Dépendance faible: estimation et théorèmes limite.Application à l'étude statistique de certains systèmes dynamiques.

Clémentine Prieur

UNIVERSITÉ PAUL SABATIER (TOULOUSE III)

Habilitation à diriger des recherches

présentée par

Clémentine PRIEUR
Laboratoire de Statistique et Probabilités
INSA Toulouse

Dépendance faible : estimation et théorèmes limite. Application à l'étude statistique de certains systèmes dynamiques.

Soutenue publiquement le 9 décembre 2006 devant le Jury composé de

Pr. Paul DOUKHAN
Pr. Fabrice GAMBOA
Pr. Béatrice LAURENT
Pr. Thomas MIKOSCH
Pr. Eric MOULINES
Pr. Emmanuel RIO
Pr. Bernard SCHMITT

RAPPORTEURS :
Pr. Mikhail GORDIN
Pr. Thomas MIKOSCH
Pr. Emmanuel RIO

UNIVERSITÉ PAUL SABATIER (TOULOUSE III)

Habilitation à diriger des recherches

présentée par

Clémentine PRIEUR
Laboratoire de Statistique et Probabilités
INSA Toulouse

Dépendance faible : estimation et théorèmes limite. Application à l'étude statistique de certains systèmes dynamiques.

Soutenue publiquement le 9 décembre 2006 devant le Jury composé de

Pr. Paul DOUKHAN
Pr. Fabrice GAMBOA
Pr. Béatrice LAURENT
Pr. Thomas MIKOSCH
Pr. Eric MOULINES
Pr. Emmanuel RIO
Pr. Bernard SCHMITT

RAPPORTEURS :
Pr. Mikhail GORDIN
Pr. Thomas MIKOSCH
Pr. Emmanuel RIO

Remerciements

Je tiens ici à exprimer ma reconnaissance envers tous ceux qui m'ont accompagnée ces dernières années. Je remercie tout d'abord Paul Doukhan qui, le premier, m'a formée à la recherche, sachant me guider tout en me laissant prendre l'indépendance indispensable à un chercheur.

Je suis très touchée de l'honneur que m'ont fait Mikhail Gordin, Thomas Mikosch et Emmanuel Rio en acceptant de rapporter mon travail. Les différents échanges que j'ai eus avec eux ont été très enrichissants.

Merci à Béatrice Laurent d'avoir accepté de présenter mon dossier de recherche. C'est avec elle aussi que j'ai découvert les joies et les devoirs de l'encadrement doctoral.

Je remercie vivement Fabrice Gamboa, Eric Moulines et Bernard Schmitt d'avoir accepté de faire partie du jury.
Dès le début de ma thèse, Bernard Schmitt a accordé de l'attention à mes travaux sur l'estimation fonctionnelle des systèmes dynamiques. Je l'en remercie sincèrement.
Merci à Fabrice Gamboa, avec qui travail et bonne humeur sont souvent associés.

Viviane Baladi n'a malheureusement pas pu faire partie de ce jury. Je profite de l'occasion pour lui exprimer ma profonde gratitude pour les nombreux échanges que nous avons eus.

Un grand merci également à Chichi, qui m'a si chaleureusement accueillie à Caracas en janvier 2006, et avec qui j'ai beaucoup échangé durant son séjour dans mon bureau à l'INSA au printemps dernier.

Je voudrais exprimer ma gratitude envers mes différents coauteurs. J'exprime une pensée plus particulière pour Jérôme Dedecker avec qui j'ai vraiment apprécié travailler ces dernières années.

Je n'oublie pas mes nombreux collègues, et bien souvent amis du département Génie Mathématique et Modélisation de l'INSA Toulouse, ainsi que du Laboratoire de Statistique et Probabilités.

Enfin, merci à tous ceux que je n'ai pas cités, mais qui un jour ou l'autre m'ont aidée à franchir des étapes.

Je garde mes derniers remerciements pour Christophe qui, depuis plusieurs années déjà, m'encourage et me soutient dans mes nombreux projets.

A Christophe, Domitille et Gaétane
A celle ou celui que nous attendons

Résumé : Le thème central des travaux présentés est l'étude des suites faiblement dépendantes non α-mélangeantes au sens de Rosenblatt (1956). La notion de mélange classique est affaiblie afin d'établir des inégalités ainsi que des théorèmes limite pour différentes classes de processus comme par exemple certains systèmes dynamiques, des chaînes de Markov non irréductibles, ou encore des fonctions de processus linéaires non mélangeants. Les résultats obtenus sont ensuite appliqués au domaine de la statistique non paramétrique.

Deux autres thématiques sont abordées dans ce manuscrit : d'une part l'étude de principes de grandes déviations (notamment pour le processus de records généralisés), et d'autre part l'estimation adaptative de fonctionnelles linéaires.

AMS 2000 Subject Classification : 37A50, 60F05, 60F10, 60F17, 60G10, 62G07, 62G08.

Mots clés : dépendance faible, processus stationnaires, théorèmes limite, inférence non paramétrique, estimation adaptative, grandes déviations.

Abstract : The major part of the presented work is devoted to new concepts of dependence extending and generalizing the classical concepts of mixing of a stationary sequence. These concepts allow the statement of various inequalities and limit theorems for different classes of processes such as dynamical systems, non irreducible Markov chains, or Bernoulli shifts. These results are then applied to derive various applications in the field of non parametric statistic.

Two other fields are also studied in the present manuscript : on one hand the study of large deviation principles (in particular for generalized records for triangular arrays of exchangeable variables), and the other hand the problem of adaptive estimation of linear functionals by Gaussian model selection.

AMS 2000 Subject Classification : 37A50, 60F05, 60F10, 60F17, 60G10, 62G07, 62G08.

Key words : weak dependence, stationary processes, limit theorems, non parametric inference, adaptive estimation, large deviations.

Table des matières

Chapitre 1

Curriculum Vitæ

Clémentine PRIEUR, née COULON

INSA Toulouse
Génie Mathématique et Modélisation (GMM)
Laboratoire de Statistique et Probabilités
135, avenue de Rangueil
31077 Toulouse cedex 4
Tél : ++33 / 5 61 55 93 17, Fax : ++33 / 5 61 55 93 20

e-mail : `clementine.prieur@insa-toulouse.fr`
page-web : `www.lsp.ups-tlse.fr/Fp/Prieur`

Née le 06/02/1976 à Bron (69)
Nationalité française
Mariée, deux enfants

Formation

- **Septembre 1999-Décembre 2001** Thèse de Doctorat à l'Université de Cergy-Pontoise.
  Sujet : *Applications statistiques de suites faiblement dépendantes et de systèmes dynamiques.*
  Jury de thèse : G. Fayolle (président), J. Diebolt, E. Rio (rapporteurs), P. Doukhan (directeur), M. Benaïm, P. Bertail, G. Grégoire, D. Talay (examinateurs).
  Mention Très honorable avec Félicitations du Jury.
- **1999** *Agrégation externe de Mathématiques.*
- **1996-1999** *Magistère de l'Ecole Normale Supérieure de Cachan.*
- **1997-1998** *Diplôme d'Etudes Approfondies (DEA) Modélisation Stochastique et Statistique de l'Université Paris XI-Orsay.*

Parcours professionnel

- Maître de conférences à l'INSA Toulouse depuis le 01/09/02.
- Allocataire Moniteur Normalien à l'Université de Cergy-Pontoise du 01/09/00 au 31/08/02.
- Ancienne élève de l'Ecole Normale Supérieure de Cachan.

Membre du groupe MAS de la SMAI.

Bénéficiaire de la prime d'encadrement doctoral et de recherche (PEDR) 2005-2009.

Enseignements Mes enseignements sont détaillés au chapitre 5.

Encadrements d'activités de recherche

- **Projets de fin d'étude**
  - 2003-2006 Responsable des projets de fin d'étude à l'INSA Toulouse, GMM.
  - 2005-2006 Aude Coutelier et Sophie Martin *Méthode non paramétrique d'homo-généisation des températures quotidiennes* Projet de fin d'étude (INSA Toulouse GMM), co-encadré avec O. Mestre (Météo France Toulouse). Résultats présentés à la conférence **5th Seminar for Homogenization and Quality Control of Climatological Databases** (Budapest, mai 2006).
- **Mémoires de MR2**
  - 2004-2005 Jordi Ordoñez Adellach *Systèmes dynamiques et grandes déviations* Mémoire de MR2 (Toulouse 3), co-encadré avec B. Bercu.
  - 2005-2006 Matthieu Lerasle *Intervalles de confiance adaptatifs* Mémoire de MR2 (Orsay), co-encadré avec B. Laurent.
- **Thèses de Doctorat**
  - 2006 Matthieu Lerasle *Régions de confiance adaptatives dans des modèles non paramétriques* Thèse co-encadrée avec B. Laurent.

Ces encadrements sont détaillés au chapitre 5.

Rédaction d'un livre En collaboration avec J. Dedecker, P. Doukhan, G. Lang, J.R. León et S. Louhichi, je participe à la rédaction d'un livre sur la dépendance faible. Cet ouvrage, intitulé *Weak dependence, examples and applications*, doit être publié chez Springer.

Programmes de coopérations industriels et institutionnels

- Membre du réseau européen Physical & Engineering Sciences : Advanced Mathematical Methods for Finance (AMaMeF)
  `http ://www.lsp.ups-tlse.fr/Recherche/Projets/index.html#amamef`
- Collaboration en cours avec Météo France sur les problèmes d'homogénéisation de séries quotidiennes (en collaboration avec O. Mestre, Météo France Toulouse).
- Collaboration avec N. Lopes Garcia, J. Enrique Garcia (Campinas, Brésil) et A. Galves (São Paolo, Brésil) *The identification of regular and transient regions in the speech sonority*, visite à Campinas en mai et juin 2006.
- Contrat avec M. Mouret du *Laboratoire Matériaux et Durabilité des Constructions* (en collaboration avec J.M. Bardet, Paris 1) sur l'étude des propriétés rhéologiques du ciment à l'état frais, 2003.

Ces programmes sont détaillés au chapitre 6.

Jury de concours

- Participation aux entretiens d'entrée à l'INSA en 3ème année, session 2003.
- Membre du jury de l'agrégation externe de mathématiques, épreuve de modélisation depuis 2003.

Autres responsabilités collectives

- **Organisation de congrès internationaux**
  - Membre du comité d'organisation du 27ème congrès EMS (European Meeting of Statisticians), Toulouse, 20-24 juillet 2009.
  - Membre du comité d'organisation du *Workshop on insider models*, conférence *AMAMEF* (Adavanced Mathematical Methods for Finance), Toulouse, 25-27 janvier 2007.
    `www.lsp.ups-tlse.fr/Fp/Baudoin/amamefindex.html`
  - Organisation de la session *Etude statistique des systèmes dynamiques* aux journées MAS à Lille (septembre 2006).
  - Membre du comité d'organisation des journées Planification d'expériences et analyse d'incertitudes pour les gros codes numériques : Approches Stochastiques,
    Toulouse, 2-3 février 2006.
  - Membre du comité d'organisation du second colloque Autosimilarité et applications à l'INSA Toulouse du 20 au 24 juin 2005.
    `http ://www.lsp.ups-tlse.fr/Autosim05/`
  - Membre du comité d'organisation de la Journée (internationale) Dépendance et Longue Mémoire à l'INSA Toulouse le 22 novembre 2002.
    `www.lsp.ups-tlse.fr/Azais/journee/lonmem2.html`

- **Participation à des commissions**
  - Membre élu de la commission de spécialistes de la 26$^{\text{ème}}$ section de l'INSA Toulouse.
  - Membre extérieur nommé de la commission de spécialistes de la 26$^{\text{ème}}$ section de l'Université Paul Sabatier.
- **Rapporteur d'articles** pour *ESAIM/Prob. Stat.*, *International Journ. Statist. and Systems*, *Stochastic Process. Appl.*, *Ann. Inst. H. Poincaré Probab. Statist.*, ...

Séminaires, mobilité

- 2006 (2 semaines) : Invitation à l'Université Centrale de Caracas (Vénézuéla).
- 2006 (6 semaines) : Séjour à l'Université d'Etat de Campinas (Brésil).
- Présentation de séminaires, dans des universités ou instituts de recherche français : Ensae, Rennes, Rouen, Nice, Amiens, Pau, Paris I, Marne-la-Vallée, Dijon ..., ou étrangers : Santander (Espagne), Caracas (Vénézuéla), Campinas (Brésil), São Paolo (Brésil) ...
- Participation à de nombreuses conférences nationales et internationales (voir liste jointe au chapitre 4).

Liste des publications Elle est présentée au chapitre 3.

Chapitre 2

Synthèse des travaux

Ce document présente une synthèse de mes travaux de recherche dont le thème central est l'étude des processus stochastiques faiblement dépendants. L'enjeu principal de ces travaux a été d'obtenir différents théorèmes limite, ainsi que des inégalités exponentielles, tout en ayant à disposition une grande variété de modèles. En particulier, un accent important a été mis sur l'étude des propriétés statistiques des systèmes dynamiques. L'idée consistant à ramener l'étude de ces derniers à celle de processus faiblement dépendants a permis notamment de résoudre des problèmes d'estimation fonctionnelle pour certains systèmes dynamiques.

La réalisation de ces travaux a nécessité des développements techniques importants. Il a également fallu introduire de nouvelles mesures de dépendance pour appréhender plus de modèles. Le contrôle des termes de covariance, qui apparaissent de façon naturelle lorsqu'on travaille avec des processus dépendants, a constitué un des points clés des différentes études.

Plus récemment, j'ai élargi mes thèmes de recherche aux deux sujets suivants : d'une part l'étude de principes de grandes déviations (notamment pour le processus de records généralisés), et d'autre part l'estimation adaptative de fonctionnelles linéaires.

Dans la première partie (partie 2.1) de ce document de synthèse, je présente mes différents travaux sur les suites faiblement dépendantes. Cette partie comporte trois sous-parties portant respectivement sur les problèmes d'estimation fonctionnelle, sur les inégalités exponentielles, avec comme application une loi du logarithme itéré, et enfin sur l'étude du processus empirique. La deuxième partie (partie 2.2) présente une classe de systèmes dynamiques dont nous avons ramené l'étude à celle d'une chaîne de Markov faiblement dépendante. Les deux parties qui suivent (parties 2.3 et 2.4) sont indépendantes des deux premières parties. La partie 2.3 est consacrée à l'étude du comportement asymptotique (principe de Donsker et grandes déviations) du processus des records généralisés. La partie 2.4 porte sur le problème de l'estimation adaptative d'une fonctionnelle linéaire dans un modèle gaussien. Enfin, la partie 2.5 expose en guise de conclusion les perspectives offertes par l'ensemble de ces travaux.

Dans un soucis de lisibilité du présent manuscrit, certains résultats ne seront pas énoncés dans toute leur généralité.

Les références **[Pri-*]**, $* \in \{0, 1, \dots, 18\}$, renvoient à mes publications présentées au chapitre 3.

2.1 Dépendance faible

En modélisation, il est souvent peu réaliste de supposer l'indépendance. C'est pourquoi, depuis plusieurs années déjà, s'est développée une vaste littérature sur les processus dépendants. Les différentes notions de mélange font sans aucun doute partie des conditions de dépendance faible les plus étudiées. Les trois principales classes de mélange sont le mélange fort (ou α-mélange), le β-mélange et le mélange uniforme (ou ϕ-mélange) introduits respectivement par Rosenblatt [52], Rozanov & Volkonskii [54] et Ibragimov [32]. Malheureusement, les conditions de mélange sont souvent difficiles à vérifier. On sait même que certains processus bien classiques ne sont pas mélangeants. Citons le célèbre exemple d'Andrews [1] :

si $(\varepsilon_i)_{i\geq 1}$ est une suite de variables aléatoires i.i.d. (indépendantes et identiquement distribuées), de marginale $\mathcal{B}(1/2)$, alors la solution stationnaire $(X_i)_{i\geq 0}$ de l'équation

$$X_n = \frac{1}{2}(X_{n-1} + \varepsilon_n)\,,\ X_0 \textit{ indépendante de } (\varepsilon_i)_{i\geq 1} \tag{2.1.1}$$

n'est pas fortement mélangeante.

De là est venue l'idée d'introduire des notions de dépendance moins restrictives, couvrant plus de modèles, pour lesquelles nous avons su développer une théorie asymptotique intéressante. Ces différents coefficients de dépendance vont nous permettre en particulier d'étudier des séries temporelles classiques en économétrie (voir les différents exemples dans **[Pri-1, Pri-18]**). Commençons par introduire différentes notions de dépendance faible utilisées dans la suite de ce manuscrit. Le premier type de dépendance, sur lequel porte les articles **[Pri-7, Pri-8, Pri-11]** écrits pendant ma thèse, est défini ci-dessous.

Définition 1 *Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. On définit*

$$\mathcal{F} = \cup_{u\in\mathbb{N}^*}\mathcal{F}_u \quad \textit{et} \quad \mathcal{G} = \cup_{u\in\mathbb{N}^*}\mathcal{G}_u\,,$$

où $\mathcal{F}_u$ et $\mathcal{G}_u$ sont deux classes de fonctions de $\mathbb{R}^u$ dans $\mathbb{R}$. Soient X et Y deux variables aléatoires à valeurs dans $\mathbb{R}^u$ et $\mathbb{R}^v$ respectivement. Si ψ est une fonction de $\mathcal{F}\times\mathcal{G}$ dans $\mathbb{R}_+$, on définit le coefficient de $(\mathcal{F}, \mathcal{G}, \psi)$-dépendance $\varepsilon(X, Y)$ par

$$\varepsilon(X,Y) = \sup_{f\in\mathcal{F}_u\,,\, g\in\mathcal{G}_v} \frac{|\mathrm{Cov}(f(X), g(Y))|}{\psi(f,g)}\,.$$

Soit $(X_n)_{n\in\mathbb{Z}}$ une suite de variables aléatoires réelles. Soit $\Gamma(u, v, k)$ l'ensemble des indices (i, j) dans $\mathbb{Z}^u\times\mathbb{Z}^v$ tels que $i_1 < \cdots < i_u \leq i_u + k \leq j_1 < \cdots < j_v$. Le coefficient de dépendance

$\varepsilon(k)$ est alors défini par

$$\varepsilon(k) = \sup_{u,v} \sup_{(i,j)\in\Gamma(u,v,k)} \varepsilon((X_{i_1},\ldots,X_{i_u}),(X_{j_1},\ldots,X_{j_v}))\,.$$

La suite $(X_n)_{n\in\mathbb{Z}}$ est dite $(\mathcal{F},\mathcal{G},\psi)$-dépendante si la suite $(\varepsilon(k))_{k\in\mathbb{N}}$ tend vers zéro.

Remarque 1

- *Cette définition s'étend sans difficulté au cas où la suite $(X_n)_{n\in\mathbb{Z}}$ est à valeurs dans un espace polonais $\mathcal{X}$, ainsi qu'au cas où l'ensemble des indices (i,j) est un espace métrique quelconque* **[Pri-18]**.
- *Cette notion de dépendance, qui apparaît pour la première fois dans les articles* [8, 24]*, rend explicite l'indépendance asymptotique entre le "passé" et le "futur". Le "passé" est progressivement oublié. Pour la suite initiale $(X_i)_{i\in\mathbb{Z}}$, le "passé" et le "futur" sont définis à partir de marginales finidimensionnelles.*

Les articles **[Pri-7, Pri-8, Pri-11]** se placent dans les deux cas particuliers suivants :

ζ-dépendance

On suppose que les X_i sont dans $\mathbb{L}^1(\mathbb{P})$. Sur $\mathbb{R}^u$, on définit la distance $d_1(x,y)=\sum_{i=1}^u |x_i-y_i|$. On prend pour $\mathcal{F}_u=\mathcal{G}_u$ l'ensemble des fonctions de $\mathbb{R}^u$ dans $\mathbb{R}$ qui sont Lipschitziennes par rapport à la distance d_1. Si $f\in\mathcal{F}^u$, on définit $d_f := u$ et $\mathrm{Lip}\,(f) := \sup_{x\neq y}\frac{|f(x)-f(y)|}{d_1(x,y)}$ la constante de Lipschitz de f. Alors, le coefficient ζ correspond au choix $\psi(f,g)=\min(d_f,d_g)\mathrm{Lip}\,(f)\mathrm{Lip}\,(g)$.

θ-dépendance

On suppose encore que les X_i sont dans $\mathbb{L}^1(\mathbb{P})$. On prend pour $\mathcal{F}_u$ la classe des fonctions bornées de $\mathbb{R}^u$ dans $\mathbb{R}$, et pour $\mathcal{G}_u$ la classe des fonctions de $\mathbb{R}^u$ dans $\mathbb{R}$ qui sont Lipschitziennes par rapport à la distance d_1. Alors, le coefficient θ correspond au choix $\psi(f,g)=d_g\|f\|_\infty\mathrm{Lip}\,(g)$.

Notons que dans la définition de θ, aucune hypothèse de régularité n'est faite sur les fonctions de l'ensemble $\mathcal{F}$. Lorsque la suite $(X_n)_{n\in\mathbb{Z}}$ est adaptée par rapport à la filtration croissante $(\mathcal{M}_i)_{i\in\mathbb{Z}}=(\sigma(X_j,\, j\le i))_{i\in\mathbb{Z}}$, il est souvent recommandé d'utiliser θ plutôt que ζ. On dit alors qu'on travaille dans un cadre causal.

Dans les articles **[Pri-11, Pri-8]**, nous donnons différentes classes d'exemples pour lesquelles nous écrivons des bornes explicites pour les coefficients $\zeta(k)$ et $\theta(k)$.

En fait, le coefficient θ appartient à une classe plus étendue de coefficients de dépendance, dont l'étude a été systématisée dans **[Pri-4]**. Cette classe de coefficients fait intervenir les espérances conditionnelles par rapport à la filtration du passé $\mathcal{M}_i := \sigma(X_j,\, j\le i)$. Commençons par définir les espaces fonctionnels suivants :

- $\Lambda := \cup_{u\in\mathbb{N}^*}\Lambda_u$, où Λ_u désigne l'ensemble des fonctions $f:\mathbb{R}^u\to\mathbb{R}$, Lipschitziennes par rapport à la distance d_1,

– $\Lambda^{(1)} := \{f \in \Lambda\,,\ \mathrm{Lip}\,(f) \leq 1\}$.

Définition 2 *Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, et $\mathcal{M}$ une sous-tribu de $\mathcal{A}$. Soit X une variable aléatoire à valeurs dans $\mathbb{R}^u$, avec $u \in \mathbb{N}^*$. On suppose que X est $\mathbb{L}^1$-intégrable, c'est-à-dire qu'il existe un x_0 dans $\mathbb{R}^u$ tel que la variable aléatoire réelle $d_1(X, x_0)$ soit dans $\mathbb{L}^1(\mathbb{P})$. On définit alors*

$$\theta(\mathcal{M}, X) = \sup\left\{\|\mathbb{E}(g(X)|\mathcal{M}) - \mathbb{E}(g(X))\|_1\ ,\ g \in \Lambda^{(1)}\right\}\ .$$

Soit $(X_i)_{i\in\mathbb{Z}}$ une suite de variables aléatoires réelles dans $\mathbb{L}^1(\mathbb{P})$. On définit alors

$$\theta_r(k) = \max_{l\leq r} \frac{1}{l} \sup_{(i,j)\in\Gamma(1,l,k)} \theta(\mathcal{M}_i, (X_{j_1}, \ldots, X_{j_l}))\ .$$

On peut montrer que $\theta(k) = \theta_\infty(k)$. Dans la définition 2 ci-dessus, on peut inverser l'ordre de la norme $\|.\|_1$ et du supremum. On obtient alors un autre coefficient

$$\tau(\mathcal{M}, X) = \left\|\sup\left\{\int g(x)\mathbb{P}_{X|\mathcal{M}}(dx) - \int g(x)\mathbb{P}_X(dx)\,,\ g \in \Lambda^{(1)}\right\}\right\|_1\ .$$

Ce coefficient a été introduit pour la première fois par Rüschendorf [55] dans le cas où Ω est un espace polonais puis généralisé au cas Ω quelconque dans **[Pri-6]**. Ce coefficient présente une propriété de couplage intéressante dont nous parlerons dans le paragraphe 2.1.2.

Pour définir les coefficients θ et τ, nous nous sommes limités à la classe de fonctions $\Lambda^{(1)}$. Introduisons maintenant les trois mesures de dépendance suivantes, faisant intervenir une classe de fonctions construite à partir des fonctions à variation bornée (voir Définition 3 ci-dessous).

Définition 3 *Une fonction $h : \mathbb{R} \to \mathbb{R}$ est dans l'ensemble BV s'il existe une mesure signée finie dh telle que $h(x) = h(0) + dh([0, x[)$ si $x \geq 0$ et $h(x) = h(0) - dh([x, 0[)$ si $x \leq 0$. On définit BV_1 comme l'ensemble des fonctions h de BV telles que $\|dh\| \leq 1$, où $\|dh\|$ désigne la variation totale de dh.*

Voici alors les définitions des coefficients dits de mélange faible.

Définition 4 *Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, $X = (X_1, \ldots, X_r)$ une variable aléatoire à valeurs dans $\mathbb{R}^r$, avec $r \in \mathbb{N}^*$, et $\mathcal{M}$ une sous-tribu de $\mathcal{A}$. Si f est une fonction dans BV_1, on pose $f^{(i)}(x) = f(x) - \mathbb{E}(f(X_i))$. On définit alors*

1. $\tilde{\alpha}(\mathcal{M}, X) = \sup_{f_1,\ldots,f_r\in BV_1} \left\|\mathbb{E}\left(\prod_{i=1}^r f_i^{(i)}(X_i)\Big|\mathcal{M}\right) - \mathbb{E}\left(\prod_{i=1}^r f_i^{(i)}(X_i)\right)\right\|_1.$
2. $\tilde{\beta}(\mathcal{M}, X) = \left\|\sup_{f_1,\ldots,f_r\in BV_1}\left|\int \prod_{i=1}^r f_i^{(i)}(x_i)\left(\mathbb{P}_{X|\mathcal{M}} - \mathbb{P}_X\right)(dx)\right|\right\|_1.$

3. $\tilde{\phi}(\mathcal{M}, X) = \sup_{f_1,\dots,f_r \in BV_1} \left\| \mathbb{E}\left(\prod_{i=1}^{r} f_i^{(i)}(X_i)|\mathcal{M}\right) - \mathbb{E}\left(\prod_{i=1}^{r} f_i^{(i)}(X_i)\right) \right\|_\infty .$

Remarque 2 *Lorsque* $r = 1$, *le coefficient* $\tilde{\alpha}$ *a été introduit par Rio* [50, equation 1.10c], *alors que les coefficients* $\tilde{\beta}$ *et* $\tilde{\Phi}$ *ont été introduits dans* **[Pri-4]**. *Dans* **[Pri-1]**, *nous avons généralisé les définitions de* $\tilde{\beta}$ *et de* $\tilde{\phi}$ *au cas où* $r \in \mathbb{N}^*$ *est quelconque.*

Ces coefficients sont plus faibles que les coefficients de mélange usuels correspondants. On montre facilement que $\tilde{\alpha}(\mathcal{M}, X) \leq 2\alpha(\mathcal{M}, \sigma(X))$, $\tilde{\beta}(\mathcal{M}, X) \leq \beta(\mathcal{M}, \sigma(X))$ *et* $\tilde{\Phi}(\mathcal{M}, X) \leq \Phi(\mathcal{M}, \sigma(X))$.

Pour une suite de variables aléatoires réelles, on a la définition 5 suivante :

Définition 5 *Soit* $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé. Soit* $(X_i)_{i\in\mathbb{Z}}$ *une suite de variables aléatoires réelles. Alors, pour* $r \in \mathbb{N}^*$ *et* $k \geq 0$, *on définit*

$$\tilde{\alpha}_r(k) = \max_{1\leq l\leq r} \sup_{(i,j)\in\Gamma(1,l,k)} \tilde{\alpha}(\mathcal{M}_i, (X_{j_1}, \dots, X_{j_l})) .$$

On définit de la même façon les coefficients $\tilde{\beta}_r(k)$ *et* $\tilde{\Phi}_r(k)$.

Remarque 3 *L'intérêt de ces coefficients par rapport aux coefficients de mélange classiques est qu'ils permettent d'étudier plus de modèles. En particulier, pour l'exemple d'Andrews (2.1.1), nous montrons que* $\tilde{\Phi}_1(i) \leq 2^{-i}$. *Pour plus d'exemples, nous renvoyons aux articles* **[Pri-1, Pri-4]**. *L'exemple des systèmes dynamiques sera développé dans la section 2.2.*

En utilisant entre autres la propriété de couplage du coefficient τ (voir **[Pri-6]**, [55]), nous avons établi le lemme de comparaison suivant :

Lemme 1 [Pri-4] *Soient* $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé,* X *une variable aléatoire à valeurs réelles et* $\mathcal{M}$ *une sous-tribu de* $\mathcal{A}$.

1. *On a les inégalités suivantes* $\tilde{\alpha}(\mathcal{M}, X) \leq \tilde{\beta}(\mathcal{M}, X) \leq \tilde{\phi}(\mathcal{M}, X)$.
2. *Soit* Q_X *l'inverse généralisée de la fonction de queue* $t \to \mathbb{P}(|X| > t)$ *: si* $u \in [0, 1]$, $Q_X(u) = \inf\{t \in \mathbb{R} : \mathbb{P}(|X| > t) \leq u\}$. *On a l'inégalité*

$$\tau(\mathcal{M}, X) \leq 2 \int_0^{\tilde{\alpha}(\mathcal{M},X)} Q_X(u) du \,.$$

3. *On suppose de plus que* X *a une fonction de répartition* F *continue de module de continuité* w. *Soit* g *définie par* $g(x) = xw(x)$. *Alors*

$$\tilde{\beta}(\mathcal{M}, X) \leq \frac{2\tau(\mathcal{M}, X)}{g^{-1}(\tau(\mathcal{M}, X))} \,. \tag{2.1.2}$$

En particulier, si F est Höldérienne, c'est-à-dire si $|F(x) - F(y)| \leq C|x - y|^\kappa$ pour un $\kappa \in]0, 1]$ et un $C > 0$, alors

$$\tilde{\beta}(\mathcal{M}, X) \leq 2C^{1/(\kappa+1)} \left(\tau(\mathcal{M}, X)\right)^{\kappa/(\kappa+1)} .$$

Si X a une densité bornée par K, on obtient la borne

$$\tilde{\beta}(\mathcal{M}, X) \leq 2\sqrt{K\tau(\mathcal{M}, X)} . \tag{2.1.3}$$

2.1.1 Estimation fonctionnelle

Les articles **[Pri-11, Pri-4, Pri-7, Pri-15]** traitent de l'estimation fonctionnelle pour des suites faiblement dépendantes. Dans les articles **[Pri-11, Pri-4]**, nous nous sommes intéressés au problème de l'estimation de la densité marginale d'une suite sationnaire. Dans **[Pri-7, Pri-15]**, nous avons étudié des problèmes de rupture dans un modèle de régression non paramétrique.

Estimation de la densité

Dans ce paragraphe, nous considérons une suite de variables aléatoires réelles $(X_i)_{i\geq 0}$, stationnaire, de densité marginale f inconnue. Soit $K : \mathbb{R} \to \mathbb{R}$ telle que $\int K(x)dx < \infty$. Soit (h_n) une suite de réels positifs qui converge vers zéro et telle que (nh_n) diverge vers l'infini. Pour estimer f, nous considérons dans un premier temps l'estimateur à noyau classique :

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

L'estimateur recentré et renormalisé est défini par :

$$Z_n(x) := Z_{n,h_n}(x) = \sqrt{nh_n}\left(\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x)\right) .$$

Nous obtenons alors le théorème suivant :

Théorème 1 [Pri-11] *On suppose que la suite $(X_i)_{i\geq 0}$ est θ_2-dépendante avec la condition suivante :*

$$\exists\, 0 < a < \frac{1}{3} \text{ tel que } \sum_{k=1}^{\infty} \theta_2(k)^a < \infty .$$

On suppose de plus que la densité marginale f est continue et que les densités marginales jointes $f_k(x, y)$ des couples (X_0, X_k) existent pour tout $k > 0$ et vérifient $\sup_{k>0} \sup_{(x,y)\in\mathbb{R}^2} f_k(x, y) < \infty$.

Alors les répartitions finidimensionnelles $(Y_n(x_1), \ldots, Y_n(x_l))$ du processus

$$Y_n(x) := Z_n(x)/\sqrt{f(x)\int K^2(t)dt}$$

convergent en loi vers une Gausienne $\mathcal{N}(0, I_l)$ *dès que* $\Pi_{i=1}^{l} f(x_i) \neq 0$.

Ce résultat reste valide si la suite $(X_i)_{i\geq 0}$ *est* ζ*-dépendante avec la condition suivante :*

$$\exists\, 0 < a < \frac{1}{4} \text{ tel que } \sum_{k=1}^{\infty} \zeta(k)^a < \infty \,.$$

Remarque 4

- *Ce théorème est un corollaire d'un théorème limite centrale pour des tableaux triangulaires faiblement dépendants établi dans* **[Pri-11]**.
- *L'étude du comportement du biais* $\mathbb{E}\hat{f}_n(x) - f(x)$ *est classique. Elle dépend uniquement de la régularité de* f *et non des propriétés de dépendance de la suite.*
- *La condition en* ζ *peut être affaiblie en considérant dans la définition 1* $\sup_{u\in\mathbb{N}^*,\, v\in\{1,2\}} \sup_{(i,j)\in\Gamma(u,v,k)}$ *pour définir* $\zeta(k)$*,* *c'est-à-dire en prenant au plus deux points dans le futur* **[Pri-11]**.
- *La démonstration de ce théorème repose principalement sur la méthode de Lindeberg-Rio initiée par Rio dans le cas* α*-mélangeant* [48].
- *Ce résultat améliore les résultats antérieurs de Doukhan & Louhichi* [25] *obtenus par la méthode des blocs de Bernstein. Dans* [25]*, les conditions de dépendance étaient :*
  - $\theta_\infty(k) = (k^{-a})$ *pour un* $a > \max\{9, 3/2(1+\delta^{-1})\}$ *si* $h_n \sim n^{-\delta}$*, ou*
  - $\zeta(k) = (k^{-a})$ *pour un* $a > 12$.

Par la suite, nous nous sommes intéressés à l'étude du MISE (Erreur Quadratique Moyenne Intégrée), pour des suites $\tilde{\beta}$-dépendantes. Avant d'énoncer les résultats obtenus dans ce cadre, il est important de citer les inégalités de covariance suivantes :

Proposition 1 [Pri-4] *Soit* $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé. Soient* X *et* Y *deux variables aléatoires réelles, et* h *une fonction de l'espace* BV*. Si* Y*,* $h(X)$ *et* $Yh(X)$ *sont intégrables, alors*

$$\mathrm{Cov}(Y, h(X)) = -\int \mathrm{Cov}(Y, \mathbb{1}_{X\leq t})dh(t) \,.$$

Soient $\mathcal{M}$ *une sous-tribu de* $\mathcal{A}$*, et* $b(\mathcal{M}, X) = \sup_{t\in\mathbb{R}} |F_{X|\mathcal{M}}(t) - F_X(t)|$*. Si* Y *est* $\mathcal{M}$*-mesurable, alors on a les inégalités suivantes :*

$$|\mathrm{Cov}(Y, h(X))| \leq \|dh\|\, \mathbb{E}(|Y| b(\mathcal{M}, X)) \leq \|dh\|\, \|Y\|_1\, \tilde{\phi}(\mathcal{M}, X) \,.$$

Remarque 5 *La première inégalité est à comparer avec l'inégalité de Delyon* [22] *(voir aussi Viennet* [58, Lemma 4.1]*) dans laquelle apparaissent deux variables* $b_1(\mathcal{M}, X)$ *et* $b_2(\mathcal{M}, X)$ *ayant même moyenne* $\beta(\mathcal{M}, \sigma(X))$*. La principale différence est que notre inégalité n'est pas symétrique, puisque le coefficient* $\tilde{\beta}(\mathcal{M}, X)$ *ne l'est pas.*

Nous obtenons alors le résultat suivant pour l'estimateur à noyau de la densité :

Proposition 2 [Pri-4] *Si K est dans BV, alors*

$$nh_n \int \mathrm{Var}(\hat{f}_n(x))dx \leq \int K^2(x)dx + 2\left(\sum_{k=1}^{n-1} \tilde{\beta}_1(k)\right) \|dK\| \int |K(x)|dx \; .$$

Pour un estimateur par projection plutôt qu'un estimateur à noyau, nous obtenons la proposition suivante :

Proposition 3 [Pri-4] *Soit $(\varphi_i)_{1\leq i\leq m}$ un système orthonormal de $\mathbb{L}^2(\mathbb{R},\lambda)$ (λ étant la mesure de Lebesgue). On supose que chaque φ_i appartient à BV. Soit $(X_i)_{i\geq 0}$ une suite stationnaire de variables aléatoires réelles. On définit*

$$Y_{j,n} = \frac{1}{n}\sum_{k=1}^{n} \varphi_j(X_k) \quad et \quad \hat{f}_n = \sum_{j=1}^{m} Y_{j,n}\varphi_j \; .$$

On a alors

$$n\int \mathrm{Var}(\hat{f}_n(x))dx \leq \sup_{x\in\mathbb{R}}\left(\sum_{j=1}^{m}\varphi_j^2(x)\right) + 2\left(\sum_{k=1}^{n-1}\tilde{\beta}_1(k)\right)\sup_{x\in\mathbb{R}}\left(\sum_{j=1}^{m}\|d\varphi_j\|\,|\varphi_j(x)|\right) \; .$$

Remarque 6

- *Pour des suites β-mélangeantes, Viennet* [58] *obtient des résultats optimaux pour le MISE sous la condition*
$$\sum_{k>0}\beta(\sigma(X_0),\sigma(X_k)) < \infty \; . \tag{2.1.4}$$
En utilisant les propositions 2 et 3, nous remplaçons dans **[Pri-4]** *l'hypothèse (2.1.4) par*
$$\sum_{k>0}\tilde{\beta}(\sigma(X_0),X_k) < \infty \; . \tag{2.1.5}$$
Pour l'estimateur à noyau, il suffit de supposer que le noyau K est dans BV et est Lebesgue intégrable. Pour l'estimateur par projection, il faut que la base (φ_i) soit bien localisée (voir **[Pri-4]** *pour plus de détails), car notre inégalité de variance est moins précise que celle de Viennet.*
- *L'hypothèse (2.1.5) est bien moins restrictive que l'hypothèse (2.1.4). Nous obtenons en particuler des exemples non mélangeants* **[Pri-4]**.

Estimation de ruptures

Dans **[Pri-7, Pri-15]**, nous avons étudié le problème de l'estimation d'une rupture dans un modèle de régression non paramétrique sous des conditions de dépendance. Il existe des travaux concernant la détection de ruptures dans un modèle de régression avec dépendance sur le bruit (voir par exemple [3]). Ici, nous supposons que la dépendance est sur les observations (X_i). Nous considérons le modèle de régression suivant :

$$Y_i = m(X_i) + \sigma(X_i)\varepsilon_i\,,\ i = 1, \ldots, n\,,$$

où $(X_i)_{i\in\mathbb{Z}}$ est stationnaire et faiblement dépendante. La suite $(\varepsilon_i)_{i\in\mathbb{Z}}$ est i.i.d. et indépendante de la suite $(X_i)_{i\in\mathbb{Z}}$. La fonction de régression $m(\cdot)$ est supposée régulière sauf en un point où la fonction elle-même présente un saut. L'objectif des travaux écrits dans **[Pri-7, Pri-15]** est d'estimer l'instant de rupture ainsi que l'amplitude de la rupture. Pour cela, nous avons suivi la démarche de Grégoire & Hamrouni [29] du cas indépendant, en l'adaptant à des observations X_i dépendantes. Pour estimer l'amplitude du saut, nous avons utilisé une méthode non paramétrique qui repose sur des estimateurs de régression linéaire locale.

Rentrons un peu plus dans les détails. Soient K un noyau et h_n une fenêtre (suite de réels positifs qui tend vers zéro). L'estimateur de régression linéaire locale est obtenu en résolvant le problème de minimisation suivant :

$$\min_{a_x,b_x} \sum_{i=1}^{n} (Y_i - a_x - b_x(x - X_i))^2 K\left(\frac{x - X_i}{h_n}\right)\,.$$

Sous certaines hypothèses, ce problème de minimisation (qui correspond à un problème des moindres carrés pondérés dans le cas indépendant) admet une unique solution $(\hat{a}_x, \hat{b}_x)$. L'estimateur de régression linéaire locale est alors donné par la définition 6 ci-dessous (voir aussi [57]).

Définition 6 *L'estimateur de régression linéaire locale de $m(x)$ est défini par*

$$\hat{m}(x) = \hat{a}_x = \frac{\sum_{i=1}^{n} \omega_i(x) Y_i}{\sum_{i=1}^{n} \omega_i(x)}\,,$$

où, pour $i = 1, \ldots, n$,

$$\omega_i(x) = K\left(\frac{x - X_i}{h_n}\right)(S_2(x) - (x - X_i)S_1(x))$$

avec, pour $l = 0, 1, 2$,

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

Dans **[Pri-7]**, nous avons supposé que la suite $(X_i)_{i\in\mathbb{Z}}$ était θ-dépendante et que la fonction de régression m ne sautait qu'en un seul instant ι connu.

Soient $m^+(\iota) := \lim_{t\to\iota, t>\iota} m(t)$ et $m^-(\iota) := \lim_{t\to\iota, t<\iota} m(t)$. Nous considérons deux noyaux continuement différentiables $K^+(\cdot)$ et $K^-(\cdot)$ tels que :

- $K^+ : [-1, 0] \to \mathbb{R}_+$,
- $K^- : [0, 1] \to \mathbb{R}_+$ avec $K^-(x) = K^+(-x)$.

Soient, pour $l \in \mathbb{N}$, $K_l^+ = \int_{-1}^0 x^l K^+(x)dx$ et $K_l^- = \int_0^1 x^l K^-(x)dx$. Normalisons les noyaux K^+ et K^- par $K_2 K_0 - K_1^2 = 1$.

Nous avons alors estimé le saut éventuel au point ι, $\gamma(\iota) = m^+(\iota) - m^-(\iota)$, en utilisant un estimateur à droite pour $m^+(\iota)$ et un autre à gauche pour $m^-(\iota)$. Ces estimateurs $\hat{m}^+(\iota)$ et $\hat{m}^-(\iota)$ sont construits respectivement à partir de K^+ et de K^- comme dans la définition 6.

Par ailleurs, nous supposons dans toute la suite de ce paragraphe que :
- $\mathbb{E}\varepsilon_0 = 0$, $\mathbb{E}(\varepsilon_0^2) = 1$, $\mathbb{E}|\varepsilon_0|^3 < \infty$.
- X_0 prend ses valeurs dans $[0,1]$. X_0 admet une densité f strictement positive et continue.
- La fonction de régression $m(\cdot) = \mathbb{E}(Y_0|X_0 = \cdot)$ est deux fois continuement différentiable dans chacun des intervalles $[0,\iota]$ et $[\iota, 1]$.
- La variance conditionnelle $\sigma^2(\cdot) = \mathrm{Var}(Y_0|X_0 = .)$ est continuement différentiable.
- L'instant de rupture ι est dans $]0,1[$.

Soit $V^+ = \int_{-1}^0 (K_2^+ - xK_1^+)^2 (K^+(x))^2 dx$. Nous obtenons alors, en développant une variante de la méthode de Lindeberg-Rio, le théorème limite centrale suivant pour l'estimateur renormalisé $\sqrt{nh_n}(\hat{\gamma}(\iota) - \gamma(\iota))$ où $\hat{\gamma}(\iota) = \hat{m}^+(\iota) - \hat{m}^-(\iota)$.

Théorème 2 [Pri-7] *On suppose que la suite $(X_i)_{i\in\mathbb{Z}}$ est θ-dépendante avec $\theta(k) = k^{-a}$ pour un $a \geq 3$. Alors, si $nh_n \to \infty$ et $nh_n^5 \to 0$,*

$$\sqrt{nh}\left(\widehat{m}^-(\iota) - m^-(\iota)\, ,\, \widehat{m}^+(\iota) - m^+(\iota)\right)'$$

converge en loi vers une Gaussienne de dimension 2 :

$$\mathcal{N}\left(0, (V_+ \sigma^2(\iota)/f(\iota))\, I_2\right)$$

où I_2 est la matrice identité.

Cependant, le cas où la rupture est localisée ne correspond qu'à certains cas pratiques, par exemple où un expérimentateur produit lui-même une discontinuité dans les paramètres (intervention chirurgicale, démarrage d'un nouveau traitement, etc). Dans **[Pri-15]**, nous avons étudié le cas plus général où l'instant de rupture est inconnu. Nous nous sommes placés sous des conditions de $\tilde{\phi}_1$-dépendance, ce qui nous a permis de contrôler les termes de covariance en utilisant l'inégalité de covariance de la proposition 1. Pour estimer l'instant ι, nous avons utilisé l'estimateur suivant

$$\hat{\iota} = \inf\{t \in \kappa\, ;\ \hat{\gamma}(t) = \sup_{x\in\kappa} \hat{\gamma}(x)\}\ ,$$

où κ est un compact inclus dans $]0,1[$. Afin d'étudier le comportement asymptotique de $\hat{\iota}$ et de $\hat{\gamma}(\hat{\iota})$, nous avons introduit le processus de déviation locale suivant :

$$\mathcal{Z}_n(z) = a(n, h_n)\left(\hat{\gamma}(\iota + \frac{h_n}{b(n,h_n)} z) - \hat{\gamma}(\iota)\right), \quad z \in [-M, M]\ .$$

Avant d'énoncer le théorème principal, nous allons introduire quelques hypothèses et notations.

Nous faisons les hypothèses suivantes :

- $\mathbf{H_1}$ $\lim_{n\to\infty} h_n = 0$ et $\lim_{n\to\infty} nh_n = \infty$.
- $\mathbf{H_2}$ $\lim_{n\to\infty} \frac{h_n}{b(n,h_n)} = 0$, $\lim_{n\to\infty} a(n,h_n) = \infty$ et $\lim_{n\to\infty} b(n,h_n) = \infty$.
- $\mathbf{H_3}$
  - $K_+(0) > 0$.
  - $K_+(-1) = 0$.
  - $0 < \lim_{n\to\infty} \frac{a(n,h_n)}{b(n,h_n)} = L_4 < \infty$.
  - $\lim_{n\to\infty} \frac{a(n,h_n)}{nh_n} = L_5 < \infty$.

Lorsque $\mathbf{H_3}$ est vérifiée, nous notons

$$\lim_{n\to\infty} \frac{a^2(n,h_n)}{nh_n b(n,h_n)} = L_4 \times L_5 = L_6 < \infty\,.$$

Soit $M^{\pm}(x) := (K_2^{\pm} - xK_1^{\pm})K^{\pm}(x)$. Nous définissons
- $\lambda_1 = L_5\, M^+(0)$,

- si $L_5 > 0$, $\lambda_2 = \frac{L_6}{L_5^2} f(\iota)$,

- $\lambda_3 = L_4 M^+(0) f(\iota) = \lambda_1\lambda_2$.

Soit $y := \frac{z}{b(n,h_n)}$. Nous définissons, pour tout z,
- $\varphi^{\pm}_{n,z}(t) := \frac{a(n,hn)}{nh_n}\left(M^{\pm}\left(\frac{\iota-t}{h_n}+y\right) - M^{\pm}\left(\frac{\iota-t}{h_n}\right)\right)(m(t) - m^{\pm}(\iota))$,

- $\varphi_{n,z}(t) := \varphi^+_{n,z}(t) - \varphi^-_{n,z}(t)$,

- $\tilde{\varphi}_{n,z}(t) := \varphi_{n,z}(t) - \mathbb{E}\left(\varphi_{n,z}(X_1)\right)$,

- $\psi^{\pm}_{n,z}(t) := \frac{a(n,hn)}{nh_n}\left(M^{\pm}\left(\frac{\iota-t}{h_n}+y\right) - M^{\pm}\left(\frac{\iota-t}{h_n}\right)\right)\sigma(t)$,

- $\psi_{n,z}(t) := \psi^+_{n,z}(t) - \psi^-_{n,z}(t)$, et

- pour tout i dans $\mathbb{N}$, $\tilde{U}_{i,n,z} := |\tilde{\varphi}_{n,z}(X_i)| + |\psi_{n,z}(X_i)|$.

Enonçons maintenant le résultat principal :

Théorème 3 [Pri-15] *Supposons* $\mathbf{H_1}$, $\mathbf{H_2}$ *et* $\mathbf{H_3}$ *vérifiées. Supposons que* $(X_i)_{i\in\mathbb{Z}}$ *est* $\tilde{\phi}_1$*-dependante avec* $\sum_{i=1}^{+\infty}\tilde{\phi}_1(i) < +\infty$*. Alors, si pour tout couple* (z_1,z_2) *et pour tout* $k\in\mathbb{N}^*$,

$$n\,\mathrm{Cov}\left(\tilde{U}_{0,n,z_1}, \tilde{U}_{k,n,z_2}\right) \xrightarrow[n\to+\infty]{} 0\,, \tag{2.1.6}$$

on a

$$\mathcal{Z}_n \Rightarrow \mathcal{Z}\,, \qquad \textit{dans} \quad \mathcal{D}\left([-M,M]\right) \tag{2.1.7}$$

où

$$\mathcal{Z}(z) = \frac{\lambda_3}{f(\iota)}\gamma(\iota)|z| + \frac{\lambda_1}{f(\iota)}\mathcal{N}(z)\,, \tag{2.1.8}$$

avec $\mathcal{N}(z)$ *défini par*

$$\mathcal{N}(z) = \begin{cases} \sum_{i=1}^{N_z^+}(-\gamma(\iota) - 2\sigma(\iota)\varepsilon_i^+) \ \textit{si}\ z \geq 0\,, \\ \sum_{i=1}^{N_{-z}^-}(-\gamma(\iota) + 2\sigma(\iota)\varepsilon_i^-) \ \textit{si}\ z < 0\,. \end{cases} \tag{2.1.9}$$

Les suites (ε_i^+) *et* (ε_i^-) *sont indépendantes et construites avec une suite de variables aléatoires i.i.d., de même loi que la variable d'erreur du modèle* ε*. De plus* N_z^+ *et* N_z^- *sont des processus de Poisson homogènes indépendants de paramètre* λ_2*, indépendants des suites* (ε_i^+) *et* (ε_i^-)*.*

Le processus de déviation locale a donc la même limite que dans le cas indépendant. La preuve repose principalement sur l'utilisation de la proposition 1, ainsi que sur des résultats de convergence de sommes de variables aléatoires dépendantes vers des lois infiniment divisibles de variance finie [17]. La condition (2.1.6) n'est pas très restrictive. Dans **[Pri-15]**, nous vérifions cette condition pour quelques exemples classiques. Du théorème 3, nous déduisons également un théorème limite centrale pour $\frac{b(n,h_n)}{h_n}(\hat{\iota}-\iota)$ ainsi que pour $\sqrt{nh_n}(\hat{\gamma}(\hat{\iota})-\gamma(\iota))$. En choisissant $b(n,h_n) = nh_n$, nous obtenons pour la convergence de $\hat{\iota}$ la vitesse n^{-1} dès que $nh_n \to \infty$.

Remarque 7 *L'hypothèse de* $\tilde{\phi}_1$*-dépendance est souvent plus forte que l'hypothèse de* θ*-dépendance. En particulier, si* X *est une variable aléatoire réelle bornée par* M*, on sait par le lemme 1 ([Proposition 2,* **Pri-4]***) que* $\theta(\mathcal{M},X) \leq 2\,M\,\tilde{\phi}(\mathcal{M},X)$*.*

2.1.2 Inégalités exponentielles, loi du logarithme itéré

Cette partie est consacrée aux inégalités exponentielles que nous avons montrées pour des suites faiblement dépendantes.

Dans **[Pri-6]**, nous donnons une inégalité de type Bennett, ainsi qu'une inégalité de type Fuk-Nagaev pour les sommes partielles de variables aléatoires τ-dépendantes.

Pour montrer ces inégalités exponentielles, commençons par énoncer le lemme de couplage suivant :

Lemme 2 [55], **[Pri-6]** *Soient* $(\Omega,\mathcal{A},\mathbb{P})$ *un espace probabilisé,* $\mathcal{M}$ *une sous-tribu de* $\mathcal{A}$ *et* X *une variable aléatoire réelle de* $\mathbb{L}^1(\mathbb{P})$*. On suppose qu'il existe une variable aléatoire* δ *uniformément distribuée sur* $[0,1]$*, indépendante de la tribu engendrée par* X *et* $\mathcal{M}$*. Alors il existe une variable aléatoire* X^**, mesurable par rapport à* $\mathcal{M}\vee\sigma(X)\vee\sigma(\delta)$*, indépendante de* $\mathcal{M}$ *et de même loi que* X*, telle que*

$$\|X - X^*\|_1 = \tau(\mathcal{M},X)\,.$$

Remarque 8

- *On déduit du lemme de couplage de Berbee* [7] *et du lemme 2 ci-dessus que le coefficient* β *de Rozanov et Volkonskii* [54] *et le coefficient* τ *présentent tous les deux une propriété d'optimalité : ils réalisent l'infimum de* $\mathbb{E}\, d_0(X, Y)$ *où* Y *est indépendante de* $\mathcal{M}$ *et de même loi que* X *pour les distances respectives* $d_0(x, y) = \mathbb{I}_{x \neq y}$ *et* $d_0(x, y) = |x - y|$*.*
- *La démonstration du lemme 2 repose sur la transformation par quantile de Major* [43]*.*
- *Le lemme 2 peut-être généralisé au cas où* X *prend ses valeurs dans un espace polonais (ou même encore plus général). Nous renvoyons à* [55, Proposition 6] *pour le cas où* X *est à valeurs dans un espace polonais* $(\mathcal{X}, d_0)$ *et où* Ω *est également polonais. Lorsque* Ω *est quelconque et pour* X *prenant ses valeurs dans un espace plus général, nous renvoyons à* **[Pri-5, Pri-6, Pri-13]**.

Le couplage est un outil important pour l'étude des propriétés statistiques d'une suite de variables aléatoires dépendantes. Par exemple, le lemme de couplage 2 permet de remplacer une suite de variables aléatoires τ-dépendante par une suite i.i.d., indépendante du passé de la suite initiale. Le prix à payer peut alors être contrôlé en terme des coefficients $\tau(k)$ de la suite initiale (voir Définition 7 ci-dessous).

Définition 7 *Soit* $(X_i)_{i\in\mathbb{Z}}$ *une suite de variables aléatoires réelles dans* $\mathbb{L}^1(\mathbb{P})$*. On définit*

$$\tau_r(k) = \max_{l \leq r} \frac{1}{l} \sup_{(i,j)\in\Gamma(1,l,k)} \tau(\mathcal{M}_i, (X_{j_1}, \ldots, X_{j_l}))\,,$$

avec $\mathcal{M}_i = \sigma(X_j,\, j \leq i)$*. On définit aussi* $\tau(k) = \tau_\infty(k)$*.*

La propriété de couplage du lemme 2 nous a permis de retrouver les inégalités de type Bennett (Proposition 4) et de type Fuk-Nagaev (Proposition 5) que Rio [50] avait montrées en mélange fort.

Proposition 4 [Pri-6] *Soit* $(X_i)_{i\geq 1}$ *une suite de variables aléatoires réelles, bornée par* M*. Soit* $\mathcal{M}_i = \sigma(X_k,\, 1 \leq k \leq i)$*. On définit* $S_k = \sum_{i=1}^{k}(X_i - \mathbb{E}(X_i))$ *et* $\overline{S}_n = \max_{1\leq k\leq n} |S_k|$*. Pour* q *dans* $\mathbb{N}^*$*, soit* v_q *dans* $\mathbb{R}_+$ *tel que*

$$v_q \geq \|X_{q[n/q]+1} + \cdots + X_n\|_2^2 + \sum_{i=1}^{[n/q]} \|X_{(i-1)q+1} + \cdots + X_{iq}\|_2^2\,.$$

Soit h *la fonction définie par* $h(x) = (1 + x)\ln(1 + x) - x$*. Alors,*

1. *pour tout* $\lambda > 0$*,*

$$\mathbb{P}(|S_n| \geq 3\lambda) \leq 4 \exp\Big(-\frac{v_q}{(qM)^2} h\Big(\frac{\lambda q M}{v_q}\Big)\Big) + \frac{n}{\lambda}\,\tau_q(q + 1)\,,$$

2. pour tout $\lambda \geq Mq$,

$$\mathbb{P}(\overline{S}_n \geq (\mathbb{1}_{q>1}+3)\lambda) \leq 4\exp\Big(-\frac{v_q}{(qM)^2}h\Big(\frac{\lambda qM}{v_q}\Big)\Big) + \frac{n}{\lambda}\tau_q(q+1)\,.$$

Avant d'énoncer l'inégalité de type Fuk-Nagaev, nous avons besoin de quelques notations.

Notation 1 *Soit* X *une variable aléatoire réelle. Soit* $H_{|X|}(x) = \mathbb{P}(|X| > x)$*. On note alors*

- $Q_{|X|}$ *l'inverse généralisée de* $H_{|X|}$.
- $G_{|X|}$ *l'inverse de* $x \to \int_0^x Q_{|X|}(u)du$.

Pour toute suite décroissante $(\delta_i)_{i\geq 0}$ *de réels positifs , on définit* $\delta^{-1}(u) = \sum_{i\geq 0}\mathbb{1}_{u<\delta_i} = \inf\{k \in \mathbb{N} : \delta_k \leq u\}$.

Proposition 5 [Pri-6] *Soit* $(X_i)_{i\geq 1}$ *une suite de variables aléatoires centrées, de carré intégrable. On définit* $(\mathcal{M}_i)_{i\geq 1}$ *et* $\overline{S}_n$ *comme à la proposition 4. Soient* X *une variable aléatoire à valeurs dans* $\mathbb{R}_+^*$ *telle que* $Q_X \geq \sup_{k\geq 1} Q_{|X_k|}$ *et*

$$s_n^2 = \sum_{i=1}^n\sum_{j=1}^n |\mathrm{Cov}(X_i, X_j)|\,.$$

Soient $R = ((\tau/2)^{-1} \circ G_X^{-1})Q_X$ *et* $S = R^{-1}$*. Alors, pour tout* $\lambda > 0$ *et tout* $r \geq 1$,

$$\mathbb{P}(\overline{S}_n \geq 5\lambda) \leq 4\Big(1+\frac{\lambda^2}{rs_n^2}\Big)^{-r/2} + \frac{4n}{\lambda}\int_0^{S(\lambda/r)} Q_X(u)du\,. \tag{2.1.10}$$

De cette inégalité, nous déduisons une loi du logarithme itéré pour une suite τ-dépendante.

Théorème 4 [Pri-6] *Soit* $(X_i)_{i\in\mathbb{Z}}$ *une suite stationnaire de variables aléatoires réelles, centrées, de carré intégrable, satisfaisant*

$$\sum_{k=1}^{\infty}\int_0^{\tau(k)/2} Q\circ G(u)du < \infty\,.$$

Soit $S_n = X_1 + \cdots + X_n$*. Alors* $n^{-1}\mathrm{Var}(S_n)$ *converge vers* σ^2*, et il existe une suite* $(Y_i)_{i\in\mathbb{N}}$ *de variables aléatoires indépendantes, de loi (éventuellement dégénérée)* $\mathcal{N}(0,\sigma^2)$*, telle que*

$$\sum_{i=1}^n (X_i - Y_i) = o(\sqrt{n\ln\ln n}) \text{ } p.s.$$

Enfin, pour clore ce paragraphe sur les inégalités exponentielles, nous allons présenter une inégalité de type Hoeffding pour les sommes partielles d'une suite $\tilde{\phi}$-dépendantes.

Proposition 6 [Pri-4] *Soient $(X_i)_{i\geq 0}$ une suite de variables aléatoires réelles et $\mathcal{M}_i$ la tribu du passé avant i. Pour toute fonction h de BV, on définit*

$$S_n(h)=\sum_{i=1}^{n} h(X_i) \quad et \quad b_{i,n}=\left(\sum_{k=0}^{n-i}\tilde{\phi}_1(k)\right)\|dh\|\,\|h(X_i)-\mathbb{E}(h(X_i))\|_{p/2}\,.$$

Alors, pour tout $p\geq 2$,

$$\|S_n(h)-\mathbb{E}(S_n(h))\|_p\leq\left(2p\sum_{i=1}^{n} b_{i,n}\right)^{1/2}\leq\|dh\|\left(2p\sum_{k=0}^{n-1}(n-k)\tilde{\phi}_1(k)\right)^{1/2}$$

et

$$\mathbb{P}(|S_n(h)-\mathbb{E}(S_n(h))|>x)\leq e^{1/e}\exp\left(\frac{-x^2}{4e\|dh\|^2\sum_{k=0}^{n-1}(n-k)\tilde{\phi}_1(k)}\right)\,.$$

Remarque 9

- *Pour démontrer la proposition 6, nous sommes partis d'une inégalité de type Burkhölder donnée dans* [16].
- *En appliquant la méthode de différences de martingales, comme dans* [14], *on peut montrer*

$$\mathbb{P}(|S_n(h)-\mathbb{E}(S_n(h))|>x)\leq 2\exp\left(\frac{-x^2}{2\|dh\|^2\sum_{i=1}^{n}\left(1+2\sum_{k=1}^{n-i+1}\tilde{\phi}_1(k)\right)^2}\right)\,.$$

Cette inégalité et celle de la proposition 6 sont de même type dès que $\sum_{k>0}\tilde{\phi}_1(k)$ est finie.

2.1.3 Processus empirique

Les articles **[Pri-1, Pri-8]** portent sur l'étude du processus empirique pour des suites faiblement dépendantes. Soit $(X_i)_{i\in\mathbb{Z}}$ une suite stationnaire de variables aléatoires à valeurs dans $\mathbb{R}^d$, de fonction de répartition marginale F. Si F_n désigne la fonction de répartition empirique, on définit pour tout t dans $\mathbb{R}^d$, $F_n(t)=n^{-1}\sum_{i=1}^{n}\mathbb{1}_{X_i\leq t}$. Les articles **[Pri-1, Pri-8]** donnent des conditions suffisantes, en termes de dépendance, pour la convergence du processus rencentré et renormalisé $\{\sqrt{n}(F_n(t)-F(t)),\, t\in\mathbb{R}^d\}$.

Pour le cas $d=1$, nous avons obtenu le résultat suivant :

Théorème 5 [Pri-8] *Soit $(X_i)_{i\in\mathbb{Z}}$ une suite stationnaire réelle, θ-dépendante. On suppose que la fonction de répartition marginale associée F est Lipschitzienne et que $\theta(r)=\mathcal{O}\left((r+1)^{-2-2\sqrt{2}-\nu}\right)$ pour un $\nu>0$. Alors $\{\sqrt{n}(F_n(t)-F(t)),\, t\in\mathbb{R}\}$ converge en loi dans l'espace de Skorohod $\mathcal{D}(\mathbb{R})$ vers le processus Gaussien de fonction de covariance définie par*

$$\Gamma(s,t)=\sum_{k\in\mathbb{Z}}\mathrm{Cov}(\mathbb{1}_{X_0\leq s},\mathbb{1}_{X_{|k|}\leq t})\,.$$

Dans **[Pri-1]**, nous nous sommes placés sous des conditions de $\tilde{\beta}_2$ et $\tilde{\phi}_2$-dépendance en dimension d quelconque. Introduisons les deux conditions suivantes :

$(\mathcal{C}_1)$ Il existe $\varepsilon > 0$ tel que $\tilde{\phi}_2(k) = \mathcal{O}\left(k^{-1-\varepsilon}\right)$.
$(\mathcal{C}_2)$ Il existe $\varepsilon > 0$ tel que $\tilde{\beta}_2(k) = \mathcal{O}\left(k^{-2d-\varepsilon}\right)$.
Nous obtenons alors le théorème suivant :

Théorème 6 [Pri-1] *On suppose que $(\mathcal{C}_1)$ ou $(\mathcal{C}_2)$ est vérifiée. Alors $\{\sqrt{n}(F_n(t) - F(t)),\, t \in \mathbb{R}^d\}$ converge faiblement dans $l^\infty(\mathbb{R}^d)$ vers un processus Gaussien tendu de fonction de covariance définie par*

$$\Gamma(s,t) = \sum_{k \in \mathbb{Z}} \mathrm{Cov}(\mathbb{I}_{X_0 \le s}, \mathbb{I}_{X_{|k|} \le t}) .$$

Pour obtenir un résultat de convergence en loi pour le processus empirique, il faut démontrer deux points essentiels : la tension et la convergence des répartitions finidimensionnelles. La tension est souvent plus difficile à obtenir. L'idée est d'utiliser une inégalité de moment de type Rosenthal adéquate, puis d'appliquer un critère de tension. Dans **[Pri-8]**, nous commençons par montrer une nouvelle inégalité de moment pour les sommes partielles d'une suite θ-dépendante (Proposition 7 ci-dessous), puis nous appliquons le critère de Shao & Yu [56] pour conclure.

Proposition 7 [Pri-8] *Soit r un réel > 2. Soit $(X_n)_{n\in\mathbb{Z}}$ une suite de variables aléatoires θ-dépendante, bornée par 1. Pour $n \ge 1$, on définit $S_n = X_1 + X_2 + \ldots + X_n$ et $S_0 = X_0 = 0$. Alors il existe une constante C_r telle que*

$$\mathbb{E}|S_n|^r \le C_r\left(s_n^r + M_{r,n}\right) ,$$

où $M_{r,n} = n\sum_{i=0}^{n-1}(i+1)^{r-2}\theta(i)$, et $s_n^2 = M_{2,n} = n\sum_{i=0}^{n-1}\theta(i)$.

Dans **[Pri-1]**, nous avons adapté la méthode d'Andrews & Pollard [2] pour démontrer un nouveau critère de tension (voir Proposition 8 ci-dessous). Nous avons utilisé ensuite les inégalités de moment données dans [15] et ce critère pour conclure.

Avant d'énoncer le critère de tension obtenu dans **[Pri-1]**, commençons par rappeler la notion de nombre de crochets.

Définition 8 *Soit Q une mesure finie sur un espace mesurable $\mathcal{X}$. Pour toute fonction mesurable f de $\mathcal{X}$ dans $\mathbb{R}$, soit $\|f\|_{Q,r} = Q(|f|^r)^{1/r}$. Si $\|f\|_{Q,r}$ est finie, on dit que f est dans L^r_Q. Soit $\mathcal{F}$ un sous-ensemble de L^r_Q. Le nombre de crochets $\mathcal{N}_{Q,r}(\varepsilon, \mathcal{F})$ est le plus petit entier N pour lequel il existe des fonctions $f_1^- \le f_1, \ldots, f_N^- \le f_N$ de $\mathcal{F}$ telles que pour tout entier $1 \le i \le N$, on ait $\|f_i - f_i^-\|_{Q,r} \le \varepsilon$, et pour toute fonction f de $\mathcal{F}$ il existe un entier $1 \le i \le N$ tel que $f_i^- \le f \le f_i$.*

Nous pouvons maintenant énoncer le critère de tension suivant :

Proposition 8 [Pri-1] *Soit $(X_i)_{i\geq 0}$ une suite stationnaire de variables aléatoires à valeurs dans un espace mesurable $\mathcal{X}$, de marginale P. Soit P_n la mesure empirique $P_n = n^{-1}\sum_{i=1}^n \delta_{X_i}$, et soit Z_n le processus empirique défini par $Z_n = \sqrt{n}(P_n - P)$. Soit Q une mesure finie sur $\mathcal{X}$ telle que $Q-P$ soit une mesure positive. Soient $\mathcal{F}$ une classe de fonctions de $\mathcal{X}$ dans $\mathbb{R}$ et $\mathcal{G} = \{f-l, (f,l) \in \mathcal{F}\times\mathcal{F}\}$. On suppose qu'il existe $r \geq 2$, $p \geq 1$ et $q > 2$ tels que pour toute fonction g de $\mathcal{G}$, on ait*

$$\|Z_n(g)\|_p \leq C(\|g\|_{Q,1}^{1/r} + n^{1/q-1/2}),$$

où C ne dépend ni de g ni de n. Si de plus

$$\int_0^1 x^{(1-r)/r}(\mathcal{N}_{Q,1}(x,\mathcal{F}))^{1/p}dx < \infty \quad et \quad \lim_{x\to 0} x^{p(q-2)/q}\mathcal{N}_{Q,1}(x,\mathcal{F}) = 0,$$

alors

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\mathbb{E}\Big(\sup_{g\in\mathcal{G},\|g\|_{Q,1}\leq\delta}|Z_n(g)|^p\Big) = 0.$$

Remarque 10

- *Le théorème 5 améliore la condition $\theta(r) = \mathcal{O}(r^{-(5+\nu)})$ donnée dans [24]. Pour montrer la tension, nous calculons un moment d'ordre $2+\sqrt{2}$ alors que dans [24], les auteurs vont jusqu'à l'ordre 4. L'inégalité de moment qu'ils utilisent à la place de notre proposition 7 ne permet en effet que les moments d'ordre entier.*
- *Dans* **[Pri-1]***, nous montrons que la condition $\mathcal{C}_3$ ci-dessous implique la condition $\mathcal{C}_2$.*
  ($\mathcal{C}_3$) Chaque composante de X_1 a une densité bornée et il existe $\varepsilon > 0$ tel que $\tau_\infty(k) = \mathcal{O}\left(k^{-4d-\varepsilon}\right)$.
  Bien que le coefficient θ soit plus faible que le coefficient τ_∞, nous obtenons pour tous les exemples étudiés dans **[Pri-8]** *les mêmes bornes pour τ_∞* **[Pri-4, Pri-1]** *que pour θ. Pour $d=1$, le théorème 6 donne donc une condition plus faible, au moins pour les modèles présentés dans* **[Pri-8]***, que le théorème 5.*
- *Pour démontrer le théorème 6, nous avons appliqué la proposition 8 avec un choix de la mesure Q inspiré par [50] :*

$$Q(dx) = \left(1 + 4\sum_{k=1}^{+\infty} b_k(x)\right) P(dx),$$

  où P est la loi de X_0 et où $b_k(X_0) = b(\sigma(X_0), X_k)$, b étant défini comme à la proposition 1. Pour plus de détails, nous renvoyons à **[Pri-1]***.*

2.2 Systèmes dynamiques

Dans cette section, nous allons nous intéresser de façon plus particulière à certains systèmes dynamiques de dimension 1. Un système dynamique, même simple et déterministe, peut présenter des comportements "imprévisibles". On parle de comportement chaotique. En choisissant deux conditions initiales voisines, il est possible qu'au bout d'un temps fini, on observe deux orbites complètement différentes. C'est cette observation qui a motivé l'étude des propriétés statistiques des systèmes dynamiques. Nous allons voir que l'étude de certains systèmes dynamiques n'est pas si différente de celle de suites faiblement dépendantes. La plupart des résultats présentés dans cette section sont rappelés dans l'article de synthèse **[Pri-12]**. Commençons par introduire une classe de systèmes dynamiques que nous noterons $\mathcal{SD}$ par la suite.

Description de la classe $\mathcal{SD}$

Soient $I = [0, 1]$, T une application de I dans lui-même. On définit $X_i = T^i$. Si μ est invariante par T, la suite $(X_i)_{i\geq 0}$ de variables aléatoires de (I, μ) dans I est strictement stationnaire. On note $\|g\|_{1,\lambda}$ la norme $\mathbb{L}^1$ par rapport à la mesure de Lebesgue λ restreinte à I, et $\|\nu\| = |\nu|(I)$ la variation totale de ν.

Inégalités de covariance. Dans nombre de cas intéressants, on peut montrer que, pour toute fonction h de BV et toute fonction k de $\mathbb{L}^1(I, \mu)$,

$$|\mathrm{Cov}(h(X_0), k(X_n))| \leq a_n \|k(X_n)\|_1 (\|h\|_{1,\lambda} + \|dh\|) \,, \tag{2.2.11}$$

pour une suite décroissante a_n qui tend vers zéro lorsque n tend vers l'infini. Dans **[Pri-4]**, nous remarquons que (2.2.11) implique

$$\begin{aligned} |\mathrm{Cov}(h(X_0), k(X_n))| &= |\mathrm{Cov}(h(X_0) - h(0), k(X_n))| \\ &\leq a_n \|k(X_n)\|_1 (\|h - h(0)\|_{1,\lambda} + \|dh\|) \,. \end{aligned}$$

Comme $\|h - h(0)\|_{1,\lambda} \leq \|dh\|$, nous obtenons

$$|\mathrm{Cov}(h(X_0), k(X_n))| \leq 2a_n \|k(X_n)\|_1 \|dh\| \,, \tag{2.2.12}$$

ce qui implique $\tilde{\phi}(\sigma(X_n), X_0) \leq 2a_n$ **[Pri-4**, Lemma 4**]**.

Chaîne de Markov associée. On définit l'opérateur $\mathcal{L}$ de $\mathbb{L}^1(I, \lambda)$ dans $\mathbb{L}^1(I, \lambda)$ *via* l'égalité

$$\int_0^1 \mathcal{L}(h)(x) k(x) \lambda(dx) = \int_0^1 h(x)(k \circ T)(x) \lambda(dx)$$

où $h \in \mathbb{L}^1(I, \lambda)$ et $k \in \mathbb{L}^\infty(I, \lambda)$. L'opérateur $\mathcal{L}$ est appelé l'opérateur de Perron-Frobenius associé à T. On suppose que μ est absolument continue par rapport à la mesure de Lebesgue, de densité f_μ. Soit I^* le support de μ. On choisit alors une version de f_μ telle que $f_\mu > 0$ sur I^* et $f_\mu = 0$ sur $(I^*)^c$. On peut toujours choisir $\mathcal{L}$ tel que $\mathcal{L}(f_\mu h)(x) = \mathcal{L}(f_\mu h)(x) \mathbb{1}_{f_\mu(x)>0}$. On définit alors le noyau Markovien associé à T par

$$K(h)(x) = \frac{\mathcal{L}(f_\mu h)(x)}{f_\mu(x)} \mathbb{1}_{f_\mu(x)>0} + \mu(h) \mathbb{1}_{f_\mu(x)=0}. \tag{2.2.13}$$

Il est ensuite facile de vérifier que $(X_0, X_1, \ldots, X_n)$ a même loi que $(Y_n, Y_{n-1}, \ldots, Y_0)$ où $(Y_i)_{i\geq 0}$ est une chaîne de Markov stationnaire, de mesure invariante μ et de noyau de transition K. Cette correspondance "système dynamique-chaîne de Markov", déjà mentionnée dans des travaux de Gordin (1968), est clairement expliquée dans [6]. Nous montrons alors le lemme suivant :

Lemme 3 [Pri-1] *Soit $(Y_i)_{i\geq 0}$ une chaîne de Markov à valeurs réelles, de noyau de transition K. On suppose qu'il existe une constante C telle que*

$$\textit{pour toute fonction } f \textit{ dans } BV \textit{ et pour tout } n > 0, \quad \|dK^n(f)\| \leq C\|df\|\,. \tag{2.2.14}$$

Alors, pour tout $j > i \geq 0$, $\tilde{\phi}(\sigma(Y_k), (Y_{k+i}, Y_{k+j})) \leq (1 + C)\tilde{\phi}(\sigma(Y_k), Y_{k+i})$.

Nous montrons alors (voir **[Pri-1]**) que si on a (2.2.11) et (2.2.14), alors pour tout $n \geq j > i \geq 0$,

$$\tilde{\phi}(\sigma(X_k, k \geq n), (X_{n-i}, X_{n-j})) \leq (1 + C)\tilde{\phi}(\sigma(X_k, k \geq n), X_{n-i}) \leq 2(1 + C)a_i\,.$$

Propriété de trou spectral. Dans de nombreux cas intéressants, on fait l'analyse spectrale de $\mathcal{L}$ dans l'espace de Banach BV muni de la norme $\|h\|_v = \|dh\| + \|h\|_{1,\lambda}$ en utilisant le théorème de Ionescu-Tulcea et Marinescu (voir Lasota & Yorke [34]). On suppose que 1 est valeur propre simple de $\mathcal{L}$ et que les autres valeurs propres du spectre sont contenues dans un disque fermé de rayon strictement plus petit que 1. On dit qu'on a une propriété de trou spectral. Il existe alors une unique probabilité T-invariante μ, absolument continue par rapport à la mesure de Lebesgue. Sa densité f_μ est dans BV, et

$$\mathcal{L}^n(h) = \lambda(h)f_\mu + \Psi^n(h) \quad \text{avec } \Psi(f_\mu) = 0 \text{ et } \quad \|\Psi^n(h)\|_v \leq D\rho^n\|h\|_v \tag{2.2.15}$$

pour un $0 \leq \rho < 1$ et un $D > 0$. Si de plus

$$\left\| \frac{1}{f_\mu} \mathbb{1}_{f_\mu > 0} \right\|_v = \gamma < \infty\,, \tag{2.2.16}$$

nous montrons dans **[Pri-1]** que les propriétés (2.2.11) et (2.2.14) sont satisfaites. Nous en déduisons donc que les coefficients $\tilde{\phi}_2(i)$ de la chaîne $(Y_i)_{i\geq 0}$ satisfont

$$\exists\, C > 0 \text{ telle que } \; \tilde{\phi}_2(i) \leq C\,\rho^i\,.$$

Application : "transformations dilatantes".

1. **β-transformations** $T(x) = \beta x - [\beta x]$ avec $\beta > 1$.
2. **Applications linéaires par morceaux** I est la réunion finie d'intervalles disjoints $(I_k)_{1\leq k\leq n}$, et $T(x) = a_k x + b_k$ sur I_k, avec $|a_k| > 1$.

3. $T(x) = a(x^{-1} - 1) - [a(x^{-1} - 1)]$ avec $a > 0$. Pour $a = 1$, on parle de **transformation de Gauss**.

Plus généralement, les transformations décrites dans [11, Section 2.1] vérifient (2.2.15).

Pour obtenir les propriétés statistiques des systèmes dynamiques de la classe $\mathcal{SD}$, on peut donc travailler sur la chaîne associée. Cette chaîne est une suite de variables aléatoires faiblement dépendantes qu'on peut étudier en utilisant la section 2.1 précédente. Dans la suite, nous donnons quelques exemples.

Estimation de la densité invariante. Dans **[Pri-9]**, un théorème limite centrale est donné pour l'estimateur à noyau de la densité invariante. Dans cet article, nous ne travaillons pas avec la chaîne de Markov associée. La méthode utilisée est une adaptation de la méthode de Lindeberg-Rio [48]. Le contrôle des covariances n'est pas formalisé en terme de coefficient de dépendance. C'est seulement dans **[Pri-4]** que le coefficient $\tilde{\phi}$ est introduit, et que le lien avec les systèmes dynamiques est formalisé. Les résultats sur le MISE (voir section 2.1) pour des suites $\tilde{\beta}$-dépendantes s'appliquent alors aux systèmes de la classe $\mathcal{SD}$ en rappelant que $\tilde{\beta}(\mathcal{M}, X) \leq \tilde{\phi}(\mathcal{M}, X)$ (voir Lemme 1, section 2.1).

Estimation de ruptures. Grâce au contrôle de $\tilde{\phi}_1$, nous pouvons également appliquer les résultats sur l'estimation de ruptures dans un modèle de régression (Théorème 3, **[Pri-15]**).

Périodogramme intégré.

Dans **[Pri-2]**, nous nous sommes intéressés à l'estimation de la densité spectrale d'un système chaotique défini pour tout $t \in \mathbb{N}$ par

$$\xi_t = \varphi(T^t(X_0)) = \varphi(X_t) \ , \tag{2.2.17}$$

où T est dans la classe $\mathcal{SD}$ et $\varphi : [0,1] \to \mathbb{R}$ est dans l'espace $BV([0,1])$. Le processus $(\xi_t)_{t\geq 0}$ est supposé centré. Avant d'énoncer le résultat obtenu, commençons par donner quelques notations et définitions.

Soit μ invariante par T. On définit le covariogramme de $(\xi_t)_{t\in\mathbb{N}}$ par $\gamma(t) = \mathbb{E}_\mu(\xi_0 \xi_t)$ pour tout t dans $\mathbb{N}$.

La densité spectrale de $(\xi_t)_{t\in\mathbb{N}}$ est donnée, pour tout λ sur le tore $\mathbb{T} = [-\pi, \pi[$, par

$$f(\lambda) = \frac{1}{2\pi} \sum_{t\in\mathbb{Z}} \gamma(|t|) e^{-it\lambda} \ .$$

Les cumulants d'ordre 4 de $(\xi_t)_{t\in\mathbb{N}}$ sont définis, pour tout $(r, s, t) \in \mathbb{Z}^3$, par

$$\kappa(r,s,t) = \mathbb{E}_\mu[\xi_0\xi_r\xi_s\xi_t] - \mathbb{E}_\mu[\xi_0\xi_r]\mathbb{E}_\mu[\xi_s\xi_t] - \mathbb{E}_\mu[\xi_0\xi_s]\mathbb{E}_\mu[\xi_r\xi_t] - \mathbb{E}_\mu[\xi_0\xi_t]\mathbb{E}_\mu[\xi_r\xi_s] \ .$$

Par stationnarité de $(\xi_t)_{t\in\mathbb{N}}$, cette définition s'étend à $\mathbb{Z}^3$.

Le périodogramme associé à $(\xi_t)_{t\in\mathbb{N}}$ est défini, pour tout $\lambda \in \mathbb{T}$, par

$$I_n(\lambda) = \frac{1}{2\pi n}\left|\sum_{t=1}^{n} \xi_t e^{-it\lambda}\right|^2 .$$

Soit $(\gamma_n(t))$ les covariances empiriques données, pour tout $0 \le t \le n-1$, par

$$\gamma_n(t) = \frac{1}{n}\sum_{k=1}^{n-t} \xi_k \xi_{t+k}$$

et $\gamma_n(t) = 0$ si $t \ge n$. On peut montrer que, pour tout $\lambda \in \mathbb{T}$

$$I_n(\lambda) = \frac{1}{2\pi}\sum_{t\in\mathbb{Z}} \gamma_n(|t|)e^{-it\lambda}.$$

A partir de là, on définit le périodogramme intégré par :

$$\mathcal{I}_n(g) = \int_{-\pi}^{\pi} g(\lambda) I_n(\lambda)\, d\lambda$$

où g est dans l'ensemble de fonctions

$$\mathcal{G} = \Big\{ g : \mathbb{T} \to \mathbb{R}\,,\ 2\pi\text{-periodiques et continues avec } \ g \in \mathbb{L}^2(\mathbb{T}) \Big\}.$$

Le théorème 7 **[Pri-2]** ci-dessous donne la convergence presque sûre de $\mathcal{I}_n(g)$ vers

$$\mathcal{I}(g) = \int_{-\pi}^{\pi} g(\lambda) f(\lambda)\, d\lambda$$

ainsi qu'un théorème limite centrale faisant apparaître la densité bispectrale

$$f_4(\lambda,\mu,\nu) = \frac{1}{(2\pi)^3}\sum_{(r,s,t)\in\mathbb{Z}^3} \kappa(r,s,t)e^{-i(r\lambda+s\mu+t\nu)} .$$

Théorème 7 [Pri-2] *Pour tout g dans $\mathcal{G}$,*

$$\lim_{n\to\infty} \mathcal{I}_n(g) = \mathcal{I}(g) \qquad\qquad p.s.$$

De plus, si

$$\sigma^2(t) = \mathbb{E}[(\xi_0\xi_t - \gamma(t))^2] + 2\sum_{s=1}^{\infty} \mathbb{E}[(\xi_s\xi_{t+s} - \gamma(t))(\xi_0\xi_t - \gamma(t))] > 0 \ ,$$

alors on a la convergence des répartitions finidimensionnelles de $\{\sqrt{n}(\mathcal{I}_n(g) - \mathcal{I}(g)),\, g \in \mathcal{G}\}$ vers celles du processus Gaussien centré $\{Z(g),\, g \in \mathcal{G}\}$ de fonction de covariance donnée, pour tout $g_1, g_2 \in \mathcal{G}$, par

$$\Gamma(g_1,g_2) = 4\pi\int_{-\pi}^{\pi} g_1(\lambda)g_2(\lambda)f^2(\lambda)d\lambda + 2\pi\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} g_1(\lambda)g_2(\mu)f_4(\lambda,-\mu,\mu)d\lambda d\mu \ .$$

Remarque 11

- *Dans* [53]*, Rosenblatt montre un théorème limite centrale pour le périodogramme intégré associé à une suite* (X_n) *strictement stationnaire et ergodique. La démonstration qu'il en donne repose sur des idées de Gordin* [26]*. Pour démontrer le théorème 7, nous travaillons sur la chaîne de Markov associée à* $(X_t)_{t\in\mathbb{N}}$*. Nous montrons que les conditions de type projectif données dans* [53] *sont vérifiées.*
- *On peut étendre ce résultat à des suites* $(\xi_t)_{t\in\mathbb{N}}$ *plus générales. Dans* **[Pri-2]***, le théorème 3 donne un critère projectif, qui, s'il est vérifié, implique la convergence du périodogramme intégré. Ce critère s'écrit en terme de coefficients de dépendance.*
- *Toujours en passant par la chaîne de Markov associée, et en appliquant cette fois-ci les résultats de* [59]*, nous donnons dans* **[Pri-2]** *le comportement asymptotique des transformées de Fourier de* $(\xi_t)_{t\in\mathbb{N}}$ *définies pour tout* θ *dans* $\mathbb{R}$ *par* $S_n(\theta) = \sum_{t=1}^n g(\xi_t)e^{it\theta}$*, où* $g : [0,1] \to \mathbb{R}$ *est une fonction de* $BV([0,1])$*.*
- *Dans* [27]*, Gordin et Lifshits étudient le périodogramme dans le cas de chaînes ayant un opérateur de transition normal. L'hypothèse de normalité exclut les chaînes associées aux systèmes dynamiques de la classe* $\mathcal{SD}$*.*

Pour les trois problèmes ci-dessus (estimation de la densité invariante, estimation d'une rupture, estimation de la densité spectrale), nous pouvons construire l'estimateur en prenant $X_0 \sim p(x)\lambda(dx)$, où $p \in BV(I)$ et λ désigne la mesure de Lebesgue sur I. Nous prenons alors $X_i = T^i X_0$. Pour la démonstration, nous commençons par montrer le résultat pour $X_0 \sim \mu$ où μ est la loi invariante, absolument continue par rapport à λ, puis nous montrons la convergence du cas non stationnaire vers les cas stationnaire **[Pri-9, Pri-2]**.

Processus empirique. Le contrôle des coefficients $\tilde{\phi}_2(k)$ permet d'appliquer le théorème 6 pour le processus empirique **[Pri-1]**.

Inégalités exponentielles. Les inégalités des propositions 4 et 5 s'appliquent. Nous retrouvons également dans **[Pri-4]** une inégalité de concentration pour des fonctions Lipschitziennes donnée dans [13]. Nous travaillons pour cela avec le coefficient φ défini de la manière suivante :

Définition 9 *Soient* $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé,* $\mathcal{M}$ *une sous-tribu de* $\mathcal{A}$*, et* X *une variable aléatoire à valeurs dans* $\mathbb{R}^r$*. On suppose que* $d_1(0, X)$ *est bornée. On définit*

$$\varphi(\mathcal{M}, X) = \sup\{\|\mathbb{E}(f(X)|\mathcal{M}) - \mathbb{E}(f(X))\|_\infty \,,\ f \in \Lambda^{(1)}\} \,.$$

On définit alors $\varphi_r(k)$ *et* $\varphi_\infty(k)$ *comme pour le coefficient* θ *(voir section 2.1).*

Remarque 12

Ce coefficient est une version uniforme du coefficient τ*.*

Il a été introduit par Rio [49]*.*

Nous montrons alors dans **[Pri-4]** que pour certains systèmes de la classe $\mathcal{SD}$ (décrits dans **[Pri-4**, Section 4.3**]** ou dans [13]),

$$\exists\, C > 0\,,\; 0 < \rho < 1 \;\text{ tels que }\; \varphi_\infty(k) \le C\,\rho^k\,.$$

Nous noterons $\mathcal{CMS}$ cette sous-classe de $\mathcal{SD}$. Nous appliquons alors une inégalité exponentielle de Rio [51], et nous retrouvons l'inégalité de concentration montrée dans [13] de manière différente.

Proposition 9 [13], **[Pri-4]** *Soit $\Lambda_{L_1,\ldots,L_n}$ l'ensemble des fonctions f de $\mathbb{R}^n$ dans $\mathbb{R}$ telles que*

$$|f(x_1,\ldots,x_n) - f(y_1,\ldots,y_n)| \le L_1|x_1 - y_1| + \cdots + L_n|x_n - y_n|\,.$$

On suppose que T est dans $\mathcal{CMS}$. Alors il existe $C > 0$ telle que pour tout $x > 0$,

$$\mathbb{P}(f(X_1,\ldots,X_n) - \mathbb{E}(f(X_1,\ldots,X_n)) \ge x) \le \exp\left(\frac{-2x^2}{C\,(L_1^2 + \cdots + L_n^2)}\right)\,.$$

Les deux parties qui suivent portent sur deux sujets indépendants, auxquels je me suis intéressée plus récemment.

2.3 Processus des records généralisés

Dans cette section, nous présentons les résultats obtenus dans **[Pri-3]** sur le processus des records généralisés. Nous renvoyons à [23] pour les rappels concernant la théorie des grandes déviations.

Nous considérons un tableau triangulaire $X = (X_k^{(n)})_{n\in\mathbb{N}^*,\,1\le k\le n}$ de variables aléatoires continues et échangeables. Par échangeables, nous entendons que pour tout n dans $\mathbb{N}^*$, la loi du vecteur aléatoire $X^{(n)} = (X_k^{(n)})_{k=1,\ldots,n}$ est invariante par permutation des indices $\{1, 2, \ldots, n\}$. Nous définissons alors, pour $n \in \mathbb{N}^*$ et $k \in \{1, \ldots, n\}$, le rang relatif $R_k^{(n)}$ de $X_k^{(n)}$ comme le nombre de coordonnées du vecteur $(X_j^{(n)})_{j=1,\ldots,k}$ plus grandes que $X_k^{(n)}$:

$$R_k(n) = \sum_{j=1}^{k} \mathbb{1}_{X_j^{(n)} \ge X_k^{(n)}}\,. \tag{2.3.18}$$

En procédant comme dans [47], nous montrons que :

$$\mathbb{P}\left(R_k^{(n)} = j\right) = \frac{1}{k},\;\; (n \in \mathbb{N}^*,\, k \in \{1,\cdots,n\},\, j \in \{1,\cdots,k\}),$$
$$R_1^{(n)},\ldots,R_n^{(n)} \text{ sont des variables aléatoires indépendantes.}$$

Nous remarquons que la loi de $R_1^{(n)}, \ldots, R_n^{(n)}$ est la même que dans le cas d'un tableau de variables aléatoires i.i.d.. Nous en déduisons que tout résultat asymptotique obtenu reposant sur $(R_j^{(n)})_{1\leq j\leq n, n\in\mathbb{N}}$ pour un tableau i.i.d. est encore valable pour un tableau échangeable.

Dans la suite, comme pour $k \in \{1, \cdots, n\}$ et $n \in \mathbb{N}^*$, la loi de $(R_1^{(n)}, \ldots, R_k^{(n)})$ ne dépend que de k, nous écrirons R_k au lieu de $R_k^{(n)}$. Nous introduisons alors les processus suivants, pour $t \in [0, 1]$:

$$Z_n^\alpha(t) \;=\; \frac{\alpha}{n^\alpha} \sum_{j=1}^{[nt^{1/\alpha}]} \mathbb{1}_{\{R_j \leq [j^\alpha]\}} \quad (0 < \alpha < 1)$$

et

$$Z_n^0(t) \;=\; \frac{1}{\log n} \sum_{j=1}^{[n^t]} \mathbb{1}_{\{R_j = 1\}},$$

où $[\cdot]$ désigne la partie entière.

A renormalisation près, le processus Z_n^0 correspond au nombre de records observés avant $[n^t]$. Le processus Z_n^0 saute aux endroits où une valeur maximale est atteinte. De façon similaire, le processus Z_n^α est le nombre de records "généralisés" observés avant $[nt^{1/\alpha}]$. Ce processus a été étudié par Deheuvels & Nevzorov [21] pour le cas i.i.d. seulement, mais dans le contexte plus général où la suite $\{[j^\alpha], j \in \mathbb{N}^*\}$ est remplacée par une suite $\{k_j, j \in \mathbb{N}^*\}$. L'intérêt de prendre en compte non seulement les records, mais aussi les k_j plus grandes observations, est de perdre moins de données puisque dans la définition du processus des records généralisés, plus d'observations apparaissent.

Dans **[Pri-3]**, nous étudions les propriétés asymptotiques de ces processus. Nous démontrons un principe de grandes déviations fonctionnel pour Z_n^α $(0 \leq \alpha < 1)$. Donnons brièvement une idée de la démonstration lorsque $0 < \alpha < 1$. Nous utilisons en fait une astuce qui consiste à remplacer la suite $(Y_j^\alpha)_{j\in\mathbb{N}^*} := (\mathbb{1}_{R_j \leq [j^\alpha]})_{j\in\mathbb{N}^*}$ par une suite de même loi, construite à partir du processus de Poisson standard. Ceci nous permet ensuite de montrer que le processus $Z_n^\alpha(\cdot)$ est exponentiellement équivalent au processus renormalisé

$$M_n^\alpha(t) = \frac{\alpha}{n^\alpha} N^*_{[nt^{1/\alpha}]}\,,\; t \in [0, 1]\,,$$

si $(N_t^*)_{t\geq 0}$ désigne le processus de Poisson d'intensité $I(x) = x^{\alpha-1}$, $x > 0$. La fonction de taux que nous obtenons *in fine* est la même que celle associée au principe de grandes déviations pour un processus de Poisson standard, à un facteur α près. Avant d'énoncer le résultat principal, précisons quelques notations.

Soit $\mathcal{AC}$ le sous-ensemble des fonctions absolument continues de $\mathbb{L}^\infty([0, 1])$ défini par

$$\mathcal{AC} = \left\{ \phi \in \mathbb{L}^\infty([0, 1]) : \; \sum_{l=1}^{k} |t_l - s_l| \to 0, s_l < t_l \leq s_{l+1} < t_{l+1} \Rightarrow \sum_{l=1}^{k} |\phi(t_l) - \phi(s_l)| \to 0 \right\}.$$

Rappelons que toute ϕ dans $\mathcal{AC}$ est différentiable presque partout. Nous noterons ϕ' sa dérivée, qui est dans $\mathbb{L}^1([0,1])$. Nous définissons $\mathcal{AC}_0^+$ comme l'ensemble des fonctions positives de $\mathcal{AC}$ tendant vers 0 en 0. Nous définissons maintenant pour $\phi \in \mathbb{L}^\infty([0,1])$,

$$J(\phi) = \begin{cases} \displaystyle\int_0^1 [\phi'(t)\log\phi'(t) - \phi'(t) + 1]dt, & \text{si } \phi \in \mathcal{AC}_0^+ \\ +\infty & \text{sinon.} \end{cases} \tag{2.3.19}$$

Remarque 13 *La fonction J joue un rôle important pour le processus de Poisson standard. Soit $(N_t)_{t\geq 0}$ le processus de Poisson standard. Lorsque T tend vers $+\infty$, $\{T^{-1}N_{tT},\, t \in [0,1]\}$ converge en probabilité vers le processus déterministe $\{t,\, t \in [0,1]\}$. De plus, ce processus satisfait un principe de grandes déviations dans $\mathcal{AC}_0^+$ de bonne fonction de taux J [37]. D'autre part, il est facile de montrer que pour ϕ dans $\mathcal{AC}_0^+$, $J(\phi) = 0$ ssi $\phi(t) \equiv t$.*

Nous obtenons le principe de grandes déviations suivant :

Théorème 8 [Pri-3] *Pour $0 < \alpha < 1$ (resp. $\alpha = 0$), (Z_n^α) satisfait dans $\mathbb{L}^\infty([0,1])$ un principe de grandes déviations de bonne fonction de taux J/α (resp. J) et de vitesse (n^α) (resp. $(\log n)$).*

En utilisant des arguments similaires, nous montrons également un principe de type Donsker pour les processus recentrés et renormalisés

$$\tilde{Z}_n^\alpha(t) = \sqrt{\frac{n^\alpha}{\alpha}}\,\{Z_n^\alpha(t) - t\} \qquad (0 < \alpha < 1,\ t \in [0,1])$$

et

$$\tilde{Z}_n^0(t) = \sqrt{\log n}\,\left\{Z_n^0(t) - t\right\}, \quad t \in [0,1]\ .$$

Nous retrouvons alors partiellement (seulement pour $0 \leq \alpha < 2/3$), mais avec un argument simple, les résultats obtenus dans le cas i.i.d. par [21]. Pour $\alpha \geq 2/3$, l'approximation de type Poisson ne fonctionne pas.

Théorème 9 [Pri-3] *Soit $\alpha \in [0, \frac{2}{3}[$. Lorsque n tend vers l'infini, $\{\tilde{Z}_n^\alpha(t), t \in [0,1]\}$ converge en loi vers le mouvement Brownien standard sur $[0,1]$.*

2.4 Estimation adaptative

Dans **[Pri-16]**, nous nous sommes intéressées au problème de l'estimation adaptative d'une fonctionnelle linéaire.

Nous nous plaçons dans le cadre du modèle gaussien :

$$Y(t) = \langle s, t\rangle + \frac{1}{\sqrt{n}} L(t)\ ,\ t \in \mathbb{H}, \tag{2.4.20}$$

où $\mathbb{H}$ est un espace de Hilbert séparable, et L est un processus gaussien centré isonormal, c'est-à-dire qui vérifie

$$\mathrm{Cov}(L(t), L(t')) = \langle t, t' \rangle \,,$$

et $s \in \mathbb{H}$ est inconnu.

Ce cadre général couvre le modèle de bruit blanc, le modèle de suite gaussienne et le modèle de régression gaussienne.

Dans **[Pri-16]**, nous construisons un nouvel estimateur adaptatif pour $T(s)$, où T est une fonctionnelle linéaire. Dans [12], Cai et Low proposent un estimateur adaptatif de $T(s)$ pour s appartenant à une classe convexe $\mathcal{F}$. Leur estimateur est construit à partir d'une série de tests sur une suite ordonnée de classes $\mathcal{F}_i \subset \mathcal{F}$. D'autre part, il existe différents travaux [33, 42, 41] présentant des approches adaptatives de type noyau. Notre construction **[Pri-16]** repose sur l'idée de sélection de modèle. On considère une collection au plus dénombrable $(S_m,\, m \in \mathcal{M})$ de sous-espaces linéaires de $\mathbb{H}$. Pour tout $m \in \mathcal{M}$, on définit l'estimateur par projection $\hat{s}_m$ de s sur S_m. Etant donnée une base orthonormale $(\phi_\lambda,\, \lambda \in \Lambda_m)$ de S_m,

$$\hat{s}_m = \sum_{\lambda \in \Lambda_m} Y(\phi_\lambda)\phi_\lambda \,.$$

En fait, $\hat{s}_m$ ne dépend pas du choix de la base puisqu'on peut facilement vérifier que

$$\hat{s}_m = \mathrm{argmin}_{v \in S_m}(\|v\|^2 - 2Y(v)) \,.$$

Il paraît alors naturel d'estimer $T(s)$ par $T(\hat{s}_m)$. Soit s_m la projection orthogonale de s sur S_m. Comme T est linéaire,

$$\mathbb{E}(T(\hat{s}_m)) = T(s_m) \,.$$

On peut donc donner la décomposition biais-variance pour le risque quadratique de l'estimateur $T(\hat{s}_m)$:

$$\mathbb{E}\left((T(\hat{s}_m) - T(s))^2\right) = (T(s_m) - T(s))^2 + \mathbb{E}\left((T(\hat{s}_m) - T(s_m))^2\right) \,.$$

On souhaite alors trouver un estimateur, parmi la collection $(T(\hat{s}_m),\, m \in \mathcal{M})$, qui minimise le critère

$$(T(s_m) - T(s))^2 + \mathbb{E}\left((T(\hat{s}_m) - T(s_m))^2\right) \,.$$

Le terme de variance se calcule simplement, en utilisant les propriétés du processus isonormal L.

$$\mathbb{E}\left((T(\hat{s}_m) - T(s_m))^2\right) = \frac{\sigma^2}{n} \sum_{\lambda \in \Lambda_m} T^2(\phi_\lambda) \,.$$

Par contre, le terme de biais $(T(s_m) - T(s))^2$ dépend de la quantité inconnue s.

La sélection de modèle par critère pénalisé a été introduite par Barron, Birgé et Massart [5]. Elle a été ensuite mise en pratique dans différents contextes (estimation de s pour le modèle (2.4.20) [10], estimation de fonctionnelles quadratiques de s [35]). Cette méthode consiste à minimiser en $m \in \mathcal{M}$ un critère faisant intervenir l'estimation du biais plus un terme de pénalité $\mathrm{pen}(m)$. Dans notre cas, le terme de biais ne peut être ni simplifié par ajout d'un terme indépendant de m, ni estimé. C'est pourquoi nous avons introduit un nouveau critère.

Supposons que $\mathcal{M}$ est un sous-ensemble de $\mathbb{N}$. Nous adaptons alors, à la manière de Birgé [9], la méthode de Lepski [38, 39, 40] à notre approche par sélection de modèles. Nous cherchons à minimiser

$$\sup_{j \geq m}(T(s_m) - T(s_j))^2 + \frac{\sigma^2}{n} \sum_{\lambda \in \Lambda_m} T^2(\phi_\lambda) \, .$$

Nous définissons, pour tout $m \in \mathcal{M}$,

$$\sigma_m^2 = \frac{\sigma^2}{n} \sum_{\lambda \in \Lambda_m} T^2(\phi_\lambda)$$

et

$$\mathrm{pen}(m) = \left(2 + \frac{1}{\delta}\right) x_m \sigma_m^2 \, ,$$

où δ est une constante strictement positive et $(x_m,\, m \in \mathcal{M})$ est une suite de réels positifs.

Nous posons, pour tout $j,\, m \in \mathcal{M}$,

$$\sigma_{j,m}^2 = \frac{\sigma^2}{n}\,\mathbb{E}\left((\sum_{\lambda \in \Lambda_m} T(\phi_\lambda)L(\phi_\lambda) - \sum_{\lambda \in \Lambda_j} T(\phi_\lambda)L(\phi_\lambda))^2 \right)$$

et

$$H(j,m) = \left(2 + \frac{1}{\delta}\right) x_{j,m}\, \sigma_{j,m}^2 \, ,$$

où $(x_{j,m},\, (j,m) \in \mathcal{M}^2)$ est une suite de réels positifs.

Nous posons alors, pour tout $m \in \mathcal{M}$,

$$\widehat{\mathrm{Crit}}(m) = \sup_{j \geq m,\, j \in \mathcal{M}} \left((T(\hat{s}_m) - T(\hat{s}_j))^2 - H(j,m) \right) + \mathrm{pen}(m) \, ,$$

et nous définissons

$$\hat{m} = \inf \left\{ m \in \mathcal{M},\, \widehat{\mathrm{Crit}}(m) \leq \inf_{j \in \mathcal{M}} \widehat{\mathrm{Crit}}(j) + \frac{1}{n} \right\} \, .$$

Dans **[Pri-16]**, nous donnons une borne supérieure pour le risque quadratique de l'estimateur $T(\hat{s}_{\hat{m}})$ (Théorème 10 ci-dessous).

Théorème 10 [Pri-16] *On observe le processus gaussien $\{Y(t),\, t \in \mathbb{H}\}$, où $Y(t)$ est défini par (2.4.20). Soit T une fonctionnelle linéaire définie sur $\mathcal{S} \subset \mathbb{H}$. Pour tout $m \in \mathcal{M}$, on définit*

$$\mathrm{Crit}(m) = \sup_{j \geq m}(T(s_j) - T(s_m))^2 + \mathrm{pen}(m)$$

et

$$\Gamma(m) = (T(s_m) - T(s))^2 + \sigma_m^2 + \sup_{j \leq m} x_{m,j}\, \sigma_{m,j}^2 \,.$$

Soit m_{opt} défini par

$$m_{opt} = \inf \left\{ m \in \mathcal{M} / \operatorname{Crit}(m) \leq \inf_{l \in \mathcal{M}} \operatorname{Crit}(l) + \frac{1}{n} \right\} \,.$$

Alors, il existe une constante $C(\delta) \geq 0$, dépendant uniquement de δ, telle que

$$\begin{aligned} \mathbb{E}\left((T(\hat{s}_{\hat{m}}) - T(s))^2\right) &\leq C(\delta)\left(\operatorname{Crit}(m_{opt}) + \Gamma(m_{opt})\right) \\ &+ C(\delta)\left(\sum_{m \in \mathcal{M}} e^{-x_m}\, \sigma_m^2 + \sum_{j \geq m_{opt}} e^{-x_{j,m_{opt}}}\, \sigma_{j,m_{opt}}^2 + \tfrac{1}{n}\right) \,. \end{aligned}$$

Ce théorème permet entre autres de donner des résultats de type minimax pour l'estimation adaptative d'une fonction (ou de sa dérivée $r^{\text{ième}}$) en un point. Nous considérons le modèle de bruit blanc gaussien

$$Y(u) = \int_0^u s(x)dx + \frac{\sigma}{\sqrt{n}} W(u)\,,\ u \in [0,1]\,,$$

où $s \in \mathbb{H} = \mathbb{L}^2([0,1])$ et W est un mouvement Brownien standard.

Soit $r \geq 0$. Nous supposons que $s^{(r)}$ existe et que $s^{(r)} \in \mathcal{C}([0,1])$, l'ensemble des fonctions continues sur $[0,1]$. Soit $x_0 \in [0,1]$. Dans **[Pri-16]**, nous avons considéré le problème de l'estimation de $T(s) = s^{(r)}(x_0)$.

Soit $\{S_j,\ j \geq 0\}$ une analyse multirésolution d'ondelette père φ et d'ondelette mère ψ. Nous renvoyons à [31] pour les rappels sur les analyses multirésolutions. On définit

$$\varphi_{j,k}(x) = 2^{j/2}\varphi(2^j x - k),\ x \in [0,1],\ j \geq 0 \text{ et } k \in \mathbb{Z}$$

et

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k),\ x \in [0,1],\ j \geq 0 \text{ et } k \in \mathbb{Z}\,.$$

Pour $J \geq 0$, soit s_J la projection orthogonale de s sur S_J, $s_J = \sum_{k \in \mathbb{Z}} \langle s, \varphi_{J,k} \rangle \varphi_{J,k}$.

Pour $\alpha > 0$ et $1 \leq q \leq +\infty$, $B_{\infty}^{\alpha,q}([0,1])$ désigne l'espace de Besov usuel, muni de la norme $\|\cdot\|_{\alpha,\infty,q}$ définie par exemple dans [31, Definition 9.2].

Nous faisons sur φ et ψ des hypothèses techniques que nous ne réécrirons pas ici (voir les conditions **(i)** à **(iv)** dans **[Pri-16**, Section 3.1**]**).

Nous obtenons alors le résultat suivant :

Corollaire 1 [Pri-16] *Soient d_n un entier tel que $2^{d_n} \geq n/\ln(n)$ et $\mathcal{M} = \{0, \ldots, d_n\}$. Soit $x_m = (\ln(2) + \varepsilon)(1 + 2r)(m \vee 1)$, et $x_{j,m} = (\ln(2) + \varepsilon)(1 + 2r)(j \vee 1)$. On définit $\hat{m}$ comme au théorème 10.*

Alors il existe une constante C dépendant uniquement de δ, ε et α telle que, si $r < \alpha \leq r+1$, et si $n/\ln(n) \geq \max(L^{1/\alpha}, L^{-2})$,

$$\sup_{s \in B_{\infty}^{\alpha;q}([0,1]),\, \|s\|_{\alpha,\infty,q} \leq L} \mathbb{E}\left((\hat{s}_{\hat{m}}^{(r)}(x_0) - s^{(r)}(x_0))^2 \right) \leq C \left(L^{\frac{2+4r}{1+2\alpha}} \left(\frac{\ln(n)}{n} \right)^{\frac{2(\alpha-r)}{2\alpha+1}} + \frac{1}{n} \right) .$$

Remarque 14

- *Nous retrouvons les vitesses minimax adaptatives établies par Lepski* [38]. *C'est-à-dire qu'on perd un facteur logarithme pour le risque quadratique par rapport à la vitesse minimax sur les espaces considérés.*
- *Le corollaire 1 s'étend en dimension $d > 1$ (voir* **[Pri-16**, Corollary 2**]***).*

Sur les simulations présentées dans **[Pri-16]**, nous comparons notre méthode d'estimation à la procédure de type C_p de Mallow présentée dans [4]. Nous considérons le modèle de régression gaussienne de dimension finie

$$y_i = s(i/n) + \sigma\varepsilon_i \quad i = 1, \ldots, n\,,$$

où $(\varepsilon_1, \ldots, \varepsilon_n)$ sont des variables aléatoires gaussiennes centrées réduites, indépendantes. Nous choisissons $\sigma = 0.2$ et $n = 256$. Nous estimons alors $s(x)$ pour différentes fonctions s définies sur $[0, 1]$ et pour tout x dans $\{i/n,\ 1 \leq i \leq n\}$. Si la fonction s est irrégulière sur un sous-intervalle de $[0, 1]$, et plus régulière ailleurs, alors la procédure de [4] va avoir tendance à sélectionner un modèle de grande dimension, de manière à capturer les irrégularités de s. Notre procédure présente l'avantage de sélectionner un modèle différent pour chaque x dans $\{i/n,\ 1 \leq i \leq n\}$. Elle s'adapte ponctuellement aux irrégularités de s.

Pour conclure ce paragraphe, notons que notre procédure présente l'avantage d'être facilement programmable.

2.5 Perspectives

Pour conclure ce document de synthèse, donnons quelques perspectives de recherche.

- **Propriétés statistiques de l'application LSV** (en collaboration avec J. Dedecker, Paris 6)

  L'application LSV (Liverani-Saussol-Vaienti [44]) $T : [0, 1] \to [0, 1]$ est définie par

  $$T(x) = \begin{cases} x(1 + 2^{\kappa}x^{\kappa}) & \text{si } 0 \leq x \leq 1/2 \\ 2x - 1 & \text{si } 1/2 < x \leq 1 \end{cases}$$

  Cette application présente un point fixe neutre en zéro, c'est-à-dire $T(0) = 0$ et $|T'(0)| = 1$. On sait que si T a un point fixe neutre, alors T a une mesure invariante μ absolument

continue par rapport à la mesure de Lebesgue. De plus, si au voisinage de ce point fixe neutre, $T(x)$ est équivalente à $x+Cx^{1+\gamma}$, $0<\gamma<1$, alors la mesure μ est finie [46]. Nous nous placerons toujours dans ce cas, qui correspond à un paramètre κ dans l'intervalle ouvert $]0,1[$. Nous nous intéressons aux propriétés statistiques du système dynamique associé à T : $X_0 \sim \mu$, $X_n = T^n X_0$. La densité invariante de ce modèle explose en 0.
En utilisant des estimées de Young [60], nous pouvons montrer que dans ce cas, la suite des coefficients de dépendance $\tilde{\alpha}(\sigma(X_n), X_0)$ décroît vers zéro avec une vitesse polynomiale. On n'a pas pour ce modèle de propriété de trou spectral de l'opérateur de Perron-Frobenius, et Young obtient ses estimées en travaillant sur une tour. Nous montrons que pour tout $\varepsilon > 0$, il existe $A, B > 0$ telles que

$$\frac{A}{n^{\frac{1-\kappa}{\kappa}}} \le \tilde{\alpha}(\sigma(X_n), X_0) \le \frac{B}{n^{\frac{1-\kappa}{\kappa}-\varepsilon}} .$$

Nous pouvons alors utiliser les résultats connus pour les suites $\tilde{\alpha}$-dépendantes [18, 19] pour montrer un théorème limite centrale (tlc) pour les sommes partielles $\sum_{i=0}^{n-1} f \circ T^i$, où $f : [0,1] \to \mathbb{R}$ n'est pas forcément bornée. Dans le cas où f est bornée, Gouëzel donne des conditions pour avoir le tlc ainsi que des résultats de type Berry-Esseen en utilisant la théorie du renouvellement [28].
Nous espérons par ailleurs montrer la décroissance polynomiale de $\tilde{\alpha}(\sigma(X_n), (X_{n-i}, X_{n-j}))$ lorsque $n \ge j \ge i$ et $i \to +\infty$ et de $\tilde{\alpha}(\sigma(X_n), (X_{n-i}, X_{n-j}, X_{n-k}))$ lorsque $n \ge k \ge j \ge i$ et $i \to +\infty$. Ceci nous permettrait d'obtenir ensuite le comportement du processus empirique et une vitesse dans le tlc en utilisant **[Pri-1]** et [20].
Parallèlement, nous pensons réaliser des simulations pour des fonctions f correspondant aux cas limites pour le tlc. Nous espérons vérifier que dans ce cas, nous n'avons pas le tlc.

- **Systèmes dynamiques échantillonnés** (en collaboration avec N. Guillotin-Plantard, Lyon 1)

Soit $(E, \mathcal{E}, \mu)$ un espace probabilisé et $T : E \to E$ une application préservant μ. On suppose que T est mesurable, inversible et ergodique. Soit $(S_k)_{k\ge 0}$ définie par

$$S_0 = 0\,, \text{ et pour tout } k \ge 1\,,\ S_k = X_1 + \cdots + X_k\,,$$

où $(X_k)_{k\ge 1}$ est une suite i.i.d. de variables aléatoires de carré intégrable, définie sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ et à valeurs dans $\mathbb{Z}$. On suppose que la marche aléatoire $(S_n)_{n\ge 0}$ est fortement apériodique et transiente. On sait alors que pour tout $x \in \mathbb{Z}$, la fonction de Green

$$G(0,x) = \sum_{k=0}^{\infty} \mathbb{P}(S_k = x)$$

est finie. Nous nous intéressons alors au comportement de

$$\sum_{k=0}^{n} f \circ T^{S_k(\omega)} ,$$

pour $f \in \mathbb{L}^1(\mu)$. Guillotin-Plantard et Schneider [30, Theorem 3.2] ont montré le résultat suivant :
Supposons qu'il existe une tribu $\mathcal{F}_0$ telle que $\mathcal{F}_0 \subset T^{-1}\mathcal{F}_0$. Alors, pour toute $f \in \mathbb{L}^2(\mu)$ telle que f est $\mathcal{F}_0$-mesurable et $\mathbb{E}_\mu(f|T\mathcal{F}_0) = 0$, pour presque tout $\omega \in \Omega$,

$$\frac{1}{\sigma\sqrt{n}} \sum_{k=0}^{n} f(T^{S_k(\omega)}) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1)$$

avec $\sigma^2 = \mathbb{E}_\mu(f^2)(2G(0,0) - 1)$.
Leur démonstration repose sur la méthode des différences de martingales. Nous nous intéressons maintenant au cas où (T, μ) est un système dynamique de la classe $\mathcal{SD}$. Nous cherchons alors pour quelles fonctions f le théorème limite centrale est encore vrai. En remarquant que pour tout $\omega \in \Omega$,

$$\sum_{k=0}^{n} f(T^{S_k(\omega)}) = \sum_{x\in\mathbb{Z}} N_n(x)(\omega) f(T^x) ,$$

où $N_n(x) = \sum_{i=0}^{n} \mathbb{I}_{S_i = x}$, nous pouvons étudier le problème plus général suivant : *Soit $(\xi_i)_{i\in\mathbb{Z}}$ une suite faiblement dépendante. Soit $(a_{n,i})_{(n,i)\in\mathbb{N}\times\mathbb{Z}}$ une suite d'entiers naturels. Quelles sont les conditions pour avoir le théorème limite centrale pour les sommes $\sum_{i\in\mathbb{Z}} a_{n,i}\,\xi_i$* ? Dans un premier temps, notre idée est d'adapter les résultats de Peligrad & Utev [45] sur le théorème limite centrale pour les processus linéaires. Leur approche repose sur les coefficients de mélange fort. Nous espérons ensuite ramener ces conditions à des conditions de mélange faible. Les résultats obtenus dans **[Pri-11]** sur le théorème limite centrale pour des tableaux triangulaires sous des hypothèses de θ_2 ou ζ-dépendance ne semblent pas suffire pour déduire un tlc pour les systèmes dynamiques de la classe $\mathcal{SD}$ échantillonnés. Nous pouvons tout de même essayer une approche similaire, de type Lindeberg-Rio, mais en bornant plus finement les termes de covariance.

- **Estimation de la densité spectrale : efficacité de l'estimateur**

Dans **[Pri-2]**, nous avons étudié les propriétés de convergence du périodogramme intégré. Il nous paraît intéressant de nous pencher sur la question de l'efficacité de cet estimateur dans notre cadre.

- **Tests d'indépendance et d'échangeabilité** (en collaboration avec A.L. Fougères, Paris X et F. Gamboa, Toulouse 3)

  Une question classique en statistique est de savoir si une suite stationnaire est indépendante ou pas. Il existe différents tests d'indépendance. En regardant de plus près, nous remarquons que la plupart d'entre eux ne testent pas en réalité l'indépendance, mais l'échangeabilité de la suite. C'est par exemple le cas de la statistique de Kendall (voir [36]) définie par :

$$\tau_n = 1 - \frac{4N}{n(n-1)},$$

  où $N = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathbb{1}_{X_i > X_j}$ désigne le nombre de paires de discordances. On peut encore écrire N en fonction des rangs relatifs R_k définis par (2.3.18) : $N = \sum_{k=1}^{n} R_k - n$. Nous montrons alors (voir **[Pri-3]**) que la statistique de Kendall τ_n est asymptotiquement équivalente à

$$2 - \frac{4}{n} \sum_{k=1}^{n} \frac{R_k}{k}.$$

  Le test d'indépendance classique construit à partir de τ_n repose donc sur le comportement en moyenne des R_k/k. La statistique $n^\alpha/\alpha Z_n^\alpha(1)$ repose sur le comportement des queues de distribution des R_k/k. Il semble donc intéressant de chercher à définir une hypothèse alternative qui serait plus difficile à détecter en regardant le comportement en moyenne des R_k/k qu'en regardant le comportement de leurs queues.

- **Grandes et moyennes déviations conditionnelles pour une somme de variables aléatoires discrètes** (en collaboration avec F. Gamboa et T. Klein, Toulouse 3)

  Il s'agit d'un travail en préparation avancée. L'idée est de montrer des grandes et moyennes déviations conditionnelles pour une somme de variables aléatoires discrètes.
  Pour $n \in \mathbb{N}^*$, soit $(X^{(n)}, Y^{(n)})$ un vecteur aléatoire tel que $X^{(n)} \in \mathbb{N}$. Si $\mathrm{Supp}(X^{(n)})$ désigne le support de $X^{(n)}$, on définit

$$m_{X^{(n)}} = \sup\{m \in \mathbb{N}, \exists\, b \in \mathbb{N}, \mathrm{Supp}(X^{(n)}) \subset m\mathbb{N} + b\}.$$

  Nous supposons que $m_{X^{(n)}} = 1$ et que $(X^{(n)}, Y^{(n)})$ converge en loi vers (X, Y) où X est à valeurs entières positives. Soit maintenant $\left((X_i^{(n)}, Y_i^{(n)})\right)_{i\in\mathbb{N}^*}$ une suite i.i.d. de vecteurs aléatoires de même loi que $(X^{(n)}, Y^{(n)})$, et indépendante de (X, Y). Pour $n \in \mathbb{N}^*$ et $q_n \in \mathbb{N}^*$, nous posons

$$S_n = X_1^{(n)} + \cdots + X_{nq_n}^{(n)}$$

  et

$$T_n = Y_1^{(n)} + \cdots + Y_{nq_n}^{(n)}.$$

Soient p, q, p_n, $q_n \in \mathbb{N}^*$ tels que p/q soit dans l'intérieur de l'enveloppe convexe de X et tels que $p_n/q_n \to p/q$. Alors nous montrons un principe de grandes déviations pour $T_n/(nq_n)$ conditionnellement à l'évènement $S_n = np_n$. Nous montrons également un principe de déviations modérées conditionnel pour T_n, correctement recentré et renormalisé. Nous souhaitons ensuite appliquer ces résultats à des problèmes combinatoires comme le remplissage d'urnes par exemple.

- **Estimation adaptative et images** (en collaboration avec B. Laurent)

  Nous aimerions mettre en pratique la procédure d'estimation adaptative que nous avons introduite dans **[Pri-16]** pour débruiter des images. Il nous semble en effet intéressant de comparer notre procédure avec d'autres outils dans ce domaine.

Bibliographie

Les références dont l'étiquette dans ce mémoire est de la forme **[Pri-*]** sont données au chapitre 3. Je ne présente ici que les autres références.

[1] Andrews D.W.K. (1984). Nonstrong mixing autoregressive processes. *J. Appl. Probab.* 21, p. 930-934.

[2] Andrews D.W.K. and Pollard D. (1994). An introduction to functional central limit theorems for dependent stochastic processes. *Int. Stat. Rev.* 62, p. 119-132.

[3] Antoch J., Hušková M. and Prášková Z. (1997). Effect of dependence on statistics for determination of change. *J. Statist. Plann. Inference* 60, 2, p. 291-310.

[4] Baraud Y. (2000). Model selection for regression on a fixed design. *Probab. Theory Related Fields* 117, p. 467-493.

[5] Barron A.R., Birgé L. and Massart P. (1999). Risk bounds for model selection via penalization. *Probab. Theory Related Fields* 113, p. 301-415.

[6] Barbour A.D., Gerrard R.M. and Reinert G. (2000). Iterates of expanding maps. *Probab. Theory Relat. Fields* 116, p. 151-180.

[7] Berbee H.C.P. (1979). Random walks with stationary increments and renewal theory. *Math. Cent. Tracts.* Amsterdam.

[8] Bickel P. and Bühlmann P. (1999). A new mixing notion and functional central limit theorems for a sieve bootstrap in time series. *Bernoulli* 5, 3, p. 413-446.

[9] Birgé L. (2001). An alternative point of view on Lepski's method. *State of the art in probability and statistics (Leiden, 1999). IMS Lecture Notes Monogr. Ser.* 36, p. 113-133. Inst. Math. Statist., Beachwood, OH.

[10] Birgé L. and Massart P. (2001). Gaussian model selection. *J. Eur. Math. Soc.* 3, 3, p. 203-268.

[11] Broise A. (1996). Transformations dilatantes de l'intervalle et théorèmes limites. Études spectrales d'opérateurs de transfert et applications. *Astérisque* 238, p. 1-109.

[12] Cai T. and Low M. (2005). On adaptive estimation of linear functionals. *Ann. Statist.* 33, 5, p. 2311-2343.

[13] Collet P., Martinez S. and Schmitt B. (2002). Exponential inequalities for dynamical measures of expanding maps of the interval. *Probab. Theory. Relat. Fields* 123, p. 301-322.

[14] Deddens J., Peligrad M. and Yang T. (1987). On strong consistency of kernel estimators under dependence assumptions. *Mathematical statistics and probability theory* B, p. 33-41. (Bad Tatzmannsdorf, 1986) Reidel, Dordrecht.

[15] Dedecker J. (2001). Exponential inequalities and functional central limit theorems for random fields. *ESAIM Probab. & Stat.* `http ://www.emath.fr/ps/` 5, p. 77-104.

[16] Dedecker J. and Doukhan P. (2003). A new covariance inequality and applications. *Stochastic Process. Appl.* 106, p. 63-80.

[17] Dedecker J. and Louhichi S. (2005). Conditional convergence to infinitely divisible distributions with finite variance. *Stochastic processes and their applications* 115, 5, p. 737-768.

[18] Dedecker J., Merlevède F. and Volny D. (2006). On the weak invariance principle for non adapted sequences under projective criteria.
`https ://hal.ccsd.cnrs.fr/ccsd-00022125`

[19] Dedecker J. and Rio E. (2000). On the functional central limit theorem for stationary processes. *Ann. Inst. H. Poincaré Probab. Statist.* 36, p. 1-34.

[20] Dedecker J. and Rio E. (2005). *Preprint* 138.
`http ://www.math.uvsq.fr/laboratoire/activites/dossierprepub/index.html`

[21] Deheuvels P. and Nevzorov V.B. (1994). Limit laws for $K-$record times. *Journal of Statistical Planning and Inference* 38, p. 279-307. North-Holland.

[22] Delyon B. (1990). Limit theorems for mixing processes. *Tech. Report* 546 IRISA, Rennes I.

[23] Dembo A. and Zeitouni O. (1998). *Large deviations techniques and applications.* Springer-Verlag, New York, second edition.

[24] Doukhan P. and Louhichi S. (1999). A new weak dependence condition and applications to moment inequalities. *Stochastic Process. Appl.* 84, p. 313-342.

[25] Doukhan P. and Louhichi S. (2001). Functional estimation for weakly dependent stationary time series. *Scand. J. Statist.* 28, p. 325-342.

[26] Gordin, M.I. (1969). The central limit theorem for stationary processes. *Dokl. Akad. Nauk SSSR* 188, p. 739-741.

[27] Gordin M.I. and Lifshits B. (1985). Fourth Vilnius conference. *Abstracts of Communications* 1, p. 182-183.

[28] Gouëzel S. (2005). Berry-Esseen theorem and local limit theorem for non uniformly expanding maps. *Ann. Inst. H. Poincaré Probab. Statist.* 41, 6, p. 997-1024.

[29] Grégoire G. and Hamrouni Z. (2002). Change point estimation by local linear smoothing. *Journal of Multivariate Analysis* 83, 1, p. 56-83.

[30] Guillotin-Plantard N. and Schneider D. (2003). Limit theorems for sampled dynamical systems. *Stochastics and Dynamics* 3, 4, p. 477-497.

[31] Härdle W., Kerkyacharian G., Picard D. and Tsybakov A. (1998). Wavelets, Approximation, and Statistical Applications. *Lecture Notes in Stat.* 129, Springer-Verlag, New York.

[32] Ibragimov I.A. (1962). Some limit theorems for stationary processes. *Theory Probab. Appl.* 7, p. 349-382.

[33] Klemelä J. and Tsybakov A. (2001). Sharp adaptive estimation of linear functionals. *Ann. Statist.* 29, 6, p. 1567-1600.

[34] Lasota A. and Yorke J.A. (1974). On the existence of invariant measures for piecewise monotonic transformations. *Trans. Amer. Math. Soc.* 186, p. 481-488.

[35] Laurent B. and Massart P. (2000). Adaptive estimation of a quadratic functional by model selection. *Ann. Statist.* 28, 5, p. 1302-1338.

[36] Lehmann E.L. (1986). *Testing Statistical Hypotheses.* John Wiley & Sons, New York.

[37] Léonard C. (2000). Large deviations for Poisson random measures and processes with independent increments. *Stochastic Process. Appl.* 85, p. 93-121.

[38] Lepski O. (1990). On a problem of adaptive estimation in gaussian white noise. *Theory Probab. Appl.* 35, p. 454-466.

[39] Lepski O. (1991). Asymptotically minimax adaptive estimation. I. Upper bounds. Optimally adaptive estimates. *Theory Probab. Appl.* 36, 4, p. 682-697.

[40] Lepski O. (1992). Asymptotically minimax adaptive estimation. II. Schemes without optimal adaptation. *Theory Probab. Appl.* 37, 3, p. 433-448.

[41] Lepski O., Mammen E. and Spokoiny V. (1997). Optimal spatial adaptation to inhomogeneous smoothness : an approach based on kernel estimates with variable bandwidth selectors. *Ann. Statist.* 25 , 3, p. 929-947.

[42] Lepski O. and Spokoiny V. (1997). Optimal pointwise adaptive methods in nonparametric estimation. *Ann. Statist.* 25, 6, p. 2512-2546.

[43] Major P. (1978). On the invariance principle for sums of identically distributed random variables. *J. Multivariate Anal.* 8, p. 487-517.

[44] Liverani C., Saussol B. and Vaienti S. (1999). A probabilistic approach to intermittency. *Ergodic Theory and Dynamical Systems* 19, p. 671-683.

[45] Peligrad M. and Utev S. (1997). Central limit theorem for linear processes. *Ann. Probab.* 25, 1, p. 443-456.

[46] Pianigiani G. (1980). First return map and invariant measures. *Israel J. Math.* 35, p. 32-48.

[47] Resnick S.I. (1987). *Extreme Values, Regular Variation, and Point Processes.* Springer-Verlag, New York.

[48] Rio E. (1995). About the Lindeberg method for strongly mixing sequences. *ESAIM-PS* 1, p. 35-61.

[49] Rio E. (1996). Sur le théorème de Berry-Esseen pour les suites faiblement dépendantes. *Probab. Theory Relat. Fields* 104, p. 255-282.

[50] Rio E. (2000). *Théorie asymptotique des processus aléatoires faiblement dépendants.* Mathématiques & Applications 31, Springer-Verlag, Berlin.

[51] Rio E. (2000). Inégalités de Hoeffding pour les fonctions lipschitziennes de suites dépendantes. *C. R. Acad. Sci. Paris Série I* 330, p. 905-908.

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

[53] Rosenblatt M. (1985). *Stationary sequences and random fields.* Birkhäuser, Boston.

[54] Rozanov Y.A. and Volkonskii V.A. (1959). Some limit theorems for random functions I. *Theory Probab. Appl.* 4, p. 178-197.

[55] Rüschendorf L. (1985). The Wasserstein Distance and Approximation Theorems. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 70, p. 117-129.

[56] Shao Q.M. and Yu H. (1996). Weak convergence for weighted empirical processes of dependent sequences. *Ann. Probab.* 24, p. 2098-2127.

[57] Tsybakov A. (2004). *Introduction à l'estimation non-paramétrique.* Mathématiques & Applications 41, Springer-Verlag, Berlin.

[58] Viennet G. (1997). Inequalities for absolutely regular sequences : application to density estimation. *Probab. Theory Relat. Fields* 107, p. 467-492.

[59] Wu W.B. (2005). Fourier transforms of stationary processes. *Proc. Amer. Math. Soc.* 133, 1, p. 285-293.

[60] Young L.S. (1999). Recurrence times and rates of mixing. *Israel J. Math.* 110, p. 153-188.

Chapitre 3

Bibliographie personnelle

Les références **[Pri-1]** à **[Pri-16]** peuvent être téléchargées au format pdf à partir de leur étiquette dans ce mémoire. Ainsi la référence **[Pri-1]** peut être téléchargée à l'aide du lien :

`http ://www.lsp.ups-tlse.fr/Fp/Prieur/Prieur-Publi/Pri-1.pdf`

[Pri-0]. Prieur C. (2001). Applications statistiques de suites faiblement dépendantes et de systèmes dynamiques. Thèse de doctorat de l'Université de Cergy-Pontoise.

Articles dans des revues internationales avec comité de lecture

[Pri-1]. Dedecker J. et Prieur C. (2006). An empirical central limit theorem for dependent sequences. To appear in *Stochastic Process. Appl.*.

[Pri-2]. Bercu B. et Prieur C. (2006). Spectral properties of chaotic processes. *Stochastics and Dynamics* 6, 3, p. 355-371.

[Pri-3]. Fougères A.L., Gamboa F. et Prieur C. (2005). Some asymptotic results for the number of generalized records. *Extremes* 8, p.27-41.

[Pri-4]. Dedecker J. et Prieur C. (2005). New dependence coefficients. Examples and applications to statistics. *Probab. Theory and Relat. Fields* 132, p. 203-236.

[Pri-5]. Dedecker J. et Prieur C. (2004). Couplage pour la distance minimale. *C.R. Acad. Sci. Paris Sér. I Math.* 338, p. 805-808.

[Pri-6]. Dedecker J. et Prieur C. (2004). Coupling for τ-dependent sequences and applications. *Journal of Theoretical Probability* 17, 4, p. 861-885.

[Pri-7]. Ango Nze P. et Prieur C. (2001). Estimation d'une rupture en dépendance faible. *C.R. Acad. Sci. Paris Sér. I Math.* 335, p. 267-270.

[Pri-8]. Prieur C. (2002). An Emprical Functional Central Limit Theorem For Weakly Dependent Sequences. *Probab. Math. Statist.* 22, 2, p. 259-287.

[Pri-9]. Prieur C. (2001). Density Estimation For One-Dimensional Dynamical Systems. *ESAIM/Prob. Stat.* 5, p. 51-76.

[Pri-10]. Prieur C. (2001). Estimation de la densité invariante de systèmes dynamiques en dimension 1. *C.R. Acad. Sci. Paris Sér. I Math.* 332, p. 761-764.

[Pri-11]. Coulon-Prieur C. et Doukhan P. (2000). A triangular central limit theorem under a new weak dependence condition. *Statist. Probab. Lett.* 47, p. 61-68.

Chapitres de livres

[Pri-12]. Prieur C. (2006). Recent results on weak dependence for causal sequences, Statistical applications and dynamical systems' example. *Lecture Notes in Stat.* 187, p. 87-104.

[Pri-13]. Dedecker J., Prieur C. et Raynaud De Fitte P. (2006). Parametrized Kantorovich-Rubinštein theorem and application to the coupling of random variables. *Lecture Notes in Stat.* 187, p. 105-121.

Articles dans des actes de conférences internationales

[Pri-14]. Prieur C. (2002). Density Estimation for One-Dimensional Dynamical Systems. Proceedings of the Seminar on Stability Problems for Stochastic Models, Part I (Eger, 2001). *Journal of Mathematical Sciences (New York)* 111, 3, p. 3601-3612. Choisi par un comité de sélection.

Articles soumis

[Pri-15]. Prieur C. (2005). Change Point Estimation by Local Linear Smoothing under a Weak Dependence Condition. *Prépublications du LSP Toulouse* LSP-2005-07. Soumis à *Math.*

Methods of Statist..

[Pri-16]. Laurent B., Ludeña C. et Prieur C. (2006). Adaptive estimation of linear functionals by model selection. *Prépublications du LSP Toulouse* LSP-2006-05. Soumis à *Ann. Statist.*.

Rapports de contrats

[Pri-17]. Mouret M., Bardet J.M. et Prieur C. (2003). Propriétés rhéologiques de la pâte de ciment, avec le Laboratoire Matériaux et Durabilité des Constructions (Toulouse). Rapport disponible sur (`www.lsp.ups-tlse.fr/Fp/Prieur`).

Livre en préparation

[Pri-18]. Dedecker J., Doukhan P., Lang G., León J.R., Louhichi S. et Prieur C.. *Weak dependence, examples and applications.* En préparation avancée. Publication prévue chez *Springer*.

Chapitre 4

Interventions en conférences nationales ou internationales

- *Journées MAS*, Lille, septembre 2006. Organisatrice de la session Etude Statistique des Systèmes Dynamiques.

- *26th European Meeting of Statisticians*, Torun (Pologne), juillet 2006. Spectral properties of chaotic processes.

- *5th Seminar for Homogenization and Quality Control of Climatological Databases*, Budapest (Hongrie), mai 2006. Daily temperature homogenisation based on non-parametric kernel regression.

- *Propriétés Stochastiques des Systèmes Dynamiques et Milieux aléatoires*, Roscoff, mai 2005.

- *Rencontres Statistiques Mathématiques Santander-Toulouse-Valladolid*, Bonascre (Pyrénées françaises), février 2005. Recent results on weak dependence. Statistical applications.

- *Statistics for Dependent Data*, Paris/Malakoff, janvier 2005. Recent results on weak dependence for causal sequences.

- *Journées MAS*, Nancy, septembre 2004. Some asymptotic results for record processes and applications.

- *Barcelona Conference on Asymptotic Statistics*, Barcelone (Espagne), septembre 2003. Some asymptotic results for record processes and applications.

- *Kolmogorov and Contemporary Mathematics*, Moscou (Russie), juin 2003. Coupling for τ-dependent random variables and applications.

- *Rencontres Statistiques Mathématiques Santander-Toulouse-Valladolid*, Santander (Espagne), mai 2003.

- 21^{st} *International Seminar On Stability Problems for Stochastic Models*, Eger (Hongrie), 2001. Density Estimation for One-Dimensional Dynamical Systems.

- *Journées MAS*, Rennes, septembre 2000. Functional estimation of dynamical systems.

- *International conference on dynamical systems*, mai 2000, Porto, Portugal. Functional estimation of dynamical systems.

- *Workshop on Transfer Operators*, mars 2000, Paderborn, Allemagne.

Chapitre 5

Enseignements et encadrements

5.1 Enseignements

Pendant ces dernières années, j'ai eu la chance d'avoir des enseignements variés. J'ai en effet pu enseigner différentes disciplines des mathématiques dans des filières plus ou moins appliquées.

- Pendant ma thèse, j'ai enseigné (dans le cadre de vacations puis d'un monitorat) à l'Université Cergy-Pontoise. J'ai assuré pendant cette période (septembre 1999 à juillet 2002) les travaux dirigés du cours de statistique pour les étudiants de maîtrise ingénierie mathématique. Cet enseignement m'a donné l'occasion de tutorer le stage de plusieurs étudiants. Durant la même période, j'ai participé à la préparation à l'épreuve de modélisation de l'agrégation. L'épreuve de modélisation n'en était alors qu'à ses débuts. A l'aide du logiciel MATLAB, j'ai préparé les agrégatifs de l'Université Cergy-Pontoise aux leçons de statistique. Toujours dans le cadre de mon monitorat, j'ai donné des travaux dirigés d'analyse aux élèves de deuxième année du DEUG Sciences du Vivant.

- Depuis septembre 2002, je suis maître de conférences à l'INSA Toulouse. L'INSA est une Grande Ecole d'Ingénieurs, sous tutelle du Ministère de la Jeunesse, de l'Education Nationale et de la Recherche. La vocation de cet établissement est de former des ingénieurs adaptés à l'environnement socio-économique. Les élèves ingénieurs de l'INSA sont acteurs de leurs choix d'orientation. Le nouveau cursus, progressivement mis en place depuis septembre 2002, se compose pour le cycle ingénieur (il existe aussi une formation continue) d'une année de tronc commun (année 1), de deux années de préorientation (années 2 et 3), et enfin de deux années de spécialisation (années 4 et 5). A Toulouse, nous disposons d'une spécialisation GMM (Génie Mathématique et Modélisation) dont le rôle est de former des ingénieurs mathématiciens. Cette orientation comporte deux filières : la filière MCS (Modélisation et Calcul Scientifique) et la filière MSS (Modélisation Stochastique et Statistique). Les élèves de la filière MSS peuvent encore choisir de se spécialiser en finance ou en biostatistique.

Depuis mon arrivée à l'INSA, j'ai pu enseigner à différents niveaux du cursus et dans différentes orientations. Voici la liste des différents cours (C), travaux dirigés (TD), travaux pratiques (TP) que j'ai donnés.

- **Tronc commun :**
  Mathématiques et Raisonnements (TD année 1), Algèbre linéaire (C-TD-TP année 2).
- **Préorientation :**
  Probabilités et Statistique (C-TD année 3, préorientation ICBE Ingénierie Chimique, Biochimique et Environnementale).
- **Orientation Génie Mathématique et Modélisation :**
  Statistique Inférentielle (C-TD-TP année 4), Initiation SAS (TP année 4), Méthodologie Statistique (C année 5 et étudiants du Master Recherche Mathématiques et Applications de Toulouse 3).

J'ai également été responsable des projets de cinquième année de la spécialisation Génie Mathématique et Modélisation de septembre 2003 à juillet 2006. Cette responsabilité consiste à contacter des chercheurs ou des industriels afin qu'ils proposent des sujets adaptés à notre orientation, puis à gérer le bon déroulement des différents projets, et enfin à organiser les soutenances qui ont lieu en anglais.

5.2 Encadrements

Depuis mon arrivée à l'INSA Toulouse, j'ai encadré plusieurs étudiants.

- J'ai commencé par encadrer des étudiants pour leur projet de troisième année à l'école. J'ai proposé différents sujets portant sur des notions classiques en probabilités et statistique (méthode de Monte-Carlo, estimation non paramétrique de la densité, processus de Galton-Watson, ...). J'ai apprécié cette approche différente des étudiants, souvent plus motivés qu'en cours magistral.

- J'ai ensuite co-encadré (2005-2006) avec Olivier Mestre deux étudiantes en cinquième année spécialisation GMM, Aude Coutelier et Sophie Martin. Olivier Mestre travaille à l'Ecole Nationale de la Météorologie située à Toulouse. Le projet portait sur le problème d'homogénéisation de données de températures quotidiennes par une méthode non paramétrique. Nous avons pu ensuite développer les résultats obtenus et les présenter lors de la conférence internationale *5th Seminar for Homogenization and Quality Control of*

Climatological Databases qui s'est déroulée à Budapest en mai 2006.
Ce travail devrait donner lieu à la rédaction d'un article.

- J'ai également co-encadré deux mémoires de Master Recherche.
  - 2004-2005 Jordi Ordoñez Adellach (Toulouse) *Systèmes dynamiques et grandes déviations*. Co-encadré avec B. Bercu (Bordeaux 1).
  - 2005-2006 Matthieu Lerasle (Paris-Sud Orsay) *Intervalles de confiance adaptatifs*. Co-encadré avec B. Laurent (INSA Toulouse).

  Grâce à ces deux co-encadrements, j'ai pu encadrer des étudiants sur deux sujets qui me tiennent à coeur : l'étude des propriétés statistiques des systèmes dynamiques d'une part, et l'estimation adaptative d'autre part.

- Matthieu Lerasle continue en thèse, toujours sous la direction de B. Laurent (INSA Toulouse) et de moi-même (co-direction 50 %-50 %), sur le sujet suivant : *Régions de confiance adaptatives dans des modèles non paramétriques*. Dans un premier temps, Matthieu s'intéresse à des fonctionnelles linéaires dans le cadre du modèle gaussien :

$$Y(t) = \langle s, t \rangle + \frac{1}{\sqrt{n}} L(t) \ , \ t \in \mathbb{H} \, , \tag{5.2.1}$$

  où $\mathbb{H}$ est un espace de Hilbert séparable, L est un processus gaussien centré isonormal, et $s \in \mathbb{H}$ est inconnu. Il travaillera ensuite sur l'estimation adaptative de la densité invariante de systèmes dynamiques en dimension 1.

Chapitre 6

Coopérations industrielles et institutionnelles

Dans ce chapitre, nous allons présenter plus en détail les différents programmes de coopération listés au chapitre 1.

- Je suis membre du Réseau européen Physical & Engineering Sciences : Advanced Mathematical Methods for Finance (AMaMeF)
  Le développement de modèles structurels d'évaluation de la dette risquée est un enjeu essentiel pour les établissements financiers à la fois :
  - dans leur gestion des risques de portefeuilles formés de titres différents émis par une même entreprise,
  - dans leurs stratégies d'arbitrage (au sens non technique) entre les différents titres de l'entreprise, ce qui suppose une bonne maîtrise des risques pris et une bonne compréhension des corrélations entre les fluctuations de prix,
  - dans le contrôle de leurs engagements vis-à-vis d'une contrepartie.
  Actuellement les modèles structurels utilisés pour évaluer la dette des entreprises sont essentiellement fondés sur des processus gaussiens. Ces modèles ne rendent pas compte de manière convenable de la structure des spreads de crédit observée. L'objectif est d'améliorer ces modèles par la prise en compte de la dynamique de la dette, l'endogénéisation du seuil de défaut, la prise en compte d'asymétries d'information, la modélisation de la dynamique de la variable d'état (valeur de l'entreprise) à l'aide de processus de diffusion mixtes brownien-Poisson, de mouvements browniens fractionnaires ou de processus α-stables.
  Dans le cadre de ce réseau, je cherche à transposer les techniques que j'utilise en dépendance pour les problèmes de calibration (en particulier de termes de variance).

- Collaboration en cours avec O. Mestre (Ecole Nationale de la Météorologie)

Une série d'observations météorologiques est dite homogène lorsque les conditions de mesure n'ont pas varié au cours du temps. L'absence d'homogénéité des longues séries instrumentales en climatologie est un problème connu depuis longtemps. Les principales causes de ruptures sont les changements d'emplacement du site de mesure, les changements de capteurs ou d'abris météorologiques. Ces changements vont se traduire par autant de ruptures dans les séries de données. Or ces ruptures artificielles peuvent être du même ordre de grandeur que les phénomènes qu'on cherche à mettre en évidence dans les séries climatiques : tendances, cycles, ... Les méthodes d'homogénéisation reposent sur deux phases distinctes. La première phase consiste à détecter les ruptures. La deuxième phase consiste à estimer et à corriger les biais introduits par les changements. En utilisant des techniques d'estimation à noyau, nous avons obtenu en collaboration avec H. Caussinus des résultats prometteurs pour la deuxième phase. Olivier Mestre et moi-même avons également encadré deux étudiantes de l'INSA, qui ont travaillé sur des données de températures maximales et minimales quotidiennes de Toulouse-Blagnac. Pour corriger les biais sur cette série, les étudiantes disposaient d'une deuxième série, dite de référence, de températures relevées cette fois-ci à Toulouse-Francazal. Tous ces résultats ont été présentés lors de la conférence internationale *5th Seminar for Homogenization and Quality Control of Climatological Databases* qui s'est déroulée à Budapest en mai 2006. Ces travaux devraient donner lieu à la rédaction d'un article.

- Collaboration avec N. Lopes Garcia, J. Enrique Garcia (Campinas, Brésil) et A. Galves (São Paolo, Brésil)
  En mai et juin 2006, j'ai visité pendant 6 semaines l'Université de Campinas au Brésil. J'ai alors collaboré avec N. Lopes Garcia, J. Enrique Garcia et A. Galves sur le problème de l'identification automatique des langues. Le but est de reconnaître la langue parlée par un locuteur inconnu, parmi un ensemble fini de langues, pour des énoncés d'une durée limitée. La sonorité vocale semble constituer une donnée importante dans l'identification des différentes langues. La sonorité vocale semble présenter deux régimes : un régime régulier et un régime très chaotique. Le problème est de trouver une méthode automatique permettant de détecter les changements de régime. Pendant mon séjour, nous avons appliqué mes travaux sur l'estimation de ruptures pour des données dépendantes **[Pri-15]** afin d'estimer et de détecter les ruptures de régime. Les résultats obtenus doivent maintenant être analysés par des linguistes.

- Etude des propriétés rhéologiques du ciment à l'état frais.
  En 2003, j'ai travaillé en collaboration avec J.M. Bardet (Paris 1) sur un contrat avec M. Mouret du *Laboratoire Matériaux et Durabilité des Constructions*. Nous avons étudié les propriétés rhéologiques de différentes pâtes de ciment en fonction de leurs constituants,

le but étant de connaître, par des techniques statistiques, les constituants ou les groupes de constituants ayant une influence sur les différentes variables rhéologiques.
Plus précisément, nous disposions de 26 résultats d'expériences consistant en la mesure de diverses propriétés rhéologiques, qui sont des valeurs numériques réelles :
– le seuil de cisaillement,
– la viscosité maximale,
– et la viscosité minimale.
Les 26 pâtes ayant permis d'obtenir ces résultats avaient été obtenues en faisant varier, de façon purement quantitative, leurs constituants qui sont :
– la quantité de ciment,
– la quantité de filler calcaire,
– la quantité d'eau,
– la quantité d'agent de cohésion,
– la quantité de super-plastifiant.
Le but de l'étude était de déterminer par des méthodes statistiques quelles étaient les variables explicatives agissant véritablement sur les variables à expliquer et comment. Pour cela, nous avons commencé par décrire les variables en présence, puis, au vu des descriptions obtenues, nous avons essayé de modéliser à l'aide de modèles linéaires les différentes variables d'intérêt, l'échec d'une telle modélisation nous conduisant finalement à utiliser des arbres de régression, qui nous ont fourni des résultats convaincants.
Il est à noter que nous n'avions que 26 données pour 5 (ou 4 si on prend en compte leurs relations linéaires) variables explicatives. Les résultats obtenus sont donc à considérer avec précaution. De même, les variables explicatives ont été choisies en fonction de certaines conditions expérimentales et suivant un modèle de régression polynômiale : cela entraîne une relative impossibilité de transposer les résultats obtenus à un haut degré de généralité.
Ces résultats ont été en partie présentés lors des *XXII^{èmes} Rencontres Universitaires de Génie Civil 2004*.

Résumé : Le thème central des travaux présentés est l'étude des suites faiblement dépendantes non α-mélangeantes au sens de Rosenblatt (1956). La notion de mélange classique est affaiblie afin d'établir des inégalités ainsi que des théorèmes limite pour différentes classes de processus comme par exemple certains systèmes dynamiques, des chaînes de Markov non irréductibles, ou encore des fonctions de processus linéaires non mélangeants. Les résultats obtenus sont ensuite appliqués au domaine de la statistique non paramétrique.

Deux autres thématiques sont abordées dans ce manuscrit : d'une part l'étude de principes de grandes déviations (notamment pour le processus de records généralisés), et d'autre part l'estimation adaptative de fonctionnelles linéaires.

AMS 2000 Subject Classification : 37A50, 60F05, 60F10, 60F17, 60G10, 62G07, 62G08.

Mots clés : dépendance faible, processus stationnaires, théorèmes limite, inférence non paramétrique, estimation adaptative, grandes déviations.

Abstract : The major part of the presented work is devoted to new concepts of dependence extending and generalizing the classical concepts of mixing of a stationary sequence. These concepts allow the statement of various inequalities and limit theorems for different classes of processes such as dynamical systems, non irreducible Markov chains, or Bernoulli shifts. These results are then applied to derive various applications in the field of non parametric statistic.

Two other fields are also studied in the present manuscript : on one hand the study of large deviation principles (in particular for generalized records for triangular arrays of exchangeable variables), and the other hand the problem of adaptive estimation of linear functionals by Gaussian model selection.

AMS 2000 Subject Classification : 37A50, 60F05, 60F10, 60F17, 60G10, 62G07, 62G08.

Key words : weak dependence, stationary processes, limit theorems, non parametric inference, adaptive estimation, large deviations.