}^4 \ln T_n)^{-1/2} \cdot T_n^{-(M \ln(1/\varrho) - a - \frac{1}{2})} + c' \cdot h_{T_n}^{-1} \cdot T_n^{-\left(\frac{\eta^2}{4(1+\kappa)^2 d_2} - a\right)}. \end{aligned}$$

La condition (1.19) implique alors

$$\begin{aligned} P\left(\sup_{x \in D_T} \Psi_2(T_n) \left| \hat{f}_{T_n}^H(x) - \mathbb{E} \hat{f}_{T_n}^H(x) \right| > \eta\right) \\ \leq c \cdot \frac{1}{(\ln T_n)^{7/2}} \cdot T_n^{-(M \ln(1/\varrho) - a - \frac{3}{2})} + c' \cdot \frac{1}{(\ln T_n)^{3/2}} \cdot T_n^{-\left(\frac{\eta^2}{4(1+\kappa)^2 d_2} - a - \frac{1}{2}\right)}, \end{aligned}$$

et pour  $\eta > 2(1+\kappa)\sqrt{(a + \frac{3}{2})d_2}$  et  $M > (a + \frac{5}{2})/\ln(1/\varrho)$ , on peut conclure avec le lemme de Borel-Cantelli. ■

### Preuve du corollaire 1.11

La preuve est une conséquence directe du théorème 1.8, du lemme 1.6 et de la relation :

$$\sup_{x \in D_T} \Psi_j(T_n) \left| \hat{f}_{T_n}^H(x) - f(x) \right| \leq \sup_{x \in D_T} \Psi_j(T_n) \left| \hat{f}_{T_n}^H(x) - \mathbb{E} \hat{f}_{T_n}^H(x) \right| + \sup_{x \in \mathbb{R}^d} h_{T_n}^{-1} \left| \mathbb{E} \hat{f}_{T_n}^H(x) - f(x) \right|.$$

Les choix du paramètre  $h_{T_n}$  sont donnés dans la preuve du corollaire 1.9. ■

### Preuve du corollaire 1.12

Partons de la décomposition :

$$\begin{aligned} \sup_{x \in \mathbb{R}^d} \Psi_j(T_n) \left| \hat{f}_{T_n}^H(x) - f(x) \right| &\leq \sup_{\|x\| \leq T_n^a} \Psi_j(T_n) \left| \hat{f}_{T_n}^H(x) - f(x) \right| \\ &\quad + \sup_{\|x\| > T_n^a} \Psi_j(T_n) \hat{f}_{T_n}^H(x) + \sup_{\|x\| > T_n^a} \Psi_j(T_n) f(x). \end{aligned}$$

L'étude du premier terme est donnée par le corollaire 1.11. Pour un  $\varepsilon > 0$ , on a

$$\left\{ \sup_{0 \leq t \leq T_n} \|X_t\| \leq T_n^a - \varepsilon, \|x\| > T_n^a \right\} \subset \left\{ \forall t \in [0, T_n], \|x - X_t\| > \varepsilon \right\}.$$

Remarquons qu'il existe un réel positif  $T_N$  tel que pour tout  $T_n \geq T_N$  :  $h_{T_n} < \varepsilon$  et

$$\left\{ \sup_{0 \leq t \leq T_n} \|X_t\| \leq T_n^a - \varepsilon, \|x\| > T_n^a \right\} \subset \left\{ \sup_{\|x\| > T_n^a} \hat{f}_{T_n}^H(x) = 0 \right\}.$$

Ainsi pour tout  $\eta > 0$  et  $T_n \geq T_0$  :

$$\begin{aligned} P\left(\sup_{\|x\|>T_n^a} \Psi_j(T_n) \hat{f}_{T_n}^H(x) \geq \eta\right) &\leq P\left(\sup_{0 \leq t \leq T_n} \|X_t\| > T_n^a - \varepsilon\right) \\ &\leq P\left(\max_{0 \leq j \leq \lfloor T_n \rfloor} \sup_{j \leq t < j+1} \|X_t\| > T_n^a - \varepsilon\right). \end{aligned}$$

Nous utilisons maintenant l'inégalité de Markov et l'hypothèse de stricte stationnarité  $H_3^*$  :

$$P\left(\sup_{\|x\|>T_n^a} \Psi_j(T_n) \hat{f}_{T_n}^H(x) \geq \eta\right) \leq \lfloor T_n \rfloor \cdot \frac{\mathbb{E}(\sup_{0 \leq t < 1} \|X_t\|)}{T_n^a - \varepsilon} = \mathcal{O}\left(\frac{1}{T_n^{a-1}}\right).$$

On peut alors utiliser le lemme de Borel-Cantelli pour  $a > 2$  :

$$\limsup_{T_n \uparrow \infty} \sup_{\|x\|>T_n^a} \Psi_j(T_n) \hat{f}_{T_n}^H(x) = 0 \quad p.s., \quad j = 1, 2.$$

D'autre part, on a

$$\Psi_j(T_n) \sup_{\|x\|>T_n^a} f(x) \leq \frac{\Psi_j(T_n)}{T_n^a} \cdot \|x\| \sup_{\|x\|>T_n^a} f(x).$$

Comme  $\lim_{\|u\| \rightarrow \infty} \|u\| f(u) = 0$ ,

$$\|x\| \sup_{\|x\|>T_n^a} f(x) \xrightarrow{T_n \uparrow \infty} 0.$$

De plus, pour tout  $a > 2$  avec les choix précédents de  $h_{T_n}$ , on a  $\Psi_j(T_n) T_n^{-a} \rightarrow 0$  quand  $T_n \uparrow \infty$ , ce qui termine la preuve.  $\blacksquare$

### Preuve du lemme 1.7

Sans perte de généralité, on admettra ici que  $x \in \pi_{T_j}^+ := \pi_{T_j} \cap [c_j, c_{j+1}[$ . Dans ce cas,

$$\hat{f}_T^{PF}(x) = \left(\frac{x - c_j}{h_T}\right) \hat{f}_{j+1} + \left(\frac{c_{j+1} - x}{h_T}\right) \hat{f}_j.$$

D'où

$$\begin{aligned} \mathbb{E}(\hat{f}_T^{PF}(x)) - f(x) &= \left(\frac{x - c_j}{h_T}\right) \frac{1}{h_T} \int_{\pi_{T_{j+1}}} f(u) du + \left(\frac{c_{j+1} - x}{h_T}\right) \frac{1}{h_T} \int_{\pi_{T_{j+1}}} f(u) du \\ &\quad - \left(\frac{c_{j+1} - x}{h_T} + \frac{x - c_j}{h_T}\right) f(x) \\ &= \left(\frac{x - c_j}{h_T}\right) \frac{1}{h_T} \int_{\pi_{T_{j+1}}} (f(u) - f(x)) du \\ &\quad + \left(\frac{c_{j+1} - x}{h_T}\right) \frac{1}{h_T} \int_{\pi_{T_{j+1}}} (f(u) - f(x)) du. \end{aligned}$$

En utilisant l'hypothèse lipschitzienne de continuité  $R^*(i)$ , on a alors

$$\begin{aligned} \left|\mathbb{E}(\hat{f}_T^{PF}(x)) - f(x)\right| &\leq \frac{1}{2} \frac{1}{h_T} \int_{\pi_{T_{j+1}}} |f(u) - f(x)| du + \frac{1}{h_T} \int_{\pi_{T_j}} |f(u) - f(x)| du \\ &\leq \frac{1}{2h_T} \int_{\pi_{T_{j+1}}} \underbrace{K_f \|u - x\|}_{\leq C'h_T} du + \frac{1}{h_T} \int_{\pi_{T_j}} \underbrace{K_f \|u - x\|}_{\leq C'h_T} du \leq C''h_T, \end{aligned}$$

où  $C'$  et  $C''$  désignent des constantes positives, ce qui prouve que

$$\limsup_{T \rightarrow \infty} h_T^{-1} \cdot \left| \mathbb{E} \hat{f}_T^{PF}(x) - f(x) \right| < \infty,$$

et *a fortiori* que

$$\limsup_{T \rightarrow \infty} h_T^{-1} \cdot \sup_{x \in \Lambda} \left| \mathbb{E} \hat{f}_T^{PF}(x) - f(x) \right| < \infty,$$

pour  $\Lambda = \nu_x$ ,  $D$  et  $\mathbb{R}$ .

Si on admet que  $f \in \mathcal{C}_b^2(\Lambda)$ , nous pouvons améliorer l'ordre de convergence du biais. Un développement de Taylor de  $f$  à l'ordre 2 au voisinage de  $m_j$  donne

$$f(x) = f(m_j) + (x - m_j)f'(m_j) + \frac{1}{2}(x - m_j)^2 f''(m_j + \theta(x - m_j)), \quad \theta \in ]0, 1[,$$

et on a

$$\mathbb{E}(\hat{f}_j) = f(m_j) - \frac{f'(m_j)}{2} h_T + \frac{1}{h_T} \int_{\pi_{Tj}} \frac{1}{2} (u - m_j)^2 f''(m_j + \theta(u - m_j)) du,$$

$$\mathbb{E}(\hat{f}_{j+1}) = f(m_j) + \frac{f'(m_j)}{2} h_T + \frac{1}{h_T} \int_{\pi_{Tj+1}} \frac{1}{2} (u - m_j)^2 f''(m_j + \theta(u - m_j)) du.$$

D'où

$$\begin{aligned} \mathbb{E}(\hat{f}_T^{PF}(x)) - f(x) &= \left( \frac{x - c_j}{h_T} \right) \frac{1}{h_T} \int_{\pi_{Tj+1}} \frac{1}{2} (u - m_j)^2 f''(m_j + \theta(u - m_j)) du \\ &\quad + \left( \frac{c_{j+1} - x}{h_T} \right) \frac{1}{h_T} \int_{\pi_{Tj}} \frac{1}{2} (u - m_j)^2 f''(m_j + \theta(u - m_j)) du \\ &\quad - \frac{1}{2} (x - m_j)^2 f''(m_j + \theta(x - m_j)). \end{aligned}$$

De là,

$$\left| \mathbb{E}(\hat{f}_T^{PF}(x)) - f(x) \right| \leq \left[ \frac{1}{2} + 1 + 1 \right] \times \frac{1}{2} \left( \frac{h_T}{2} \right)^2 \sup_{x \in \Lambda} |f''(x)|,$$

et

$$h_T^{-2} \cdot \left| \mathbb{E}(\hat{f}_T^{PF}(x)) - f(x) \right| \leq \frac{5}{16} \sup_{x \in \Lambda} |f''(x)|. \quad \blacksquare$$

### Preuve du corollaire 1.13

Pour  $x \in \pi_{Tj}^+ := \pi_{Tj} \cap [c_j, c_{j+1}[$ , on pose maintenant

$$Z_t := \left( \frac{x - c_j}{h_T} \right) \frac{1}{h_{T_n}} (\mathbb{1}_{\pi_{Tj+1}}(X_t) - p_{j+1}) + \left( \frac{c_{j+1} - x}{h_T} \right) \frac{1}{h_{T_n}} (\mathbb{1}_{\pi_{Tj}}(X_t) - p_j),$$

et on redéfinit  $(2\lfloor q_T \rfloor + 1)$  blocs comme suit :

$$S'_{T_n}(x) := \sum_{k=1}^{\lfloor q_T \rfloor + 1} V_{T_n}(2k - 1), \quad S''_{T_n}(x) := \sum_{k=1}^{\lfloor q_T \rfloor} V_{T_n}(2k),$$

où  $V_{T_n}(k) := \int_{(k-1)r_T}^{kr_T} Z_t dt$ ,  $k = 1, \dots, 2\lfloor q_T \rfloor$ , et  $V_{T_n}(2\lfloor q_T \rfloor + 1) = \int_{2\lfloor q_T \rfloor r_T}^{2q_T r_T} Z_t dt$ , de telle sorte que

$$\frac{1}{T_n} (S'_{T_n}(x) + S''_{T_n}(x)) = \hat{f}_{T_n}^{PF}(x) - \mathbb{E} \hat{f}_{T_n}^{PF}(x).$$

Pour tout  $\epsilon > 0$  et tout  $\kappa > 0$ , on a

$$\begin{aligned} P \left( |S'_{T_n}(x)| > \frac{T_n \epsilon}{2} \right) &\leq P \left( \left| \sum_{k=1}^{\lfloor q_T \rfloor + 1} V_{T_n}(2k-1) - W_{T_n}(2k-1) \right| > \frac{T_n \epsilon \kappa}{2(1+\kappa)} \right) \\ &\quad + P \left( \left| \sum_{k=1}^{\lfloor q_T \rfloor + 1} W_{T_n}(2k-1) \right| > \frac{T_n \epsilon}{2(1+\kappa)} \right) =: \text{I} + \text{II}. \end{aligned}$$

La majoration

$$\max_{1 \leq k \leq 2\lfloor q_T \rfloor} |V_{T_n}(k)| \leq \frac{r_T}{h_{T_n}} \quad \text{et} \quad |V_{T_n}(2\lfloor q_T \rfloor + 1)| \leq 2 \frac{r_T}{h_{T_n}}, \quad (1.21)$$

est encore valable et comme  $X$  est géométriquement fortement mélangeant, nous avons

$$\text{I} \leq \frac{4(1+\kappa)}{\kappa} \alpha_0 \frac{e^{-r_T \log(1/\varrho)}}{\epsilon h_{T_n}} \times (1 + o(1)).$$

Avec (1.21) et l'inégalité de Bernstein, on a encore

$$\text{II} \leq 2 \exp \left( - \frac{T_n^2 \epsilon^2}{8(1+\kappa)^2 \sum_{k=1}^{\lfloor q_T \rfloor + 1} \mathbb{E}(W_{T_n}^2(2k-1)) + \frac{8}{3}(1+\kappa) \frac{r_T}{h_{T_n}} T_n \epsilon} \right).$$

Nous pouvons ensuite reprendre les mêmes majorations de  $\mathbb{E}(W_n^2(2k-1)) = \mathbb{E}(V_{T_n}^2(2k-1))$  que celles établies dans l'étude de l'histogramme. En effet,

$$\begin{aligned} \mathbb{E}(V_{T_n}^2(2k-1)) &= \left( \frac{x - c_j}{h_T} \right)^2 \cdot \frac{1}{h_{T_n}^2} \iint_{[0, r_T]^2} \text{Cov}(\mathbb{1}_{\pi_{T_{j+1}}}(X_s), \mathbb{1}_{\pi_{T_{j+1}}}(X_t)) ds dt \\ &\quad + \left( \frac{c_{j+1} - x}{h_T} \right)^2 \cdot \frac{1}{h_{T_n}^2} \iint_{[0, r_T]^2} \text{Cov}(\mathbb{1}_{\pi_{T_j}}(X_s), \mathbb{1}_{\pi_{T_j}}(X_t)) ds dt \\ &\quad + 2 \left( \frac{x - c_j}{h_T} \right) \left( \frac{c_{j+1} - x}{h_T} \right) \cdot \frac{1}{h_{T_n}^2} \iint_{[0, r_T]^2} \text{Cov}(\mathbb{1}_{\pi_{T_j}}(X_s), \mathbb{1}_{\pi_{T_{j+1}}}(X_t)) ds dt. \end{aligned}$$

Sous les conditions  $H_1$  d'“optimalité”, on a

$$\begin{aligned} r_T^{-1} h_{T_n} \cdot \mathbb{E}(V_{T_n}^2(2k-1)) &\leq \left\{ \left( \frac{x - c_j}{h_T} \right)^2 + \left( \frac{c_{j+1} - x}{h_T} \right)^2 + 2 \left( \frac{x - c_j}{h_T} \right) \left( \frac{c_{j+1} - x}{h_T} \right) \right\} \times 2u_0 f(x) \times (1 + o(1)) \\ &\leq 2u_0 f(x) \times (1 + o(1)). \end{aligned}$$

Sous les conditions  $H_2$  de “suroptimalité” avec  $H_2(ii)$ , on a

$$\begin{aligned} r_T^{-1} \cdot \mathbb{E}(V_{T_n}^2(2k-1)) &= \left\{ \left( \frac{x - c_j}{h_T} \right)^2 + \left( \frac{c_{j+1} - x}{h_T} \right)^2 + 2 \left( \frac{x - c_j}{h_T} \right) \left( \frac{c_{j+1} - x}{h_T} \right) \right\} \\ &\quad \times 2 \int_0^\infty g_u(x, x) du \times (1 + o(1)) \\ &= 2 \int_0^\infty g_u(x, x) du \times (1 + o(1)). \end{aligned}$$

Sous les conditions  $H_2$  de “suroptimalité” avec  $H_2(ii)'$ , on a

$$\begin{aligned} & r_T^{-1} \cdot \mathbb{E}(V_{T_n}^2(2k-1)) \\ & \leq \left\{ \left( \frac{x-c_j}{h_T} \right)^2 + \left( \frac{c_{j+1}-x}{h_T} \right)^2 + 2 \left( \frac{x-c_j}{h_T} \right) \left( \frac{c_{j+1}-x}{h_T} \right) \right\} \times 2\varphi(x) \times (1+o(1)) \\ & \leq 2\varphi(x) \times (1+o(1)). \end{aligned}$$

La suite de la preuve est ensuite rigoureusement identique à celle donnée pour l'histogramme. ■

### Preuve du corollaire 1.14

Sous l'hypothèse  $R^*(ii)'$  et avec l'inégalité triangulaire, on a

$$\begin{aligned} \min \left\{ \Psi_j(T_n), h_{T_n}^{-2} \right\} \cdot \sup_{x \in \Lambda} \left| \hat{f}_{T_n}^{PF}(x) - f(x) \right| & \leq \Psi_j(T_n) \cdot \sup_{x \in \Lambda} \left| \hat{f}_{T_n}^{PF}(x) - \mathbb{E} \hat{f}_{T_n}^{PF}(x) \right| \\ & \quad + h_{T_n}^{-2} \cdot \sup_{x \in \Lambda} \left| \mathbb{E} \hat{f}_{T_n}^{PF}(x) - f(x) \right|. \end{aligned}$$

De là,

$$\sup_{x \in \Lambda} \left| \hat{f}_{T_n}^{PF}(x) - f(x) \right| = \mathcal{O} \left( \max \left\{ \Psi_j^{-1}(T_n), h_{T_n}^2 \right\} \right) \quad p.s.,$$

et les vitesses du polygone de fréquences découlent ensuite des choix  $h_{T_n} \sim (T_n^{-1} \ln T_n)^{1/5}$  pour  $j = 1$  (cas optimal) et  $h_{T_n} = o \left( (T_n^{-1} \ln T_n)^{1/4} \right)$  pour  $j = 2$  (cas suroptimal). ■

## Chapitre 2

# Régressogramme en temps continu

**Résumé.** – Dans ce chapitre, nous étudions les propriétés du régressogramme dont la construction est analogue à celle de l’histogramme pour estimer la fonction de régression associée à un processus bivarié à temps continu. Nous proposons une version “temps continu” du régressogramme et donnons les vitesses optimale et suroptimale de l’erreur ponctuelle en moyenne quadratique. L’estimation de la régression est notamment d’une grande utilité en théorie de la prévision et pour les nombreuses applications qui en découlent.

### 2.1 Cadre d’étude

Soit  $Z = \{Z_t = (X_t, Y_t), t \in \mathbb{R}\}$  un processus stochastique bivarié défini sur un espace de probabilité  $(\Omega, \mathcal{A}, P)$  et à valeurs dans  $\mathbb{R}^d \times \mathbb{R}^{d'}$ ; chaque  $Z_t$  ayant la même loi  $\mu'$  de densité  $f_Z(x, y)$  par rapport à la mesure de Lebesgue. Nous introduisons aussi une fonction borélienne  $m$  de  $\mathbb{R}^{d'}$  dans  $\mathbb{R}$  telle que  $(\omega, t) \mapsto m^2(Y_t(\omega))$  soit intégrable par rapport à la mesure produit  $P \otimes \lambda_T$  pour tout  $T$  positif ( $\lambda_T$  désignant la mesure de Lebesgue sur  $[0, T]$ ); cette fonction est connue et choisie par le statisticien. Nous nous proposons de construire un estimateur de la fonction de régression définie par

$$r : x \mapsto \mathbb{E}(m(Y_0) | X_0 = x),$$

qui minimise la distance entre  $m(Y_0)$  et  $X_0$  au sens des moindres carrés. Plus précisément, l’étude suivante porte sur l’estimation de la fonction  $r$  par le régressogramme à partir de l’observation complète d’une trajectoire du processus  $Z$  sur  $[0, T]$ . La connaissance du paramètre de régression  $r$  est notamment utile pour estimer les valeurs futures de  $m(Y)$  sachant que  $X = x$  et pour comprendre le lien entre les variables  $X$  et  $Y$ . Dans notre étude, nous aurons besoin des paramètres fonctionnels suivants :

$$f(x) = \int_{\mathbb{R}^{d'}} f_Z(x, y) dy, \quad x \in \mathbb{R}^d;$$
$$g(x) = \int_{\mathbb{R}^{d'}} m(y) f_Z(x, y) dy, \quad x \in \mathbb{R}^d,$$

tels que

$$r(x) = g(x)/f(x) \text{ si } f(x) > 0 \quad \text{et} \quad r(x) = \mathbb{E}(m(Y_0)) \text{ si } f(x) = 0,$$

$f$  étant la fonction de densité de  $X_t$  que l'on supposera ensuite continue et strictement positive sur un voisinage d'un point  $x$  de  $\mathbb{R}^d$ . Rappelons aussi pour la suite que le processus  $Z$  est fortement mélangeant relativement au coefficient défini par

$$\alpha(\mathcal{B}, \mathcal{C}) := \sup_{B \in \mathcal{B}, C \in \mathcal{C}} |P(B \cap C) - P(B)P(C)|,$$

si

$$\alpha_Z(u) := \sup_{t \in \mathbb{R}} \alpha\left(\sigma(Z_s, s \leq t), \sigma(Z_s, s \geq t+u)\right) \downarrow_{u \rightarrow \infty} 0,$$

ou relativement au coefficient de 2-alpha-mélange plus faible introduit par Bosq (1998) si

$$\alpha_Z^{(2)}(u) := \sup_{t \in \mathbb{R}} \alpha\left(\sigma(Z_t), \sigma(Z_{t+u})\right) \downarrow_{u \rightarrow \infty} 0.$$

**Construction de l'estimateur.** À partir de la partition  $\Pi_T$  de l'espace  $\mathbb{R}^d$  définie dans le chapitre 1, nous supposons pour tout  $x = (x_1, \dots, x_d)' \in \mathbb{R}^d$  l'existence d'un indice  $j$  tel que  $x \in \pi_{Tj}$ . Nous introduisons alors les estimateurs de type histogramme basés sur une observation de  $X$  sur  $[0, T]$  pour les paramètres fonctionnels  $f$  et  $g$  :

$$\begin{aligned} \hat{f}_T(x) &= \sum_j \left[ \frac{1}{Th_T^d} \int_0^T \mathbf{1}_{\pi_{Tj}}(X_t) dt \right] \mathbf{1}_{\pi_{Tj}}(x) =: \sum_j \hat{f}_j \mathbf{1}_{\pi_{Tj}}(x); \\ \hat{g}_T(x) &= \sum_j \left[ \frac{1}{Th_T^d} \int_0^T m(Y_t) \mathbf{1}_{\pi_{Tj}}(X_t) dt \right] \mathbf{1}_{\pi_{Tj}}(x) =: \sum_j \hat{g}_j \mathbf{1}_{\pi_{Tj}}(x). \end{aligned}$$

**Définition 2.1.** Au point  $x \in \mathbb{R}^d$  et relativement à la partition  $\Pi_T$ , le régressogramme est défini par

$$\begin{aligned} \hat{r}_T(x) &= \frac{\hat{g}_T(x)}{\hat{f}_T(x)} = \sum_j \frac{\hat{g}_j}{\hat{f}_j} \mathbf{1}_{\pi_{Tj}}(x) =: \sum_j \hat{r}_j \mathbf{1}_{\pi_{Tj}}(x) \quad \text{si } \hat{f}_T(x) > 0; \\ &= \frac{1}{T} \int_0^T m(Y_t) dt \quad \text{si } \hat{f}_T(x) = 0. \end{aligned}$$

La régression est ainsi estimée sur l'intervalle de temps  $[0, T]$  par la fonction en escalier qui, sur chaque élément  $\pi_{Tj}$ , est constante et égale à la valeur moyenne des  $m(Y_t)$  tels que  $X_t \in \pi_{Tj}$  divisée par le temps moyen passé par le processus  $X$  dans  $\pi_{Tj}$ . Dans la suite, pour s'affranchir du cas de figure où  $\hat{f}_T(x)$  s'annule, on étudiera une version modifiée, notée  $\hat{r}_T^\epsilon$ , de  $\hat{r}_T$  définie par

$$\hat{r}_T^\epsilon(x) = \frac{\hat{g}_T(x)}{\hat{f}_T(x) + \epsilon_T}, \quad x \in \mathbb{R}^d,$$

où  $\epsilon_T \downarrow_{T \rightarrow \infty} 0^+$ .

**Remarque 2.1.** Observons que l'on peut aussi bien écrire  $\hat{r}_T^\epsilon$  sous la forme

$$\hat{r}_T^\epsilon(x) = \int_0^T m(Y_t) p_{tT}^\epsilon(x) dt, \quad x \in \mathbb{R}^d, \quad (2.1)$$

où

$$p_{tT}^\epsilon(x) = \sum_j \left[ \frac{\frac{1}{Th_T^d} \mathbb{1}_{\pi_{Tj}}(X_t)}{\frac{1}{Th_T^d} \int_0^T \mathbb{1}_{\pi_{Tj}}(X_t) dt + \epsilon_T} \right] \mathbb{1}_{\pi_{Tj}}(x).$$

**Remarque 2.2 (Application à la prévision d'un processus à temps continu).** Soit  $\{\xi_t, t \in \mathbb{R}\}$  un processus markovien à valeur dans  $\mathbb{R}^d$  strictement stationnaire, géométriquement fortement mélangeant et dont une trajectoire est observée sur l'intervalle  $[0, T]$ . Pour prédire la valeur d'une variable aléatoire réelle de carré intégrable  $\zeta_{T+H} = m(\xi_{T+H})$ ,  $0 < H < T$  ( $H$  s'appelle alors l'*horizon* de la prévision), nous considérons le processus associé

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

et nous utilisons le régressogramme basé sur l'observation  $\{Z_t, 0 \leq t \leq S\}$  où  $S = T - H$ . Le prédicteur non paramétrique de  $\zeta_{T+H}$  est alors de la forme

$$\hat{\zeta}_{T+H} = \sum_j \left[ \frac{\int_0^S m(\xi_{t+H}) \mathbb{1}_{\pi_{Tj}}(\xi_t) dt}{\int_0^S \mathbb{1}_{\pi_{Tj}}(\xi_t) dt} \right] \mathbb{1}_{\pi_{Tj}}(\xi_{S+H}).$$

Des cas d'intérêt sont obtenus, par exemple, avec  $m(x) = x$  et  $d = 1$  (prédiction usuelle) ou bien avec  $m(x) = \mathbb{1}_B(x)$  ( $B \in \mathcal{B}_{\mathbb{R}^d}$ ) et  $d \geq 1$  (prédiction d'une zone d'alarmes). Une différence avec le temps discret est que l'on peut envisager ici une prévision à court terme, ce qui se traduit mathématiquement par la condition  $H \rightarrow 0^+$ . Les applications sont très variées : prévision de la consommation d'électricité, de l'évolution des cours de la bourse au court d'une séance, du processus de comptage d'un processus ponctuel... Pour un aperçu complet de la théorie de la prédiction non paramétrique en temps discret et en temps continu, on pourra se référer aux ouvrages de Bosq (2005) et de Bosq et Blanke (2007). L'application de la prédiction pour des processus à temps continu est également discutée avec des exemples par Ferraty et Vieu (2006), p. 195–201.

**Erreur en moyenne quadratique du régressogramme.** On se propose d'étudier l'erreur ponctuelle en moyenne quadratique (EMQ) associée au régressogramme qui est définie par

$$\text{EMQ}(\hat{r}_T(x)) = \mathbb{E}(\hat{r}_T(x) - r(x))^2.$$

Concernant l'existence de ce critère, notons que celle-ci peut être remise en question dans la mesure où l'estimateur a un dénominateur qui est une variable aléatoire. Pour lever cette difficulté technique, dans le cadre i.i.d. de l'estimation de la densité, de la fonction de hasard et de la régression, Marron et Härdle (1986) introduisent une mesure alternative de l'erreur quadratique intégrée pour des estimateurs qui sont des fonctions rationnelles de delta-familles ; ces résultats ont ensuite été étendus par Vieu (1991) au cadre fortement mélangeant pour des estimateurs à noyau.

Une autre approche utilisée par Chèze-Payaud (1994b) fait intervenir la décomposition :

$$\hat{r}_T(x) - r(x) = (\hat{r}_T(x) - r(x)) \frac{f(x) - \hat{f}_T(x)}{f(x)} + r(x) \frac{f(x) - \hat{f}_T(x)}{f(x)} + \frac{\hat{g}_T(x) - g(x)}{f(x)},$$

pour étudier en temps continu la convergence uniforme presque sûre, la convergence ponctuelle en moyenne quadratique et la normalité asymptotique de l'estimateur de Nadaraya-Watson dans les cas optimal et suroptimal. Cette décomposition est encore utilisée par Bosq et Chèze-Payaud (1999) pour déterminer l'erreur ponctuelle en moyenne quadratique à partir de valeurs discrétisées du processus  $Z$ . Notons que les résultats en temps continu s'obtiennent aussi à l'aide d'une décomposition plus simple introduite par Bosq (1998) (chapitre 5) de la forme :

$$\hat{r}_T(x) - r(x) = \hat{r}_T(x) - \frac{\mathbb{E}\hat{g}_T(x)}{\mathbb{E}\hat{f}_T(x)} + \frac{\mathbb{E}\hat{g}_T(x)}{\mathbb{E}\hat{f}_T(x)} - r(x),$$

où

$$\hat{r}_T(x) - \frac{\mathbb{E}\hat{g}_T(x)}{\mathbb{E}\hat{f}_T(x)} = \hat{r}_T(x) \frac{\mathbb{E}\hat{f}_T(x) - \hat{f}_T(x)}{\mathbb{E}\hat{f}_T(x)} + \frac{\hat{g}_T(x) - \mathbb{E}\hat{g}_T(x)}{\mathbb{E}\hat{f}_T(x)},$$

et

$$r(x) - \frac{\mathbb{E}\hat{g}_T(x)}{\mathbb{E}\hat{f}_T(x)} = r(x) \frac{\mathbb{E}\hat{f}_T(x) - f(x)}{\mathbb{E}\hat{f}_T(x)} + \frac{g(x) - \mathbb{E}\hat{g}_T(x)}{\mathbb{E}\hat{f}_T(x)}.$$

De la même façon pour l'estimateur  $\hat{r}_T^\epsilon$ , on pourra écrire

$$\hat{r}_T^\epsilon(x) - r(x) = \hat{r}_T^\epsilon(x) - \frac{\mathbb{E}\hat{g}_T(x)}{\mathbb{E}\hat{f}_T(x) + \epsilon_T} + \frac{\mathbb{E}\hat{g}_T(x)}{\mathbb{E}\hat{f}_T(x) + \epsilon_T} - r(x),$$

où

$$\hat{r}_T^\epsilon(x) - \frac{\mathbb{E}\hat{g}_T(x)}{\mathbb{E}\hat{f}_T(x) + \epsilon_T} = \hat{r}_T^\epsilon(x) \frac{\mathbb{E}\hat{f}_T(x) - \hat{f}_T(x)}{\mathbb{E}\hat{f}_T(x) + \epsilon_T} + \frac{\hat{g}_T(x) - \mathbb{E}\hat{g}_T(x)}{\mathbb{E}\hat{f}_T(x) + \epsilon_T}, \quad (2.2)$$

et

$$r(x) - \frac{\mathbb{E}\hat{g}_T(x)}{\mathbb{E}\hat{f}_T(x) + \epsilon_T} = r(x) \frac{\mathbb{E}\hat{f}_T(x) - f(x) + \epsilon_T}{\mathbb{E}\hat{f}_T(x) + \epsilon_T} + \frac{g(x) - \mathbb{E}\hat{g}_T(x)}{\mathbb{E}\hat{f}_T(x) + \epsilon_T}. \quad (2.3)$$

Pour l'étude en temps continu du régressogramme, nous adopterons cette dernière décomposition afin d'obtenir les vitesses de convergence optimale et suroptimale de l'erreur ponctuelle en moyenne quadratique. Pour ce faire, en supposant l'existence de la densité jointe  $f_{(X_s, X_t)}$  du couple  $(X_s, X_t)$  pour  $s \neq t$ , on introduit les fonctions de dépendance associées aux couples  $(X_s, X_t)$  et  $(Z_s, Z_t)$  :

$$\begin{aligned} g_{s,t} &= f_{(X_s, X_t)} - f \otimes f, & s \neq t; \\ g_{s,t}^* &= f_{(Z_s, Z_t)} - f_Z \otimes f_Z, & s \neq t, \end{aligned}$$

telles que  $g_{s,t} = g_{0,|t-s|} =: g_{|t-s|}$  et  $g_{s,t}^* = g_{0,|t-s|}^* =: g_{|t-s|}^*$ . On posera alors

$$h(x', x'') = \int_0^\infty |g_u(x', x'')| du,$$

et

$$H(x', x'') = \int_0^\infty |G_u(x', x'')| du,$$

où

$$G_u(x', x'') = \iint_{\mathbb{R}^{2d'}} m(y') m(y'') g_u^*(x', y'; x'', y'') dy' dy'', \quad u > 0.$$

## 2.2 Convergence ponctuelle en moyenne quadratique

Désignons par  $\nu_x$  un voisinage ouvert fixé du point  $x$ . Les principales hypothèses portant sur les paramètres fonctionnels et sur le processus sont les suivantes :

### Hypothèses $\widehat{R}$

- (i)  $\inf_{u \in \nu_x} f(u) > 0$  et  $\sup_{u \in \nu_x} f(u) \leq M_1^x < \infty$ ;
- (ii)  $\exists K_f^x > 0$ ,  $|f(y) - f(z)| \leq K_f^x \|y - z\|$  pour  $(y, z) \in \nu_x^2$ ;
- (iii)  $\exists K_g^x > 0$ ,  $|g(y) - g(z)| \leq K_g^x \|y - z\|$  pour  $(y, z) \in \nu_x^2$ .

### Hypothèses $\widehat{H}_0$

- (i)  $\exists M_Y > 0$ ,  $\sup_{0 \leq t \leq T < \infty} \mathbb{E}(m^2(Y_t) | \mathcal{F}_T) < M_Y^2$  p.s., où  $\mathcal{F}_T = \sigma\{X_t, 0 \leq t \leq T\}$ ;
- (ii)  $\exists \lambda, \kappa > 0$ ,  $\mathbb{E}(\exp(\lambda |m(Y_0)|^\kappa)) < \infty$ .

Hypothèses  $\widehat{H}_1$  Il existe un borélien de  $\mathbb{R}^2$  :  $\Gamma = \{(s, t) \in \mathbb{R}^2, |t - s| \leq u_0, u_0 > 0\}$  tel que

- (i)  $g_{s,t}$  existe pour  $(s, t) \notin \Gamma$ ;
- (ii)  $\sup_{(y,z) \in \nu_x^2} f_{u|0}(z|y) \leq M_2^x < \infty$  pour  $u \geq u_0$ ;
- (iii)  $Z$  est arithmétiquement fortement mélangeant de coefficient  $\alpha_Z^{(2)} : \alpha_Z^{(2)}(u) \leq a_0 u^{-\varrho}$  pour  $u \geq u_0$ ,  $a_0 > 0$  et  $\varrho > 2$ .

Hypothèses  $\widehat{H}_2$  ( $d = 1$ )  $h$  et  $H$  existent, sont bornées et continues en  $(x, x)$ .

**Remarque 2.3.** La condition  $\widehat{H}_0(i)$  introduite par Bosq (1998) est clairement satisfaite si le processus  $|m(Y_t)|$  est borné pour tout  $t \in \mathbb{R}$ ; ou encore, par exemple, si les processus  $(m(Y_t))$  et  $(X_t)$  sont indépendants. Une autre possibilité intéressante concerne le modèle

$$m(Y_t) = r(X_t) + \varepsilon_t, \quad t \in \mathbb{R},$$

où  $r(X_t)$  est borné,  $(\varepsilon_t)$  est un processus stationnaire de carré intégrable et où  $(X_t)$  et  $(\varepsilon_t)$  sont indépendants. La condition  $\widehat{H}_0(ii)$  d'existence du moment exponentiel de  $|m(Y_0)|$  est aussi utilisée dans Bosq et Blanke (2007) et nous retrouvons la condition locale  $\widehat{H}_2$  dans Bosq (1998). Enfin, les hypothèses  $\widehat{H}_1$  coïncident avec les hypothèses  $H_1$  du premier chapitre pour l'obtention de la vitesse optimale de l'histogramme (en portant la condition de mélange fort sur le processus bivarié  $Z$ ).

Le théorème suivant établit les vitesses optimale et suroptimale du régressogramme :

**Théorème 2.1.** (i) Cas "optimal" : si les conditions  $\widehat{R}$ ,  $\widehat{H}_0$  et  $\widehat{H}_1$  sont vérifiées, alors les choix  $\varepsilon_T = o(h_T)$  et  $h_T = c_T \cdot T^{-\frac{1}{d+2}}$  tel que  $c_T \rightarrow c$ ,  $0 < c < \infty$ , entraînent

$$\limsup_{T \rightarrow \infty} T^{\frac{2}{d+2}} \cdot \text{EMQ}(\widehat{r}_T^\varepsilon(x)) \leq \frac{2c^2}{f^2(x)} (C_f^2(x)r^2(x) + C_g^2(x)) + \frac{8u_0 M_Y^2}{c^d f(x)}, \quad x \in \mathbb{R}^d,$$

où  $C_f^2(x)$  (respectivement  $C_g^2(x)$ ) est une constante positive qui dépend à la fois de  $x$  et de  $f$  (respectivement  $g$ );

(ii) Cas “suroptimal” ( $d = 1$ ) : si les conditions  $\widehat{R}$ ,  $\widehat{H}_0(i)$  et  $\widehat{H}_2$  sont vérifiées et si  $\epsilon_T = o(h_T)$  et  $h_T = o(T^{-\frac{1}{2}})$ , alors

$$\limsup_{T \rightarrow \infty} T \cdot \text{EMQ}(\hat{r}_T^\epsilon(x)) \leq \frac{4}{f^2(x)} \left( M_Y^2 \int_0^\infty |g_u(x, x)| du + \int_0^\infty |G_u(x, x)| du \right), x \in \mathbb{R}.$$

**Remarque 2.4.** Les vitesses du régressogramme sont les mêmes que celles obtenues pour l’histogramme au chapitre 1 avec les mêmes choix du paramètre  $h_T$ . Ces résultats sont de plus similaires à ceux présentés par Bosq (1998) (théorèmes 5.1 et 5.2, p. 132 et 135) pour l’estimateur de Nadaraya-Watson sauf pour la vitesse optimale qui est de l’ordre  $T^{-4/(d+4)}$  au lieu de  $T^{-2/(d+2)}$  en raison du biais plus important du régressogramme.

## 2.3 Discussion

Dans ce chapitre, nous avons établi les vitesses optimale et suroptimale de l’erreur  $L^2$  du régressogramme en temps continu. Cette étude est moins exhaustive que celle donnée pour l’estimation de la densité mais on pourra ensuite compléter cette étude avec la convergence presque sûre ainsi que la normalité asymptotique en partant des techniques utilisées par Chèze-Payaud (1994b) et Bosq (1998) avec l’estimateur de Nadaraya-Watson. Il serait en outre très intéressant de chercher à améliorer les vitesses du régressogramme par une interpolation linéaire; ce qui revient à construire un estimateur analogue au polygone de fréquences pour la régression. Un tel estimateur n’a semble-t-il pas encore été examiné sinon dans le cadre d’une comparaison numérique avec l’estimateur de Nadaraya-Watson proposée par Simonoff et Hurvich (1993) sur des données liées au diabète sous l’appellation de “regressygon” !

## 2.4 Preuves

Pour la suite, nous aurons besoin de la relation

$$\mathbb{E} \left[ \hat{r}_T^\epsilon(x) (\hat{f}_T(x) - \mathbb{E} \hat{f}_T(x)) \right]^2 \leq M_Y^2 \text{Var}(\hat{f}_T(x)). \quad (2.4)$$

Nous reprenons la preuve telle qu’elle est donnée dans Bosq (1998) (p. 132) pour l’estimateur  $\hat{r}_T$ . À partir de la forme (2.1) du régressogramme et en utilisant l’inégalité de Cauchy-Schwarz :

$$\begin{aligned} \mathbb{E} \left[ \hat{r}_T^\epsilon(x) (\hat{f}_T(x) - \mathbb{E} \hat{f}_T(x)) \right]^2 &= \mathbb{E} \left( \mathbb{E} \left[ (\hat{r}_T^\epsilon(x))^2 (\hat{f}_T(x) - \mathbb{E} \hat{f}_T(x))^2 \middle| \mathcal{F}_T \right] \right) \\ &= \mathbb{E} \left( (\hat{f}_T(x) - \mathbb{E} \hat{f}_T(x))^2 \mathbb{E} \left[ \left( \int_0^T m(Y_t) p_{tT}^\epsilon(x) dt \right)^2 \middle| \mathcal{F}_T \right] \right) \\ &= \mathbb{E} \left( (\hat{f}_T(x) - \mathbb{E} \hat{f}_T(x))^2 \iint_{[0, T]^2} p_{sT}^\epsilon(x) p_{tT}^\epsilon(x) \mathbb{E} [m(Y_s) m(Y_t) | \mathcal{F}_T] ds dt \right) \\ &\leq \mathbb{E} \left( (\hat{f}_T(x) - \mathbb{E} \hat{f}_T(x))^2 \left( \int_0^T p_{tT}^\epsilon(x) \mathbb{E} [m^2(Y_t) | \mathcal{F}_T]^{1/2} dt \right)^2 \right). \end{aligned}$$

En remarquant que  $\int_0^T p_{tT}^\epsilon(x) dt < 1$ , on conclut avec l'hypothèse  $\widehat{H}_0(i)$  :

$$\mathbb{E} \left[ \widehat{r}_T^\epsilon(x) (\widehat{f}_T(x) - \mathbb{E}\widehat{f}_T(x)) \right]^2 \leq M_Y^2 \mathbb{E} (\widehat{f}_T(x) - \mathbb{E}\widehat{f}_T(x))^2.$$

Nous aurons aussi besoin pour démontrer le théorème 2.1 de connaître l'ordre de la convergence ponctuelle du terme de biais pour  $\widehat{g}_T$ .

**Lemme 2.1.** *Si la condition  $\widehat{R}(iii)$  est satisfaite, alors on a*

$$\limsup_{T \rightarrow \infty} h_T^{-1} \cdot |\mathbb{E}(\widehat{g}_T(x)) - g(x)| \leq C_g(x) < \infty, \quad x \in \mathbb{R}^d,$$

où  $C_g(x)$  est une constante positive qui dépend à la fois de  $x$  et de  $g$ .

### Preuve du lemme 2.1

Pour  $x \in \pi_{Tj}$ , on a

$$\begin{aligned} \mathbb{E}(\widehat{g}_j) &= \frac{1}{Th_T^d} \mathbb{E} \int_0^T m(Y_t) \mathbf{1}_{\pi_{Tj}}(X_t) dt = \frac{1}{h_T^d} \mathbb{E} \left( \mathbf{1}_{\pi_{Tj}}(X_0) \mathbb{E}[m(Y_0) | X_0] \right) \\ &= \frac{1}{h_T^d} \int_{\pi_{Tj}} \mathbb{E}[m(Y_0) | X_0 = u] f(u) du = \frac{1}{h_T^d} \int_{\pi_{Tj}} r(u) f(u) du. \end{aligned}$$

D'où

$$\mathbb{E}(\widehat{g}_j) - g(x) = \frac{1}{h_T^d} \int_{\pi_{Tj}} (g(u) - g(x)) du.$$

À partir d'un réel  $T$  tel que  $\pi_{Tj} \subset \nu_x$ , on utilise la condition lipschitzienne de continuité sur  $g$  donnée par  $\widehat{R}(iii)$  :

$$|\mathbb{E}(\widehat{g}_j) - g(x)| \leq \frac{1}{h_T^d} \int_{\pi_{Tj}} |g(u) - g(x)| du \leq \frac{1}{h_T^d} \int_{\pi_{Tj}} K_g^x \underbrace{\|u - x\|}_{\leq Ch_T} du \leq C_g(x) h_T,$$

où  $C > 0$  et où  $C_g(x)$  désigne une constante positive qui dépend à la fois de  $x$  et de  $g$ , ce qui prouve que

$$\limsup_{T \rightarrow \infty} h_T^{-1} \cdot |\mathbb{E}(\widehat{g}_j) - g(x)| \leq C_g(x) < \infty. \quad \blacksquare$$

### Preuve du théorème 2.1

Partie (i) : partant de la décomposition (2.2), de l'inégalité élémentaire " $(a + b)^2 \leq 2a^2 + 2b^2$ " et de la relation (2.4), on a

$$\begin{aligned} &\mathbb{E} \left( \widehat{r}_T^\epsilon(x) - \frac{\mathbb{E}\widehat{g}_T(x)}{\mathbb{E}\widehat{f}_T(x) + \epsilon_T} \right)^2 \\ &\leq \frac{2}{(\mathbb{E}\widehat{f}_T(x) + \epsilon_T)^2} \left[ \mathbb{E} \left( \widehat{r}_T^\epsilon(x) (\widehat{f}_T(x) - \mathbb{E}\widehat{f}_T(x)) \right)^2 + \mathbb{E} \left( \widehat{g}_T(x) - \mathbb{E}\widehat{g}_T(x) \right)^2 \right] \\ &\leq \frac{2}{(\mathbb{E}\widehat{f}_T(x) + \epsilon_T)^2} \left[ M_Y^2 \text{Var}(\widehat{f}_T(x)) + \text{Var}(\widehat{g}_T(x)) \right]. \end{aligned} \quad (2.5)$$

D'après l'étude de l'histogramme effectuée dans le chapitre 1 avec les conditions  $\widehat{H}_1$  :

$$\limsup_{T \rightarrow \infty} Th_T^d \cdot \text{Var}(\widehat{f}_T(x)) \leq 2u_0 f(x). \quad (2.6)$$

L'étude de  $\text{Var}(\widehat{g}_T(x))$  est similaire :

$$\begin{aligned} Th_T^d \cdot \text{Var}(\widehat{g}_j) &= \frac{1}{Th_T^d} \left\{ \iint_{[0,T]^2 \cap \Gamma} \text{Cov}\left(m(Y_s)\mathbb{1}_{\pi_{T_j}}(X_s), m(Y_t)\mathbb{1}_{\pi_{T_j}}(X_t)\right) dsdt \right. \\ &\quad \left. + \iint_{[0,T]^2 \cap \Gamma^c} \text{Cov}\left(m(Y_s)\mathbb{1}_{\pi_{T_j}}(X_s), m(Y_t)\mathbb{1}_{\pi_{T_j}}(X_t)\right) dsdt \right\} =: \frac{1}{Th_T^d} (I_1 + I_2). \end{aligned}$$

Majoration de  $I_1$  avec l'inégalité de Cauchy-Schwarz :

$$\begin{aligned} I_1 &\leq \text{Var}(m(Y_0)\mathbb{1}_{\pi_{T_j}}(X_0)) \iint_{[0,T]^2 \cap \Gamma} dsdt \leq \mathbb{E}(m^2(Y_0)\mathbb{1}_{\pi_{T_j}}(X_0)) \iint_{[0,T]^2 \cap \Gamma} dsdt \\ &\leq \mathbb{E}\left(\mathbb{1}_{\pi_{T_j}}(X_0)\mathbb{E}[m^2(Y_0)|\mathcal{F}_T]\right) \iint_{[0,T]^2 \cap \Gamma} dsdt \leq M_Y^2 \cdot h_T^d f(\xi_j) \cdot (2u_0T - u_0^2), \quad (2.7) \end{aligned}$$

où  $\xi_j \in \pi_{T_j}$ .

Majoration de  $I_2$  sous l'hypothèse de stationnarité sur la densité jointe du couple  $(Z_s, Z_t)$ . Posons  $W_t := m(Y_t)\mathbb{1}_{\pi_{T_j}}(X_t)$  et montrons que la seconde intégrale est négligeable devant la première. Comme la variable  $m(Y_0)$  est non nécessairement bornée, nous suivons une idée de Mack et Silverman (1982) (voir aussi Chèze-Payaud (1994b)) pour décomposer la seconde intégrale comme suit

$$\begin{aligned} I_2 &= \iint_{[0,T]^2 \cap \Gamma^c} \left[ \begin{aligned} &\text{Cov}\left(W_s\mathbb{1}_{|m(Y_s)| \leq \alpha_0(\ln T)^{1/\kappa}}, W_t\mathbb{1}_{|m(Y_t)| \leq \alpha_0(\ln T)^{1/\kappa}}\right) \\ &+ 2\text{Cov}\left(W_s\mathbb{1}_{|m(Y_s)| > \alpha_0(\ln T)^{1/\kappa}}, W_t\mathbb{1}_{|m(Y_t)| \leq \alpha_0(\ln T)^{1/\kappa}}\right) \\ &+ \text{Cov}\left(W_s\mathbb{1}_{|m(Y_s)| > \alpha_0(\ln T)^{1/\kappa}}, W_t\mathbb{1}_{|m(Y_t)| > \alpha_0(\ln T)^{1/\kappa}}\right) \end{aligned} \right] dsdt \\ &=: I_{21} + I_{22} + I_{23}, \end{aligned}$$

pour un  $\alpha_0$  strictement positif dont les valeurs sont déterminées dans la suite. La constante  $\kappa$  est celle donnée par la condition  $\widehat{H}_0(ii)$ .

$$\begin{aligned} |I_{21}| &\leq 2Th_T^{-2d\gamma} \int_{u_0}^{h_T^{-d+\varepsilon}} \iint_{\pi_{T_j}^2} (f_u(y, z) + f(y)f(z)) dydz du \\ &+ 2T \int_{h_T^{-d+\varepsilon}}^{\infty} \left| \text{Cov}\left(W_0\mathbb{1}_{|m(Y_0)| \leq \alpha_0(\ln T)^{1/\kappa}}, W_u\mathbb{1}_{|m(Y_u)| \leq \alpha_0(\ln T)^{1/\kappa}}\right) \right| du =: 2T(I_{21a} + I_{21b}), \end{aligned}$$

avec  $\varepsilon$  réel positif dont on précisera les valeurs ensuite. Avec les conditions  $\widehat{R}(i)$ ,  $\widehat{R}(ii)$  et  $\widehat{H}_1(ii)$ , pour  $T$  assez grand tel que  $\pi_{T_j} \subset \nu_x$ , on a

$$\begin{aligned} I_{21a} &\leq \alpha_0^2(\ln T)^{2/\kappa} \int_{u_0}^{h_T^{-d+\varepsilon}} \iint_{\pi_{T_j}^2} f(y)(f_{u|0}(z|y) + f(z)) dydz du \\ &\leq \alpha_0^2(\ln T)^{2/\kappa} \cdot (h_T^{-d+\varepsilon} - u_0) \cdot h_T^d (M_1^x + M_2^x) \cdot h_T^d f(\xi_j) \\ &\leq \alpha_0^2(\ln T)^{2/\kappa} \left[ h_T^{d+\varepsilon} \cdot (M_1^x + M_2^x) f(\xi_j) \right]. \end{aligned}$$

En utilisant l'inégalité de Davydov (cf. lemme 3.3) pour un couple  $(p, q) \in ]1, \varrho - 1] \times \left[1 + \frac{\varrho}{\varrho - 2}, \infty\right[$  tel que  $\frac{2}{q} + \frac{1}{p} = 1$ , on a

$$I_{21b} \leq \alpha_0^2 (\ln T)^{2/\kappa} \int_{h_T^{-d+\varepsilon}}^{\infty} 2p \cdot 2^{1/p} \cdot \|\mathbb{1}_{\pi_{Tj}}(X_0)\|_q \|\mathbb{1}_{\pi_{Tj}}(X_u)\|_q \cdot (\alpha_Z^{(2)}(u))^{1/p} du,$$

soit avec l'hypothèse de mélange arithmétique  $\widehat{H}_1(iii)$  :

$$I_{21b} \leq \alpha_0^2 (\ln T)^{2/\kappa} \left[ h_T^{\frac{1}{p}\{(d-\varepsilon)(\varrho-p)+(p-1)d\}} \cdot \frac{2p^2(2a_0)^{1/p}}{\varrho-p} f(\xi_j)^{1-\frac{1}{p}} \right].$$

D'où

$$\frac{1}{Th_T^d} |I_{21}| \leq \alpha_0^2 (\ln T)^{2/\kappa} \left[ h_T^\varepsilon \cdot 2(M_1^x + M_2^x) f(\xi_j) + h_T^{\frac{1}{p}\{(d-\varepsilon)(\varrho-p)-d\}} \cdot \frac{4p^2(2a_0)^{1/p}}{\varrho-p} f(\xi_j)^{1-\frac{1}{p}} \right].$$

En prenant  $\varepsilon \in ]0, d\left(1 - \frac{1}{\varrho-p}\right)[$  avec  $1 < p < \varrho - 1$  et pour le choix du paramètre  $h_T$  donné dans l'énoncé du théorème, on en déduit alors que

$$\frac{1}{Th_T^d} I_{21} = o(1).$$

Pour l'étude de  $I_{22}$ , on applique l'inégalité de Cauchy-Schwarz :

$$|I_{22}| \leq 2T^2 \left\{ \mathbb{E} \left( m^2(Y_0) \mathbb{1}_{\pi_{Tj}}(X_0) \mathbb{1}_{|m(Y_0)| \leq \alpha_0 (\ln T)^{1/\kappa}} \right) \right\}^{1/2} \\ \times \left\{ \mathbb{E} \left( m^2(Y_0) \mathbb{1}_{\pi_{Tj}}(X_0) \mathbb{1}_{|m(Y_0)| > \alpha_0 (\ln T)^{1/\kappa}} \right) \right\}^{1/2}.$$

On applique encore une fois l'inégalité de Cauchy-Schwarz :

$$|I_{22}| \leq 2T^2 \left\{ \mathbb{E}(m^2(Y_0)) \right\}^{1/2} \times \left\{ \mathbb{E}(m^4(Y_0)) \right\}^{1/4} \left\{ P(\lambda |m(Y_0)|^\kappa > \lambda \alpha_0^\kappa (\ln T)) \right\}^{1/4}.$$

De là, on obtient avec l'inégalité de Markov :

$$|I_{22}| \leq 2T^{\frac{8-\lambda\alpha_0^\kappa}{4}} \left\{ \mathbb{E}(m^2(Y_0)) \right\}^{1/2} \times \left\{ \mathbb{E}(m^4(Y_0)) \right\}^{1/4} \left\{ \mathbb{E}(\exp(\lambda |m(Y_0)|^\kappa)) \right\}^{1/4}.$$

On en déduit alors que

$$\frac{1}{Th_T^d} I_{22} = o(1) \quad \text{avec la condition } T^{8-\lambda\alpha_0^\kappa} \rightarrow 0,$$

qui est vérifiée pour un choix de  $\alpha_0$  suffisamment grand.

L'étude du terme  $I_{23}$  est ensuite similaire à celle du terme  $I_{22}$ .

Il nous reste à étudier le terme de "pseudo-biais" donné par (2.3). On a

$$\left( r(x) - \frac{\mathbb{E}\hat{g}_T(x)}{\mathbb{E}\hat{f}_T(x) + \varepsilon_T} \right)^2 \leq \frac{2r^2(x) \left( \mathbb{E}\hat{f}_T(x) - f(x) + \varepsilon_T \right)^2 + 2 \left( g(x) - \mathbb{E}\hat{g}_T(x) \right)^2}{\left( \mathbb{E}\hat{f}_T(x) + \varepsilon_T \right)^2}.$$

Sous les conditions de continuité lipschitzienne sur  $f$  et  $g$  pour  $T$  assez grand, on a montré (lemme 2.1) que

$$\left(\mathbb{E}\hat{g}_T(x) - g(x)\right)^2 \leq C_g^2(x)h_T^2,$$

et pour le choix  $\epsilon_T = o(h_T)$ , on déduit du lemme 1.2 que

$$\begin{aligned} \left(\mathbb{E}\hat{f}_T(x) - f(x) + \epsilon_T\right)^2 &= \left(\mathbb{E}\hat{f}_T(x) - f(x)\right)^2 + 2\epsilon_T\left(\mathbb{E}\hat{f}_T(x) - f(x)\right) + \epsilon_T^2 \\ &\leq C_f^2(x)h_T^2 + o(h_T^2). \end{aligned}$$

D'où si  $\epsilon_T = o(h_T)$ ,

$$\limsup_{T \rightarrow \infty} h_T^{-2} \cdot \left(r(x) - \frac{\mathbb{E}\hat{g}_T(x)}{\mathbb{E}\hat{f}_T(x) + \epsilon_T}\right)^2 \leq \frac{2}{f^2(x)} (C_f^2(x)r^2(x) + C_g^2(x)). \quad (2.8)$$

Finalement, en combinant les expressions (2.5), (2.6), (2.7) et (2.8), avec le choix  $h_T = c_T \cdot T^{-\frac{1}{d+2}}$ , tel que  $c_T \rightarrow c$ ,  $0 < c < \infty$ , on en déduit que

$$\limsup_{T \rightarrow \infty} T^{\frac{2}{d+2}} \cdot \text{EMQ}(\hat{r}_T^\epsilon(x)) \leq \frac{2c^2}{f^2(x)} (C_f^2(x)r^2(x) + C_g^2(x)) + \frac{8u_0M_Y^2}{c^d f(x)}.$$

Partie (ii) : nous reprenons la décomposition (2.1) et l'expression (2.5) :

$$\mathbb{E}\left(\hat{r}_T^\epsilon(x) - \frac{\mathbb{E}\hat{g}_T(x)}{\mathbb{E}\hat{f}_T(x) + \epsilon_T}\right)^2 \leq \frac{2}{(\mathbb{E}\hat{f}_T(x) + \epsilon_T)^2} \left[M_Y^2 \text{Var}(\hat{f}_T(x)) + \text{Var}(\hat{g}_T(x))\right].$$

Sous la condition donnée par  $\hat{H}_2$  par convergence dominée, on a d'une part :

$$T \cdot \text{Var}(\hat{f}_T(x)) \leq 2 \int_0^\infty |g_u(x, x)| du \times (1 + o(1)).$$

De la même manière, on étudie la variance de  $\hat{g}_T(x)$  :

$$\begin{aligned} \text{Var}(\hat{g}_T(x)) &= \frac{1}{T^2 h_T^{2d}} \iint_{[0, T]^2} \text{Cov}\left(m(Y_s)\mathbb{1}_{\pi_{T_j}}(X_s), m(Y_t)\mathbb{1}_{\pi_{T_j}}(X_t)\right) ds dt \\ &= \frac{2}{T h_T^{2d}} \int_0^T \left(1 - \frac{u}{T}\right) \text{Cov}\left(m(Y_0)\mathbb{1}_{\pi_{T_j}}(X_0), m(Y_u)\mathbb{1}_{\pi_{T_j}}(X_u)\right) du, \end{aligned}$$

où

$$\begin{aligned} &\text{Cov}\left(m(Y_0)\mathbb{1}_{\pi_{T_j}}(X_0), m(Y_u)\mathbb{1}_{\pi_{T_j}}(X_u)\right) \\ &= \iint_{\pi_{T_j}^2} \iint_{\mathbb{R}^{2d'}} m(y_1)m(y_2)g_u^*(x_1, y_1; x_2, y_2) dy_1 dy_2 dx_1 dx_2 = \iint_{\pi_{T_j}^2} G_u(x_1, x_2) dx_1 dx_2. \end{aligned}$$

Sous la condition  $\hat{H}_2$  par convergence dominée, on a alors

$$T \cdot \text{Var}(\hat{g}_T(x)) \leq 2 \int_0^\infty |G_u(x, x)| du \times (1 + o(1)).$$

On conclut ensuite en choisissant  $\epsilon = o(h_T)$  et  $h_T = o(T^{-1/2})$  :

$$\limsup_{T \rightarrow \infty} T \cdot \text{EMQ}(\hat{r}_T^\epsilon(x)) \leq \frac{4}{f^2(x)} \left(M_Y^2 \int_0^\infty g_u(x, x) du + \int_0^\infty G_u(x, x) du\right).$$

■

## Chapitre 3

# On the $L_2$ -rates of histograms and frequency polygons from sampled data

**Abstract.** – In this chapter, we study some asymptotic properties for nonparametric density estimators based upon  $n$  discretized observations  $X_{t_1}, \dots, X_{t_n}$  of a weakly stationary continuous-time process. More precisely, we establish that histograms and frequency polygons can attain the same optimal  $L_2$ -rates as in the i.i.d. case. Moreover, if sample paths are sufficiently irregular, a “parametric” rate is even possible by using a suitable “high frequency” sampling design.

### 3.1 Preliminaries and notations

Let  $X^T = \{X_t, 0 \leq t \leq T\}$  be a measurable  $\mathbb{R}^d$ -valued,  $d \geq 1$ , continuous-time process on the probability space  $(\Omega, \mathcal{F}, P)$ , where the  $X_t$ 's have a common distribution  $\mu$  admitting a density  $f$  with respect to the Lebesgue measure  $\lambda_d$  over  $\mathbb{R}^d$ . We suppose that the joint density  $f_{(X_s, X_t)}$  of  $(X_s, X_t)$  does exist for all  $s \neq t$  such that  $f_{(X_s, X_t)} = f_{(X_0, X_{|t-s|})} =: f_{|t-s|}$ , which is a quite weak stationarity condition. We also denote by  $g_u$  the function defined for all  $u > 0$  as  $g_u := f_u - f \otimes f$  where  $(f \otimes f)(y, z) = f(y)f(z)$ . Some required asymptotic independence conditions on the process (including  $\alpha$ -mixing condition) will be specified later with the assumptions. Our purpose is to estimate the function  $f$  from  $n$  observations collected up to time  $T$  by using one of the sampling procedures described below.

#### 3.1.1 Sampling schemes

Let  $\mathcal{T}^n = \{t_k, 0 \leq k \leq n\}$  be a strictly increasing sequence of points in time – or event arrival times – such that  $0 = t_0 < t_1 < \dots < t_n =: T_n$  and  $T_n \rightarrow \infty$  as  $n \rightarrow \infty$ . Then the first two random schemes defined in Masry (1983) are the following :

**[RS] Renewal sampling.** The set of times for observations  $\mathcal{T}^n \equiv \mathcal{T}_1^n$  is a renewal type process on  $[0, \infty)$  such that

$$t_0 = 0 \quad \text{and} \quad t_k = \sum_{j=1}^k \tau_j, \quad 1 \leq k \leq n,$$

where  $\tau^n = \{\tau_k, 1 \leq k \leq n\}$  is a sequence of positive and i.i.d. random variables – or inter-arrival times – generated by a given probability density function  $g(t) > 0$  with finite mean  $\delta$ . Let  $g_k$  be the  $k$ th fold convolution of  $g$  with itself,  $g_k(t)$  is the density function of  $t_k$  and we define the renewal density  $h$  by  $h(t) := \sum_{k=1}^{\infty} g_k(t)$ . Here and below, the function  $h$  is supposed to be bounded by a positive constant  $\bar{h}_0$ . Note that [RS] has been used in many works dealing with density kernel estimation. Cite among others those of Masry (1983), Stoyanov et Robinson (1991), Vilar et Vilar (2000) or Wu (1997) for the optimal rates of uniform almost sure convergence. For relevant literature on renewal processes we may refer to Cox (1962) or to Karlin et Taylor (1981) (Chapter 5).

**Remark 3.1.** In general, we have  $h(u) \rightarrow \delta^{-1}$  as  $u \rightarrow \infty$ , but the boundedness of  $h$  is a condition which holds for a large class of sequences  $\tau^n$ . For example, when  $\tau^n$  has a Gamma density of type  $r$ , *i.e.*,

$$g(t) = \frac{(r/\delta)(rt/\delta)^{r-1} \exp(-rt/\delta)}{(r-1)!}, \quad r \in \mathbb{N}^*, \quad \delta > 0, \quad t \geq 0,$$

with mean  $\delta$  and variance  $\delta^2/r$ , the probability law of the renewal process may be readily computed. Thus, if  $r = 1$ ,  $\mathcal{T}_1^n$  is a Poisson process and  $h(t) \equiv \delta^{-1}$ ,  $t \geq 0$ . If  $r = 2$ , we have  $h(t) = \delta^{-1}(1 - \exp(-4t/\delta))$ . In both cases, the choice  $\bar{h}_0 = \delta^{-1}$  is appropriate (see Masry (1983)). The detailed calculus of  $h$  is given for instance in Parzen (1999) (Chapter 5). Recall the example in Aït-Sahalia et Mykland (2003), a histogram distribution of the time between trades is plotted with a shape close to an exponential reference density.

**[JS] Jittered sampling.** First, we suppose that the process is regularly observed with a period  $\delta > 0$ . This sequence  $\mathcal{T}^n \equiv \mathcal{T}_2^n$  is then contaminated by an additive noise due to the plausible imperfections of the measurement recording system :

$$t_0 = Z_0 \quad \text{and} \quad t_k = k\delta + Z_k, \quad 1 \leq k \leq n,$$

where  $Z^n = \{Z_k, 0 \leq k \leq n\}$  denotes an i.i.d. random sample from a symmetrical probability density function  $g_J(z)$  over  $[-\delta/2, \delta/2]$ . In contrast with [RS], observe that [JS] is only partially random since  $t_k$  has a deterministic component. Such a design is still used for instance in Masry (1983) and in Vilar et Vilar (2000) where kernel density estimators and MISE are considered.

Finally, we introduce a periodic scheme examined by Bosq (1995, 1997) where the sampling step  $\delta_n$  is  $n$ -decreasing in a deterministic manner.

**[HFS] High frequency sampling.** In order to represent the occurrence of high frequency observations during a long time, the sampling instants in  $\mathcal{T}^n \equiv \mathcal{T}_3^n$  are defined as

$$t_{0,n} = 0 \quad \text{and} \quad t_{k,n} = k\delta_n, \quad 1 \leq k \leq n,$$

where  $\delta_n > 0$  and  $\delta_n \rightarrow 0^+$ ,  $T_n = n\delta_n \rightarrow \infty$ ,  $n \rightarrow \infty$ . In the sequel, we will give minimal thresholds  $\delta_n^*$  over which our estimators converge with the optimal rates of the i.i.d. case. The knowledge of  $\delta_n^*$  will also help us to minimize the costs of estimation without altering the rates. To explain, observe that two situations may occur in applications. First, if the time of observation is  $[0, T_n]$ , with a fixed large enough  $T_n$ , then the use of  $\delta_n^*$  allows to select the minimal number  $n^*$  of points in  $[0, T_n]$  to estimate  $f$ . On the other hand, consider that a maximal and large enough sample size  $n$  is available, we can deduce from  $\delta_n^*$  a minimal time  $T_n^* = n\delta_n^*$  of observation (see Blanke et Pumo (2003)). Furthermore, we will emphasize the convenience of such a framework to sample a continuous-time process. Thus, under the CL's conditions *i.e.*  $\int_0^\infty \|g_u\|_\infty du < \infty$  and  $g_u(\cdot, \cdot)$  is continuous at  $(x, x)$  for each  $u > 0$ , Bosq (1997) provides the exact pointwise asymptotic quadratic error of kernel estimators which is of order  $T_n^{-1}$  from a minimal sample size  $n^*$ . In that situation, the sampling scheme is said to be *admissible*. Concerning admissible sampling in nonparametric density estimation, let us cite other works by Leblanc (1995) for wavelets estimators, by Biau (2003) for spacial kernel estimators and, more recently, by Comte et Merlevède (2005) and Blanke (2006), respectively, for projection and adaptative kernel estimators.

**Remark 3.2.** The observations drawn from [RS] and [JS] are irregularly spaced in time, but in order to compare with [HFS], we may observe that the “long-term” expected inter-arrival time between two consecutive instants is equal to  $\delta$ . In other words, both designs tend to become themselves periodic with a fixed step  $\delta$  as  $n \rightarrow \infty$ . If  $\mathcal{T}^n$  is random, it is also assumed that the processes  $X^T$  and  $\mathcal{T}^n$  are independent. In the sequel, our estimators will be based upon the stochastic model  $(X_{t_1}, \dots, X_{t_n})$ , where all observations are random variables in the sense of being measurable with respect to the product  $\sigma$ -algebra generated by  $X^T$  and  $\mathcal{T}^n$ .

### 3.1.2 Mean integrated squared error

The global accuracy of density estimators can be measured by the *mean integrated squared error* which is the expected squared distance between a density estimator  $\hat{f}_n$  and the true density  $f$  integrated over  $\mathbb{R}^d$  :

$$\text{MISE}(\hat{f}_n) = \mathbb{E} \int_{\mathbb{R}^d} \left( \hat{f}_n(x) - f(x) \right)^2 dx.$$

It is also the sum of the *integrated squared bias* (ISB) and the *integrated variance* (IV) :

$$\text{ISB}(\hat{f}_n) = \int_{\mathbb{R}^d} \left( \mathbb{E}(\hat{f}_n(x)) - f(x) \right)^2 dx \quad \text{and} \quad \text{IV}(\hat{f}_n) = \int_{\mathbb{R}^d} \mathbb{E} \left( \hat{f}_n(x) - \mathbb{E}(\hat{f}_n(x)) \right)^2 dx.$$

Notations. Let us fix the following usual notations :  $\mathcal{C}^k(\mathbb{R}^d)$  denotes the set of  $k$ -times continuously differentiable functions and  $L_k(\mathbb{R}^d)$  the set of functions with integrable  $k$ th power over  $\mathbb{R}^d$  such that  $\|f\|_k = \left( \int_{\mathbb{R}^d} f^k(x) dx \right)^{1/k}$ .

## 3.2 Histogram

Since any bounded measurable functions can be approximated by piecewise constant functions, it could be appropriate to make choice of a piecewise constant estimator. That is why we primarily examine the *histogram*, which is the oldest and most popular nonparametric estimator. Because of its simplicity, histogram is still widely used in presentation, research and practice by statisticians. Obviously, theoretical properties have been also extensively studied in the i.i.d. case and we refer *e.g.* to Bosq et Lecoutre (1987) (Chapter 6, Section A) as well as to Scott (1992) (Chapter 3) for a background material. Recently, Lejeune (2006) derived in continuous-time both optimal and full rates of MISE and asymptotic normality under CL's conditions. In this section we aim at extending some of these results for observations collected at discretized instants according to the sequences  $\mathcal{T}_i^n$ ,  $i = 1, 2, 3$ .

### 3.2.1 Definitions and assumptions

Prior to the definition of our estimator, we introduce a partition of  $\mathbb{R}^d$ , say  $\Pi_n$ , into hypercubes of volume  $h_n^d$  such that  $n \rightarrow \infty$ ,  $h_n \rightarrow 0^+$ ,  $nh_n^d \rightarrow \infty$  :

$$\Pi_n = \{\pi_{nj}, j \in \mathbb{Z}^d\},$$

and

$$\pi_{nj} = \prod_{k=1}^d [b_{jk}, b_{j_{k+1}}[ = \prod_{k=1}^d \left[ c_{jk} - \frac{h_n}{2}, c_{jk} + \frac{h_n}{2} \right], \quad j = (j_1, \dots, j_d) \in \mathbb{Z}^d,$$

where  $b_j = (b_{j_1}, \dots, b_{j_d})' \in \mathbb{R}^d$ ,  $b_{j_{k+1}} - b_{j_k} = h_n$  and  $c_{j_k} = \frac{b_{j_k} + b_{j_{k+1}}}{2}$ . Here  $h_n$  is a smoothing parameter commonly referred to as the *bin width*. From now on, we will suppose for any  $x \in \mathbb{R}^d$  the existence of an index  $j(x, n)$  in  $\mathbb{Z}^d$  such that  $x \in \pi_{j(x, n)}$  ( $=: \pi_{nj}$ ). The histogram estimator introduced below has a unique value, denoted by  $\hat{f}_j$ , over each hypercube  $\pi_{nj}$  of  $\Pi_n$ , which explains its high computational advantage.

**Definition 3.1.** *Given the partition  $\Pi_n$ , the histogram estimator of  $f$  is defined as*

$$\hat{f}_n^H(x) = \sum_j \left[ \frac{1}{nh_n^d} \sum_{k=1}^n \mathbb{1}_{\pi_{nj}}(X_{t_k}) \right] \mathbb{1}_{\pi_{nj}}(x) = \sum_j \hat{f}_j \mathbb{1}_{\pi_{nj}}(x), \quad x \in \mathbb{R}^d,$$

where  $\mathbb{1}_A$  stands for the indicator function of the set  $A$ .

Let denote by  $\|\cdot\|$  any norm of  $\mathbb{R}^d$ . In order to state our main results, it is convenient to recall the notion of generalized Riemann integral in the following sense (see Bertrand-Retali (1978)).

**Definition 3.2.** *Let  $g$  be a real function defined on  $\mathbb{R}^d$ . If  $(g_n^-)$  and  $(g_n^+)$  are two sequences of step functions on  $\Pi_n$ , Riemann-integrable on  $\mathbb{R}^d$ , and such that*

$$g_n^- \leq g_{n+1}^- \leq g \leq g_{n+1}^+ \leq g_n^+ \quad \text{and} \quad \lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} g_n^-(x) dx = \lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} g_n^+(x) dx = I,$$

then we say that  $I$  is the generalized Riemann integral (GRI) of  $g$ . In this case,  $g$  is Lebesgue-integrable on  $\mathbb{R}^d$ ,  $\int_{\mathbb{R}^d} g(x)dx = I$  and

$$\forall \varepsilon > 0, \quad g_\varepsilon(x) = \sup_{t \in \mathbb{R}^d} \{g(t); \|t - x\| \leq \varepsilon\}$$

is Lebesgue-integrable on  $\mathbb{R}^d$  and  $\int_{\mathbb{R}^d} g_\varepsilon(x)dx$  remains bounded as  $\varepsilon \rightarrow 0$ .

Note that in the sequel any ‘‘integrable’’ function will be more exactly ‘‘Riemann-integrable’’. Let  $\mathcal{A}$  and  $\mathcal{B}$  be two sub- $\sigma$ -algebras of  $\mathcal{F}$ , we introduce the classical strong mixing coefficient defined as

$$\alpha(\mathcal{A}, \mathcal{B}) := \sup_{A \in \mathcal{A}, B \in \mathcal{B}} |P(A \cap B) - P(A)P(B)|.$$

These are now the main assumptions over processes.

### Assumptions $A_0$

- (i)  $f \in \mathcal{C}^2(\mathbb{R}^d)$  so that all the partial derivatives and their squares exist and admit a GRI;
- (ii)  $f$  is continuous at  $x$  and  $\|f\|_\infty = \sup_{y \in \mathbb{R}^d} f(y) < \infty$ .

### Assumptions $A_1$

- (i) There exists a positive integrable function  $k(\cdot)$  defined on  $\mathbb{R}^d$  such that for all  $y \in \mathbb{R}^d$ ,  $\sup_{z \in \mathbb{R}^d} |g_u(y, z)| \leq \pi_0 k(y)$ ,  $\pi_0 > 0$ ,  $u \geq u_0 > 0$ ;
- (ii)  $X^T$  is an arithmetically strongly mixing (ASM) process :

$$\exists \rho > 2, a_0 > 0 : \forall u \geq u_1 > u_0, \alpha_X^{(2)}(u) := \alpha(\sigma(X_0), \sigma(X_u)) \leq a_0 u^{-\rho},$$

where  $\sigma(X)$  denotes the  $\sigma$ -algebra of events generated by a random variable  $X$ .

### Assumptions $A'_1$

- (i) There exists a positive integrable function  $\varphi(\cdot)$  defined on  $\mathbb{R}^d$  such that for all  $y \in \mathbb{R}^d$ ,  $\sup_{z \in \mathbb{R}^d} f_u(y, z) \leq \varphi(y)u^{-\gamma_0}$ ,  $\gamma_0 > 0$ ,  $0 < u \leq u_0$ ;
- (ii) There exists a positive integrable function  $k(\cdot)$  defined on  $\mathbb{R}^d$  such that for all  $y \in \mathbb{R}^d$ ,  $\sup_{z \in \mathbb{R}^d} |g_u(y, z)| \leq k(y)\pi(u)$ ,  $u \geq u_0$ , where  $\pi(\cdot)$  is a bounded and ultimately decreasing function which satisfies  $\int_{u_1}^\infty \pi(u)du < \infty$ ,  $u_1 > u_0$ .

The assumptions above are classical in nonparametric estimation with dependent data.  $A_0$  displays some constraints of regularity on the true density  $f$  and its two continuous derivatives. The condition  $A_0(i)$  is specific to the bias treatment, it was previously introduced by Lecoutre (1985) to study the multivariate histogram.

The following conditions should take into account the local behavior of sample paths as well as the properties of asymptotic independence of processes (respectively described with the behavior of  $g_u$  for  $u$  near the origin and for  $u$  large). Let  $f_{u|0}(\cdot|x_0)$  denote the conditional density of  $X_u$  given  $X_0 = x_0$ .  $A_1(i)$  is a mild condition on  $g_u$  for intermediate

values of  $u$ . Since  $f$  is bounded, it slightly weakens the one :  $\sup_{(y,z) \in \mathbb{R}^{2d}} f_{u|0}(z|y) \leq C$ ,  $u \geq u_0$ ,  $0 < C < \infty$ , made by Masry (1983), Carbon *et al.* (1997), and Lejeune (2005).

$A'_1(i)$  appears to be less usual in density estimation, but it is a typical condition for continuous-time framework since it should be linked with the local regularity of sample paths. Roughly speaking, it allows us to control the explosive behavior of the joint densities  $f_u(\cdot, \cdot)$  in a neighborhood of  $u = 0$ . Here  $A'_1(i)$  is needed with [HFS] to obtain a full rate  $T_n^{-1}$  for some  $\gamma_0 < 1$ , which characterizes “irregular sample paths” ; while  $\gamma_0 \geq 1$  provides all intermediate rates of convergence. Assumptions  $A'_1$  are in the spirit of those made (and widely discussed) by Blanke et Pumo (2003). Blanke (2006) proposes also examples of uni- and multi-dimensional processes satisfying our conditions ; these include particularly a broad class of homogeneous diffusion processes useful in many applied works. In view of her examples the case  $\gamma_0 > 1$  seems rather adapted to the multidimensional context since we have  $\gamma_0 = d/2$  (see Section 3.4).

Other kind of assumptions, namely  $A_1(ii)$  and  $A'_1(ii)$ , ensures asymptotic independence between variables distant in time. The first one involves  $\alpha$ -mixing which is well-known to be weaker than many dependence structures as  $\phi$ ,  $\beta$  or  $\rho$ -mixing (see *e.g.* Doukhan (1994)). Especially, the 2- $\alpha$ -mixing coefficient proposed by Bosq (1998) is even weaker than strongly mixing since it does not suppose to know the past of  $X_0$  and the future of  $X_u$ . Nevertheless,  $A_1(ii)$  fails to make our sampling admissible and we shall have to consider instead the other independence condition given by  $A'_1(ii)$ .

### 3.2.2 Rates of convergence

Using each sampling design defined above, we can establish now the optimal rate of histograms. For the sake of readability, some crucial lemmas which provide upper bounds for the variances and the covariances of  $\hat{f}_n^H$  are postponed to the proofs. Let  $f'_i := \partial f / \partial x_i$  and define the *roughness* ( $\mathcal{R}$ ) of  $f'_i$  by its squared  $L_2$ -norm :  $\mathcal{R}(f'_i) := \int_{\mathbb{R}^d} f'_i(x)^2 dx$ . The following lemma due to Lecoutre (1985) furnishes an expression for the ISB of  $\hat{f}_n^H$ .

**Lemma 3.1.** *If Assumption  $A_0(i)$  is satisfied then*

$$\text{ISB}(\hat{f}_n^H) = \frac{h_n^2}{12} \mathcal{R}_d(f') + o(h_n^2),$$

where  $\mathcal{R}_d(f') := \sum_{i=1}^d \mathcal{R}(f'_i)$ .

**Remark 3.3.** It is noteworthy that the bias of  $\hat{f}_n^H$  only depends on the bin width and the true unknown density  $f$ , not on the data. In particular the dependence structure does not matter : since each random variable  $X_{t_k}$  has the same distribution, one has

$$\mathbb{E} \hat{f}_n^H(x) = \frac{1}{nh_n^d} \sum_{k=1}^n P(X_{t_k} \in \pi_{nj}) = \frac{1}{h_n^d} P(X_0 \in \pi_{nj}) = \frac{1}{h_n^d} \int_{\pi_{nj}} f(x) dx, \quad x \in \pi_{nj},$$

and the proof is then identical to those given by Lecoutre (1985) with independent observations. Concerning the variance term, our results will be affected by the scheme being used.

**[RS]-[JS]** – Let us denote by  $\lceil x \rceil$  the smallest integer not less than the real  $x$ . The first part of the next theorem gives an asymptotic upper bound for IV. Consequently, from an ad hoc choice of the bin width  $h_n$  which balances both ISB and IV terms, we infer that histograms can achieve the same optimal rate  $n^{-2/(d+2)}$  of convergence to  $f$  as in the i.i.d. case.

**Theorem 3.1.** (i) Under  $A_0(ii)$  and  $A_1$  and if  $f^{1-\frac{1}{p}} \in L_1(\mathbb{R}^d)$  for  $1 < p < \rho - 1$ , then

$$\limsup_{n \rightarrow \infty} nh_n^d \cdot \text{IV}(\hat{f}_n^H) \leq 1 + C;$$

(ii) If in addition  $A_0(i)$  holds then the choice  $h_n = cn^{-\frac{1}{d+2}}$ ,  $0 < c < \infty$ , yields

$$\limsup_{n \rightarrow \infty} n^{\frac{2}{d+2}} \cdot \text{MISE}(\hat{f}_n^H) \leq \frac{c^2}{12} \mathcal{R}_d(f') + \frac{1}{c^d} (1 + C),$$

where  $C = 2u_0\bar{h}_0$  for [RS] and  $C = 2 \lceil \frac{u_0}{\delta} \rceil$  for [JS].

**Remark 3.4.** If  $p = \rho - 1$  the rates in Theorem 3.1 remain valid but with larger asymptotic constants (see proofs). Thus for instance, if  $\rho \geq 3$ , one may choose  $p = 2$  provided that  $f^{1/2} \in L_1(\mathbb{R}^d)$ .

**[HFS]** – Now the high frequency model is interesting to understand the connection between both discrete and continuous-time frameworks. Here the period  $\delta_n$  is now a function of the sampling size  $n$  so that all observations can be as close in time as desired provided  $n$  large enough. Within this new setup the local condition  $A'_1(i)$  on the joint density of  $(X_0, X_u)$  is also needed for the small values of  $u$  and some positive  $\gamma_0$  supposed to be known. In this framework the previous optimal rate of order  $n^{-2/(d+2)}$  is still preserved. Moreover, if  $A'_1(i)$  holds with  $\gamma_0 < 1$ , a “parametric” rate is even possible in a sense made precise below.

**Theorem 3.2.** According to the value of  $\gamma_0$  we assume that  $\delta_n \geq \delta_n^*(\gamma_0)$  defined as

$$\delta_n^*(\gamma_0) := d_1 h_n^d \mathbf{1}_{\{\gamma_0 < 1\}} + d_2 h_n^d \ln(h_n^{-d}) \mathbf{1}_{\{\gamma_0 = 1\}} + d_3 h_n^{d/\gamma_0} \mathbf{1}_{\{\gamma_0 > 1\}}, \quad 0 < d_1, d_2, d_3 < \infty. \quad (3.1)$$

(i) Then under  $A_0(ii)$  and  $A'_1$

$$\limsup_{n \rightarrow \infty} nh_n^d \cdot \text{IV}(\hat{f}_n^H) \leq 1 + C_{\gamma_0};$$

(ii) If in addition  $A_0(i)$  holds with  $h_n = cn^{-\frac{1}{d+2}}$ ,  $0 < c < \infty$ , then

$$\limsup_{n \rightarrow \infty} n^{\frac{2}{d+2}} \cdot \text{MISE}(\hat{f}_n^H) \leq \frac{c^2}{12} \mathcal{R}_d(f') + \frac{1}{c^d} (1 + C_{\gamma_0}),$$

where  $C_{\gamma_0}$  is a positive constant which depends upon  $\gamma_0$  (see its explicit form in proofs).

**Remark 3.5.** Using  $A_0$  with either  $A_1$  or  $A'_1$ , our results in Theorems 3.1 and 3.2 are similar to those derived with independent variables by Scott (1979), Freedman et Diaconis (1981), and therefore extended by Lecoutre (1985) to the  $d$ -dimensional setup. Thus we recover (in limsup) the same optimal rate  $n^{-2/(d+2)}$  in terms of MISE but also the same asymptotic constant up to the additional constant  $C$  or  $C_{\gamma_0}$ , which is a non negligible remainder of the covariance term. Furthermore, note that if  $\delta_n \gg \delta_n^*(\gamma_0)$ , we can remove  $C_{\gamma_0}$  in Theorem 3.2 to get the exact limiting constant of the i.i.d. case with  $h_n = cn^{-1/(d+2)}$ .

**Remark 3.6 (Optimal bin width).** For the purpose of several applications, one has now at disposal an asymptotically optimal bin width of the form  $h_n = cn^{-1/3}$  ( $d = 1$ ), but with non specified factor  $c$ . The important problem of finding an optimal value of  $c$  is not studied here; however many practical formulas are available in the literature by minimizing asymptotically the MISE in the independent case. Numerous solutions arise from the fact that  $c$  is explicitly a functional of the unknown density and its derivative. Among possibilities, if one may assume that the shape of  $f$  is not far from a Gaussian distribution, Scott (1979) proposes taking  $c \simeq 3.49\hat{\sigma}$ , with  $\hat{\sigma}$  denoting some variance estimate based upon observations; and Freedman et Diaconis (1981) suggest taking  $c$  to be twice the inter-quartile range of the data. Further discussion is beyond the scope of this chapter and, for a complete review of existing bin width selection procedures, we refer *e.g.* to Birgé et Rozenholc (2006) and the references therein.

Assume now that the sample paths of the underlying process are sufficiently irregular, corresponding to some value of  $\gamma_0 < 1$ , the following corollary follows as a straightforward consequence of Theorem 3.2.

**Corollary 3.1.** *If  $A_0$  and  $A'_1$  are satisfied with  $\gamma_0 < 1$  then both choices  $h_n = o(T_n^{-1/2})$  and  $\delta_n = d_0 h_n^d$ ,  $0 < d_0 < \infty$ , lead to*

$$\limsup_{n \rightarrow \infty} T_n \cdot \text{MISE}(\hat{f}_n^H) \leq d_0 + 2 \left\{ \frac{\|\varphi\|_1 u_0^{1-\gamma_0}}{1-\gamma_0} + u_0 \|f\|_\infty + \left[ (u_1 - u_0) \sup_{u \in [u_0, u_1]} \pi(u) + \int_{u_1}^\infty \pi(u) du \right] \|k\|_1 \right\}.$$

**Remark 3.7.** The rate of order  $T_n^{-1}$  exhibited in Corollary 3.1 deserves some words of explanation. Indeed, to understand, it suffices to observe that both optimal choices  $\delta_n^*$  ( $\gamma_0 < 1$ ) and  $h_n$  in Theorem 3.2 simply lead to  $T_n \simeq n^{2/(d+2)}$ : this rate is therefore well parametric with respect to  $T_n$  but yet optimal with respect to  $n$ . Furthermore, the interesting point is that here the optimal rate is obtained with the smallest choice of  $\delta_n$ , which enables to optimize the time of observation of the process. In continuous-time, let us recall that both optimal and parametric rates are given in Lejeune (2006).

In the following proposition we state the exact limit of the pointwise variance of  $\hat{f}_n^H(x)$  associated with admissible sampling. It is in fact a direct transposition from kernel to histogram estimators of a result by Bosq (1997) (Proposition 7.1. (i)). The proof will be therefore omitted.

**Proposition 3.1.** *Let  $x \in \mathbb{R}^d$  and assume that*

(i)  $\|g_u\|_\infty \leq \pi(u)$  where  $(1 + \pi(u))\pi(u)$  is integrable over  $[0, +\infty)$  and  $u\pi(u)$  is bounded and ultimately decreasing. Furthermore  $g_u(\cdot, \cdot)$  is continuous at  $(x, x)$ ;

(ii)  $\sup_{(y,z) \in \mathbb{R}^{2d}} \left| \sum_{r=1}^{\infty} \delta_n g_r \delta_n(y, z) - \int_0^\infty g_u(y, z) du \right| \rightarrow 0$  as  $\delta_n \downarrow 0^+$ ,

then

$$\lim_{n \rightarrow \infty} T_n \cdot \text{Var}(\hat{f}_n^H(x)) = 2 \int_0^\infty g_u(x, x) du.$$

### 3.3 Frequency polygon

Given a histogram, the *frequency polygon* results from a natural smoothing to get a continuous estimator, but its use should be justified also by theoretical reason since we immediately improve the weak order  $h_n^2$  inherent to the bias of histograms. The main properties are gathered in Scott (1992) (Chapter 4) within the i.i.d. setup. The dependent case was then treated by Carbon *et al.* (1997), and recently extended to the random fields by Carbon (2006). In continuous-time, Lejeune (2005, 2006) established both optimal and full rates of MISE and asymptotic normality; the extension to the random fields is done in a submitted work by Bensaïd et Dabo-Niang (2007). Particularly, all these investigations owe much to the surprising result of Scott (1985) that, frequency polygons and kernel density estimators share the same rate of convergence  $n^{-4/5}$  in terms of MISE. For the sake of simplicity, we shall confine attention to the real case ( $d = 1$ ).

Notations. For convenience,  $f'$  and  $f''$  denote the first and second derivatives of  $f$  and we define the roughness of  $f''$  by  $\mathcal{R}(f'') := \int_{\mathbb{R}} f''(x)^2 dx$ .

#### 3.3.1 Definition and assumptions

The frequency polygon is simply constructed by connecting the mid-points of the histogram heights with straight lines.

**Definition 3.3.** *One describes the frequency polygon  $\hat{f}_n^{FP}$  by the expression*

$$\hat{f}_n^{FP}(x) = \sum_j \left[ \left( \frac{x - c_j}{h_n} \right) \hat{f}_{j+1} + \left( \frac{c_{j+1} - x}{h_n} \right) \hat{f}_j \right] \mathbb{1}_{[c_j, c_{j+1}[}(x), \quad x \in \mathbb{R}.$$

There exists in the literature some alternative definitions in modifying the method of interpolation as the *edge frequency polygon* introduced by Jones *et al.* (1998) or its extended form by Dong et Zheng (2001). These estimators share the same order  $n^{-4/5}$  of MISE – but with different asymptotic constants – for densities admitting two continuous derivatives. Unfortunately, they are very complicated to define and handle in higher dimensions (see *e.g.* Scott (1985) for the two-dimensional setup and Scott (1992) for the  $d$ -dimensional setup).

Other modified histogram density estimators have been investigated to improve the competitiveness of standard histograms. Among them, the *piecewise linear histogram* given

by Beirlant *et al.* (1999) has noteworthy the same rate of order  $n^{-2/5}$  ( $d = 1$ ) as kernel estimators with respect to the mean integrated absolute error (against  $n^{-1/3}$  for standard histograms) and, unlike frequency polygons, it admits a unique and simple multivariate extension but with disadvantage of possible negative values. A generalized estimate was therefore proposed by Berline *et al.* (2002) to circumvent this problem without altering the  $n^{-2/5}$ -rate. Comparisons between all these estimates are beyond the scope of this chapter and we refer to the graphical and numerical examples performed in the last two cited papers with independent data.

Before proceeding to the frequency polygon study we need to reformulate slightly our conditions  $A_0$  on  $f$ .

### Assumptions $A'_0$

- (i)  $f \in \mathcal{C}^2(\mathbb{R})$ ,  $f(x) > 0$  for all  $x$  almost everywhere,  $f'' \in L_1(\mathbb{R})$  and  $f, f'' \in L_2(\mathbb{R})$ ;
- (ii)  $|f''(x) - f''(y)| \leq l_0|x - y|^\nu$ ,  $l_0 > 0$ ,  $\nu \in ]0, 1]$ , for  $(x, y) \in \mathbb{R}^2$ ;
- (iii)  $f$  is continuous at  $x$  and  $\|f\|_\infty < \infty$ .

### 3.3.2 Rates of convergence

The next lemma gives the ISB contribution.

**Lemma 3.2.** *If Assumptions  $A'_0(i)(ii)$  are satisfied then*

$$\text{ISB}(\hat{f}_n^{FP}) = \frac{49}{2880} \mathcal{R}(f'') h_n^4 + o(h_n^4).$$

**Remark 3.8.** The nice order  $h_n^4$  is much better compared with histograms and familiar for more sophisticated density estimators as kernel estimators. The proof is similar to those written in Scott (1985) except that, here, we use the Hölderian continuity condition on  $f''$ , namely  $A'_0(ii)$ , and consequently, we do not need to suppose the existence of the third derivative of  $f$ . As emphasized with histograms the bias term still does not depend on the sampling scheme.

**[RS]-[JS]** – Using our earlier analysis on histograms and a new suitable choice of  $h_n$  we have the optimal rate for frequency polygons. Note that constants  $C$  and  $C_{\gamma_0}$  are unchanged.

**Theorem 3.3.** *(i) Under  $A'_0(iii)$  and  $A_1$  and if  $f^{1-\frac{1}{p}} \in L_1(\mathbb{R})$  for  $1 < p < \rho - 1$ , then*

$$\limsup_{n \rightarrow \infty} n h_n \cdot \text{IV}(\hat{f}_n^{FP}) \leq \frac{2}{3} + C;$$

*(ii) If in addition  $A'_0(i)(ii)$  hold then the choice  $h_n = c n^{-\frac{1}{5}}$ ,  $0 < c < \infty$ , yields*

$$\limsup_{n \rightarrow \infty} n^{\frac{4}{5}} \cdot \text{MISE}(\hat{f}_n^{FP}) \leq \frac{49}{2880} c^4 \mathcal{R}(f'') + \frac{1}{c} \left( \frac{2}{3} + C \right).$$

[HFS] – Finally, taking into account the local regularity of sample paths through high frequency observations, we recover both optimal and “parametric” rates of the quadratic error.

**Theorem 3.4.** *According to the values of  $\gamma_0$ , we consider optimal choices  $\delta_n^*(\gamma_0)$  given by (3.1).*

(i) *Then under  $A'_0(iii)$  and  $A'_1$*

$$\limsup_{n \rightarrow \infty} nh_n \cdot \text{IV}(\hat{f}_n^{FP}) \leq \frac{2}{3} + C_{\gamma_0};$$

(ii) *If in addition  $A'_0(i)$  hold with  $h_n = cn^{-\frac{1}{5}}$ ,  $0 < c < \infty$ , then*

$$\limsup_{n \rightarrow \infty} n^{\frac{4}{5}} \cdot \text{MISE}(\hat{f}_n^{FP}) \leq \frac{49}{2880} c^4 \mathcal{R}(f'') + \frac{1}{c} \left( \frac{2}{3} + C_{\gamma_0} \right).$$

Moreover any choice of  $\delta_n$  satisfying  $\delta_n/\delta_n^*(\gamma_0) \rightarrow \infty$  as  $n \rightarrow \infty$  allows to remove  $C_{\gamma_0}$ .

**Remark 3.9.** In both Theorems 3.3 and 3.4 we exhibit (in limsup) the same  $n^{-4/5}$ -consistency obtained in Scott (1985) with i.i.d. observations, in Carbon *et al.* (1997), and in Lejeune (2005). This rate is also optimal in some minimax sense (see Bosq (1998), Chapter 4). If we take  $p = \rho - 1$  in Theorem 3.3 the rates are still valid up to a decayed constant. For a practical bin width value, among possibilities, Scott suggests taking  $c \simeq 2.15\hat{\sigma}$  in accordance with the Gaussian rule.

Now, in the situation of irregular sample paths (*i.e.*  $\gamma_0 < 1$ ), frequency polygons have also a “parametric” rate.

**Corollary 3.2.** *If  $A'_0$  and  $A'_1$  are satisfied with  $\gamma_0 < 1$  then both choices  $h_n = o(T_n^{-1/4})$  and  $\delta_n = d_0 h_n$ ,  $0 < d_0 < \infty$ , lead to*

$$\begin{aligned} \limsup_{n \rightarrow \infty} T_n \cdot \text{MISE}(\hat{f}_n^{FP}) \leq & \frac{2d_0}{3} + 2 \left\{ \frac{\|\varphi\|_1 u_0^{1-\gamma_0}}{1-\gamma_0} + u_0 \|f\|_\infty \right. \\ & \left. + \left[ (u_1 - u_0) \sup_{u \in [u_0, u_1]} \pi(u) + \int_{u_1}^\infty \pi(u) du \right] \|k\|_1 \right\}. \end{aligned}$$

For completeness, the exact limit of the pointwise variance of  $\hat{f}_n^{FP}(x)$  follows straightforwardly from Proposition 3.1. The proof will be therefore omitted.

**Proposition 3.2.** *Under conditions of Proposition 3.1 one has*

$$\lim_{n \rightarrow \infty} T_n \cdot \text{Var}(\hat{f}_n^{FP}(x)) = 2 \int_0^\infty g_u(x, x) du, \quad x \in \mathbb{R}.$$

### 3.4 Examples

A few examples of processes satisfying our conditions are given in this section. These include a certain class of  $d$ -dimensional homogeneous diffusion processes defined as solutions of the stochastic differential equation

$$dX_t = S(X_t)dt + \sigma(X_t)dW_t, \quad X_0, \quad 0 \leq t \leq T, \quad (3.2)$$

where  $\{W_t, 0 \leq t \leq T\}$  is a standard  $d$ -dimensional Wiener process,  $S : \mathbb{R}^d \rightarrow \mathbb{R}^d$  is a bounded Borel function,  $\sigma$  is a bounded continuous non-degenerate matrix  $d \times d$  function and  $X_0$  is an initial value of  $X_t$ .

**Case 1 :**  $\gamma_0 = 1/2$ ,  $d = 1$ . The usefulness of unidimensional diffusion processes is well-known for instance in financial, biological or engineering applications subject to random fluctuations. Here we assume that  $X^T$  is a Markov diffusion process with transition density  $p_u(z|y)$  so that  $f_u(y, z) = f(y)p_u(z|y)$ . The transition density is typically very hard to derive explicitly from the forward and backward Kolmogorov equations; but Leblanc (1997) (Proposition 11) gives conditions on  $S$  and  $\sigma$  to get strictly stationary and geometrically  $\beta$ -mixing processes from Eq. (3.2) which satisfy  $p_u(z|y) \leq Cu^{-1/2}$ ,  $0 < u \leq 1$ . One may find also this inequality in Veretennikov (1999) if  $\sigma$  is non-degenerate, and  $S$  and  $\sigma$  are bounded and of class  $C^\alpha$ -Hölder continuous. This implies in particular that  $A'_1(i)$  is fulfilled with  $\gamma_0 = 1/2$  and, for instance, in the following situations.

**Example 3.1 (Ornstein-Uhlenbeck process).** Let the observed process  $X^T$  be from (3.2) with

$$S(x) = \theta(\mu - x) \quad \text{and} \quad \sigma(x) \equiv \sigma, \quad x \in \mathbb{R}, \quad (\theta, \sigma) > 0, \quad \mu \in \mathbb{R}.$$

The invariant Gaussian distribution has mean value  $\mu$  and variance  $\sigma^2/2\theta$  and, for any  $(y, z) \in \mathbb{R}^2$ , the Gaussian transition density is given by

$$p_u(z|y) = \sqrt{\frac{\theta}{\pi\gamma_u^2}} \exp \left\{ -\frac{\theta[z - \mu - (y - \mu)e^{-\theta u}]^2}{\gamma_u^2} \right\},$$

where  $\gamma_u^2 := \sigma^2(1 - e^{-2\theta u})$ . Then  $\gamma_u^2 \geq \sigma^2(1 - e^{-\theta a})u$ , for any  $0 < u \leq 1$ , guarantees the desired inequality. In the finance literature this process is often referred to as the Vasicek's Model (Vašíček (1977)) for modelling the dynamics of a short-time interest rate.

**Example 3.2 (Double Exponential process).** Consider the diffusion process with a discontinuous drift function

$$S(x) = -\theta \operatorname{sgn}(x) \quad \text{and} \quad \sigma(x) \equiv 1, \quad x \in \mathbb{R}, \quad \theta > 0.$$

It has an invariant density  $f(x) = \theta \exp(-2\theta|x|)$  and there exists a constant  $C_\theta$  (see Leblanc (1997)) such that

$$p_u(z|y) \leq \frac{C_\theta}{\sqrt{2\pi u}}, \quad (y, z) \in \mathbb{R}^2, \quad 0 < u \leq 1.$$

**Case 2 :**  $\gamma_0 > 1$ ,  $d \geq 1$ ,  $\sigma \equiv 1$ . If the drift  $S$  is a bounded vector field on  $\mathbb{R}^d$ , it is shown in Qian *et al.* (2003) that the choice  $\gamma_0 = d/2$ ,  $d \geq 2$ , does work. This result has been extended to the case of unbounded drift *with at most linear growth*, which is important in financial mathematics (see Qian et Zheng (2004)).

**Case 3 :**  $\gamma_0 = 1$ ,  $d \geq 1$ . Leaving diffusion processes out, it is well-established (see *e.g.* Blanke et Bosq (1997)) that unidimensional and mean-square differentiable Gaussian processes fulfil  $A'_1(i)$  with  $\gamma_0 = 1$ . Combining the previous cases, one may also construct examples in a multidimensional setting. For instance, if  $X_1^T$  and  $X_2^T$  are two real and independent Gaussian processes with  $\gamma_1 = \gamma_2 = 1/2$ ,  $A'_2(i)$  holds provided that  $\gamma_0 = \gamma_1 + \gamma_2 = 1$ .

### 3.5 Simulations

Simulations are performed with R to illustrate our previous results with comparisons between histogram, frequency polygon and density kernel estimators. To do so, we focus on the real Gaussian case for two stationary processes : Ornstein-Uhlenbeck (non differentiable,  $\gamma_0 = 0.5$ ) and CAR(2) (differentiable,  $\gamma_0 = 1$ ). The study is then similar to those made by Bosq et Blanke (2007) (Chapter 8, Section 3) with kernel type estimators.

**Ornstein-Uhlenbeck process :** we first return to the possible solutions  $X^T$  of the stochastic differential equation

$$dX_t = -\theta X_t dt + \sigma dW_t, \quad X_0, \theta > 0, 0 \leq t \leq T,$$

that is

$$X_t = \sigma \int_0^t e^{-\theta(t-s)} dW_s.$$

Then trajectories are obtained by iterating the following exact autoregressive formulation :

$$X_{(i+1)\delta} = e^{-\theta\delta} X_{i\delta} + Z_{(i+1)\delta}, \quad X_0 \sim \mathcal{N}(0, \sigma^2/2\theta), \quad \delta > 0, 0 \leq i \leq n,$$

where  $\{Z_{i\delta}, 0 \leq i \leq n\}$  is a Gaussian white noise of variance  $\sigma^2(1 - e^{-2\theta\delta})/2\theta$ . We choose  $\theta = 1$  and  $\sigma = \sqrt{2}$  such that the  $X_t$ 's have the invariant distribution  $\mathcal{N}(0, 1)$ .

**CAR(2) process :** let  $X^T$  be a continuous-time autoregressive process of order 2 (CAR(2)) defined as the solution of the following linear stochastic differential equation

$$X_t^{(2)} - \alpha_2 X_t^{(1)} - \alpha_1 X_t = \sigma W_t^{(1)}, \quad 0 \leq t \leq T, \quad (3.3)$$

where  $X_\bullet^{(i)}$  denotes the  $i$ th mean-square derivative of  $X_\bullet$  and  $\{W_t, 0 \leq t \leq T\}$  is a standard Wiener process, independent of  $X_0$ ,  $X_0^{(1)}$  and  $X_0^{(2)}$ . Note that the Ornstein-Uhlenbeck process is not more than a CAR(1) process (with  $\alpha_2 = 0$  and  $X_t^{(2)}$  replaced by  $X_t^{(1)}$ ). We are interested only in stationary solutions of (3.3). For instance, the choices  $\alpha_1 = -1/2$ ,  $\alpha_2 = -2$  and  $\sigma = \sqrt{2}$  leads to a Gaussian process which is stationary, mean-square differentiable ( $\gamma_0 = 1$ ) and with marginal distribution  $\mathcal{N}(0, 1)$ . Then trajectories follow from

the exact discrete formulation :

$$\begin{pmatrix} X_{(i+1)\delta} \\ X_{(i+1)\delta}^{(1)} \end{pmatrix} = e^{\mathbb{A}\delta} \begin{pmatrix} X_{i\delta} \\ X_{i\delta}^{(1)} \end{pmatrix} + \begin{pmatrix} Z_{(i+1)\delta}^{(1)} \\ Z_{(i+1)\delta}^{(2)} \end{pmatrix}, \quad \delta > 0, \quad 0 \leq i \leq n,$$

with

$$X_0 \sim \mathcal{N}(0, 1), \quad X_0^{(1)} \sim \mathcal{N}\left(0, \frac{1}{2}\right),$$

where  $\mathbb{A}$ , a matrix  $2 \times 2$ , and  $Z$ , the covariance of  $\left(Z_{(i+1)\delta}^{(1)}, Z_{(i+1)\delta}^{(2)}\right)'$ , have explicit but rather complicated expressions that can be computed from *e.g.* Tsai et Chan (2000).

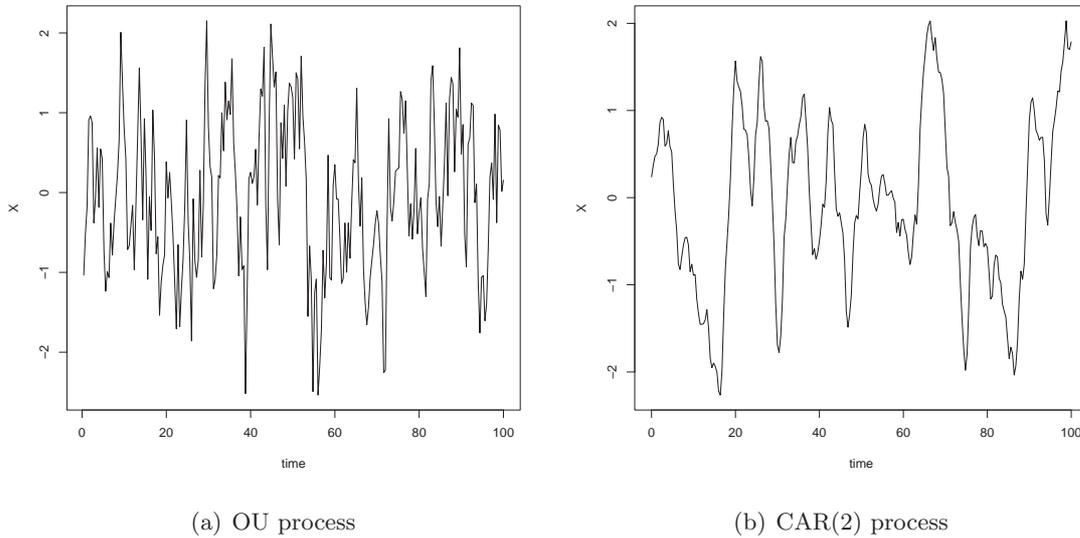


FIG. 3.1 – Simulated sample paths with  $n = 250$  and  $\delta = 0.4$

Furthermore, we shall study the effect of various  $\delta$  upon efficiency in estimation. For our purpose, we simulate  $N_r$  of Ornstein-Uhlenbeck (OU) or CAR(2) processes and compute an approximation of the average squared error (ASE) defined by

$$\text{ASE}(\delta) := \frac{1}{N_r} \sum_{r=1}^{N_r} \left( \frac{1}{n} \sum_{i=1}^n \left( \hat{f}_n(x_{i\delta,r}) - f(x_{i\delta,r}) \right)^2 \right),$$

where  $f$  is the  $\mathcal{N}(0, 1)$  density and  $\hat{f}_n(x_{i\delta,r})$  is the (Gaussian) estimate, associated to the  $j$ th simulated sampled path and based on  $n$   $\delta$ -equidistant observations. In particular, in the case of strongly mixing processes with kernel density estimates, Vieu (1991) showed (Theorem 2.1) that ASE may be a good approximation of MISE since one has

$$\lim_{n \rightarrow \infty} \sup_{h_n \in [b_1 n^{-u}, b_2 n^{-v}]} \frac{|ASE(\hat{f}_n(x)) - MISE(\hat{f}_n)|}{ASE(\hat{f}_n(x))} = 1 \quad p.s.,$$

where  $0 < v \leq 1/5 \leq u < 2/9$  and where  $b_1, b_2$  are positive constants. Observe that  $[b_1 n^{-u}, b_2 n^{-v}]$  contains asymptotic minimizers of ASE and MISE of order  $n^{-1/5}$ .

From a simulated sample path of the OU process (with  $\tau_n = 10^{-2}$ ), we construct the histogram and the frequency polygon from  $n = 1000$  observations with  $h_n = 0.3$ , which is the optimal bin width for histogram (with the “Scott rule”). We select deliberately a high sampling size to get a somewhat robust estimation, that allows us to put ahead the effect of the sampling step  $\delta_n$ . We test the maximal sampling rate  $\delta_n = 0.01$  ( $T_n = 10$ ) and one close to the threshold  $\delta_n^*(\gamma_0)$  with  $\delta_n = 0.35$  ( $T_n = 350$ ). Figure 3.2 shows that estimation is not consistent in the first case, but that approximation turns to be good for  $\delta_n$  equal to 0.35. From a simulated sample path of the CAR(2) process, it should be noticed that  $\delta_n = 0.35$  (corresponding to a good choice in the OU case) is not adapted for this smooth process whereas  $\delta_n = 0.35$  ( $T_n = 750$ ) furnishes a much better approximation of the  $\mathcal{N}(0, 1)$  density.

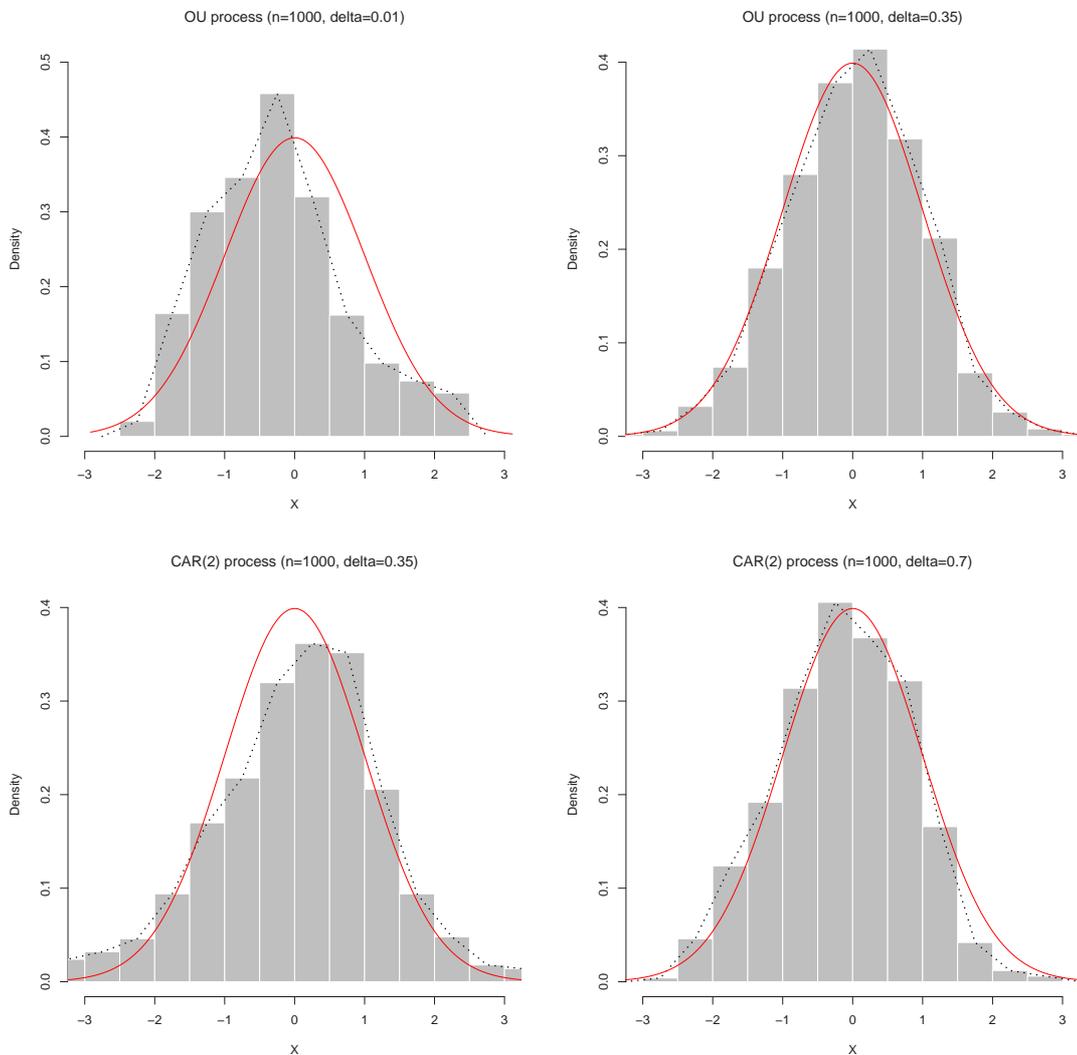


FIG. 3.2 – Frequency polygon (dotted line) and  $\mathcal{N}(0, 1)$  density (solid line)

For both processes with 250 observations, Figure 3.3 clearly establishes that choices of

over-small  $\delta$  yield to inconsistent estimation. This phenomenon is intensified for the regular process CAR(2). Finally, we note that for larger  $\delta$ , a stabilization occurs around the ASE obtained with 250 i.i.d. observations (bottom dotted line on the figure). For the sake of simplicity, note that the ASE of frequency polygon has been computed with the optimal bin width  $h_n = 0.5$  relative to histogram. Nevertheless, the hierarchy between histogram, frequency polygon and Gaussian kernel estimate appears sufficiently explicit. With the best choice  $h_n = 0.65$ , the frequency polygon should be even much closer to the kernel estimator.

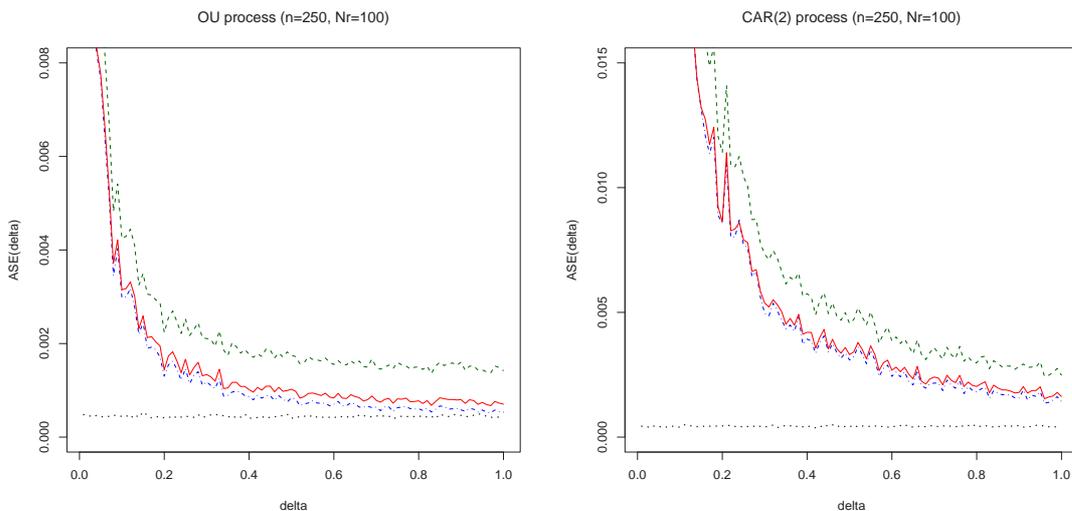


FIG. 3.3 – ASE of histogram (dashed line), frequency polygon (solid line) and Gaussian kernel (dotdash line)

### 3.6 Discussion

In this work we derive the optimal  $L_2$ -rates of two computationally advantageous density estimators in a specific setup where observations are discretely sampled from a continuous-time process. For practical considerations, we have used three different sampling procedures letting the statistician more options in accordance with the features of real data. Therefore our main results state that all designs either random or deterministic lead to the optimal rates  $n^{-2/(d+2)}$  for histograms and  $n^{-4/5}$  ( $d = 1$ ) for frequency polygons with respect to the MISE convergence, which are those derived in the i.i.d. case. Especially, we have focused on a “high frequency” sampling because of some parallels with the idealized continuous-time framework. This setting enlightens the inner dependence between the local properties of sample paths and the rates of convergence, which might be explained as follows : irregular sample paths carry much more information than regular ones where the correlation between two successive variables is much stronger. Consequently, we infer that more the paths are irregular ( $\gamma_0 < 1$ ) more the time of observation would be shortened, in

order to obtain a good behavior of both estimators. Numerical studies also well confirm these theoretical results.

### 3.7 Proofs

Throughout this section, we detail the proofs of Theorems 3.1, 3.2, 3.3 and 3.4. In order to do this, some auxiliary lemmas are necessary to derive upper bound expressions for the variance of  $\hat{f}_n^H(x)$ ,  $x \in \pi_{nj}$ , which will depend on the sampling scheme being used. Now denote  $\sigma(X)$  the  $\sigma$ -algebra of events generated by a random variable  $X$  and  $\|X\|_q = (\mathbb{E}|X|^q)^{1/q}$  with  $1 \leq q < \infty$ .  $X \in L_q(P)$  means also that  $\|X\|_q < \infty$ . We recall the following useful covariance inequality as written in Bosq (1998) (p. 21).

**Lemma 3.3 (Davydov's inequality).** *Let  $X \in L_q(P)$  and  $Y \in L_r(P)$  such that  $q > 1$ ,  $r > 1$  and  $\frac{1}{r} + \frac{1}{q} = 1 - \frac{1}{p}$ , then*

$$|\text{Cov}(X, Y)| \leq 2p \left[ 2\alpha(\sigma(X), \sigma(Y)) \right]^{1/p} \|X\|_q \|Y\|_r.$$

#### 3.7.1 Histogram

- **Variance bounds with random sampling**

**Lemma 3.4 ([RS]).** *If  $A_0(ii)$  and  $A_1$  hold then we obtain for  $1 < p \leq \rho - 1$  :*

$$\begin{aligned} nh_n^d \cdot \text{Var}(\hat{f}_j) &\leq f(\xi_j)(1 - h_n^d f(\xi_j))(1 + 2u_0 \bar{h}_0) + 2\pi_0 \bar{h}_0 k(\dot{\xi}_j) \cdot h_n^\varepsilon \\ &\quad + \frac{4p^2(2a_0)^{1/p} \bar{h}_0}{\rho - p} f(\xi_j)^{1-\frac{1}{p}} \cdot h_n^{\frac{1}{p}\{(d-\varepsilon)(\rho-p)-d\}}, \end{aligned} \quad (3.4)$$

with  $0 \leq \varepsilon \leq d \left(1 - \frac{1}{\rho-p}\right)$  and  $(\xi_j, \dot{\xi}_j) \in \pi_{nj}^2$ .

**Lemma 3.5 ([JS]).** *Under the same conditions as in Lemma 3.4 and  $1 < p \leq \rho - 1$  :*

$$\begin{aligned} nh_n^d \cdot \text{Var}(\hat{f}_j) &\leq f(\xi_j)(1 - h_n^d f(\xi_j)) \left(1 + 2 \left\lceil \frac{u_0}{\delta} \right\rceil\right) + 2\pi_0 k(\dot{\xi}_j) \left(h_n^\varepsilon - \left\lceil \frac{u_0}{\delta} \right\rceil h_n^d\right) \\ &\quad + \frac{4p^2(2a_0)^{1/p}}{(\rho - p)\delta^{\frac{p}{p}}} f(\xi_j)^{1-\frac{1}{p}} \cdot h_n^{\frac{1}{p}\{(d-\varepsilon)(\rho-p)-d\}} (1 - 2h_n^{d-\varepsilon})^{1-\frac{p}{p}}, \end{aligned} \quad (3.5)$$

with  $0 \leq \varepsilon \leq d \left(1 - \frac{1}{\rho-p}\right)$  and  $(\xi_j, \dot{\xi}_j) \in \pi_{nj}^2$ .

For further use the proofs are written in a more general context studying the covariance.

**Proof of Lemma 3.4** For any  $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$ , we suppose the existence of two indexes  $j_1(x, n)$  and  $j_2(y, n)$  in  $\mathbb{Z}^d$  such that  $x \in \pi_{j_1(x, n)} (= \pi_{nj_1})$  and  $y \in \pi_{j_2(y, n)} (= \pi_{nj_2})$ . Thus

$$\hat{f}_n^H(x) = \hat{f}_{j_1} = \frac{1}{nh_n^d} \sum_{k=1}^n \mathbb{1}_{\pi_{nj_1}}(X_{t_k}), \quad \hat{f}_n^H(y) = \hat{f}_{j_2} = \frac{1}{nh_n^d} \sum_{k=1}^n \mathbb{1}_{\pi_{nj_2}}(X_{t_k}),$$

and

$$\begin{aligned} nh_n^d \cdot \text{Cov}(\hat{f}_{j_1}, \hat{f}_{j_2}) &= \frac{1}{nh_n^d} \sum_{k=1}^n \text{Cov}(\mathbb{1}_{\pi_{nj_1}}(X_{t_k}), \mathbb{1}_{\pi_{nj_2}}(X_{t_k})) \\ &\quad + \frac{2}{nh_n^d} \sum_{p=1}^{n-1} \sum_{q=p+1}^n \text{Cov}(\mathbb{1}_{\pi_{nj_1}}(X_{t_p}), \mathbb{1}_{\pi_{nj_2}}(X_{t_q})) =: V_n + C_n. \end{aligned}$$

Set  $p_k := P(X_0 \in \pi_{nk})$ ,  $k \in \mathbb{Z}^d$ . The ‘‘variance’’ term  $V_n$  is easy to compute.

$$V_n = \frac{1}{nh_n^d} \sum_{k=1}^n \text{Cov}(\mathbb{1}_{\pi_{nj_1}}(X_0), \mathbb{1}_{\pi_{nj_2}}(X_0)) = \frac{1}{h_n^d} (P(X_0 \in \pi_{nj_1}, X_0 \in \pi_{nj_2}) - p_{j_1} p_{j_2}).$$

Since  $f$  is continuous there exists at least one point  $\xi_j \in \pi_{nj}$  such that  $\int_{\pi_{nj}} f(x) dx = h_n^d f(\xi_j)$ . Then if  $j_1 \neq j_2$ , we get

$$V_n = -\frac{1}{h_n^d} p_{j_1} p_{j_2} = -h_n^d f(\xi_{j_1}) f(\xi_{j_2}),$$

where  $(\xi_{j_1}, \xi_{j_2}) \in \pi_{nj_1} \times \pi_{nj_2}$ . Otherwise if  $j_1 = j_2 = j$ :

$$V_n = \frac{1}{h_n^d} p_j (1 - p_j) = f(\xi_j) (1 - h_n^d f(\xi_j)).$$

Now let us turn to the ‘‘covariance’’ term  $C_n$ . By weak stationarity and since  $t_p - t_q \stackrel{d}{=} t_{p-q}$ , where ‘‘ $\stackrel{d}{=}$ ’’ means equality in distribution, we have

$$\begin{aligned} C_n &= \frac{2}{nh_n^d} \sum_{r=1}^{n-1} \sum_{p=1}^{n-r} \text{Cov}(\mathbb{1}_{\pi_{nj_1}}(X_0), \mathbb{1}_{\pi_{nj_2}}(X_{t_{p+r}-t_p})) \\ &= \frac{2}{h_n^d} \sum_{r=1}^{n-1} \left(1 - \frac{r}{n}\right) \int_0^\infty \text{Cov}(\mathbb{1}_{\pi_{nj_1}}(X_0), \mathbb{1}_{\pi_{nj_2}}(X_u)) g_r(u) du \\ &=: C_{n,1} + C_{n,2} + C_{n,3}, \end{aligned}$$

where

$$\begin{aligned} C_{n,1} &:= \frac{2}{h_n^d} \sum_{r=1}^{n-1} \left(1 - \frac{r}{n}\right) \int_0^{u_0} \text{Cov}(\mathbb{1}_{\pi_{nj_1}}(X_0), \mathbb{1}_{\pi_{nj_2}}(X_u)) g_r(u) du, \\ C_{n,2} &:= \frac{2}{h_n^d} \sum_{r=1}^{n-1} \left(1 - \frac{r}{n}\right) \int_{u_0}^{h_n^{-d+\varepsilon}} \text{Cov}(\mathbb{1}_{\pi_{nj_1}}(X_0), \mathbb{1}_{\pi_{nj_2}}(X_u)) g_r(u) du, \\ C_{n,3} &:= \frac{2}{h_n^d} \sum_{r=1}^{n-1} \left(1 - \frac{r}{n}\right) \int_{h_n^{-d+\varepsilon}}^\infty \text{Cov}(\mathbb{1}_{\pi_{nj_1}}(X_0), \mathbb{1}_{\pi_{nj_2}}(X_u)) g_r(u) du, \end{aligned}$$

for some  $0 \leq \varepsilon < d$  to be specified later. Let us recall that  $h(u) = \sum_{r=1}^\infty g_r(u)$ , one seeks to upper bound each covariance term. First, by Cauchy-Schwarz inequality and Fubini’s theorem

$$\begin{aligned} |C_{n,1}| &\leq \frac{2}{h_n^d} \sqrt{\text{Var}(\mathbb{1}_{\pi_{nj_1}}(X_0))} \sqrt{\text{Var}(\mathbb{1}_{\pi_{nj_2}}(X_0))} \int_0^{u_0} h(u) du \\ &\leq 2u_0 \bar{h}_0 \sqrt{f(\xi_{j_1}) f(\xi_{j_2}) (1 - h_n^d f(\xi_{j_1})) (1 - h_n^d f(\xi_{j_2}))}. \end{aligned}$$

Then  $A_1(i)$  and Fubini imply

$$\begin{aligned} |C_{n,2}| &\leq \frac{2}{h_n^d} \int_{u_0}^{h_n^{-d+\varepsilon}} \iint_{\pi_{nj_1} \times \pi_{nj_2}} \sup_{y \in \mathbb{R}^d} |g_u(x, y)| dx dy h(u) du \\ &\leq 2\pi_0 \bar{h}_0 k(\dot{\xi}_{j_1}) \cdot h_n^\varepsilon, \end{aligned}$$

where  $\dot{\xi}_{j_1} \in \pi_{nj_1}$ .

Now, it is clear that for  $n$  large enough we have  $h_n^{-d+\varepsilon} \geq u_1$ . So using Davydov's inequality (cf. Lemma 3.3) with mixing condition  $A_1(ii)$  and Fubini, and for any  $(p, q) \in ]1, \rho - 1] \times \left[1 + \frac{\rho}{\rho-2}, \infty\right]$  such that  $\frac{2}{q} + \frac{1}{p} = 1$  :

$$\begin{aligned} |C_{n,3}| &\leq \frac{2}{h_n^d} \int_{h_n^{-d+\varepsilon}}^{\infty} 2p2^{1/p} \|\mathbf{1}_{\pi_{nj_1}}(X_0)\|_q \|\mathbf{1}_{\pi_{nj_2}}(X_u)\|_q (\alpha_X^{(2)}(u))^{1/p} h(u) du \\ &\leq 4p(2a_0)^{1/p} \bar{h}_0 \sqrt{f(\xi_{j_1})^{1-\frac{1}{p}} f(\xi_{j_2})^{1-\frac{1}{p}}} \cdot h_n^{-\frac{d}{p}} \int_{h_n^{-d+\varepsilon}}^{\infty} u^{-\frac{\rho}{p}} du \\ &\leq \frac{4p^2(2a_0)^{1/p} \bar{h}_0}{\rho - p} \sqrt{f(\xi_{j_1})^{1-\frac{1}{p}} f(\xi_{j_2})^{1-\frac{1}{p}}} \cdot h_n^{\frac{1}{p}\{(d-\varepsilon)(\rho-p)-d\}}. \end{aligned}$$

Finally, setting  $k_1 := 2u_0 \bar{h}_0$ ,  $k_2 := 2\pi_0 \bar{h}_0$  and  $k_3 := \frac{4p^2(2a_0)^{1/p} \bar{h}_0}{\rho-p}$ , one has

$$\begin{aligned} nh_n^d \cdot \text{Cov}(\hat{f}_{j_1}, \hat{f}_{j_2}) &\leq -h_n^d f(\xi_{j_1}) f(\xi_{j_2}) + k_1 \sqrt{f(\xi_{j_1}) f(\xi_{j_2}) (1 - h_n^d f(\xi_{j_1})) (1 - h_n^d f(\xi_{j_2}))} \\ &\quad + k_2 k(\dot{\xi}_{j_1}) \cdot h_n^\varepsilon + k_3 \sqrt{f(\xi_{j_1})^{1-\frac{1}{p}} f(\xi_{j_2})^{1-\frac{1}{p}}} \cdot h_n^{\frac{1}{p}\{(d-\varepsilon)(\rho-p)-d\}}. \quad (3.6) \end{aligned}$$

We deduce the lemma by taking  $j_1 = j_2 = j$  with the appropriate expression of  $V_n$ . It turns out that the covariance is a  $\mathcal{O}\left(\frac{1}{nh_n^d}\right)$  for any choice of  $\varepsilon$  pertaining to  $\left[0, d\left(1 - \frac{1}{\rho-p}\right)\right]$ . ■

**Proof of Lemma 3.5** Here the calculus of  $V_n$  is exactly the same as in the proof of Lemma 3.4. In fact, the delicate point will consist again in bounding  $C_n$ . To do so, we give the common probability density function, say  $\Delta_Z$ , of all random variables  $\{Z_j - Z_i, i < j\}$ . Since the variables  $\{Z_i, 0 \leq i \leq n\}$  are supposed to be independent and symmetrically distributed, we have  $\Delta_Z(t) = (g_J * g_J)(t) = \int_{\mathbb{R}} g_J(t-y) g_J(y) dy$  with support over  $[-\delta, \delta]$ . Let us denote by  $\lfloor x \rfloor$  the largest integer less than or equal to the real  $x$ . Set  $r_0 := \lceil u_0/\delta \rceil$  and  $r_n^1 := \lfloor h_n^{-d+\varepsilon} \rfloor$  for some  $0 \leq \varepsilon < d$  to be specified later. Now weak stationarity implies

$$\begin{aligned} C_n &= \frac{2}{nh_n^d} \sum_{r=1}^{n-1} \sum_{p=1}^{n-r} \text{Cov}(\mathbf{1}_{\pi_{nj_1}}(X_0), \mathbf{1}_{\pi_{nj_2}}(X_{t_p+r-t_p})) \\ &= \frac{2}{h_n^d} \sum_{r=1}^{n-1} \left(1 - \frac{r}{n}\right) \int_{-\delta}^{\delta} \text{Cov}(\mathbf{1}_{\pi_{nj_1}}(X_0), \mathbf{1}_{\pi_{nj_2}}(X_{r\delta+t})) \Delta_Z(t) dt \\ &=: C_{n,1} + C_{n,2} + C_{n,3}, \end{aligned}$$

where

$$\begin{aligned} C_{n,1} &:= \frac{2}{h_n^d} \sum_{r=1}^{r_0} \left(1 - \frac{r}{n}\right) \int_{-\delta}^{\delta} \text{Cov}(\mathbb{1}_{\pi_{nj_1}}(X_0), \mathbb{1}_{\pi_{nj_2}}(X_{r\delta+t})) \Delta_Z(t) dt, \\ C_{n,2} &:= \frac{2}{h_n^d} \sum_{r=r_0+1}^{r_n^1} \left(1 - \frac{r}{n}\right) \int_{-\delta}^{\delta} \text{Cov}(\mathbb{1}_{\pi_{nj_1}}(X_0), \mathbb{1}_{\pi_{nj_2}}(X_{r\delta+t})) \Delta_Z(t) dt, \\ C_{n,3} &:= \frac{2}{h_n^d} \sum_{r=r_n^1+1}^{n-1} \left(1 - \frac{r}{n}\right) \int_{-\delta}^{\delta} \text{Cov}(\mathbb{1}_{\pi_{nj_1}}(X_0), \mathbb{1}_{\pi_{nj_2}}(X_{r\delta+t})) \Delta_Z(t) dt. \end{aligned}$$

By Cauchy-Schwarz we get

$$\begin{aligned} |C_{n,1}| &\leq \frac{2}{h_n^d} \sum_{r=1}^{r_0} \sqrt{\text{Var}(\mathbb{1}_{\pi_{nj_1}}(X_0))} \sqrt{\text{Var}(\mathbb{1}_{\pi_{nj_2}}(X_0))} \int_{-\delta}^{\delta} \Delta_Z(t) dt \\ &\leq 2 \left\lceil \frac{u_0}{\delta} \right\rceil \sqrt{f(\xi_{j_1}) f(\xi_{j_2}) (1 - h_n^d f(\xi_{j_1})) (1 - h_n^d f(\xi_{j_2}))}. \end{aligned}$$

Then using  $A_1(i)$

$$\begin{aligned} |C_{n,2}| &\leq \frac{2}{h_n^d} \sum_{r=r_0+1}^{r_n^1} \int_{-\delta}^{\delta} \Delta_Z(t) \iint_{\pi_{nj_1} \times \pi_{nj_2}} \sup_{y \in \mathbb{R}^d} |g_{r\delta+t}(x, y)| dx dy dt \\ &\leq 2(r_n^1 - r_0) \pi_0 h_n^d k(\dot{\xi}_{j_1}) \int_{-\delta}^{\delta} \Delta_Z(t) dt \\ &\leq 2\pi_0 k(\dot{\xi}_{j_1}) \left( h_n^{-d+\varepsilon} - \left\lceil \frac{u_0}{\delta} \right\rceil \right) h_n^d. \end{aligned}$$

By Davydov's inequality and  $A_1(ii)$

$$|C_{n,3}| \leq \frac{2}{h_n^d} \sum_{r=r_n^1+1}^{n-1} \int_{-\delta}^{\delta} \left| \text{Cov}(\mathbb{1}_{\pi_{nj_1}}(X_0), \mathbb{1}_{\pi_{nj_2}}(X_{r\delta+t})) \right| \Delta_Z(t) dt.$$

For any  $(p, q) \in ]1, \rho-1] \times \left[1 + \frac{\rho}{\rho-2}, \infty\right]$  such that  $\frac{2}{q} + \frac{1}{p} = 1$  and since  $\alpha_X^{(2)}(\cdot)$  is decreasing, we have

$$\begin{aligned} |C_{n,3}| &\leq \frac{1}{h_n^d} \cdot 4p2^{1/p} \cdot h_n^{\frac{2d}{q}} f(\xi_{j_1})^{1/q} f(\xi_{j_2})^{1/q} \sum_{r=r_n^1+1}^{n-1} \int_{-\delta}^{\delta} \left( \alpha_X^{(2)}((r-1)\delta) \right)^{1/p} \Delta_Z(t) dt \\ &\leq h_n^{-\frac{d}{p}} \cdot \frac{4p2^{1/p}}{\delta} \sqrt{f(\xi_{j_1})^{1-\frac{1}{p}} f(\xi_{j_2})^{1-\frac{1}{p}}} \cdot \int_{(r_n^1-1)\delta}^{\infty} \left( \alpha_X^{(2)}(u) \right)^{1/p} du. \end{aligned}$$

Since  $X^T$  is ASM we deduce

$$|C_{n,3}| \leq \frac{4p^2(2a_0)^{1/p}}{(\rho-p)\delta^{\frac{\rho}{p}}} \sqrt{f(\xi_{j_1})^{1-\frac{1}{p}} f(\xi_{j_2})^{1-\frac{1}{p}}} \cdot h_n^{-\frac{d}{p}} (r_n^1 - 1)^{1-\frac{\rho}{p}}.$$

Now if  $p < \rho$  ( $\Leftrightarrow 1 - \frac{\rho}{p} < 0$ ) and since  $r_n^1 > h_n^{-d+\varepsilon} - 1$  we may write

$$|C_{n,3}| \leq \frac{4p^2(2a_0)^{1/p}}{(\rho-p)\delta^{\frac{\rho}{p}}} \sqrt{f(\xi_{j_1})^{1-\frac{1}{p}} f(\xi_{j_2})^{1-\frac{1}{p}}} \cdot h_n^{\frac{1}{p}\{(d-\varepsilon)(\rho-p)-d\}} (1 - 2h_n^{d-\varepsilon})^{1-\frac{\rho}{p}},$$

where  $(1 - 2h_n^{d-\varepsilon})^{1-\frac{\rho}{p}} \rightarrow 1$  as  $n \rightarrow \infty$ . Hence we obtain

$$|C_{n,3}| \leq \frac{4p^2(2a_0)^{1/p}}{(\rho-p)\delta^{\frac{\rho}{p}}} \sqrt{f(\xi_{j_1})^{1-\frac{1}{p}} f(\xi_{j_2})^{1-\frac{1}{p}}} \cdot h_n^{\frac{1}{p}\{(d-\varepsilon)(\rho-p)-d\}} (1 - 2h_n^{d-\varepsilon})^{1-\frac{\rho}{p}}.$$

Finally, setting  $k_4 := 2 \lceil \frac{u_0}{\delta} \rceil$ ,  $k_5 := 2\pi_0$  and  $k_6 := \frac{4p^2(2a_0)^{1/p}}{(\rho-p)\delta^{\rho/p}}$ , one has

$$\begin{aligned} nh_n^d \cdot \text{Cov}(\hat{f}_{j_1}, \hat{f}_{j_2}) &\leq -h_n^d f(\xi_{j_1}) f(\xi_{j_2}) + k_4 \sqrt{f(\xi_{j_1}) f(\xi_{j_2}) (1 - h_n^d f(\xi_{j_1})) (1 - h_n^d f(\xi_{j_2}))} \\ &+ k_5 k(\dot{\xi}_{j_1}) \left( h_n^\varepsilon - \lceil \frac{u_0}{\delta} \rceil h_n^d \right) + k_6 \sqrt{f(\xi_{j_1})^{1-\frac{1}{p}} f(\xi_{j_2})^{1-\frac{1}{p}}} \cdot h_n^{\frac{1}{p}\{(d-\varepsilon)(\rho-p)-d\}} (1 - 2h_n^{d-\varepsilon})^{1-\frac{\rho}{p}}, \end{aligned} \quad (3.7)$$

which implies the desired result. Thus the covariance is still a  $\mathcal{O}\left(\frac{1}{nh_n^d}\right)$  for any  $\varepsilon$  in  $\left[0, d\left(1 - \frac{1}{\rho-p}\right)\right]$ .  $\blacksquare$

### Proof of Theorem 3.1

[RS] – By integrating over  $\pi_{n_j}$  the right-hand side term of (3.4) and by summing up over all hypercubes, we first derive an asymptotic upper bound for IV. For some  $\varepsilon \in \left[0, d\left(1 - \frac{1}{\rho-p}\right)\right]$ ,

$$\begin{aligned} nh_n^d \cdot \int_{\pi_{n_j}} \text{Var}(\hat{f}_j) dx &\leq h_n^d \left\{ f(\xi_j) (1 - h_n^d f(\xi_j)) \{1 + k_1\} \right. \\ &\quad \left. + k_2 k(\dot{\xi}_j) \cdot h_n^\varepsilon + k_3 f(\xi_j)^{1-\frac{1}{p}} \cdot h_n^{\frac{1}{p}\{(d-\varepsilon)(\rho-p)-d\}} \right\}. \end{aligned}$$

Then since the functions  $f^{1-\frac{1}{p}}$ ,  $f$ ,  $f^2$  and  $k$  are supposed to be Riemann-integrable over  $\mathbb{R}^d$ , one may write

$$\sum_j h_n^d f^\kappa(\xi_j) = \|f^\kappa\|_1 + o(1) \quad \text{and} \quad \sum_j h_n^d k(\dot{\xi}_j) = \|k\|_1 + o(1),$$

and we obtain

$$nh_n^d \cdot \text{IV}(\hat{f}_n^H) \leq \left\{ 1 + k_1 + k_2 \|k\|_1 \cdot h_n^\varepsilon + k_3 \|f^{1-\frac{1}{p}}\|_1 \cdot h_n^{\frac{1}{p}\{(d-\varepsilon)(\rho-p)-d\}} \right\} (1 + o(1)).$$

The second part of the proof comes from the choice  $h_n = cn^{-\frac{1}{d+2}}$ . So Lemma 3.1 yields

$$\lim_{n \rightarrow \infty} n^{\frac{2}{d+2}} \cdot \text{ISB}(\hat{f}_n^H) = \frac{c^2}{12} \mathcal{R}_d(f').$$

On the other hand, if  $p = \rho - 1$  ( $\Leftrightarrow \varepsilon = 0$ ), we have

$$\limsup_{n \rightarrow \infty} n^{\frac{2}{d+2}} \cdot \text{IV}(\hat{f}_n^H) \leq \frac{1}{c^d} \left\{ 1 + k_1 + k_2 \|k\|_1 + k_3 \|f^{1-\frac{1}{p}}\|_1 \right\}.$$

Hence

$$\limsup_{n \rightarrow \infty} n^{\frac{2}{d+2}} \cdot \text{MISE}(\hat{f}_n^H) \leq \frac{c^2}{12} \mathcal{R}_d(f') + \frac{1}{c^d} \left\{ 1 + k_1 + k_2 \|k\|_1 + k_3 \|f^{1-\frac{1}{p}}\|_1 \right\}.$$

If  $p < \rho - 1$  ( $\Leftrightarrow \varepsilon > 0$ ), we improve the asymptotic constant :

$$\limsup_{n \rightarrow \infty} n^{\frac{2}{d+2}} \cdot \text{IV}(\hat{f}_n^H) \leq \frac{1}{c^d} \{1+k_1\} \quad \text{and} \quad \limsup_{n \rightarrow \infty} n^{\frac{2}{d+2}} \cdot \text{MISE}(\hat{f}_n^H) \leq \frac{c^2}{12} \mathcal{R}_d(f') + \frac{1}{c^d} \{1+k_1\}.$$

[JS] – Now, let us integrate over  $\pi_{n_j}$  the right-hand side expression in Eq. (3.5) :

$$\begin{aligned} nh_n^d \cdot \int_{\pi_{n_j}} \text{Var}(\hat{f}_j) dx &\leq h_n^d \left\{ f(\xi_j)(1 - h_n^d f(\xi_j)) \{1+k_4\} + k_5 k(\dot{\xi}_j) \left( h_n^\varepsilon - \left\lceil \frac{u_0}{\delta} \right\rceil h_n^d \right) \right. \\ &\quad \left. + k_6 f(\xi_j)^{1-\frac{1}{p}} \cdot h_n^{\frac{1}{p}\{(d-\varepsilon)(\rho-p)-d\}} (1 - 2h_n^{d-\varepsilon})^{1-\frac{p}{d}} \right\}, \end{aligned}$$

for any  $\varepsilon \in \left[0, d\left(1 - \frac{1}{\rho-p}\right)\right]$ . Then sum up over all indexes  $j$  to obtain

$$nh_n^d \cdot \text{IV}(\hat{f}_n^H) \leq \left\{ 1+k_4+k_5\|k\|_1 \left( h_n^\varepsilon - \left\lceil \frac{u_0}{\delta} \right\rceil h_n^d \right) + k_6 \|f^{1-\frac{1}{p}}\|_1 \cdot h_n^{\frac{1}{p}\{(d-\varepsilon)(\rho-p)-d\}} \right\} (1+o(1)).$$

Therefore, if  $p = \rho - 1$  ( $\Leftrightarrow \varepsilon = 0$ ), the bin width choice  $h_n = cn^{-\frac{1}{d+2}}$  entails

$$\limsup_{n \rightarrow \infty} n^{\frac{2}{d+2}} \cdot \text{IV}(\hat{f}_n^H) \leq \frac{1}{c^d} \left\{ 1+k_4+k_5\|k\|_1 + k_6 \|f^{1-\frac{1}{p}}\|_1 \right\}.$$

Hence

$$\limsup_{n \rightarrow \infty} n^{\frac{2}{d+2}} \cdot \text{MISE}(\hat{f}_n^H) \leq \frac{c^2}{12} \mathcal{R}_d(f') + \frac{1}{c^d} \left\{ 1+k_4+k_5\|k\|_1 + k_6 \|f^{1-\frac{1}{p}}\|_1 \right\}.$$

If  $p < \rho - 1$ , we get a better asymptotic constant :

$$\limsup_{n \rightarrow \infty} n^{\frac{2}{d+2}} \cdot \text{IV}(\hat{f}_n^H) \leq \frac{1}{c^d} \{1+k_4\} \quad \text{and} \quad \limsup_{n \rightarrow \infty} n^{\frac{2}{d+2}} \cdot \text{MISE}(\hat{f}_n^H) \leq \frac{c^2}{12} \mathcal{R}_d(f') + \frac{1}{c^d} \{1+k_4\}.$$

■

### • Variance bounds with high frequency sampling

The period depends now on the sample size in such manner that  $\delta_n \downarrow 0^+$  as  $n \rightarrow \infty$ . We start by giving another upper bound expression for the variance of  $\hat{f}_n^H(x)$  which depends upon  $\gamma_0$ .

**Lemma 3.6** ([HFS]). *If  $A_0(ii)$  and  $A'_1(i)(ii)$  hold, then we obtain*

$$\begin{aligned} nh_n^d \cdot \text{Var}(\hat{f}_j) &\leq f(\xi_j)(1 - h_n^d f(\xi_j)) + 2\varphi(\dot{\xi}_j) \left( \sum_{r=1}^{r_n^0} \frac{1}{r^{\gamma_0}} \right) h_n^d \delta_n^{-\gamma_0} + \left\{ 2u_0 \|f\|_\infty f(\xi_j) \right. \\ &\quad \left. + 2(u_1 - u_0 + \delta_n) k(\ddot{\xi}_j) \sup_{u \in [u_0, u_1]} \pi(u) + 2k(\ddot{\xi}_j) \int_{u_1}^\infty \pi(u) du \left( 1 + \frac{\pi(u_1)}{\int_{u_1}^\infty \pi(u) du} \delta_n \right) \right\} h_n^d \delta_n^{-1}, \end{aligned} \tag{3.8}$$

with  $(\xi_j, \dot{\xi}_j, \ddot{\xi}_j) \in \pi_{n_j}^3$  and it entails that the variance is a  $\mathcal{O}\left(\frac{1}{nh_n^d}\right)$  if

- for  $\gamma_0 < 1$ ,  $\delta_n^*(\gamma_0) \sim h_n^d$  ;

- for  $\gamma_0 = 1$ ,  $\delta_n^*(\gamma_0) \sim h_n^d \ln(1/h_n^d)$ ;
- for  $\gamma_0 > 1$ ,  $\delta_n^*(\gamma_0) \sim h_n^{d/\gamma_0}$ .

**Proof of Lemma 3.6** The calculus of  $V_n$  remains identical. Now to upper bound  $C_n$ , we have to make use of the local assumption  $A'_1(i)$ . Set  $r_n^0 := \lfloor u_0/\delta_n \rfloor$  and  $r_n^1 := \lfloor u_1/\delta_n \rfloor$ , since  $X^T$  is weakly stationary one may write

$$C_n = \frac{2}{h_n^d} \sum_{r=1}^{n-1} \left(1 - \frac{r}{n}\right) \text{Cov}(\mathbb{1}_{\pi_{nj_1}}(X_0), \mathbb{1}_{\pi_{nj_2}}(X_{r\delta_n})) =: C_{n,1} + C_{n,2},$$

where

$$C_{n,1} := \frac{2}{h_n^d} \sum_{r=1}^{r_n^0} \left(1 - \frac{r}{n}\right) \text{Cov}(\mathbb{1}_{\pi_{nj_1}}(X_0), \mathbb{1}_{\pi_{nj_2}}(X_{r\delta_n})),$$

$$C_{n,2} := \frac{2}{h_n^d} \sum_{r=r_n^0+1}^{n-1} \left(1 - \frac{r}{n}\right) \text{Cov}(\mathbb{1}_{\pi_{nj_1}}(X_0), \mathbb{1}_{\pi_{nj_2}}(X_{r\delta_n})).$$

First using  $A'_1(i)$  we get

$$\begin{aligned} |C_{n,1}| &\leq \frac{2}{h_n^d} \sum_{r=1}^{r_n^0} \iint_{\pi_{nj_1} \times \pi_{nj_2}} \left\{ \sup_{y \in \mathbb{R}^d} f_{r\delta_n}(x, y) + \|f\|_\infty f(x) \right\} dx dy \\ &\leq 2 \sum_{r=1}^{r_n^0} \int_{\pi_{nj_1}} \{ \varphi(x)(r\delta_n)^{-\gamma_0} + \|f\|_\infty f(x) \} dx \\ &\leq 2\varphi(\dot{\xi}_{j_1}) \left( \sum_{r=1}^{r_n^0} \frac{1}{r^{\gamma_0}} \right) h_n^d \delta_n^{-\gamma_0} + 2u_0 \|f\|_\infty f(\xi_{j_1}) h_n^d \delta_n^{-1}, \end{aligned}$$

where  $(\xi_{j_1}, \dot{\xi}_{j_1}) \in \pi_{nj_1}^2$ . Setting  $k_7 := 2u_0 \|f\|_\infty$ , we obtain

$$|C_{n,1}| \leq 2\varphi(\dot{\xi}_{j_1}) \left( \sum_{r=1}^{r_n^0} \frac{1}{r^{\gamma_0}} \right) h_n^d \delta_n^{-\gamma_0} + k_7 f(\xi_{j_1}) h_n^d \delta_n^{-1}.$$

Then using  $A'_1(ii)$

$$\begin{aligned} |C_{n,2}| &\leq \frac{2}{h_n^d} \sum_{r=r_n^0+1}^{n-1} \iint_{\pi_{nj_1} \times \pi_{nj_2}} \sup_{y \in \mathbb{R}^d} |g_{r\delta_n}(x, y)| dx dy \\ &\leq 2 \sum_{r=r_n^0+1}^{n-1} \int_{\pi_{nj_1}} k(x) \pi(r\delta_n) dx \\ &\leq 2h_n^d k(\ddot{\xi}_{j_1}) \left( \sum_{r=r_n^0+1}^{r_n^1} \pi(r\delta_n) + \sum_{r=r_n^1+1}^{n-1} \pi(r\delta_n) \right), \end{aligned}$$

where  $\ddot{\xi}_{j_1} \in \pi_{nj_1}$ . On the one hand, one has

$$\sum_{r=r_n^0+1}^{r_n^1} \pi(r\delta_n) \leq (r_n^1 - r_n^0) \sup_{u \in [u_0, u_1]} \pi(u) \leq (u_1 - u_0) \sup_{u \in [u_0, u_1]} \pi(u) \left( 1 + \frac{\delta_n}{u_1 - u_0} \right) \delta_n^{-1}.$$

On the other hand, the monotonicity of  $\pi(\cdot)$  implies

$$\sum_{r=r_n^1+1}^{n-1} \pi(r\delta_n) \leq \delta_n^{-1} \sum_{r=r_n^1+1}^{n-1} \delta_n \pi(r\delta_n) \leq \left\{ (u_1 - r_n^1 \delta_n) \pi((r_n^1 + 1) \delta_n) + \int_{u_1}^{\infty} \pi(u) du \right\} \delta_n^{-1}.$$

Setting  $k_8 := 2(u_1 - u_0) \sup_{u \in [u_0, u_1]} \pi(u)$  and  $k_9 := 2 \int_{u_1}^{\infty} \pi(u) du$ , we thus obtain

$$|C_{n,2}| \leq k_8 k(\ddot{\xi}_{j_1}) \left(1 + \frac{\delta_n}{u_1 - u_0}\right) h_n^d \delta_n^{-1} + k_9 k(\ddot{\xi}_{j_1}) \left(1 + \frac{\pi(u_1)}{\int_{u_1}^{\infty} \pi(u) du} \delta_n\right) h_n^d \delta_n^{-1}.$$

Thence

$$\begin{aligned} nh_n^d \cdot \text{Cov}(\hat{f}_{j_1}, \hat{f}_{j_2}) &\leq -h_n^d f(\xi_{j_1}) f(\xi_{j_2}) + 2\varphi(\dot{\xi}_{j_1}) \left( \sum_{r=1}^{r_n^0} \frac{1}{r^{\gamma_0}} \right) h_n^d \delta_n^{-\gamma_0} + k_7 f(\xi_{j_1}) h_n^d \delta_n^{-1} \\ &\quad + k_8 k(\ddot{\xi}_{j_1}) \left(1 + \frac{\delta_n}{u_1 - u_0}\right) h_n^d \delta_n^{-1} + k_9 k(\ddot{\xi}_{j_1}) \left(1 + \frac{\pi(u_1)}{\int_{u_1}^{\infty} \pi(u) du} \delta_n\right) h_n^d \delta_n^{-1}, \end{aligned} \quad (3.9)$$

which leads to the desired result. Using Eq. (3.9), we also deduce the optimal choices  $\delta_n^*(\gamma_0)$  of  $\delta_n$  *i.e.* the smallest values of  $\delta_n$  so that  $C_n$  is a  $\mathcal{O}(1)$ . These choices are given by (3.1) in accordance with the values of  $\gamma_0$ .  $\blacksquare$

### Proof of Theorem 3.2

By integrating over  $\pi_{n_j}$  the right-hand side expression in Eq. (3.8) :

$$\begin{aligned} nh_n^d \cdot \int_{\pi_{n_j}} \text{Var}(\hat{f}_j) dx &\leq h_n^d \left\{ f(\xi_j)(1 - h_n^d f(\xi_j)) + 2\varphi(\dot{\xi}_j) \left( \sum_{r=1}^{r_n^0} \frac{1}{r^{\gamma_0}} \right) h_n^d \delta_n^{-\gamma_0} + k_7 f(\xi_j) h_n^d \delta_n^{-1} \right. \\ &\quad \left. + k_8 k(\ddot{\xi}_j) \left(1 + \frac{\delta_n}{u_1 - u_0}\right) h_n^d \delta_n^{-1} + k_9 k(\ddot{\xi}_j) \left(1 + \frac{\pi(u_1)}{\int_{u_1}^{\infty} \pi(u) du} \delta_n\right) h_n^d \delta_n^{-1} \right\}. \end{aligned}$$

Then let us sum up over all indexes  $j$ . Since  $\varphi$  is Riemann-integrable, we obtain

$$nh_n^d \cdot \text{IV}(\hat{f}_n^H) \leq \left\{ 1 + 2\|\varphi\|_1 \left( \sum_{r=1}^{r_n^0} \frac{1}{r^{\gamma_0}} \right) h_n^d \delta_n^{-\gamma_0} + k_7 h_n^d \delta_n^{-1} + (k_8 + k_9) \|k\|_1 h_n^d \delta_n^{-1} \right\} (1 + o(1)). \quad (3.10)$$

According to the values of  $\gamma_0$ , we derive all asymptotic bounds with optimal choices of  $\delta_n$  :

- if  $\gamma_0 < 1$ , the choice  $\delta_n^*(\gamma_0) = d_1 h_n^d$ ,  $0 < d_1 < \infty$ , entails

$$\left( \sum_{r=1}^{r_n^0} \frac{1}{r^{\gamma_0}} \right) h_n^d \delta_n^{-\gamma_0} \leq \frac{1}{d_1} \frac{u_0^{1-\gamma_0}}{1-\gamma_0} \left[ 1 - \frac{d_1^{1-\gamma_0} \gamma_0}{u_0^{1-\gamma_0}} h_n^{d(1-\gamma_0)} \right];$$

- if  $\gamma_0 = 1$ , the choice  $\delta_n^*(\gamma_0) = d_2 h_n^d \ln(h_n^{-d})$ ,  $0 < d_2 < \infty$ , entails

$$\left( \sum_{r=1}^{r_n^0} \frac{1}{r} \right) h_n^d \delta_n^{-1} \leq \frac{1}{d_2} \left[ 1 - \ln(h_n^{-d}) \ln \left( e \frac{u_0}{d_2} \ln(h_n^{-d}) \right) \right];$$

- if  $\gamma_0 > 1$ , the choice  $\delta_n^*(\gamma_0) = d_3 h_n^{d/\gamma_0}$ ,  $0 < d_3 < \infty$ , entails

$$\left( \sum_{r=1}^{r_n^0} \frac{1}{r^{\gamma_0}} \right) h_n^d \delta_n^{-\gamma_0} \leq \frac{\gamma_0}{d_3^{\gamma_0} (\gamma_0 - 1)} \left[ 1 - \frac{1}{\gamma_0 u_0^{\gamma_0 - 1}} \delta_n^{\gamma_0 - 1} \right].$$

So setting

$$C_{\gamma_0} := \frac{1}{d_1} \left\{ \frac{2\|\varphi\|_1 u_0^{1-\gamma_0}}{1-\gamma_0} + k_7 + (k_8 + k_9)\|k\|_1 \right\} \mathbf{1}_{\{\gamma_0 < 1\}} + \frac{2\|\varphi\|_1}{d_2} \mathbf{1}_{\{\gamma_0 = 1\}} + \frac{2\|\varphi\|_1 \gamma_0}{d_3^{\gamma_0} (\gamma_0 - 1)} \mathbf{1}_{\{\gamma_0 > 1\}},$$

it follows that

$$\limsup_{n \rightarrow \infty} n h_n^d \cdot \text{IV}(\hat{f}_n^H) \leq 1 + C_{\gamma_0}.$$

To conclude, it suffices to choose  $h_n = c n^{-\frac{1}{d+2}}$  as in Theorem 3.1.

$$\limsup_{n \rightarrow \infty} n^{\frac{2}{d+2}} \cdot \text{IV}(\hat{f}_n^H) \leq \frac{1}{c^d} (1 + C_{\gamma_0}) \quad \text{and} \quad \limsup_{n \rightarrow \infty} n^{\frac{2}{d+2}} \cdot \text{MISE}(\hat{f}_n^H) \leq \frac{c^2}{12} \mathcal{R}_d(f') + \frac{1}{c^d} (1 + C_{\gamma_0}),$$

and we can also improve our asymptotic constant for any choice of  $\delta_n \gg \delta_n^*(\gamma_0)$  :

$$\limsup_{n \rightarrow \infty} n^{\frac{2}{d+2}} \cdot \text{IV}(\hat{f}_n^H) \leq \frac{1}{c^d} \quad \text{and} \quad \limsup_{n \rightarrow \infty} n^{\frac{2}{d+2}} \cdot \text{MISE}(\hat{f}_n^H) \leq \frac{c^2}{12} \mathcal{R}_d(f') + \frac{1}{c^d}.$$

■

### Proof of Corollary 3.1

Observe that Eq. (3.10) together with  $T_n = n \delta_n$  entail

$$T_n \cdot \text{IV}(\hat{f}_n^H) \leq \left\{ h_n^{-d} \delta_n + 2\|\varphi\|_1 \left( \sum_{r=1}^{r_n^0} \frac{1}{r^{\gamma_0}} \right) \delta_n^{1-\gamma_0} + k_7 + (k_8 + k_9)\|k\|_1 \right\} (1 + o(1)).$$

Looking at the right-hand side expression above, the second term is clearly divergent for  $\gamma_0 \geq 1$ . If  $\gamma_0 < 1$  then

$$\left( \sum_{r=1}^{r_n^0} \frac{1}{r^{\gamma_0}} \right) \delta_n^{1-\gamma_0} \leq \frac{u_0^{1-\gamma_0}}{1-\gamma_0} \left[ 1 - \frac{\gamma_0}{u_0^{1-\gamma_0}} \delta_n^{1-\gamma_0} \right],$$

and for the optimal choice  $\delta_n^* = d_0 h_n^d$ ,  $0 < d_0 < \infty$ , we have

$$\limsup_{n \rightarrow \infty} T_n \cdot \text{IV}(\hat{f}_n^H) \leq d_0 + \frac{2\|\varphi\|_1 u_0^{1-\gamma_0}}{1-\gamma_0} + k_7 + (k_8 + k_9)\|k\|_1.$$

Finally, using Lemma 3.1 with  $h_n = o(T_n^{-1/2})$ , the bias term is negligible and we deduce a “parametric” rate :

$$\limsup_{n \rightarrow \infty} T_n \cdot \text{MISE}(\hat{f}_n^H) \leq d_0 + \frac{2\|\varphi\|_1 u_0^{1-\gamma_0}}{1-\gamma_0} + k_7 + (k_8 + k_9)\|k\|_1.$$

■

### 3.7.2 Frequency polygon

#### Proof of Theorem 3.3

[RS] – First observe that

$$\int_{\mathbb{R}} \text{Var}(\hat{f}_n^{FP}(x)) dx = \sum_j \int_{[c_j, c_{j+1}[} \text{Var}(\hat{f}_n^{FP}(x)) dx,$$

where

$$\begin{aligned} \int_{c_j}^{c_{j+1}} \text{Var}(\hat{f}_n^{FP}(x)) dx &= \frac{1}{h_n^2} \int_{c_j}^{c_{j+1}} \left\{ (x - c_j)^2 \text{Var}(\hat{f}_{j+1}) + (c_{j+1} - x)^2 \text{Var}(\hat{f}_j) \right. \\ &\quad \left. + 2(x - c_j)(c_{j+1} - x) \text{Cov}(\hat{f}_j, \hat{f}_{j+1}) \right\} dx. \end{aligned}$$

For any  $j \in \mathbb{Z}$ , let us denote by  $\bar{V}_j$  (respectively  $\bar{C}_{j,j+1}$ ) an upper bound expression for  $nh_n \cdot \text{Var}(\hat{f}_j)$  (respectively  $nh_n \cdot \text{Cov}(\hat{f}_j, \hat{f}_{j+1})$ ), which is independent of  $x$ . We get

$$nh_n \cdot \int_{c_j}^{c_{j+1}} \text{Var}(\hat{f}_n^{FP}(x)) dx \leq \frac{h_n}{3} \{ \bar{V}_j + \bar{V}_{j+1} + \bar{C}_{j,j+1} \}. \quad (3.11)$$

Introduce now both expressions (3.4) and (3.6) in Eq. (3.11) then for  $\varepsilon \in \left[0, 1 - \frac{1}{\rho-p}\right]$ ,

$$\begin{aligned} &nh_n \cdot \int_{c_j}^{c_{j+1}} \text{Var}(\hat{f}_n^{FP}(x)) dx \\ &\leq \frac{h_n}{3} \left\{ f(\xi_j)(1 - h_n f(\xi_j)) \{1 + k_1\} + k_2 f(\xi_j) \cdot h_n^\varepsilon + k_3 f(\xi_j)^{1-\frac{1}{p}} \cdot h_n^{\frac{1}{p} \{(1-\varepsilon)(\rho-p)-1\}} \right\} \\ &+ \frac{h_n}{3} \left\{ f(\xi_{j+1})(1 - h_n f(\xi_{j+1})) \{1 + k_1\} + k_2 f(\xi_{j+1}) \cdot h_n^\varepsilon + k_3 f(\xi_{j+1})^{1-\frac{1}{p}} \cdot h_n^{\frac{1}{p} \{(1-\varepsilon)(\rho-p)-1\}} \right\} \\ &+ \frac{h_n}{3} \left\{ -h_n f(\xi_j) f(\xi_{j+1}) + k_1 \sqrt{f(\xi_j) f(\xi_{j+1}) (1 - h_n f(\xi_j)) (1 - h_n f(\xi_{j+1}))} \right. \\ &\quad \left. + k_2 f(\xi_j) \cdot h_n^\varepsilon + k_3 \sqrt{f(\xi_j)^{1-\frac{1}{p}} f(\xi_{j+1})^{1-\frac{1}{p}}} \cdot h_n^{\frac{1}{p} \{(1-\varepsilon)(\rho-p)-1\}} \right\}. \end{aligned}$$

We bound the IV of  $\hat{f}_n^{FP}$  by summing up over all indexes  $j$ . So for  $\varepsilon \in \left[0, 1 - \frac{1}{\rho-p}\right]$ ,

$$nh_n \cdot \text{IV}(\hat{f}_n^{FP}) \leq \left\{ \frac{2}{3} + k_1 + k_2 \cdot h_n^\varepsilon + k_3 \|f^{1-\frac{1}{p}}\|_1 \cdot h_n^{\frac{1}{p} \{(1-\varepsilon)(\rho-p)-1\}} \right\} (1 + o(1)).$$

Now the bin width choice  $h_n = cn^{-\frac{1}{5}}$ ,  $0 < c < \infty$ , yields first

$$\lim_{n \rightarrow \infty} n^{\frac{4}{5}} \cdot \text{ISB}(\hat{f}_n^{FP}) = \frac{49}{2880} c^A \mathcal{R}(f'').$$

On the other hand if  $p = \rho - 1$  ( $\Leftrightarrow \varepsilon = 0$ ) :

$$\limsup_{n \rightarrow \infty} n^{\frac{4}{5}} \cdot \text{IV}(\hat{f}_n^{FP}) \leq \frac{1}{c} \left\{ \frac{2}{3} + k_1 + k_2 + k_3 \|f^{1-\frac{1}{p}}\|_1 \right\}.$$

Hence

$$\limsup_{n \rightarrow \infty} n^{\frac{4}{5}} \cdot \text{MISE}(\hat{f}_n^{FP}) \leq \frac{49}{2880} c^4 \mathcal{R}(f'') + \frac{1}{c} \left\{ \frac{2}{3} + k_1 + k_2 + k_3 \|f^{1-\frac{1}{p}}\|_1 \right\}.$$

If  $p < \rho - 1$  ( $\Leftrightarrow 0 < \varepsilon < 1$ ) :

$$\limsup_{n \rightarrow \infty} n^{\frac{4}{5}} \cdot \text{IV}(\hat{f}_n^{FP}) \leq \frac{1}{c} \left\{ \frac{2}{3} + k_1 \right\}$$

and

$$\limsup_{n \rightarrow \infty} n^{\frac{4}{5}} \cdot \text{MISE}(\hat{f}_n^{FP}) \leq \frac{49}{2880} c^4 \mathcal{R}(f'') + \frac{1}{c} \left\{ \frac{2}{3} + k_1 \right\}.$$

[JS] – The outlines of the proof are unchanged. Introduce both expressions (3.5) and (3.7) in Eq. (3.11) and sum up over all indexes  $j$ , then it follows that for  $\varepsilon \in \left[0, 1 - \frac{1}{\rho-p}\right]$ ,

$$\begin{aligned} nh_n \cdot \text{IV}(\hat{f}_n^{FP}) &\leq \left\{ \frac{2}{3} + k_4 + k_5 \left( h_n^\varepsilon - \left\lceil \frac{u_0}{\delta} \right\rceil h_n \right) \right. \\ &\quad \left. + k_6 \|f^{1-\frac{1}{p}}\|_1 \cdot h_n^{\frac{1}{p}\{(1-\varepsilon)(\rho-p)-1\}} (1 - 2h_n^{1-\varepsilon})^{1-\frac{\rho}{p}} \right\} (1 + o(1)). \end{aligned}$$

Take  $h_n = cn^{-\frac{1}{5}}$ ,  $0 < c < \infty$ , then if  $p = \rho - 1$  :

$$\limsup_{n \rightarrow \infty} n^{\frac{4}{5}} \cdot \text{IV}(\hat{f}_n^{FP}) \leq \frac{1}{c} \left\{ \frac{2}{3} + k_4 + k_5 + k_6 \|f^{1-\frac{1}{p}}\|_1 \right\}.$$

Hence

$$\limsup_{n \rightarrow \infty} n^{\frac{4}{5}} \cdot \text{MISE}(\hat{f}_n^{FP}) \leq \frac{49}{2880} c^4 \mathcal{R}(f'') + \frac{1}{c} \left\{ \frac{2}{3} + k_4 + k_5 + k_6 \|f^{1-\frac{1}{p}}\|_1 \right\}.$$

If  $p < \rho - 1$  :

$$\limsup_{n \rightarrow \infty} n^{\frac{4}{5}} \cdot \text{IV}(\hat{f}_n^{FP}) \leq \frac{1}{c} \left\{ \frac{2}{3} + k_4 \right\}$$

and

$$\limsup_{n \rightarrow \infty} n^{\frac{4}{5}} \cdot \text{MISE}(\hat{f}_n^{FP}) \leq \frac{49}{2880} c^4 \mathcal{R}(f'') + \frac{1}{c} \left\{ \frac{2}{3} + k_4 \right\}. \quad \blacksquare$$

#### Proof of Theorem 3.4

[HFS] – Introduce now both expressions (3.8) and (3.9) in Eq. (3.11) and sum up over all indexes  $j$ , we get

$$nh_n \cdot \text{IV}(\hat{f}_n^{FP}) \leq \left\{ \frac{2}{3} + 2\|\varphi\|_1 \left( \sum_{r=1}^{r_n^0} \frac{1}{r^{\gamma_0}} \right) h_n \delta_n^{-\gamma_0} + k_7 h_n \delta_n^{-1} + (k_8 + k_9) \|k\|_1 h_n \delta_n^{-1} \right\} (1 + o(1)). \quad (3.12)$$

Then  $h_n = cn^{-\frac{1}{5}}$ ,  $0 < c < \infty$ , together with the optimal choices  $\delta_n^*$  of  $\delta_n$  yield

$$\limsup_{n \rightarrow \infty} n^{\frac{4}{5}} \cdot \text{IV}(\hat{f}_n^{FP}) \leq \frac{1}{c} \left\{ \frac{2}{3} + C_{\gamma_0} \right\}$$

and

$$\limsup_{n \rightarrow \infty} n^{\frac{4}{5}} \cdot \text{MISE}(\hat{f}_n^{FP}) \leq \frac{49}{2880} c^4 \mathcal{R}(f'') + \frac{1}{c} \left\{ \frac{2}{3} + C_{\gamma_0} \right\}. \quad \blacksquare$$

**Proof of Corollary 3.2**

From (3.12) with the choice  $\delta_n^* = d_0 h_n$ ,  $0 < d_0 < \infty$ , one has

$$T_n \cdot \text{IV}(\hat{f}_n^{FP}) \leq \left\{ \frac{2d_0}{3} + \frac{2\|\varphi\|_1 u_0^{1-\gamma_0}}{1-\gamma_0} + k_7 + (k_8 + k_9)\|k\|_1 \right\} (1 + o(1)).$$

Taking  $h_n = o(T_n^{-1/4})$  the bias term is negligible and we get a “parametric” rate :

$$\limsup_{n \rightarrow \infty} T_n \cdot \text{MISE}(\hat{f}_n^{FP}) \leq \frac{2d_0}{3} + \frac{2\|\varphi\|_1 u_0^{1-\gamma_0}}{1-\gamma_0} + k_7 + (k_8 + k_9)\|k\|_1.$$

■

# Bibliographie

- Aït-Sahalia, Y. et Mykland, P. A. (2003). The effects of random and discrete sampling when estimating continuous-time diffusions. *Econometrica*, **71**(2), 483–549.
- Banon, G. (1978). Nonparametric identification for diffusion processes. *SIAM J. Control Optim.*, **16**(3), 380–395.
- Banon, G. et Nguyen, H. T. (1978). Sur l'estimation récurrente de la densité et de sa dérivée pour un processus de Markov. *C. R. Acad. Sci. Paris Sér. A-B*, **286**(16), A691–A694.
- Banon, G. et Nguyen, H. T. (1981). Recursive estimation in diffusion model. *SIAM J. Control Optim.*, **19**(5), 676–685.
- Baraud, Y. et Birgé, L. (2006). Estimating the intensity of a random measure by histogram type estimators. <http://www.citebase.org/abstract?id=oai:arXiv.org:math/0608663>.
- Beirlant, J., Berliet, A., et Györfi, L. (1999). On piecewise linear density estimators. *Statist. Neerlandica*, **53**(3), 287–308.
- Bensaïd, N. et Dabo-Niang, S. (2007). Frequency polygons for continuous random fields. *personal communication*, 32 pages.
- Berliet, A., Hobza, T., et Vajda, I. (2002). Generalized piecewise linear histograms. *Statist. Neerlandica*, **56**(3), 301–313.
- Bertrand-Retali, M. (1974). Convergence uniforme stochastique d'un estimateur d'une densité de probabilité dans  $\mathbf{R}^s$ . *C. R. Acad. Sci. Paris Sér. A*, **278**, 451–453.
- Bertrand-Retali, M. (1978). Convergence uniforme d'un estimateur de la densité par la méthode du noyau. *Rev. Roumaine Math. Pures Appl.*, **23**(3), 361–385.
- Biau, G. (2003). Spatial kernel density estimation. *Math. Methods Statist.*, **12**(4), 371–390.
- Bibby, B. M., Skovgaard, I. M., et Sørensen, M. (2005). Diffusion-type models with given marginal distribution and autocorrelation function. *Bernoulli*, **11**(2), 191–220.
- Birgé, L. et Rozenholc, Y. (2006). How many bins should be put in a regular histogram. *ESAIM Probab. Stat.*, **10**, 24–45 (electronic).

- Blanke, D. (2004). Sample paths adaptive density estimation. *Math. Methods Statist.*, **13**(2), 123–152.
- Blanke, D. (2006). Adaptive sampling schemes for density estimation. *J. Statist. Plann. Inference*, **136**(9), 2898–2917.
- Blanke, D. et Bosq, D. (1997). Accurate rates of density estimators for continuous-time processes. *Statist. Probab. Lett.*, **33**(2), 185–191.
- Blanke, D. et Merlevède, F. (2000). Estimation of the asymptotic variance of kernel density estimators for continuous time processes. *Math. Methods Statist.*, **9**(3), 270–296.
- Blanke, D. et Pumo, B. (2003). Optimal sampling for density estimation in continuous time. *J. Time Ser. Anal.*, **24**(1), 1–23.
- Bosq, D. (1969). Estimation de la densité conditionnelle et de la régression. *C. R. Acad. Sci. Paris Sér. A-B*, **269**, A661–A664.
- Bosq, D. (1995). Sur le comportement exotique de l'estimateur à noyau de la densité marginale d'un processus à temps continu. *C. R. Acad. Sci. Paris Sér. I Math.*, **320**(3), 369–372.
- Bosq, D. (1997). Parametric rates of nonparametric estimators and predictors for continuous time processes. *Ann. Statist.*, **25**(3), 982–1000.
- Bosq, D. (1998). *Nonparametric statistics for stochastic processes*, volume 110 of *Lecture Notes in Statistics*. Springer-Verlag, New York, second edition. Estimation and prediction.
- Bosq, D. (2005). *Inférence et prévision en grandes dimensions*. Collection “Économie et Statistiques Avancées”. [Collection “Economics and Advanced Statistics”]. Economica, Paris.
- Bosq, D. et Blanke, D. (2004). Local superefficiency of data-driven projection density estimators in continuous time. *SORT*, **28**(1), 37–53.
- Bosq, D. et Blanke, D. (2007). *Inference and prediction in large dimensions*. personal communication.
- Bosq, D. et Chèze-Payaud, N. (1999). Optimal asymptotic quadratic error of nonparametric regression function estimates for a continuous-time process from sampled-data. *Statistics*, **32**(3), 229–247.
- Bosq, D. et Davydov, Y. (1999). Local time and density estimation in continuous time. *Math. Methods Statist.*, **8**(1), 22–45.
- Bosq, D. et Lecoutre, J.-P. (1987). *Théorie de l'estimation fonctionnelle*. Collection “Économie et Statistiques Avancées”. [Collection “Economics and Advanced Statistics”]. Economica, Paris.

- Bosq, D., Merlevède, F., et Peligrad, M. (1999). Asymptotic normality for density kernel estimators in discrete and continuous time. *J. Multivariate Anal.*, **68**(1), 78–95.
- Boyd, D. W. et Steele, J. M. (1978). Lower bounds for nonparametric density estimation rates. *Ann. Statist.*, **6**(4), 932–934.
- Carbon, M. (2005). Frequency polygons for random fields. *INSEE - Série des Documents de Travail du CREST n° 2005-04*.
- Carbon, M. (2006). Polygone des fréquences pour des champs aléatoires. *C. R. Math. Acad. Sci. Paris*, **342**(9), 693–696.
- Carbon, M. et Tran, L. T. (1996). On histograms for linear processes. *J. Statist. Plann. Inference*, **53**(3), 403–419.
- Carbon, M., Garel, B., et Tran, L. T. (1997). Frequency polygons for weakly dependent processes. *Statist. Probab. Lett.*, **33**(1), 1–13.
- Castellana, J. V. et Leadbetter, M. R. (1986). On smoothed probability density estimation for stationary processes. *Stochastic Process. Appl.*, **21**(2), 179–193.
- Chèze-Payaud, N. (1994a). Nonparametric regression and prediction for continuous-time processes. *Publ. Inst. Statist. Univ. Paris*, **38**(2), 37–58.
- Chèze-Payaud, N. (1994b). *Régression, prédiction et discrétisation des processus à temps continu*. Thèse de l'Université Pierre et Marie Curie (Paris 6).
- Collomb, G. (1978). Estimation non-paramétrique de la régression : régressogramme et méthode du noyau. *Publications du Laboratoire de Statistique et Probabilités de l'Université de Toulouse n° 07-78*, pages 1–59.
- Collomb, G. (1981). Estimation non-paramétrique de la régression : revue bibliographique. *Internat. Statist. Rev.*, **49**(1), 75–93.
- Comte, F. et Merlevède, F. (2002). Adaptive estimation of the stationary density of discrete and continuous time mixing processes. *ESAIM Probab. Statist.*, **6**, 211–238 (electronic). New directions in time series analysis (Luminy, 2001).
- Comte, F. et Merlevède, F. (2005). Super optimal rates for nonparametric density estimation via projection estimators. *Stochastic Process. Appl.*, **115**(5), 797–826.
- Cox, D. R. (1962). *Renewal theory*. Methuen & Co. Ltd., London.
- Dehay, D. et Kutoyants, Y. A. (2004). On confidence intervals for distribution function and density of ergodic diffusion process. *J. Statist. Plann. Inference*, **124**(1), 63–73.
- Deheuvels, P. (1977). Estimation non paramétrique de la densité par histogrammes généralisés. *Rev. Statist. Appl.*, **25**(3), 5–42.

- Delecroix, M. (1980). Estimation des densités d'un processus stationnaire à temps continu. *C. R. Acad. Sci. Paris Sér. A-B*, **290**(1), A45–A48.
- Dong, J. et Zheng, C. (2001). Generalized edge frequency polygon for density estimation. *Statist. Probab. Lett.*, **55**(2), 137–145.
- Doukhan, P. (1994). *Mixing*, volume 85 of *Lecture Notes in Statistics*. Springer-Verlag, New York. Properties and examples.
- Eubank, R. L. (1999). *Nonparametric regression and spline smoothing*, volume 157 of *Statistics : Textbooks and Monographs*. Marcel Dekker Inc., New York, second edition.
- Ferraty, F. et Vieu, P. (2002/2003). *Statistique fonctionnelle : modèles nonparamétriques de régression, Notes de cours de DEA*. Université Paul Sabatier (Toulouse 3), <http://www.lsp.ups-tlse.fr/staph/livres.html>.
- Ferraty, F. et Vieu, P. (2006). *Nonparametric functional data analysis*. Springer Series in Statistics. Springer, New York. Theory and practice.
- Freedman, D. et Diaconis, P. (1981). On the histogram as a density estimator :  $L_2$  theory. *Z. Wahrsch. Verw. Gebiete*, **57**(4), 453–476.
- Frenay, A. (2001). Sur l'estimation de la densité marginale d'un processus à temps continu par projection orthogonale. *Ann. I.S.U.P.*, **45**(1), 55–92.
- Gasser, T. et Müller, H.-G. (1979). Kernel estimation of regression functions. In *Smoothing techniques for curve estimation (Proc. Workshop, Heidelberg, 1979)*, volume 757 of *Lecture Notes in Math.*, pages 23–68. Springer, Berlin.
- Geffroy, J. (1980). Étude de la convergence du régressogramme. *Publ. Inst. Statist. Univ. Paris*, **25**(1-2), 41–56.
- Guillou, A. et Merlevède, F. (2001). Estimation of the asymptotic variance of kernel density estimators for continuous time processes. *J. Multivariate Anal.*, **79**(1), 114–137.
- Györfi, L., Härdle, W., Sarda, P., et Vieu, P. (1989). *Nonparametric curve estimation from time series*, volume 60 of *Lecture Notes in Statistics*. Springer-Verlag, Berlin.
- Györfi, L., Kohler, M., Krzyżak, A., et Walk, H. (2002). *A distribution-free theory of nonparametric regression*. Springer Series in Statistics. Springer-Verlag, New York.
- Hall, P. et Hannan, E. J. (1988). On stochastic complexity and nonparametric density estimation. *Biometrika*, **75**(4), 705–714.
- Härdle, W. (1990). *Applied nonparametric regression*, volume 19 of *Econometric Society Monographs*. Cambridge University Press, Cambridge.
- Jones, M. C., Samiuddin, M., Al-Harbey, A. H., et Maatouk, T. A. H. (1998). The edge frequency polygon. *Biometrika*, **85**(1), 235–239.

- Karlin, S. et Taylor, H. M. (1981). *A second course in stochastic processes*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York.
- Kutoyants, Y. A. (1998). Efficient density estimation for ergodic diffusion processes. *Stat. Inference Stoch. Process.*, **1**(2), 131–155.
- Kutoyants, Y. A. (2004). *Statistical inference for ergodic diffusion processes*. Springer Series in Statistics. Springer-Verlag London Ltd., London.
- Labrador, B. (2006). Almost sure convergence of the  $k_T$ -occupation time density estimator. *C. R. Math. Acad. Sci. Paris*, **343**(10), 665–669.
- Leblanc, F. (1993). Estimation de la densité d'un processus strictement stationnaire à temps continu par méthode d'ondelettes. *C. R. Acad. Sci. Paris Sér. I Math.*, **317**(2), 201–204.
- Leblanc, F. (1995). Discretized wavelet density estimators for continuous time stochastic processes. In *Wavelets and statistics (Villard de Lans, 1994)*, volume 103 of *Lecture Notes in Statist.*, pages 209–224. Springer, New York.
- Leblanc, F. (1997). Density estimation for a class of continuous time processes. *Math. Methods Statist.*, **6**(2), 171–199.
- Lecoutre, J.-P. (1982). *Contribution à l'estimation non paramétrique de la régression*. Thèse de l'Université Pierre et Marie Curie (Paris 6).
- Lecoutre, J.-P. (1985). The  $L_2$ -optimal cell width for the histogram. *Statist. Probab. Lett.*, **3**(6), 303–306.
- Lejeune, F.-X. (2005). Vitesses optimale et suroptimale des polygones de fréquences pour les processus à temps continu. *C. R. Math. Acad. Sci. Paris*, **341**(1), 59–62.
- Lejeune, F.-X. (2006). Propriétés des estimateurs par histogrammes et polygones de fréquences de la densité marginale d'un processus à temps continu. *Ann. I.S.U.P.*, **50**(1-2), 47–77.
- Mack, Y. P. et Silverman, B. W. (1982). Weak and strong uniform consistency of kernel regression estimates. *Z. Wahrsch. Verw. Gebiete*, **61**(3), 405–415.
- Major, P. (1973). On a non-parametric estimation of the regression function. *Studia Sci. Math. Hungar.*, **8**, 347–361.
- Marron, J. S. et Härdle, W. (1986). Random approximations to some measures of accuracy in nonparametric curve estimation. *J. Multivariate Anal.*, **20**(1), 91–113.
- Masry, E. (1983). Probability density estimation from sampled data. *IEEE Trans. Inform. Theory*, **29**(5), 696–709.

- Nadaraya, E. A. (1964). On estimating regression. *Theory Probab. Applic.*, **9**, 141–142.
- Nason, G. P. et Silverman, B. (2000). Wavelets for regression and other statistical problems. In *Smoothing and regression : Approaches, computation, and application*, Ed. M. G. Schimek, Wiley Series in Probability and Statistics, pages 159–193. John Wiley & Sons Inc., New York.
- Neveu, J. (1970). *Bases mathématiques du calcul des probabilités*. Préface de R. Fortet. Deuxième édition, revue et corrigée. Masson et Cie, Éditeurs, Paris.
- Nguyen, H. T. (1979). Density estimation in a continuous-time stationary Markov process. *Ann. Statist.*, **7**(2), 341–348.
- Nguyen, H. T. et Pham, T. D. (1980). Sur l'utilisation du temps local en statistique des processus. *C. R. Acad. Sci. Paris Sér. A-B*, **290**(3), A165–A168.
- Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.*, **33**, 1065–1076.
- Parzen, E. (1999). *Stochastic processes*, volume 24 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA. Reprint of the 1962 original.
- Pollard, D. (1984). *Convergence of stochastic processes*. Springer Series in Statistics. Springer-Verlag, New York.
- Qian, Z. et Zheng, W. (2004). A representation formula for transition probability densities of diffusions and applications. *Stochastic Process. Appl.*, **111**(1), 57–76.
- Qian, Z., Russo, F., et Zheng, W. (2003). Comparison theorem and estimates for transition probability densities of diffusion processes. *Probab. Theory Related Fields*, **127**(3), 388–406.
- Rio, E. (2000). *Théorie asymptotique des processus aléatoires faiblement dépendants*, volume 31 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer-Verlag, Berlin.
- Rosenblatt, M. (1956a). A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci. U. S. A.*, **42**, 43–47.
- Rosenblatt, M. (1956b). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.*, **27**, 832–837.
- Sabry, H. (1978). Sur l'estimation non paramétrique des fonctions de régression. *C. R. Acad. Sci. Paris Sér. A-B*, **286**(20), A941–A944.
- Scott, D. W. (1979). On optimal and data-based histograms. *Biometrika*, **66**(3), 605–610.

- Scott, D. W. (1985). Frequency polygons : theory and application. *J. Amer. Statist. Assoc.*, **80**(390), 348–354.
- 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.
- Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. Monographs on Statistics and Applied Probability. Chapman & Hall, London.
- Simonoff, J. S. et Hurvich, C. (1993). A study of the effectiveness of simple density estimation methods. *Comput. Statist.*, **8**, 259–278.
- Stone, C. J. (1983). Optimal uniform rate of convergence for nonparametric estimators of a density function or its derivatives. In *Recent advances in statistics*, pages 393–406. Academic Press, New York.
- Stoyanov, J. M. et Robinson, P. M. (1991). Semiparametric and nonparametric inference from irregular observations on continuous time stochastic processes. In *Nonparametric functional estimation and related topics (Spetses, 1990)*, volume 335 of *NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci.*, pages 553–558. Kluwer Acad. Publ., Dordrecht.
- Tran, L. T. (1994). Density estimation for time series by histograms. *J. Statist. Plann. Inference*, **40**(1), 61–79.
- Tsai, H. et Chan, K. S. (2000). A note on the covariance structure of a continuous-time ARMA process. *Statist. Sinica*, **10**(3), 989–998.
- Tukey, J. W. (1961). Curves as parameters, and touch estimation. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob.*, Vol. I, pages 681–694. Univ. California Press, Berkeley, Calif.
- Vašíček, O. A. (1977). An equilibrium characterization of the term structure. *J. Fin. Econ.*, **5**(2), 177–188.
- Veretennikov, A. Y. (1999). On Castellana-Leadbetter’s condition for diffusion density estimation. *Stat. Inference Stoch. Process.*, **2**(1), 1–9 (2000).
- Vieu, P. (1991). Quadratic errors for nonparametric estimates under dependence. *J. Multivariate Anal.*, **39**(2), 324–347.
- Vilar, J. A. et Vilar, J. M. (2000). Finite sample performance of density estimators from unequally spaced data. *Statist. Probab. Lett.*, **50**(1), 63–73.
- Wahba, G. (1990). *Spline models for observational data*, volume 59 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

Watson, G. S. (1964). Smooth regression analysis. *Sankhyā Ser. A*, **26**, 359–372.

Wu, B. (1997). Kernel density estimation under weak dependence with sampled data. *J. Statist. Plann. Inference*, **61**(1), 141–154.