



Comportements Asymptotiques des Processus Stationnaires et des Processus Empiriques dans des Systèmes Dynamiques

Olivier Durieu

► To cite this version:

Olivier Durieu. Comportements Asymptotiques des Processus Stationnaires et des Processus Empiriques dans des Systèmes Dynamiques. Mathématiques [math]. Université de Rouen, 2008. Français. NNT : . tel-00346539

HAL Id: tel-00346539

<https://theses.hal.science/tel-00346539>

Submitted on 11 Dec 2008

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

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

THÈSE

en vue de l'obtention du titre de
Docteur de l'Université de Rouen

Discipline : Mathématiques

Spécialité : Probabilités

présentée par

OLIVIER DURIEU

Sujet :

COMPORTEMENTS ASYMPTOTIQUES DES PROCESSUS
STATIONNAIRES ET DES PROCESSUS EMPIRIQUES DANS DES
SYSTÈMES DYNAMIQUES

Soutenue le 1^{er} Décembre 2008

Composition du Jury

Président	:	Claude DELLACHERIE	CNRS - Université de Rouen
Rapporteurs	:	Pierre COLLET	CNRS - École Polytechnique
		Jérôme DEDECKER	Université Pierre et Marie Curie
		Michael WOODROOFE	University of Michigan
Examineurs	:	Jean-Pierre CONZE	Université de Rennes
		Youri DAVYDOV	Université de Lille
		Herold DEHLING	Ruhr-Universität Bochum
		Florence MERLEVÈDE	Université de Marne-la-vallée
		Thierry DE LA RUE	CNRS - Université de Rouen
Directeurs de Thèse	:	Dalibor VOLNÝ	Université de Rouen
		Philippe JOUAN	Université de Rouen

Thèse préparée à l'Université de Rouen
Laboratoire de Mathématiques Raphaël Salem, UMR-CNRS 6085

reMERCIements

Mon premier MERCI va à

Dalibor Volný et Philippe Jouan, mes directeurs de thèse, qui ont su être présents et me guider durant ces trois années ;

Un grand MERCI à

Pierre Collet, Jérôme Dedecker et Michael Woodroffe, qui ont accepté, et j'en suis très honoré, de rapporter cette thèse.

Claude Dellacherie qui a bien voulu présider le jury ;

Florence Merlevède, Jean-Pierre Conze, Youri Dadyvov, Herold Dehling et Thierry de la Rue qui ont témoigné de l'intérêt pour mon travail en acceptant de faire partie de mon jury.

En dehors du laboratoire, MERCI à

Jérôme Dedecker, Herold Dehling, Loïc Hervé, Stéphane Le Borgne et François Ledrappier qui à un moment ou à un autre m'ont consacré du temps et aidé à avancer dans mon travail.

Au sein du laboratoire, MERCI à

Élise Janvresse, Thierry de la Rue et Claude Dellacherie pour l'organisation de l'atelier et leur disponibilité auprès des doctorants ;

Olivier, Nico, Vincent, Islam, Nadira, Saturnin, Éditha, Ouerdia, Manel, Ali, J-C,... , mes collègues et amis doctorants (ou post-doctorants) qui m'ont permis de travailler dans une ambiance agréable ;

Marguerite Losada, Edwige Auvray, Isabelle Lamitte et Gérard Grancher qui font tourner le laboratoire ;

tous les membres du LMRS qui n'ont pas déjà été cités.

Côté enseignement, MERCI à

Olivier Benois qui m'a épaulé durant mon monitorat ;

tous les enseignants du département de mathématiques avec qui j'ai eu plaisir à travailler, notamment Patricia, Philippe et Olivier ;

Marc Jolly qui fait tourner le département.

Côté famille, pour des raisons non mathématiques, MERCI à

Claudine, Sylvie et Yves ;

Élouann et Maël ;

Émilie.

Comportements Asymptotiques des Processus Stationnaires et des Processus Empiriques dans des Systèmes Dynamiques

Cette thèse se consacre à l'étude de théorèmes limites pour des suites de variables aléatoires stationnaires (en particulier issues d'un système dynamique). Nous nous concentrons sur deux résultats importants, notamment par leurs applications en statistiques. Nous étudions tout d'abord le comportement limite des sommes de variables aléatoires, plus précisément le théorème limite central et son principe d'invariance. Ensuite nous considérons le principe d'invariance pour les processus empiriques. Dans le cadre du principe d'invariance faible de Donsker, plusieurs résultats s'obtiennent au travers d'approximations par des martingales et plus généralement par des critères projectifs. Nous comparons quatre de ces critères et montrons leur indépendance mutuelle. Les critères étudiés sont la décomposition martingale-cobord (Gordin, 1969), la condition de Hannan (1979), le critère de Dedecker et Rio (2000) et la condition de Maxwell et Woodroffe (2000). En ce qui concerne le comportement asymptotique des processus empiriques, nous établissons un principe d'invariance dans le cas des automorphismes du tore. Cela permet de sortir du cadre hyperbolique connu et d'obtenir un premier résultat pour une transformation partiellement hyperbolique. Nous proposons également une nouvelle approche, basée sur des méthodes d'opérateurs, permettant d'établir un principe d'invariance empirique. Cette méthode s'applique en particulier aux cas où l'on a de bonnes propriétés pour une classe de fonctions ne contenant pas les fonctions indicatrices. C'est en particulier le cas de certains systèmes dynamiques dont l'opérateur de transfert admet un trou spectral. En dernier lieu, suivant une question de Burton et Denker (1987), nous nous intéressons à la classe des processus pour lesquels le théorème limite central a lieu. En référence au cadre des processus empiriques, nous étudions en particulier les suites de sommes partielles des itérées d'une fonction indicatrice.

Mots clés : Processus stationnaires ; Théorème limite central ; Principe d'invariance faible ; Approximations martingale ; Critères projectifs ; Systèmes dynamiques ; Distribution empirique ; Processus empiriques ; Inégalités de moment ; Partielle hyperbolicité ; Mélange multiple ; Généricité ; Fonctions de Morse ; Chaînes de Markov ; Forte ergodicité ; Sommes de variables aléatoires ; Théorème ergodique.

Classification AMS : 28D05 ; 37A50 ; 37D30 ; 57R45 ; 60F05 ; 60F17 ; 60G10 ; 60G30 ; 60G42 ; 60J10 ; 62G20.

Asymptotic Behaviors of Stationary Processes and Empirical Processes in Dynamical Systems

The aim of this thesis is the study of limit theorems for stationary sequences of random variables (in particular, from dynamical system). We concentrate on two results which are important by their applications in statistics. We first study the asymptotic behavior of sums of random variables, precisely the central limit theorem and its invariance principles. We also consider the invariance principle of empirical processes. For the Donsker's weak invariance principle, many results can be obtained by martingale approximations and more generally by projective criteria. We compare four of these criteria and we show that they are independent of each other. These criteria are the martingale-coboundary decomposition (Gordin, 1969), the Hannan condition (1979), the Dedecker and Rio criterion (2000) and the Maxwell-Woodroffe condition (2000). Concerning the asymptotic behavior of empirical processes, we establish an invariance principle in the case of toral automorphisms. This permits to generalize the known hyperbolic case and to find a first result for a partially hyperbolic transformation. We also propose a new approach, based on operator techniques, for establishing an invariance principle. This technique is well adapted to cases when we only have good properties for a class of functions not containing the indicators. In particular, this is the case for some dynamical systems for which the transfer operator admits a spectral gap. At the end, following a question by Burton and Denker (1987), we are interested in the class of processes for which the central limit theorem holds. To refer to the empirical processes case, we study in particular the sequences of partial sums of iterates of indicator functions.

Keywords: Stationary processes; Central limit theorem; Weak invariance principle; Martingale approximation; Projective criterion; Dynamical systems; Empirical distribution; Empirical processes; Partial hyperbolicity; Multiple mixing; Morse function; Genericity; Moment inequalities; Markov chains; Strong ergodicity; Ergodic theorem; Sums of random variables.

AMS Classification: 28D05; 37A50; 37D30; 57R45; 60F05; 60F17; 60G10; 60G30; 60G42; 60J10; 62G20.

Table des matières

Présentation des Résultats	11
1 On Projective Criteria leading to the Weak Invariance Principle	25
1.1 Introduction	26
1.1.1 Martingale-coboundary decomposition	27
1.1.2 Dedecker-Rio criterion	28
1.1.3 Maxwell-Woodroffe condition	28
1.1.4 Hannan criterion	29
1.2 Comparison of the criteria	29
1.3 A preliminary result	31
1.4 General approach	34
1.5 Some counterexamples	43
1.6 A slightly different model	52
1.7 Other counterexamples	55
2 Empirical Invariance Principle for Ergodic Torus Automorphisms ; Genericity	59
2.1 Introduction	60
2.2 Empirical Invariance Principle	61
2.2.1 Hyperbolic case	62
2.2.2 Quasi-hyperbolic case	64
2.2.3 Proof of Theorem 2.2	65
2.2.4 A counterexample	75
2.3 Regularity of the distribution functions of Morse functions	76
2.3.1 Main result	77
2.3.2 Proof of Theorem 2.16	78
3 New Techniques for Empirical Processes of Dependent Data	87
3.1 Introduction	88
3.2 Main results	89

3.3	Proof of Theorem 3.1	91
3.3.1	The bounded case	91
3.3.2	The unbounded case	100
3.4	Proof of Theorem 3.3	103
3.5	A Fourth Moment Inequality for Functionals of Stationary Processes	105
3.5.1	Fourth moment inequality for Markov chains	105
3.5.2	Proof of Theorem 3.10	107
3.5.3	Fourth moment inequality for dynamical systems	112
3.6	Applications	113
3.6.1	Uniform ergodicity	114
3.6.2	Expanding maps	114
3.6.3	Subshifts	115
3.6.4	Linear processes	115
3.6.5	Random iterative Lipschitz models	117
3.6.6	Autoregressive models	118
4	On Sums of Indicator Functions in Dynamical Systems	121
4.1	Introduction	122
4.2	Result	122
4.3	Some preliminary results	123
4.4	Proof of Theorem 4.1	132
4.5	A remark on the non-ergodic case	133
4.6	A constructive proof	134
A	Multiple Mixing for Ergodic Torus Automorphisms	147
	Bibliographie	151

Présentation des Résultats

Cette thèse est consacrée à l'étude de certains théorèmes limites classiques pour des suites de variables aléatoires stationnaires. Nous nous intéresserons notamment au principe d'invariance faible de Donsker et au principe d'invariance pour des processus empiriques. La plus grande partie du travail se fera dans le cadre des systèmes dynamiques. Ainsi, ce qui suit réunira des notions liées aux statistiques et des notions relatives à la théorie ergodique des systèmes dynamiques. Nous espérons que cette thèse pourra intéresser un public proche aussi bien de l'un que de l'autre de ces domaines.

Considérons un espace probabilisé $(\Omega, \mathcal{A}, \mu)$ ainsi qu'une transformation mesurable T , de Ω dans lui même, qui préserve la mesure μ . On a donc pour chaque ensemble $A \in \mathcal{A}$, $\mu(T^{-1}A) = \mu(A)$. L'ensemble $(\Omega, \mathcal{A}, \mu, T)$ sera appelé système dynamique. Si X désigne une variable aléatoire réelle définie sur Ω , alors la suite de variables aléatoires $(X \circ T^i)_{i \in \mathbb{N}}$ forme un processus stationnaire. Nous noterons souvent X_i la variable $X \circ T^i$. Lorsque la transformation T est bijective, on pourra aussi considérer le processus stationnaire bi-infini $(X_i)_{i \in \mathbb{Z}}$. Nous rappelons qu'il y a stationnarité si quelques soient les entiers n et k , la loi du vecteur aléatoire $(X_0, \dots, X_n)$ est la même que celle de $(X_k, \dots, X_{n+k})$. En terme de processus aléatoires stationnaires, le fait de ne considérer que des processus dans le cadre de systèmes dynamiques n'est en fait pas une restriction. En effet, à tout processus aléatoire réel $(Y_i)_{i \in \mathbb{Z}}$ correspond un système dynamique $(\Omega, \mathcal{A}, \mu, T)$ et une variable aléatoire réelle X sur Ω tels que $(X \circ T^i)_{i \in \mathbb{Z}}$ et $(Y_i)_{i \in \mathbb{Z}}$ aient la même loi.

Ici, l'intérêt principal est l'étude du comportement asymptotique du processus $(X_i)_{i \in \mathbb{Z}}$ et plus particulièrement celui de ses sommes partielles. Notons S_n les sommes de Birkhoff

$$S_n = \sum_{i=0}^{n-1} X \circ T^i, \quad n \geq 1.$$

Le théorème ergodique de Birkhoff [7] assure que si la fonction X appartient à l'espace $L^1(\mu)$, alors pour μ -presque tout ω dans Ω , $\frac{S_n(\omega)}{n}$ converge vers l'espérance conditionnelle $E(X|\mathcal{I})(\omega)$ où $\mathcal{I} = \{A \in \mathcal{A} \mid T^{-1}A = A\}$ désigne la tribu des invariants pour T . Lorsque $\mathcal{I}$

est la tribu triviale (ses éléments sont de mesure 0 ou 1), on dit que le système est ergodique. Le processus $(X \circ T^i)_{i \in \mathbb{Z}}$ sera dit ergodique si le système $(\Omega, \mathcal{A}, \mu, T)$ est ergodique.

L'ensemble des processus stationnaires et ergodiques contient évidemment celui des processus i.i.d. (indépendants identiquement distribués) pour lesquels de nombreux résultats concernant le comportement asymptotique sont déjà bien connus.

Théorème limite central.

Le résultat classique, connu sous le nom de Théorème Limite Central (TLC), assure que pour toute suite de variables aléatoires réelles i.i.d. $(X_i)_{i \in \mathbb{N}}$ d'espérance nulle et de variance finie σ^2 , la suite $\left(\frac{1}{\sigma\sqrt{n}} \sum_{i=0}^{n-1} X_i \right)_{n \geq 1}$ converge en loi vers une variable aléatoire suivant une loi normale centrée réduite. Ce type de résultat est à la base de nombreux développements notamment en statistique. Cependant, l'hypothèse d'indépendance est souvent restrictive. C'est pourquoi de nombreux travaux portent sur l'adaptation du TLC au cas dépendant. Nous nous concentrerons ici sur le cas stationnaire qui donne déjà un cadre de travail plus réaliste. D'une manière générale, on dit qu'un TLC a lieu pour une suite de variables aléatoires $(X_i)_{i \in \mathbb{N}}$ si

$$\frac{S_n}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N},$$

où $\xrightarrow{\mathcal{D}}$ désigne la convergence étroite et $\mathcal{N}$ une loi normale.

Au début des années soixante, indépendamment, Billingsley [5] et Ibragimov [58] ont obtenu un TLC pour des suites stationnaires ergodiques d'accroissements d'une martingale (ou différences de martingale). On dit qu'une suite de variables aléatoires stationnaire $(X_i)_{i \in \mathbb{Z}}$ est une suite d'accroissements d'une martingale par rapport à une filtration $(\mathcal{F}_i)_{i \in \mathbb{Z}}$ si pour tout $i \in \mathbb{Z}$, X_i est $\mathcal{F}_i$ -mesurable et $E(X_{i+1}|\mathcal{F}_i) = 0$.

Théorème (Billingsley, 1961 et Ibragimov, 1963). *Soit $(X_i)_{i \in \mathbb{Z}}$ une suite ergodique d'accroissements d'une martingale telle que $X_0 \in L^2$. Alors $\left(\frac{S_n}{\sqrt{n}} \right)_{n \geq 1}$ converge étroitement vers une loi normale.*

Ce résultat fondamental a connu plusieurs généralisations comme présentées dans le livre de Hall et Heyde [50], il ouvre également la voie à plusieurs théorèmes limites basés sur des approximations par des martingales.

Dans la suite, nous supposerons que T est bijective bi-mesurable et que $\mathcal{F}$ est une sous tribu de $\mathcal{A}$ telle que $\mathcal{F} \subset T^{-1}\mathcal{F}$. Nous poserons $\mathcal{F}_i = T^{-i}\mathcal{F}$, $i \in \mathbb{Z}$, $\mathcal{F}_{-\infty} = \bigcap_{i \in \mathbb{Z}} \mathcal{F}_i$ et $\mathcal{F}_{+\infty} = \bigvee_{i \in \mathbb{Z}} \mathcal{F}_i$. Une variable aléatoire X telle que $E(X|\mathcal{F}_{-\infty}) = 0$ et $E(X|\mathcal{F}_{+\infty}) = X$ sera dite régulière.

Le premier résultat utilisant des approximations par des martingales est dû à Gordin [44], en 1969. Il observe que si la suite $(X_i)_{i \in \mathbb{Z}}$ est suffisamment "proche" d'une différence de martingale, alors un TLC a lieu. Notons H_i l'espace des fonctions $\mathcal{F}_i$ -mesurables de carré intégrable et Q l'ensemble des fonctions Y telles qu'il existe deux entiers l et k ($l < k$) pour lesquels $Y \in H_k \ominus H_l$.

Théorème (Gordin, 1969). *Soit $(\Omega, \mathcal{A}, \mu, T)$ un système dynamique ergodique. Si $X \in L^2$ est une variable aléatoire centrée et si*

$$\inf_{Y \in Q} \limsup_{n \rightarrow \infty} \frac{1}{n} E \left(\sum_{i=0}^{n-1} (X - Y) \circ T^i \right)^2 = 0, \quad (1)$$

alors $\left(\frac{S_n}{\sqrt{n}} \right)_{n \geq 1}$ converge étroitement vers une loi normale centrée.

Notons que l'hypothèse (1) est équivalente à l'existence une fonction $Y \in H_0 \ominus H_{-1}$ telle que

$$\lim_{n \rightarrow \infty} \frac{1}{n} E \left(\sum_{i=0}^{n-1} (X - Y) \circ T^i \right)^2 = 0.$$

En particulier, si l'on peut décomposer X en

$$X = m + g - g \circ T$$

où $(m \circ T^i)_{i \in \mathbb{Z}}$ forme une suite d'accroissements d'une martingale dans L^2 par rapport à une filtration $(\mathcal{F}_i)_{i \in \mathbb{Z}}$ (i.e. $m \in H_0 \ominus H_{-1}$) et g est une fonction mesurable (on parle de décomposition martingale-cobord) alors, sous l'hypothèse d'ergodicité, on peut appliquer le théorème de Billingsley et Ibragimov au processus $(m \circ T^i)_{i \in \mathbb{Z}}$, tandis que les sommes partielles $\sum_{i=0}^{n-1} (g - g \circ T) \circ T^i = g - g \circ T^n$ sont stochastiquement bornées. On en déduit un TLC pour la suite $(X_i)_{i \in \mathbb{Z}}$. Dans le cas où X_0 est $\mathcal{F}_0$ -mesurable (on parlera du cas adapté), une telle décomposition a lieu avec m et g dans L^2 dès que la série $\sum_{i=0}^{n-1} E(X_i | \mathcal{F}_0)$ converge dans L^2 . Dans le cas non adapté, c'est à dire si X_0 n'est pas $\mathcal{F}_0$ -mesurable, si de plus la série $\sum_{i=0}^{n-1} (X_{-i} - E(X_{-i} | \mathcal{F}_0))$ converge dans L^2 , alors la même décomposition a lieu (voir Hall et Heyde [50]). Cependant, il n'est pas nécessaire que g soit dans L^2 pour que le TLC ait lieu. Plus généralement, voir Volný [85], on a $X = m + g - g \circ T$ avec $m, g \in L^p$, $p \geq 1$ de sorte que $(m \circ T^i)_{i \in \mathbb{Z}}$ forme une suite d'accroissements d'une martingale par rapport à la filtration $(\mathcal{F}_i)_{i \in \mathbb{Z}}$ si et seulement si $\sum_{i=0}^{n-1} E(X_i | \mathcal{F}_0)$ et $\sum_{i=0}^{n-1} (X_{-i} - E(X_{-i} | \mathcal{F}_0))$ convergent dans L^p .

Notons P_i l'opérateur de projection de L^2 sur l'orthocomplément $H_i \ominus H_{i-1}$, *i.e.*

$$P_i(X) = E(X|\mathcal{F}_i) - E(X|\mathcal{F}_{i-1}).$$

En conséquence d'un résultat de Heyde [55], on obtient le TLC sous la condition projective suivante.

Théorème (Heyde, 1974). *Soit $(\Omega, \mathcal{A}, \mu, T)$ un système dynamique ergodique. Si X est une variable aléatoire régulière centrée vérifiant*

$$\sum_{i \in \mathbb{Z}} \|P_i(X)\|_2 < \infty, \quad (2)$$

où $\|\cdot\|_2$ désigne la norme de l'espace $L^2(\mu)$, alors $\left(\frac{S_n}{\sqrt{n}}\right)_{n \geq 1}$ converge étroitement vers une loi normale centrée.

Plus récemment, Dedecker [15] a donné un critère projectif pour le TLC. Son travail s'applique aux champs aléatoires pour lesquels les critères précédents paraissent difficiles à étendre. Bien que le critère soit de nature projective, il n'y a pas directement ici d'approximation par une martingale. C'est une adaptation de la méthode de Lindeberg [67] qui permet d'obtenir le résultat. En dimension 1 (ce qui nous intéresse ici), le critère reste intéressant puisque, comme nous le verrons, il est disjoint des résultats précédents.

Théorème (Dedecker, 1998). *Soit $(\Omega, \mathcal{A}, \mu, T)$ un système dynamique ergodique. Soit $X \in L^2$ une variable aléatoire centrée $\mathcal{F}_0$ -mesurable et $X_i = X \circ T^i$. Si*

$$\sum_{i=0}^{n-1} X_0 E(X_i|\mathcal{F}_0) \text{ converge dans } L^1, \quad (3)$$

alors $\left(\frac{S_n}{\sqrt{n}}\right)_{n \geq 1}$ converge étroitement vers une loi normale.

En 2000, Maxwell et Woodroffe [69] ont établi un TLC pour des chaînes de Markov. Ils ont également utilisé des approximations par des martingales en passant par des solutions d'une approximation de l'équation de Poisson $h = g + Qh$, où Q est l'opérateur de la chaîne Markov. Ce travail s'inscrit à la suite de ceux de Gordin et Lifšic [46], Kipnis et Varadhan [60] et Woodroffe [89]. Bien qu'étant établi pour des chaînes de Markov dans leur article, le théorème de Maxwell et Woodroffe se transcrit en terme de système dynamique (avec les notations précédentes).

Théorème (Maxwell et Woodroffe, 2000). *Soit $(\Omega, \mathcal{A}, \mu, T)$ un système dynamique ergodique. Si $X \in L^2$ est une variable aléatoire centrée $\mathcal{F}_0$ -mesurable telle que*

$$\sum_{n=1}^{\infty} n^{-\frac{3}{2}} \|E(S_n | \mathcal{F}_0)\|_2 < \infty, \quad (4)$$

alors $\left(\frac{S_n}{\sqrt{n}}\right)_{n \geq 1}$ converge étroitement vers une loi normale.

D'après Volný [86], le théorème reste valable dans le cas non adapté en ajoutant à la condition (4) l'hypothèse

$$\sum_{n=1}^{\infty} n^{-\frac{3}{2}} \|S_n - E(S_n | \mathcal{F}_n)\|_2 < \infty. \quad (5)$$

Notons que les résultats précédents s'étendent au cas non ergodique (en considérant chaque composante ergodique du système). La loi limite devient alors un mélange de lois normales.

Ces quatre théorèmes ne représentent pas une liste exhaustive des résultats concernant le TLC pour des variables dépendantes. Dans un cadre de travail proche, on pourra citer les articles de Derriennic et Lin [25], [24], [26] ainsi que Woodroffe et Wu [92]. Évidemment d'autres approches ont également porté leurs fruits, par exemple les travaux basés sur des propriétés de trou spectral pour l'opérateur de transfert (voir Hennion et Hervé [54], Baladi [2]) sur lesquels nous reviendrons. On pourra voir aussi le travail de synthèse de Derriennic [23] sur la question du TLC en théorie ergodique.

Principe d'invariance faible.

Le terme de principe d'invariance remonte aux travaux de Erdős et Kac [42] qui observèrent que certaines fonctions des sommes partielles d'une suite de variables aléatoires i.i.d. possèdent une limite étroite qui ne dépend pas de la suite elle-même. La version fonctionnelle du TLC suivante, connue sous le nom de principe d'invariance faible (souvent nous dirons seulement principe d'invariance), a été établie par Donsker [28] en 1951.

Théorème (Donsker, 1951). *Soit $(X_i)_{i \in \mathbb{N}}$ une suite i.i.d. de variables aléatoires de moyenne nulle et variance finie σ^2 , alors le processus*

$$\left\{ \frac{1}{\sigma \sqrt{n}} (S_{[nt]} - (nt - [nt])X_{[nt]+1}), t \in [0, 1] \right\}$$

converge en loi vers une mesure de Wiener dans l'espace $C[0, 1]$ muni de la norme uniforme.

Nous rappelons que la mesure de Wiener est une loi de probabilité sur l'espace des fonctions continues $C[0, 1]$. Par définition, un processus $\{x(t), 0 \leq t \leq 1\}$ admet la mesure de Wiener comme distribution s'il possède des accroissements indépendants et si pour chaque t , $x(t)$ suit une loi normale centrée de variance t . Un tel processus est appelé mouvement brownien. On pourra voir le livre de Billingsley [6] pour la démonstration de l'existence de la mesure de Wiener. Notons également que l'on peut formuler le théorème de Donsker dans l'espace $D[0, 1]$ des fonctions càdlàg (continues à droite et limitées à gauche) muni de la topologie de Skorohod (voir Billingsley [6], pour la définition). Sous les mêmes hypothèses, on a la convergence en loi, dans $D[0, 1]$, du processus $\{\frac{1}{\sigma\sqrt{n}}S_{[nt]}, t \in [0, 1]\}$ vers un mouvement brownien.

Là aussi, de nombreuses généralisations de ce théorème au cas dépendant existent. Nous nous intéressons à celles qui constituent des prolongements des résultats précédemment mentionnés pour le TLC.

En ce qui concerne la méthode de Gordin, Heyde [56] a montré que la décomposition martingale cobord dans L^2 entraîne également le principe d'invariance. Mais, si $X = m + g - g \circ T$ avec $(m \circ T^i)_{i \in \mathbb{Z}}$ une suite d'accroissements d'une martingale, il n'est pas nécessaire que g soit dans L^2 pour avoir le résultat. En revanche, contrairement au TLC, il existe des exemples où m est dans L^2 , g est dans L^1 et le principe d'invariance n'a pas lieu (voir Volný et Samek [88]). D'après Gordin [45], la décomposition a lieu avec $m \in L^2$ et $g \in L^1$ si $\sum_{i=0}^{n-1} \|E(X_i|\mathcal{F}_0)\|_1 < \infty$, $\sum_{i=0}^{n-1} \|X_{-i} - E(X_{-i}|\mathcal{F}_0)\|_1 < \infty$ et $\liminf_{n \rightarrow \infty} \frac{1}{\sqrt{n}} E|S_n| < \infty$. Plus précisément, une condition nécessaire et suffisante pour qu'une telle décomposition ait lieu est que $\sum_{i=0}^{n-1} E(X_i|\mathcal{F}_0)$ et $\sum_{i=0}^{n-1} (X_{-i} - E(X_{-i}|\mathcal{F}_0))$ convergent dans L^1 et $\liminf_{n \rightarrow \infty} \frac{1}{\sqrt{n}} E|S_n| < \infty$, voir Volný [85]. De plus, dans le cas où $m \in L^2$, le principe d'invariance a lieu si et seulement si $\frac{1}{\sqrt{n}} \max_{i \leq n} |g \circ T^i|$ converge en probabilité vers zéro (voir Hall et Heyde [50]). Dans le cas adapté, on obtient le résultat suivant.

Théorème. *Soit $(\Omega, \mathcal{A}, \mu, T)$ un système dynamique ergodique. Soit $X \in L^2$ une variable aléatoire centrée $\mathcal{F}_0$ -mesurable et $X_i = X \circ T^i$. Si*

$$\sum_{i=0}^{n-1} E(X_i|\mathcal{F}_0) \text{ converge dans } L^1 \text{ et } \liminf_{n \rightarrow \infty} \frac{1}{\sqrt{n}} E|S_n| < \infty \quad (6)$$

alors $\{\frac{1}{\sqrt{n}}S_{[nt]}, t \in [0, 1]\}$ converge en loi vers un mouvement brownien si et seulement si $\frac{1}{\sqrt{n}} \max_{i \leq n} |g \circ T^i| \xrightarrow[n \rightarrow \infty]{} 0$ en probabilité.

En ce qui concerne la condition (2), dans le cas adapté, Hannan [51], [52] a établi qu'elle entraîne le principe d'invariance sous une condition de mélange pour T . En 2003, Dedecker et Merlevède [16] se sont affranchis de cette condition de mélange. Le cas général (non adapté) fut établi par Dedecker, Merlevède et Volný [17]. On a donc le théorème suivant.

Théorème (Dedecker, Merlevède et Volný, 2007). *Soit $(\Omega, \mathcal{A}, \mu, T)$ un système dynamique ergodique. Si $X \in L^2$ est une variable aléatoire régulière centrée vérifiant la condition (2), alors $\{\frac{1}{\sqrt{n}}S_{[nt]}, t \in [0, 1]\}$ converge en loi vers un mouvement brownien.*

Dans l'article de Dedecker et Rio [19], la condition (3) est étendue au cas du principe d'invariance.

Théorème (Dedecker et Rio, 2000). *Soit $(\Omega, \mathcal{A}, \mu, T)$ un système dynamique ergodique. Soit $X \in L^2$ une variable aléatoire centrée $\mathcal{F}_0$ -mesurable et $X_i = X \circ T^i$. Si (3) a lieu, alors $\{\frac{1}{\sqrt{n}}S_{[nt]}, t \in [0, 1]\}$ converge en loi vers un mouvement brownien.*

En 2005, Peligrad et Utev [76] ont établi une nouvelle inégalité maximale qui permet de généraliser le résultat de Maxwell et Woodroffe.

Théorème (Peligrad et Utev, 2005). *Soit $(\Omega, \mathcal{A}, \mu, T)$ un système dynamique ergodique. Si $X \in L^2$ est une variable aléatoire centrée $\mathcal{F}_0$ -mesurable telle que (4) ait lieu, alors $\{\frac{1}{\sqrt{n}}S_{[nt]}, t \in [0, 1]\}$ converge en loi vers un mouvement brownien.*

Ce résultat reste vrai dans le cas non adapté en ajoutant la condition (5), voir Volný [87]. De plus, la condition (4) est optimale dans le sens où pour toute suite décroissante $(a_n)_{n \in \mathbb{N}}$ tendant vers zéro, il existe une variable aléatoire $X \in L^2$ vérifiant (4) telle que la suite $\left(\frac{S_n}{\sqrt{n}}\right)_{n \geq 1}$ n'est pas stochastiquement bornée (voir Peligrad et Utev [76]).

Encore une fois, les résultats précédents s'étendent au cas non ergodique, le processus limite pouvant alors être un mélange de processus gaussiens.

Les quatre derniers résultats cités fournissent donc quatre critères permettant de montrer qu'un processus stationnaire vérifie un principe d'invariance faible. Dans le premier chapitre de cette thèse, nous avons voulu préciser la relation entre ces différents critères. Ces résultats faisant intervenir des conditions projectives similaires, il est intéressant de savoir si certains sont plus faibles ou plus forts que d'autres. Déjà, dans l'article de Merlevède, Peligrad et Utev [70], les performances des deux théorèmes précédents sont comparées, notamment pour des suites α -mélangeantes et des suites ρ -mélangeantes. En fait, le chapitre 1 de cette thèse permet de montrer qu'en toute généralité les quatre critères sont indépendants. Nous y établissons le théorème suivant.

Théorème (1.4). *Dans tout système dynamique d'entropie positive, quelque soit le critère parmi (6), (2), (3) et (4), il existe une variable aléatoire dans L^2 vérifiant seulement ce critère, et il existe une variable aléatoire dans L^2 vérifiant seulement les autres critères.*

Nous montrons ceci en donnant des exemples différents pour chaque critères. Il est à noter que tous nos exemples correspondent à des variables aléatoires adaptées à la filtration donnée. Les quatre conditions restent donc indépendantes si l'on se restreint au cas adapté.

Dans la suite, nous étudierons un autre théorème limite important qui concerne la fonction de répartition empirique.

Fonction de répartition empirique.

En statistiques inférentielles, il est souvent question de la fonction de répartition empirique. Le théorème de Glivenko-Cantelli montre que cette fonction peut fournir une approximation de la fonction de répartition et permet de construire des tests statistiques. Soit $(X_i)_{i \in \mathbb{N}}$ un processus aléatoire stationnaire à valeurs réelles défini sur un espace probabilisé $(\Omega, \mathcal{A}, \mu)$. On note $F(t) = \mu(X_0 \leq t)$, $t \in \mathbb{R}$, la fonction de répartition commune des X_i . Pour chaque entier n , la fonction de répartition empirique de l'échantillon $X_0, \dots, X_{n-1}$ est donnée par

$$F_n(t) = \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{1}_{\{X_i \leq t\}}, \quad t \in \mathbb{R}.$$

Théorème (Glivenko-Cantelli, 1933). *Soit $(X_i)_{i \in \mathbb{N}}$ une suite de variables aléatoires réelles i.i.d., alors avec probabilité 1,*

$$D_n := \sup_{t \in \mathbb{R}} |F_n(t) - F(t)| \xrightarrow[n \rightarrow \infty]{} 0.$$

Notons que le théorème de Glivenko-Cantelli reste vrai pour les suites stationnaires ergodiques réelles (voir par exemple Dehling et Philipp [21], Théorème 1.1).

Processus empiriques.

Le processus empirique associé à la suite $(X_i)_{i \in \mathbb{N}}$ est donné par

$$W_n(t) = \sqrt{n}(F_n(t) - F(t)) = \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} (\mathbb{1}_{\{X_i \leq t\}} - F(t)), \quad t \in \mathbb{R}.$$

Le processus aléatoire $\{W_n(t), t \in \mathbb{R}\}$ est un processus stationnaire centré. Lorsque la fonction de répartition est suffisamment régulière, en utilisant sa fonction pseudo-inverse (voir section 2.2.3), on peut se ramener au cas où F est la distribution uniforme sur $[0, 1]$. Le processus empirique associé sera alors appelé processus empirique uniforme. Le principe d'invariance pour le processus empirique dans le cas i.i.d. est encore une fois dû à Donsker [29], qui justifie une conjecture énoncée en 1949 par Doob [30].

Théorème (Donsker, 1952). *Soit $(X_i)_{i \in \mathbb{N}}$ une suite i.i.d. de variables aléatoires uniformément réparties sur $[0, 1]$, alors le processus $\{W_n(t), t \in [0, 1]\}$ converge en loi dans l'espace $D[0, 1]$ vers un pont brownien standard $W(t)_{0 \leq t \leq 1}$, i.e. un processus gaussien de moyenne nulle, à trajectoires continues et de covariances $\text{Cov}(W(s), W(t)) = s \wedge t - st$, $s, t \in [0, 1]$.*

C'est en fonction de ce résultat qu'est construit le test d'adéquation non-paramétrique basé sur la statistique D_n , développé par Kolmogorov et Smirnov (test KS). Le processus empirique joue donc un rôle central en statistique non-paramétrique. De nombreux autres résultats dans le cas indépendant ont été obtenus ainsi que de nombreuses applications. On pourra voir les ouvrages de Pollard [79] et de Van der Vaart et Wellner [83]. Encore une fois, nous allons nous intéresser aux extensions de ce résultat au cas des variables dépendantes.

Les premiers travaux dans ce sens figurent dans le livre de Billingsley [6], où un principe d'invariance empirique pour des fonctions de certaines suites φ -mélangeantes est énoncé (voir [6] Théorème 22.2). Une application remarquable concerne la transformation T de l'intervalle $[0, 1]$ définie par $Tx = 2x \mod 1$. Si X est une variable uniformément répartie sur $[0, 1]$, alors le processus empirique associé à la suite stationnaire $(X \circ T^n)_{n \in \mathbb{N}}$ converge en loi vers un processus gaussien. Le même résultat est établi pour la transformation T_1 , relative au développement en fraction continue, définie par $T_1x = \frac{1}{x} - \left\lfloor \frac{1}{x} \right\rfloor$. D'autres théorèmes pour d'autres conditions de dépendance faible ont été démontrés plus tard par plusieurs auteurs. Pour une synthèse de tous ces résultats, on pourra voir l'ouvrage de Dehling et Philipp [21] ainsi que les références s'y trouvant. Un des résultats figurant dans cet ouvrage concerne les processus absolument réguliers. Le principe d'invariance pour ce type de processus, établi par Borovkova, Burton et Dehling [8], s'applique en particulier aux suites de variables aléatoires issues d'un système dynamique hyperbolique.

Systèmes uniformément hyperboliques.

En théorie des systèmes dynamiques, on parle de système hyperbolique lorsque le système présente à la fois des propriétés de contraction et d'expansion. Soit M une variété riemannienne compacte et $f : M \rightarrow M$ un C^∞ -difféomorphisme. On dit que f est un difféomorphisme d'Anosov (ou uniformément hyperbolique) si $f(M) = M$ et si pour chaque point x de M , l'espace tangent $T_x M$ s'écrit comme somme directe

$$T_x M = E_x^s \oplus E_x^u$$

où E_x^s et E_x^u varient continûment avec x et $D_x f$ est une contraction sur E_x^s tandis que c'est une dilatation sur E_x^u (dont les constantes sont uniformes en x). Les espaces E_x^s et E_x^u sont respectivement appelés sous-espace stable et sous-espace instable. Grâce au travail de

Bowen [9], nous savons que, d'une manière générale, les systèmes d'Anosov peuvent être codés sous forme de processus absolument réguliers. Nous pouvons donc, en appliquant le théorème de Borovkova et al. [8], obtenir un principe d'invariance pour les processus empirique associés à des systèmes d'Anosov, voir section 2.2.1.

Théorème. *Soit $(M, \mathcal{M}, \mu, f)$ un système d'Anosov et $g : M \rightarrow [0, 1]$ une fonction mesurable hölderienne dont la fonction de répartition est lipschitzienne. Alors le processus empirique associé à $(g \circ f^i)_{i \in \mathbb{N}}$ converge en loi vers un processus gaussien centré $(Y(t))_{0 \leq t \leq 1}$ de covariance*

$$\text{Cov}(Y(s), Y(t)) = \sum_{i=1}^{\infty} \text{Cov}(\mathbb{1}_{[0,s]}(g \circ T^i), \mathbb{1}_{[0,t]}(g \circ T)) + \sum_{i=2}^{\infty} \text{Cov}(\mathbb{1}_{[0,s]}(g \circ T), \mathbb{1}_{[0,t]}(g \circ T^i)).$$

Systèmes partiellement hyperboliques.

En marge du cadre uniformément hyperbolique, les systèmes partiellement hyperboliques font l'objet de récentes études (voir Hasselblatt et Pesin [53]). Le terme partiellement hyperbolique signifie que la transformation possède une partie uniformément hyperbolique mais qu'il reste une direction suivant laquelle elle n'est ni une contraction ni une expansion. Une grande partie des théorèmes limites connus dans le cas hyperbolique restent valables dans le cas partiellement hyperbolique. Ce qui est en partie démontré par Dolgopyat [27]. La question qui se pose à nous est celle du principe d'invariance empirique.

Des exemples simples de difféomorphismes d'Anosov sont les automorphismes du tore qui n'ont que des valeurs propres de module différent de 1 (appelé automorphismes hyperboliques du tore, voir les travaux de Katznelson [59]). On parlera d'automorphisme quasi-hyperbolique du tore lorsque la transformation possède des valeurs propres de module 1 (à ce propos on pourra voir l'article de Lind [66] et celui de Le Borgne [62] pour la question du TLC). Dans cette situation, il n'est plus question de se ramener au cas des processus absolument réguliers. Dans le chapitre 2 de cette thèse, nous allons montrer que néanmoins un principe d'invariance empirique a lieu pour les automorphismes quasi-hyperboliques. Nous établissons le théorème suivant.

Théorème (2.2). *Soit T un automorphisme ergodique du tore $\mathbb{T}^d$ de dimension $d \geq 2$. Soit $f : \mathbb{T}^d \rightarrow \mathbb{R}$ une fonction borélienne, bornée, dont la fonction de répartition est continue et vérifiant*

$$\exists C, \xi > 0 \text{ tel que } \forall t \in \mathbb{R} \text{ et } \forall \varepsilon > 0, \quad \text{Lebesgue}(\partial_\varepsilon \{f \leq t\}) \leq C\varepsilon^\xi. \quad (*)$$

où $\partial_\varepsilon A$ désigne le bord de l'ensemble A élargi d'une distance ε . Alors le processus empirique associé à $(f \circ T^i)_{i \in \mathbb{N}}$ converge en loi vers un processus gaussien centré.

Dans la suite du chapitre 2, la condition $(*)$ est étudiée. Nous montrons qu'elle est plus générale que celle du théorème précédent car satisfaite dès que la fonction f et sa fonction de répartition sont hölderiennes. De plus, elle est vérifiée de façon générique parmi les fonctions de classe C^k , $k \geq 2$ (voir Corollaire 2.17). Pour démontrer cela nous passons par l'étude de la régularité des fonctions de répartition des fonctions de Morse définies sur une variété riemannienne compacte. Nous rappelons qu'une fonction de Morse est une fonction de classe au moins C^2 , dont les singularités sont isolées et non dégénérées. Nous obtenons le résultat général suivant.

Théorème (2.16). *Soit M une variété riemannienne compacte munie de sa mesure de probabilité naturelle. Soit $f : M \rightarrow \mathbb{R}$ une fonction de Morse et F sa fonction de répartition.*

Si $\dim M = 1$ ou 2 , la fonction de répartition F est $\frac{1}{2}$ -hölderienne et C^1 en dehors des singularités de f .

Si $\dim M \geq 3$, la fonction de répartition F est C^1 .

Technique de démonstration.

Dans le chapitre 2, nous utilisons pour montrer le théorème 2.2 une technique habituelle pour obtenir un principe d'invariance. La preuve s'effectue en deux temps. Tout d'abord on établit la convergence finie-dimensionnelle du processus puis on montre que le processus est tendu. La première partie découle en générale du TLC pour le processus $(W_n(t))_{n \geq 1}$ à t fixé. La tension est le plus souvent établie par une méthode de chaînage qui repose sur une inégalité de moment d'ordre quatre (ou plus) des accroissements du processus empirique $W_n(s) - W_n(t)$. Cette technique est utilisée dans la plupart des articles concernant le principe d'invariance pour les processus empiriques, le point difficile étant souvent l'obtention de l'inégalité de moment. En 2004, Collet, Martinez et Schmitt [13] ont démontré un principe d'invariance pour une classe d'applications dilatantes de l'intervalle. En s'appuyant sur le travail de Broise [10], qui donne un trou spectral pour l'opérateur de transfert agissant sur l'espace des fonctions à variation bornée, ils obtiennent l'inégalité de moment d'ordre quatre voulue. Pour ce faire, on utilise le fait que les fonctions indicatrices $\mathbb{1}_{[0,t]}$, $0 \leq t \leq 1$ sont des fonctions à variation bornée.

Trou spectral.

On dit que le système $(\Omega, \mathcal{A}, \mu, T)$ admet un trou spectral sur un espace de fonctions intégrables $\mathcal{B}$ si l'opérateur de transfert $P : L^1 \rightarrow L^1$, défini par la relation

$$\int P f \cdot g \, d\mu = \int f \cdot g \circ T \, d\mu, \quad f \in L^1, g \in L^\infty,$$

conserve l'espace $\mathcal{B}$ et possède sur $\mathcal{B}$ une unique valeur propre de module 1, qui est simple, tandis que le reste de son spectre est contenu dans un disque de rayon strictement inférieur à 1. Classiquement, cette propriété de trou spectral permet aussi d'obtenir un TLC (voir Rousseau-Egele [81], Guivarc'h et Hardy [48] ou Baladi [2]), ici pour les fonctions à variation bornée, duquel découle le TLC fini-dimensionnel. Cependant, il existe de nombreuses situations où un trou spectral s'obtient plus naturellement pour un espace de fonctions plus petit (voir Hennion et Hervé [54]), et en particulier ne contenant pas les indicatrices. L'exemple de Gouëzel [47] montre qu'il existe des transformations dilatantes de l'intervalle qui présentent un trou spectral sur l'espace des fonctions lipschitziennes mais pas sur celui de fonctions à variation bornée. Dans une telle situation, on ne peut pas directement en déduire une convergence finie-dimensionnelle pour le processus empirique.

Nouvelle technique.

Dans le chapitre 3, nous proposons une autre méthode, qui reste basée sur la technique de chaînage, pour obtenir un principe d'invariance. Supposons, par exemple, que notre système admet un trou spectral sur l'espace $\mathcal{L}$ des fonctions lipschitziennes. Alors, nous obtenons l'inégalité de moment d'ordre 4 suivante (voir section 3.5), qui n'est pas directement utilisable pour la technique de chaînage usuelle.

Théorème (3.13). *Pour toute fonction $f \in \mathcal{L}$ telle que $E_\mu(f) = 0$ et $\|f\|_\infty \leq 1$,*

$$E_\mu \left[\left(\sum_{i=1}^n f \circ T^i \right)^4 \right] \leq K \left[n \|f\|_1 \log^3(\|f\| + 1) + n^2 \|f\|_1^2 \log^2(\|f\| + 1) \right],$$

$$\text{où } \|f\| = \|f\|_\infty + \sup_{x \neq y} \frac{|f(x) - f(y)|}{|x - y|}.$$

Nous développons donc dans la première partie du chapitre 3 une méthode basée sur une approximation du processus empirique $\{W_n(t), t \in [0, 1]\}_{n \geq 1}$ par une suite de processus $\{W_n^{(m)}(t), t \in [0, 1]\}_{n, m \geq 1}$ construits à partir de fonctions lipschitziennes. Lorsque l'on veut approcher une fonction indicatrice par une suite de fonctions lipschitziennes, la norme lipschitzienne est d'autant plus grande que l'approximation est bonne. L'intérêt de l'inégalité ci-dessus est que la norme lipschitzienne n'apparaît qu'au travers de son logarithme. Notre méthode repose sur le théorème suivant, qui correspond à une version pour les espaces complets du théorème 4.2 de Billingsley [6].

Théorème (3.3). *Soit (S, ρ) un espace métrique séparable et complet et $X_n, X_n^{(m)}$ et $X^{(m)}$, $n, m \geq 1$ des variables aléatoires à valeurs dans S vérifiant*

$$\begin{aligned} X_n^{(m)} &\xrightarrow{\mathcal{D}} X^{(m)} \text{ quand } n \rightarrow \infty, \forall m \\ \lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} P(\rho(X_n, X_n^{(m)}) \geq \varepsilon) &= 0, \forall \varepsilon > 0. \end{aligned}$$

Alors il existe une variable aléatoire X à valeurs dans S telle que

$$X_n \xrightarrow{\mathcal{D}} X \text{ quand } n \rightarrow \infty.$$

Afin de montrer que la deuxième assertion est réalisée, nous adaptons la méthode de chaînage en utilisant seulement des fonctions lipschitziennes et notre inégalité précédente. Le résultat obtenu est un nouveau principe d'invariance, Théorème 3.1. Bien que la méthode soit nouvelle, ce théorème fournit des applications comparables, par exemples, aux résultats de Dedecker et Prieur [18] ou Wu et Shao [91]. Ces applications sont discutées en section 3.6.

Sommes de fonctions indicatrices.

Ce travail concernant la fonction de répartition empirique, nous amène à travailler avec des sommes de fonctions indicatrices. De ce fait, nous avons voulu étudier la question du TLC pour les fonctions indicatrices. Dans le chapitre 4, nous donnons des résultats concernant les sommes de fonctions indicatrices dans les systèmes dynamiques.

En 1987, Burton et Denker [11] ont montré que tout système dynamique apériodique $(\Omega, \mathcal{A}, \mu, T)$ admet un processus $(X \circ T^i)_{i \in \mathbb{N}}$ qui vérifie le TLC. Ils montrèrent également que l'ensemble des variables $X \in L^2$ qui vérifient cette propriété est dense dans l'espace L_0^2 des fonctions de carré intégrable telles que $E(X|\mathcal{I}) = 0$. En 1990, Volný [84] a montré que néanmoins, il existe un G_δ , dense dans L_0^2 , de fonctions X telles que la suite des distributions de $\left(\frac{S_n}{\sqrt{n}}\right)_{n \in \mathbb{N}}$ est dense dans l'espace des probabilités sur $\mathbb{R}$. Ce résultat reste vrai si l'on remplace la normalisation en $\frac{1}{\sqrt{n}}$ par une normalisation en $\frac{1}{a_n}$ pour n'importe quelle suite croissante $(a_n)_{n \in \mathbb{N}}$ telle que $a_n \nearrow \infty$ et $\frac{a_n}{n} \rightarrow 0$ quand n tend vers l'infini. Remarquons que l'on obtient ainsi une nouvelle preuve de la convergence arbitrairement lente dans le théorème ergodique.

Ici, nous montrons que l'on peut obtenir un résultat similaire pour les fonctions $\mathbb{1}_A - \mu(A)$, $A \in \mathcal{A}$. Sur $\mathcal{A}$, on considère la pseudo-métrique définie par la mesure de la différence symétrique. Le théorème démontré au chapitre 4 est le suivant.

Théorème (4.1). *Soit $(\Omega, \mathcal{A}, \mu, T)$ un système dynamique apériodique et ergodique et $(a_n)_{n \geq 1} \subset \mathbb{R}_+$ une suite croissante telle que $a_n \nearrow \infty$ et $\frac{a_n}{n} \rightarrow 0$ quand $n \rightarrow \infty$. Il existe un G_δ dense composé d'ensembles $A \in \mathcal{A}$ ayant la propriété suivante :*

Pour toute probabilité ν sur $\mathbb{R}$, il existe une sous-suite $(n_k)_{k \in \mathbb{N}}$ telle que

$$\frac{1}{a_{n_k}} S_{n_k}(\mathbb{1}_A - \mu(A)) \xrightarrow[k \rightarrow \infty]{\mathcal{D}} \nu.$$

On donnera également une construction explicite d'un tel ensemble, qui restera valable dans le cas non ergodique.

Commentaires.

La thèse est rédigée de manière à ce que les quatre chapitres puissent être lus de manière indépendante. L'annexe concerne une proposition du chapitre 2. L'essentiel de la thèse reprend les résultats des articles [20], [36], [37], [38], [39], [40], qui ont été conservés dans leur langue d'origine.

Chapitre 1

On Projective Criteria leading to the Weak Invariance Principle

Abstract

The aim of this chapter is to compare four projective criteria leading to the central limit theorem and to the Donsker's weak invariance principle. These criteria are

- the martingale-coboundary decomposition developed by Gordin (see [44], [45]);
- the projective criterion introduced by Dedecker and Rio (see [15], [19]);
- the Maxwell and Woodroffe condition (see [69], [76], [86], [87]);
- the Hannan condition (see [52], [16], [17]).

We prove that in every ergodic dynamical system with positive entropy, these four criteria are independent in the following sense: if we consider one of these criteria, we can find a function in L^2 which satisfies this one but not the three others or the inverse.

Keywords: Stationary process; Central limit theorem; Weak invariance principle; Martingale approximation; Projective criterion.

AMS Classification: 60F05, 60F17, 60G10, 28D05, 60G42.

1.1 Introduction

Let $(\Omega, \mathcal{A}, \mu)$ be a probability space and $T : \Omega \longrightarrow \Omega$ a bijective bimeasurable transformation preserving the measure μ (i.e. $\mu(T^{-1}A) = \mu(A)$, $\forall A \in \mathcal{A}$). We will assume that it is ergodic, i.e. $T^{-1}A = A$ implies $\mu(A) = 0$ or $\mu(A) = 1$.

Let f be a measurable function defined on Ω , then the sequence of random variables $(f \circ T^i)_{i \in \mathbb{Z}}$ is a stationary process. Note that, on the other hand, for every stationary random process $(X_i)_{i \in \mathbb{Z}}$, there exist a dynamical system $(\Omega, \mathcal{A}, \mu, T)$ and a function f on Ω such that the processes $(f \circ T^i)_{i \in \mathbb{Z}}$ and $(X_i)_{i \in \mathbb{Z}}$ have the same distribution (see, e.g., Cornfeld et al. [14] p. 178).

For all $n \geq 1$, let

$$S_n(f) = \sum_{i=0}^{n-1} f \circ T^i.$$

We say that the function f satisfies the *Central Limit Theorem* (CLT) if $\frac{1}{\sqrt{n}}S_n(f)$ converges in distribution to a normal law.

For all $n \geq 1$ and $t \in [0, 1]$, let

$$S_n(f, t) = S_{[tn]}(f) = \sum_{i=0}^{[tn]} f \circ T^i$$

where $[x]$ denotes the greatest integer that is smaller than x . We say that the function f satisfies the *Weak Invariance Principle* (also called Donsker's invariance principle) if the process $\{\frac{1}{\sqrt{n}}S_n(f, t) / t \in [0, 1]\}$ converges in distribution to some Gaussian process in the space $D[0, 1]$ of càdlàg functions equipped with the Skorohod topology (see Billingsley [6]). In the sequel, invariance principle always designates the weak invariance principle.

These two limit theorems are known to hold in the independent case and have been extensively studied also in dependent cases. Several methods of proving them have been developed. In the sequel, we restrict our attention to four of them, which present some projective conditions.

Let $\mathcal{F}$ be a sub- σ -algebra of $\mathcal{A}$ such that $T^{-1}\mathcal{F} \supset \mathcal{F}$. We denote by $\mathcal{F}_i$ the σ -algebra $T^{-i}\mathcal{F}$. The function f is called *regular* with respect to the filtration $(\mathcal{F}_i)_{i \in \mathbb{Z}}$ if

$$E(f|\mathcal{F}_{-\infty}) = 0 \text{ and } E(f|\mathcal{F}_{+\infty}) = f.$$

We assume that f is a square integrable zero-mean function and we write L^p for $L^p(\mu)$, $p \geq 1$.

1.1.1 Martingale-coboundary decomposition

This method of proving the CLT was first used by Gordin [44]. The idea is to represent f in the form

$$f = m + g - g \circ T,$$

where $(m \circ T^i)_{i \in \mathbb{Z}}$ is a martingale difference sequence. The term $g - g \circ T$ is called a coboundary and g is the transfer function. This decomposition is called a martingale-coboundary decomposition.

If $m \in L^2$, the CLT for martingale differences of Billingsley [5] and Ibragimov [58] applies. If g is measurable, the telescopic sum $\frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} (g - g \circ T) \circ T^i$ goes to zero in probability. So, if we can find the above decomposition with $m \in L^2$ and g measurable, the CLT holds for f by application of theorem 4.1 of Billingsley [6]. Moreover, in this case, if the transfer function g belongs to L^2 , we also have the invariance principle, as proved by Heyde [56] (see also Hall and Heyde [50]).

On the other hand, there exist counterexamples with $g \in L^1$ and $g - g \circ T \in L^2$ where the invariance principle does not hold, see Volný and Samek [88]. According to Hall and Heyde [50] (see also Volný and Samek [88]), if $m \in L^2$, a necessary and sufficient condition to have the invariance principle is

$$\frac{1}{\sqrt{n}} \max_{i \leq n} |g \circ T^i| \xrightarrow{n \rightarrow \infty} 0 \text{ in probability.} \quad (1.1)$$

We say that the process $(f \circ T^i)_{i \in \mathbb{Z}}$ (or the function f) admits a *martingale-coboundary decomposition in L^p* , $p \geq 1$, if m and g are in L^p . If the function f is regular with respect to $(\mathcal{F}_i)_{i \in \mathbb{Z}}$, then f admits a martingale-coboundary decomposition in L^p with the martingale difference $(m \circ T^i)_{i \in \mathbb{Z}}$ with respect to the filtration $(\mathcal{F}_i)_{i \in \mathbb{Z}}$ if and only if the series

$$\sum_{i=0}^{\infty} E(f \circ T^i | \mathcal{F}_0) \quad \text{and} \quad \sum_{i=0}^{\infty} (f \circ T^{-i} - E(f \circ T^{-i} | \mathcal{F}_0))$$

converge in L^p , see Volný [85]. Remark that if the process $(f \circ T^i)_{i \in \mathbb{Z}}$ is adapted to $(\mathcal{F}_i)_{i \in \mathbb{Z}}$, the second sum is equal to zero.

According to what precedes, the existence of a martingale-coboundary decomposition in L^p , with $p \geq 2$ implies the invariance principle. If $p = 1$, $m \in L^2$ and condition (1.1) holds, we also have the invariance principle. In the following, as we will see, we almost only study the martingale-coboundary decomposition in L^1 . For proofs in this situation, see Gordin [45] and Esseen and Janson [43]. To summarize, the martingale-coboundary

decomposition in L^1 holds if and only if

$$\sum_{i=0}^{\infty} E(f \circ T^i | \mathcal{F}_0) \text{ and } \sum_{i=0}^{\infty} (f \circ T^{-i} - E(f \circ T^{-i} | \mathcal{F}_0)) \text{ converge in } L^1 \quad (\text{M-C})$$

and in the adapted case, the second sum vanishes (so, it clearly converges).

This method gives results in various situations. An interesting example is its application to differentiable dynamical system. It is well adapted to the hyperbolic case (e.g. Liverani [68]), or the partially hyperbolic case (e.g. Le Borgne [62]).

1.1.2 Dedecker-Rio criterion

Another method is to establish a projective property developed by Dedecker [15]. He introduced this criterion to prove CLT for random fields. Dedecker and Rio [19] have shown that it gives a powerful criterion for proving the invariance principle (in dimension one). We say that the sequence $(f \circ T^i)_{i \in \mathbb{Z}}$ (or the function f) satisfies the *Dedecker-Rio criterion* if

$$\sum_{k=1}^{\infty} f E(f \circ T^k | \mathcal{F}_0) \text{ converges in } L^1. \quad (\text{D-R})$$

According to Dedecker and Rio [19], if $f \in L^2$ is adapted to the filtration $(\mathcal{F}_i)_{i \in \mathbb{Z}}$ and (D-R) holds, then f satisfies the invariance principle.

1.1.3 Maxwell-Woodroffe condition

We say that the sequence $(f \circ T^i)_{i \in \mathbb{Z}}$ (or the function f) satisfies the *Maxwell-Woodroffe condition* if

$$\sum_{n=1}^{\infty} \frac{\|E(S_n(f) | \mathcal{F}_0)\|_2}{n^{\frac{3}{2}}} < \infty. \quad (\text{M-W})$$

This criterion was first introduced by Maxwell and Woodroffe [69]. In the adapted case, they proved that the CLT holds under this condition and Peligrad and Utev [76] proved that this condition implies the weak invariance principle. In the general case, the weak invariance principle holds as soon as (M-W) and

$$\sum_{n=1}^{\infty} \frac{\|S_n(f) - E(S_n(f) | \mathcal{F}_n)\|_2}{n^{\frac{3}{2}}} < +\infty,$$

(see Volný [86], [87]).

For example of applications of the last two methods, the reader can see the survey by Merlevède, Peligrad and Utev [70].

1.1.4 Hannan criterion

Let us denote by $H_k = L^2(\mathcal{F}_k)$ the space of $\mathcal{F}_k$ -measurable functions which are square integrable and denote by P_k the orthogonal projection operator onto the space $H_k \ominus H_{k-1}$. For $f \in L^2$,

$$P_k(f) = E(f|\mathcal{F}_k) - E(f|\mathcal{F}_{k-1}).$$

Let f be a regular function for the filtration $(\mathcal{F}_i)_{i \in \mathbb{Z}}$. As a consequence of a result given by Heyde [55] (see Volný [85] Theorem 6) the CLT holds as soon as

$$\sum_{i \in \mathbb{Z}} \|P_0(f \circ T^i)\|_2 < \infty. \quad (\text{HC})$$

In the adapted case, this result and the weak invariance principle were proved by Hannan [51], [52] under the assumption that T is weakly mixing. Hannan's weak invariance principle was proved without the extra assumption by Dedecker and Merlevède [16], Corollary 3. Finally, in the general case, the weak invariance principle under (HC) is due to Dedecker, Merlevède and Volný [17], Corollary 2.

Our purpose is to compare the dependence between these four criteria. Section 1.2 contains the statement of our main result while the remainder of the chapter is devoted to its proof. Section 1.3 and Section 1.4 present a general type of a suitable function in a dynamical system. In Section 1.5, this model is used to produce specific counterexamples proving our result. In Section 1.6, we adapt our model to get the counterexamples of Section 1.7.

1.2 Comparison of the criteria

It is of interest to know whether one of the considered criteria implies another. This is the question that we propose to answer. First, note that a simple application of the Hölder inequality

$$\left\| \sum f E(f \circ T^k | \mathcal{F}_0) \right\|_1 \leq \left\| \sum E(f \circ T^k | \mathcal{F}_0) \right\|_p \|f\|_q \quad (1.2)$$

with $\frac{1}{p} + \frac{1}{q} = 1$, leads to the following remarks.

Remark 1.1. *The martingale-coboundary decomposition in L^2 implies the Dedecker-Rio criterion.*

Proof.

It follows from the convergence of $\left(\sum_{k=1}^n E(f \circ T^k | \mathcal{F}_0) \right)_{n \geq 1}$ in L^2 and the inequality (1.2) with $p = q = 2$. □

Remark 1.2. *The martingale-coboundary decomposition in L^2 also implies the Maxwell-Woodroffe condition.*

So we are interested in the martingale-coboundary decomposition in L^1 . The same kind of arguments show

Remark 1.3.

- (a) *For bounded functions, the martingale-coboundary decomposition in L^1 implies the Dedecker-Rio criterion.*
- (b) *For a function f such that $|f| > C > 0$, the inverse implication is true.*

Proof.

(a) follows from application of (1.2) with $p = 1$ and $q = \infty$. For (b), it is enough to note that $\frac{1}{f}$ is a bounded function. $\square$

We will see, by counterexamples in L^2 , that in general, the martingale-coboundary decomposition in L^1 , the Dedecker-Rio criterion, the Maxwell-Woodroffe condition and the Hannan criterion do not result from each other, even if the function verifies the CLT or the invariance principle. Clearly, for the example given by Volný and Samek [88], which satisfies the martingale-coboundary decomposition in L^1 but not the invariance principle, the three other criteria do not hold. But, we can also find counterexamples in the class of functions satisfying the invariance principle. Our main result is the following.

Theorem 1.4. *The martingale-coboundary decomposition in L^1 (M-C), the Dedecker-Rio criterion (D-R), the Maxwell-Woodroffe condition (M-W) and the Hannan criterion (HC) are independent:*

in all ergodic dynamical system with positive entropy, for each condition among the four, there exists an L^2 -function satisfying it but not the three others and there exists an L^2 -function satisfying the three others but not this one.

In particular, the four criteria are pairwise independent.

In other words, in each case, there exists a function in L^2 satisfying

- (i) (M-C) and the invariance principle but neither (D-R), (M-W) nor (HC);
- (ii) (D-R) but neither (M-C), (M-W) nor (HC);
- (iii) (M-W) but neither (M-C), (D-R) nor (HC);
- (iv) (HC) but neither (M-C), (D-R) nor (M-W);

- (v) (M-C), (D-R) and (M-W) but not (HC);
- (vi) (M-C), (D-R) and (HC) but not (M-W);
- (vii) (M-C), (M-W) and (HC) but not (D-R);
- (viii) (D-R), (M-W) and (HC) but not (M-C).

To prove Theorem 1.4, we shall give eight examples of functions in L^2 satisfying (i)-(viii). These functions will be defined in the same way, so we begin by a general construction. The first step is to choose disjoint sets having a nice property. Next section is devoted to the exposition of the construction of these sets.

1.3 A preliminary result

Let $\mathcal{C}$ be a sub- σ -algebra of $\mathcal{A}$ such that $T^{-1}\mathcal{C} = \mathcal{C}$. We assume that the measure μ restricted to $\mathcal{C}$ is non-atomic. The goal is to establish Lemma 1.7 corresponding to the construction of disjoint sets A_k quasi-invariant under a finite number of iterations of the transformation. Moreover, we want to control the measure of the A_k . First, we recall the following lemma. A proof can be found in the article by del Junco and Rosenblatt [22], as a particular case of theorem 2.2. It can also be done directly by using Rokhlin Lemma.

Lemma 1.5. *Let $N \in \mathbb{N}$, $0 < \rho < 1$ and $\varepsilon > 0$. There exists a set $A \in \mathcal{C}$ such that $\mu(A) = \rho$ and for all $i, j \in \{0, \dots, N\}$,*

$$\mu(T^{-i}A \Delta T^{-j}A) \leq \varepsilon.$$

Proof.

We choose $d \leq \frac{\varepsilon}{2(N+1)}$ such that $M = \frac{\rho}{d} \in \mathbb{N}$. We can find a Rokhlin tower $\{F, T^{-1}F, \dots, T^{-N'}F\}$ with $N' \gg N$, such that $F \in \mathcal{C}$ and $\mu(F) = d$. Set

$$A = F \cup T^{-1}F \cup \dots \cup T^{-M+1}F.$$

Thus

$$\mu(A) = Md = \rho$$

and if $i, j \in \{0, \dots, N\}$ then

$$\mu(T^{-i}A \Delta T^{-j}A) \leq (N+1)2\mu(F) \leq \varepsilon.$$

□

Remark 1.6. Let A_1, A_2, B_1, B_2 be mesurable sets and $\varepsilon_1, \varepsilon_2 > 0$. If $\mu(A_1 \triangle A_2) \leq \varepsilon_1$ and $\mu(B_1 \triangle B_2) \leq \varepsilon_2$, then $\mu((A_1 \setminus B_1) \triangle (A_2 \setminus B_2)) \leq \varepsilon_1 + \varepsilon_2$.

We are going to use this remark as well as Lemma 1.5 to show the following key lemma.

Lemma 1.7. Let $(N_k)_{k \in \mathbb{N}} \subset \mathbb{N}$ with $N_k \nearrow \infty$, $(\rho_k)_{k \in \mathbb{N}}$ be a decreasing sequence of positive reals such that $\sum_{k \in \mathbb{N}} \rho_k < 1$ and $(\varepsilon_k)_{k \in \mathbb{N}}$ be a strictly decreasing sequence of positive reals converging to zero. Choose $0 < a < 1 - \sum_{k \in \mathbb{N}} \rho_k$.

There exists a sequence of sets $(A_k)_{k \in \mathbb{N}} \subset \mathcal{C}$ such that

- (1) the sets A_k are mutually disjoint;
- (2) $a\rho_k \leq \mu(A_k) \leq \rho_k$, for all $k \geq 0$;
- (3) for all $k \geq 1$ and for all $i, j \in \{0, \dots, N_k\}$, $\mu(T^{-i}A_k \triangle T^{-j}A_k) \leq \varepsilon_k$.

Proof.

First, set $\delta_k = \varepsilon_k - \varepsilon_{k+1}$, $k \geq 0$, $\lambda = \sum_{k \in \mathbb{N}} \rho_k < 1$ and $\alpha = \frac{1 - \lambda - a}{\lambda} \in (0, 1)$.

For $k = 0$, by Lemma 1.5, there exists a set $A'_0 \in \mathcal{C}$ such that $\mu(A'_0) = \rho_0$ and $\mu(T^{-i}A'_0 \triangle T^{-j}A'_0) \leq \delta_0$ for all $i, j \in \{0, \dots, N_0\}$.

We will proceed by induction. Assume there exist sets $A'_0, A'_1, \dots, A'_k \in \mathcal{C}$ such that

- (a) for all $l \in \{0, \dots, k\}$, $\mu(A'_l) = \rho_l$;
- (b) for all $l \in \{0, \dots, k\}$, for all $i, j \in \{0, \dots, N_l\}$, $\mu(T^{-i}A'_l \triangle T^{-j}A'_l) \leq \delta_l$;
- (c) for all $i \neq j \in \{0, \dots, k\}$, $\mu(A'_i \cap A'_j) \leq (1 + \alpha)\mu(A'_i)\mu(A'_j)$.

By Birkoff's ergodic theorem, there exist a set $E \in \mathcal{C}$ with $\mu(E) \geq \frac{1}{2}$ and an integer $N \geq 1$ such that for all $j \in \{0, \dots, k\}$, for all $\omega \in E$, for all $n \geq N$,

$$\frac{1}{n} \left| S_n(\mathbb{1}_{A'_j} - \mu(A'_j))(\omega) \right| \leq \alpha \rho_j. \quad (1.3)$$

Let $\bar{N} = \max\{N_{k+1}, N\}$ and $d \leq \frac{\delta_1}{2\bar{N} + 1}$ such that $M = \frac{\rho_{k+1}}{d} \in \mathbb{N}$. There exists a measurable set $F \subset E$ which is the base of a Rokhlin tower of height at least M such that $\mu(F) = d$. We set

$$A'_{k+1} = F \cup T^{-1}F \cup \dots \cup T^{-M+1}F.$$

Then

$$\mu(A'_{k+1}) = M\mu(F) = \rho_{k+1},$$

for all $i, j \in \{0, \dots, N_{k+1}\}$,

$$\mu(T^{-i}A'_{k+1} \triangle T^{-j}A'_{k+1}) \leq \delta_{k+1}$$

and further, by (1.3), for all $i \in \{0, \dots, k\}$,

$$\begin{aligned} \mu(A'_i \cap A'_{k+1}) &= \mu(F) \int_F S_M(\mathbb{1}_{A'_i}) d\mu \\ &\leq \mu(F)(M\mu(A'_i) + \alpha\rho_i) \\ &\leq (1 + \alpha)\mu(A'_{k+1})\mu(A'_i). \end{aligned}$$

We thus get a sequence of sets $(A'_k)_{k \in \mathbb{N}}$ satisfying (a)-(c) for all $k \geq 0$.

We define

$$A_k = A'_k \setminus \bigcup_{j=k+1}^{\infty} A'_j.$$

Hence, (1) holds by construction. We have $\mu(A_k) \leq \mu(A'_k) = \rho_k$ and

$$\begin{aligned} \mu(A_k) &\geq \mu(A'_k) - \sum_{j=k+1}^{\infty} \mu(A'_k \cap A'_j) \\ &\geq \rho_k - \rho_k(1 + \alpha) \sum_{j=k+1}^{\infty} \mu(A'_j) \\ &= \rho_k(1 - \lambda - \alpha\lambda) \\ &= a\rho_k. \end{aligned}$$

So, (2) is verified.

For (3), we use the preceding remark to have, for all $k \geq 0$, for all $i, j \in \{0, \dots, N_k\}$,

$$\mu(T^{-i}A_k \triangle T^{-j}A_k) \leq \sum_{l=k}^{\infty} \delta_l = \varepsilon_k.$$

□

Note that if the ρ_k 's decrease exponentially fast, we can prove Lemma 1.7 without the help of Birkoff's ergodic theorem (see Durieu and Volný [40]).

An important feature of Lemma 1.7 is that there is no dependence between $(N_k)_{k \in \mathbb{N}}$ and $(\rho_k)_{k \in \mathbb{N}}$ and the ε_k 's can be chosen arbitrarily small.

1.4 General approach

Here, we give the general model from which the counterexamples will be deduced, proving Theorem 1.4. We will define a "pattern function" depending on sequences $(N_k)_{k \in \mathbb{N}} \subset \mathbb{N}$, $(\theta_k)_{k \in \mathbb{N}} \subset \mathbb{R}_+$, $(\rho_k)_{k \in \mathbb{N}} \subset (0, 1)$ and $(\varepsilon_k)_{k \in \mathbb{N}}$ ($\varepsilon_k \ll \rho_k$). In Section 1.5, we will see that changing the values of the sequences provides different counterexamples.

The model:

$(\Omega, \mathcal{A}, \mu, T)$ is an ergodic dynamical system with positive entropy. By Sinai's theorem, it admits a factor which is a Bernoulli shift with the same entropy (see Sinai [82]). So, it is sufficient to consider the case where $(\Omega, \mathcal{A}, \mu, T)$ is a Bernoulli shift with positive entropy.

This means there exist $l \in \mathbb{N}^*$ and $p_i > 0$, $i = 0, \dots, l$ with $\sum_{i=0}^l p_i = 1$ such that

$$\Omega = \{0, 1, \dots, l\}^{\mathbb{Z}};$$

$\mathcal{A}$ is the product σ -algebra;

μ is the product measure given by $\mu(\{\omega \in \Omega / w_0 = i\}) = p_i$, for $i = 0, \dots, l$;

T is the left shift on Ω , i.e. $(Tw)_i = w_{i+1}$.

Now, using Ornstein isomorphism theorem (see Ornstein [73]), we can see that a Bernoulli shift is isomorphic to a product of two Bernoulli shifts. In particular, our system admits two independent Bernoulli factors. We denote by $\mathcal{B}$ and $\mathcal{C}$ the T -invariant σ -algebras corresponding to them. In order to simplify some proofs, we assume that the first one is a Bernoulli $\left(\frac{1}{2}, \frac{1}{2}\right)$. The reader can check that all the upcoming proofs remain valid for another Bernoulli shift.

So, we can define a $\mathcal{B}$ -measurable function $e_0 : \Omega \longrightarrow \{-1, 1\}$ such that

$$\mu(\{e_0 = -1\}) = \mu(\{e_0 = 1\}) = \frac{1}{2}$$

and if $e_i = e_0 \circ T^i$ for $i \in \mathbb{Z}$, then $(e_i)_{i \in \mathbb{Z}}$ is an i.i.d. sequence. Of course, the sequence of random variables $(e_i)_{i \in \mathbb{Z}}$ is independent of $\mathcal{C}$. Let $\mathcal{F}_0 = \mathcal{C} \vee \sigma\{e_i, i \leq 0\}$ and $\mathcal{F}_k = T^{-k}\mathcal{F}_0 = \mathcal{C} \vee \sigma\{e_i, i \leq k\}$, $k \in \mathbb{Z}$.

We introduce three sequences with the following properties:

$(\theta_k)_{k \in \mathbb{N}} \subset (0, +\infty)$;

$(\rho_k)_{k \in \mathbb{N}} \subset (0, 1)$ a decreasing sequence such that $\sum_{k \geq 0} \rho_k < 1$;

$(N_k)_{k \in \mathbb{N}} \subset \mathbb{N}$ such that $N_{k+1} > N_k$.

These sequences being fixed, we can choose a decreasing sequence $(\varepsilon_k)_{k \in \mathbb{N}} \subset (0, 1)$ such that

$$\sum_{k=1}^{\infty} \theta_k N_k \sqrt{\varepsilon_k} < \infty, \quad (1.4)$$

which implies $\sum_{k=1}^{\infty} \theta_k N_k \varepsilon_k < \infty$.

By application of Lemma 1.7, we consider the sets $A_k \in \mathcal{C}$ corresponding to the sequences $(N_k)_{k \in \mathbb{N}}$, $(\rho_k)_{k \in \mathbb{N}}$ and $(\varepsilon_k)_{k \in \mathbb{N}}$. The function f is defined by

$$f = \sum_{k=1}^{\infty} \theta_k e_{-N_k} \mathbb{1}_{A_k}, \quad (1.5)$$

where $\mathbb{1}_A$ is the indicator function of the set A .

We consider the stationary process $(f \circ T^i)_{i \in \mathbb{Z}}$ which is adapted to the filtration $(\mathcal{F}_i)_{i \in \mathbb{Z}}$.

Proposition 1.8. *The function f belongs to L^2 if and only if $\sum_{k=1}^{\infty} \theta_k^2 \rho_k < \infty$.*

Proof.

By disjointness of the sets A_k ,

$$\|f\|_2^2 = \sum_{k=1}^{\infty} \|f_k \mathbb{1}_{A_k}\|_2^2 = \sum_{k=1}^{\infty} \theta_k^2 \mu(A_k).$$

Now, by Lemma 1.7, $a\rho_k \leq \mu(A_k) \leq \rho_k$. Thus,

$$a \sum_{k=1}^{\infty} \theta_k^2 \rho_k \leq \|f\|_2^2 \leq \sum_{k=1}^{\infty} \theta_k^2 \rho_k.$$

□

In what follows, we apply the four studied criteria to our function f . We express the fact that f satisfies one of them by conditions concerning the sequences $(N_k)_{k \in \mathbb{N}}$, $(\theta_k)_{k \in \mathbb{N}}$ and $(\rho_k)_{k \in \mathbb{N}}$.

Proposition 1.9. *The stationary process $(f \circ T^i)_{i \in \mathbb{Z}}$ admits a martingale-coboundary decomposition in $L^1(\Omega)$ if and only if $\sum_{k=1}^{\infty} \theta_k \sqrt{N_k} \rho_k < \infty$.*

Proof.

We recall that the function f admits a martingale-coboundary decomposition in L^1 if and only if (M-C) holds and here the second sum of (M-C) vanishes.

Necessary condition. We assume that $\sum_{k=1}^{\infty} \theta_k \sqrt{N_k} \rho_k = \infty$. We shall show that if $\sum_{i=1}^{\infty} E(f \circ T^i | \mathcal{F}_0)$ converges in L^1 then $E \left| \sum_{i=1}^{\infty} E(f \circ T^i | \mathcal{F}_0) \right| = \infty$, a contradiction.

For all k and i , $\mathbb{1}_{A_k} \circ T^i$ is $\mathcal{F}_0$ -measurable, so

$$\sum_{i=1}^{\infty} E(f \circ T^i | \mathcal{F}_0) = \sum_{i=1}^{\infty} \sum_{k=1}^{\infty} \theta_k E(e_{-N_k} \circ T^i | \mathcal{F}_0) \mathbb{1}_{A_k} \circ T^i.$$

We will use the fact that the measure of $A_k \triangle T^{-i} A_k$ is small when $i \leq N_k$ to simplify the summation. Note that $E(e_i | \mathcal{F}_0) = e_i$ if $i \leq 0$ and $E(e_i | \mathcal{F}_0) = 0$ if $i > 0$, so

$$\begin{aligned} \sum_{i=1}^{\infty} E(f \circ T^i | \mathcal{F}_0) &= \sum_{k=1}^{\infty} \theta_k \sum_{i=1}^{N_k} e_{-N_k+i} \mathbb{1}_{T^{-i} A_k} \\ &= \sum_{k=1}^{\infty} \theta_k \sum_{i=1}^{N_k} e_{-N_k+i} \mathbb{1}_{A_k} \\ &\quad + \sum_{k=1}^{\infty} \theta_k \sum_{i=1}^{N_k} e_{-N_k+i} (\mathbb{1}_{T^{-i} A_k \setminus A_k} - \mathbb{1}_{A_k \setminus T^{-i} A_k}). \end{aligned} \quad (1.6)$$

Note that $|\mathbb{1}_{T^{-i} A_k \setminus A_k} - \mathbb{1}_{A_k \setminus T^{-i} A_k}| = \mathbb{1}_{T^{-i} A_k \triangle A_k}$ and by construction, $\mu(T^{-i} A_k \triangle A_k) \leq \varepsilon_k$ for $i \leq N_k$. Therefore,

$$\begin{aligned} E \left| \sum_{k=1}^{\infty} \theta_k \sum_{i=1}^{N_k} e_{-N_k+i} (\mathbb{1}_{T^{-i} A_k \setminus A_k} - \mathbb{1}_{A_k \setminus T^{-i} A_k}) \right| &\leq \sum_{k=1}^{\infty} \theta_k \sum_{i=1}^{N_k} E |\mathbb{1}_{T^{-i} A_k \setminus A_k} - \mathbb{1}_{A_k \setminus T^{-i} A_k}| \\ &\leq \sum_{k=1}^{\infty} \theta_k N_k \varepsilon_k < \infty, \quad \text{by (1.4).} \end{aligned}$$

Hence, it remains to prove the L^1 -divergence of the first term in (1.6).

By disjointness of the sets A_k and by independence between the random variable e_i and the set A_k for all $i \in \mathbb{Z}$ and $k \geq 1$,

$$E \left| \sum_{k=1}^{\infty} \theta_k \sum_{i=1}^{N_k} e_{-N_k+i} \mathbb{1}_{A_k} \right| = \sum_{k=1}^{\infty} \theta_k E \left| \sum_{i=1}^{N_k} e_{-N_k+i} \right| \mu(A_k).$$

Now, by independence of the e_i 's, we can use the Marcinkiewicz-Zygmund inequality (see e.g. Theorem 8.1. in [49] or Theorem 10.3.2 in [12]). There exists a constant $A > 0$, such that

$$E \left| \sum_{i=1}^{N_k} e_{-N_k+i} \right| \geq AE \left(\sum_{i=1}^{N_k} e_{-N_k+i}^2 \right)^{\frac{1}{2}} = A\sqrt{N_k}.$$

Recall that $\mu(A_k) \geq a\rho_k$ (Lemma 1.7). So,

$$E \left| \sum_{k=1}^{\infty} \theta_k \sum_{i=1}^{N_k} e_{-N_k+i} \mathbb{1}_{A_k} \right| \geq aA \sum_{k=1}^{\infty} \theta_k \sqrt{N_k} \rho_k = \infty, \quad \text{by assumption.}$$

This concludes the proof of the necessary condition.

Sufficient condition. We assume that $\sum_{k=1}^{\infty} \theta_k \sqrt{N_k} \rho_k < \infty$.

Let $I_n = \sum_{i=1}^n E(f \circ T^i | \mathcal{F}_0)$. We will prove that $I_n \in L^1$ for all $n \geq 1$ and that the sequence $(I_n)_{n \geq 1}$ is a Cauchy sequence in L^1 . The proposition will follow from the completeness of the space L^1 .

To begin, we use the structure of the sets A_k in the same way as in the first part of the proof. We have

$$\begin{aligned} I_n &= \sum_{i=1}^n \sum_{k=1}^{\infty} E(f_k \circ T^i | \mathcal{F}_0) \mathbb{1}_{A_k} \circ T^i \\ &= \sum_{k=1}^{\infty} \theta_k \sum_{i=1}^{\min(n, N_k)} e_{-N_k+i} \mathbb{1}_{A_k} \\ &\quad + \sum_{k=1}^{\infty} \theta_k \sum_{i=1}^{\min(n, N_k)} e_{-N_k+i} (\mathbb{1}_{T^{-i}A_k \setminus A_k} - \mathbb{1}_{A_k \setminus T^{-i}A_k}) \\ &= \Gamma_n^1 + \Gamma_n^2, \end{aligned} \tag{1.7}$$

and

$$\begin{aligned} E|\Gamma_n^2| &\leq \sum_{k=1}^{\infty} \theta_k \sum_{i=1}^{\min(n, N_k)} E |e_{-N_k+i} (\mathbb{1}_{T^{-i}A_k \setminus A_k} - \mathbb{1}_{A_k \setminus T^{-i}A_k})| \\ &\leq \sum_{k=1}^{\infty} \theta_k \sum_{i=1}^{\min(n, N_k)} \mu(T^{-i}A_k \triangle A_k) \\ &\leq \sum_{k=1}^{\infty} \theta_k N_k \varepsilon_k < \infty, \quad \text{by (1.4).} \end{aligned} \tag{1.8}$$

On the other hand, by Marcinkiewicz-Zygmund inequality, there exists a constant $B > 0$ such that

$$E \left| \sum_{i=1}^{\min(n, N_k)} e_{-N_k+i} \right| \leq BE \left(\sum_{i=1}^{\min(n, N_k)} e_{-N_k+i}^2 \right)^{\frac{1}{2}}.$$

Recall that $\mu(A_k) \leq \rho_k$ (Lemma 1.7). Because e_i is independent of A_k for all $i \in \mathbb{Z}$, $k \geq 1$, we have

$$\begin{aligned} E|\Gamma_n^1| &\leq B \sum_{k=1}^{\infty} \theta_k E \left(\sum_{i=1}^{\min(n, N_k)} e_{-N_k+i}^2 \right)^{\frac{1}{2}} \rho_k \\ &\leq B \sum_{k=1}^{\infty} \theta_k \sqrt{N_k} \rho_k < \infty, \quad \text{by assumption.} \end{aligned} \quad (1.9)$$

Applying (1.8) and (1.9) to (1.7) shows that $I_n \in L^1$ for all $n \in \mathbb{N}$.

Now, we will show that $(I_n)_{n \geq 1}$ is a Cauchy sequence in L^1 .

We fix $p \in \mathbb{N}^*$. We have

$$I_{n+p} - I_n = \sum_{k=1}^{\infty} \sum_{i=\min(n, N_k)+1}^{\min(n+p, N_k)} E(f_k \circ T^i | \mathcal{F}_0) \mathbb{1}_{A_k} \circ T^i.$$

Using successively assumption (1.4), Marcinkiewicz-Zygmund inequality and the independence between the e_i 's and the sets A_k (see the calculus made before for I_n), we obtain

$$\begin{aligned} E|I_{n+p} - I_n| &\leq B \sum_{k=1}^{\infty} \theta_k E \left(\sum_{i=\min(n, N_k)+1}^{\min(n+p, N_k)} e_{-N_k+i}^2 \right)^{\frac{1}{2}} \rho_k \\ &\quad + \sum_{k=1}^{\infty} \theta_k \sum_{i=\min(n, N_k)+1}^{\min(n+p, N_k)} \mu(T^{-i} A_k \triangle A_k). \end{aligned}$$

Note that the sum $\sum_{i=\min(n, N_k)+1}^{\min(n+p, N_k)}$ is empty if $N_k \leq n$, is composed of $N_k - (n+1)$ terms if $n < N_k < n+p$ and of p terms otherwise. In the second and in the third case, the number of terms in the sum is less than N_k . So, for all $p \in \mathbb{N}^*$,

$$E|I_{n+p} - I_n| \leq B \sum_{k: N_k > n} \theta_k \sqrt{N_k} \rho_k + \sum_{k: N_k > n} \theta_k N_k \varepsilon_k. \quad (1.10)$$

By assumption and hypothesis (1.4), $\sum_{k=1}^{\infty} \theta_k \sqrt{N_k} \rho_k < \infty$ and $\sum_{k=1}^{\infty} \theta_k N_k \varepsilon_k < \infty$. Hence, both sums in (1.10) go to 0 with $n \rightarrow \infty$ uniformly for all $p \in \mathbb{N}^*$. $(I_n)_{n \geq 1}$ is thus a Cauchy sequence. $\square$

Proposition 1.10. *If the function f belongs to L^2 , the stationary process $(f \circ T^i)_{i \in \mathbb{Z}}$ verifies the Dedecker-Rio criterion if and only if $\sum_{k=1}^{\infty} \theta_k^2 \sqrt{N_k} \rho_k < \infty$.*

Proof.

It follows the idea of the proof of Proposition 1.9. So, some similar passages are given with less details.

Necessary condition. We assume that $\sum_{k=1}^{\infty} \theta_k^2 \sqrt{N_k} \rho_k = \infty$.

We shall show that if $\sum_{i=1}^{\infty} f E(f \circ T^i | \mathcal{F}_0)$ converges in L^1 , then $E \left| \sum_{i=1}^{\infty} f E(f \circ T^i | \mathcal{F}_0) \right| = \infty$. First,

$$\begin{aligned} \sum_{i=1}^{\infty} f E(f \circ T^i | \mathcal{F}_0) &= \sum_{k=1}^{\infty} f \theta_k \sum_{i=1}^{\infty} E(e_{-N_k} \circ T^i | \mathcal{F}_0) \mathbb{1}_{T^{-i} A_k} \\ &= \sum_{k=1}^{\infty} f \theta_k \sum_{i=1}^{N_k} e_{-N_k+i} \mathbb{1}_{T^{-i} A_k}. \end{aligned} \quad (1.11)$$

As in the proof of Proposition 1.9, we decompose $\mathbb{1}_{T^{-i} A_k}$ into $\mathbb{1}_{A_k} + (\mathbb{1}_{T^{-i} A_k \setminus A_k} - \mathbb{1}_{A_k \setminus T^{-i} A_k})$. Applying the Cauchy-Scharwz inequality, we obtain

$$\begin{aligned} E \left| \sum_{k=1}^{\infty} f \theta_k \sum_{i=1}^{N_k} e_{-N_k+i} (\mathbb{1}_{T^{-i} A_k \setminus A_k} - \mathbb{1}_{A_k \setminus T^{-i} A_k}) \right| &\leq \|f\|_2 \sum_{k=1}^{\infty} \theta_k \sum_{i=1}^{N_k} \|\mathbb{1}_{T^{-i} A_k \triangle A_k}\|_2 \\ &\leq \|f\|_2 \sum_{k=1}^{\infty} \theta_k N_k \sqrt{\varepsilon_k}. \end{aligned}$$

Hypothesis (1.4), the fact that f belongs to L^2 , and (1.11) show that the convergence of the integral

$$E \left| \sum_{i=1}^{\infty} f E(f \circ T^i | \mathcal{F}_0) \right|$$

is equivalent to the convergence of

$$E \left| \sum_{k=1}^{\infty} f \theta_k \left(\sum_{i=1}^{N_k} e_{-N_k+i} \right) \mathbb{1}_{A_k} \right|.$$

Now, the sets A_k being disjoint, we have

$$\begin{aligned} \sum_{k=1}^{\infty} f\theta_k \left(\sum_{i=1}^{N_k} e_{-N_k+i} \right) \mathbb{1}_{A_k} &= \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} \theta_j e_{-N_j} \theta_k \left(\sum_{i=1}^{N_k} e_{-N_k+i} \right) \mathbb{1}_{A_k} \mathbb{1}_{A_j} \\ &= \sum_{k=1}^{\infty} \theta_k^2 e_{-N_k} \left(\sum_{i=1}^{N_k} e_{-N_k+i} \right) \mathbb{1}_{A_k}. \end{aligned}$$

Using the disjointness of the sets A_k , the independence between e_i and A_k for all $i \in \mathbb{Z}$ and $k \geq 1$, the independence between the random variables e_{-N_k+i} , $i = 0, \dots, N_k$, the Marcinkiewicz-Zygmund inequality and the assumption, we obtain

$$\begin{aligned} E \left| \sum_{k=1}^{\infty} f\theta_k \left(\sum_{i=1}^{N_k} e_{-N_k+i} \right) \mathbb{1}_{A_k} \right| &= \sum_{k=1}^{\infty} \theta_k^2 E |e_{-N_k}| E \left| \sum_{i=1}^{N_k} e_{-N_k+i} \right| \mu(A_k) \quad (1.12) \\ &\geq A \sum_{k=1}^{\infty} \theta_k^2 E \left(\sum_{i=1}^{\min(n, N_k)} e_{-N_k+i}^2 \right)^{\frac{1}{2}} \mu(A_k) \\ &\geq aA \sum_{k=1}^{\infty} \theta_k^2 \sqrt{N_k} \rho_k = \infty, \end{aligned}$$

where a comes from Lemma 1.7 and $A > 0$ comes from the Marcinkiewicz-Zygmund inequality.

Sufficient condition. We assume that $\sum_{k=1}^{\infty} \theta_k^2 \sqrt{N_k} \rho_k < \infty$.

Let $J_n = \sum_{i=1}^n f E(f \circ T^i | \mathcal{F}_0)$. We shall prove that $(J_n)_{n \geq 1}$ is a Cauchy sequence in L^1 , which proves the proposition.

First, we show that $J_n \in L^1$ for all n , i.e., $E|J_n| < \infty$ for all n . Indeed,

$$J_n = \sum_{k=1}^{\infty} f\theta_k \sum_{i=1}^{\min(n, N_k)} e_{-N_k+i} \mathbb{1}_{T^{-i}A_k}.$$

So, decomposing $\mathbb{1}_{T^{-i}A_k}$ into $\mathbb{1}_{A_k} + (\mathbb{1}_{T^{-i}A_k \setminus A_k} - \mathbb{1}_{A_k \setminus T^{-i}A_k})$, using the Cauchy-Schwarz inequality and (1.4), we show that it is enough to prove the convergence of

$$E \left| \sum_{k=1}^{\infty} f\theta_k \left(\sum_{i=1}^{\min(n, N_k)} e_{-N_k+i} \right) \mathbb{1}_{A_k} \right|.$$

We repeat the calculus leading to (1.12) and we apply the Marcinkiewicz-Zygmund inequality. So, there exists $B > 0$ such that

$$\begin{aligned} E \left| \sum_{k=1}^{\infty} f \theta_k \left(\sum_{i=1}^{\min(n, N_k)} e_{-N_k+i} \right) \mathbb{1}_{A_k} \right| &= \sum_{k=1}^{\infty} \theta_k^2 E |e_{-N_k}| E \left| \sum_{i=1}^{\min(n, N_k)} e_{-N_k+i} \right| \mu(A_k) \\ &\leq B \sum_{k=1}^{\infty} \theta_k^2 \sqrt{N_k} \rho_k < \infty, \quad \text{by assumption.} \end{aligned}$$

Now, we fix $p \in \mathbb{N}^*$. By similar arguments, we can show that

$$\begin{aligned} E |J_{n+p} - J_n| &\leq B \sum_{k=1}^{\infty} \theta_k^2 E \left(\sum_{i=\min(n, N_k)+1}^{\min(n+p, N_k)} e_{-N_k+i}^2 \right)^{\frac{1}{2}} \rho_k \\ &\quad + \|f\|_2 \sum_{k=1}^{\infty} \theta_k \sum_{i=\min(n, N_k)+1}^{\min(n+p, N_k)} \|\mathbb{1}_{T^{-i}A_k \triangle A_k}\|_2. \end{aligned}$$

The same considerations about the sum $\sum_{i=\min(n, N_k)+1}^{\min(n+p, N_k)}$ as in the proof of Proposition 1.9 lead to

$$E |J_{n+p} - J_n| \leq B \sum_{k: N_k > n} \theta_k^2 \sqrt{N_k} \rho_k + \|f\|_2 \sum_{k: N_k > n} \theta_k N_k \sqrt{\varepsilon_k}.$$

By assumption and by (1.4), both sums go to 0 with $n \rightarrow \infty$, uniformly for $p \in \mathbb{N}^*$. Hence, $(J_n)_{n \geq 1}$ is a Cauchy sequence in the space L^1 , which is complete. $\square$

Proposition 1.11. *For the previously defined function f ,*

$$\sum_{n=1}^{\infty} \frac{\|E(S_n(f)|\mathcal{F}_0)\|_2}{n^{\frac{3}{2}}} < \infty \quad \text{if and only if} \quad \sum_{n=1}^{\infty} \frac{(\sum_{k=1}^{\infty} \theta_k^2 \min(n, N_k) \rho_k)^{\frac{1}{2}}}{n^{\frac{3}{2}}} < \infty.$$

Proof.

Note that

$$E(S_n(f)|\mathcal{F}_0) = \sum_{k=1}^{\infty} \theta_k \sum_{i=1}^{\min(n, N_k)} e_{-N_k+i} \mathbb{1}_{A_k} + \sum_{k=1}^{\infty} \theta_k \sum_{i=1}^{\min(n, N_k)} e_{-N_k+i} (\mathbb{1}_{T^{-i}A_k \setminus A_k} - \mathbb{1}_{A_k \setminus T^{-i}A_k})$$

and

$$\left\| \sum_{k=1}^{\infty} \theta_k \sum_{i=1}^{\min(n, N_k)} e_{-N_k+i} (\mathbb{1}_{T^{-i}A_k \setminus A_k} - \mathbb{1}_{A_k \setminus T^{-i}A_k}) \right\|_2 \leq \sum_{k=1}^{\infty} \theta_k \min(n, N_k) \sqrt{\varepsilon_k}.$$

So, by (1.4),

$$\sum_{n=1}^{\infty} \frac{\|E(S_n(f)|\mathcal{F}_0)\|_2}{n^{\frac{3}{2}}} < \infty \quad \text{if and only if} \quad \sum_{n=1}^{\infty} \frac{\|\sum_{k=1}^{\infty} \theta_k \sum_{i=1}^{\min(n, N_k)} e_{-N_k+i} \mathbb{1}_{A_k}\|_2}{n^{\frac{3}{2}}} < \infty.$$

Now, by independence, applying the Marcinkiewicz-Zygmund inequality, we can see that there exist $A, B > 0$ such that

$$aA \sum_{k=1}^{\infty} \theta_k^2 \min(n, N_k) \rho_k \leq E \left| \sum_{k=1}^{\infty} \theta_k \sum_{i=1}^{\min(n, N_k)} e_{-N_k+i} \mathbb{1}_{A_k} \right|^2 \leq B \sum_{k=1}^{\infty} \theta_k^2 \min(n, N_k) \rho_k.$$

The proposition is proved. $\square$

Proposition 1.12. *For the previously defined function f ,*

$$\sum_{i \in \mathbb{Z}} \|P_0(f \circ T^i)\|_2 < \infty \quad \text{if and only if} \quad \sum_{k \geq 1} \theta_k \sqrt{\rho_k} < \infty.$$

Proof.

First of all, the sequence $(f \circ T^i)_{i \in \mathbb{Z}}$ is adapted to the filtration and then for all $i < 0$,

$$P_0(f \circ T^i) = 0.$$

For $i \geq 0$, we have

$$\begin{aligned} P_0(f \circ T^i) &= E(f \circ T^i | \mathcal{F}_0) - E(f \circ T^i | \mathcal{F}_{-1}) \\ &= \sum_{k \geq 1} \theta_k [E(e_{-N_k+i} | \mathcal{F}_0) - E(e_{-N_k+i} | \mathcal{F}_{-1})] \mathbb{1}_{A_k} \circ T^i. \end{aligned}$$

Since the function e_j is $\mathcal{F}_0$ -measurable for $j \leq 0$ and independent of $\mathcal{F}_0$ for $j > 0$,

$$\begin{aligned} E(e_{-N_k+i} | \mathcal{F}_0) - E(e_{-N_k+i} | \mathcal{F}_{-1}) &= \begin{cases} e_0 & \text{if } i = N_k \\ 0 & \text{otherwise} \end{cases} \\ &= e_0 \mathbb{1}_{\{i=N_k\}}. \end{aligned}$$

Thus

$$\begin{aligned} P_0(f \circ T^i) &= \sum_{k \geq 1} \theta_k e_0 \mathbb{1}_{\{i=N_k\}} \mathbb{1}_{A_k} + \sum_{k \geq 1} \theta_k e_0 \mathbb{1}_{\{i=N_k\}} (\mathbb{1}_{T^{-i}A_k} - \mathbb{1}_{A_k}) \\ &= I_1(i) + I_2(i). \end{aligned}$$

For I_2 , we use the fact that $\mu(A_k \Delta T^{-i}A_k) \leq \varepsilon_k$ for $0 \leq i \leq N_k$ to get

$$\begin{aligned} \|I_2(i)\|_2 &\leq \sum_{k \geq 1} \theta_k \mathbb{1}_{\{i=N_k\}} \|e_0 \mathbb{1}_{A_k \Delta T^{-i}A_k}\|_2 \\ &\leq \sum_{k \geq 1} \theta_k \mathbb{1}_{\{i=N_k\}} \sqrt{\varepsilon_k}. \end{aligned}$$

Remark for each $i \geq 0$, there is at most one integer k such that $N_k = i$ and for each $k \geq 1$, there exists an integer i such that $i = N_k$. We deduce

$$\begin{aligned} \sum_{i \geq 0} \|I_2(i)\|_2 &\leq \sum_{i \geq 0} \sum_{k \geq 1} \theta_k \mathbb{1}_{\{i=N_k\}} \sqrt{\varepsilon_k} \\ &= \sum_{k \geq 1} \theta_k \sqrt{\varepsilon_k} \end{aligned}$$

which is finite by the assumptions.

Thus, $\sum_{i \geq 0} \|P_i(f)\|_2$ is converging if and only if $\sum_{i \geq 0} \|I_1(i)\|_2$ is converging. Now for a fixed i , since the sets A_k are disjoint and since there is at most one k such that $N_k = i$, we have

$$\begin{aligned} \|I_1(i)\|_2 &= \sqrt{\sum_{k \geq 1} \theta_k^2 \mathbb{1}_{\{i=N_k\}} \mu(A_k)} \\ &= \sum_{k \geq 1} \theta_k \mathbb{1}_{\{i=N_k\}} \sqrt{\mu(A_k)}. \end{aligned}$$

Finally,

$$\begin{aligned} \sum_{i \geq 0} \|I_1(i)\|_2 &= \sum_{i \geq 0} \sum_{k \geq 1} \theta_k \mathbb{1}_{\{i=N_k\}} \sqrt{\mu(A_k)} \\ &= \sum_{k \geq 1} \theta_k \sqrt{\mu(A_k)}. \end{aligned}$$

We can conclude the proof using $a\rho_k \leq \mu(A_k) \leq \rho_k$. □

1.5 Some counterexamples

Counterexample 1, proof of (viii)

In this section, we give an example of a function satisfying the Dedecker-Rio criterion, the Maxwell-Woodroffe condition and also the Hannan criterion but not the martingale-coboundary decomposition in L^1 . To do this, we consider the function f defined in (1.5) by the sequences

$$\rho_k = \frac{1}{4^k}, \quad N_k = 4^{2k} \text{ and } \theta_k = \frac{1}{k}, k \geq 1.$$

First,

$$\sum_{k=1}^{\infty} \theta_k^2 \rho_k = \sum_{k=1}^{\infty} \frac{1}{k^2 4^k} < \infty;$$

then, by Proposition 1.8, the function f belongs to L^2 . We have

$$\sum_{k=1}^{\infty} \theta_k \sqrt{N_k} \rho_k = \sum_{k=1}^{\infty} \frac{1}{k} = \infty,$$

hence, by Proposition 1.9, the stationary process $(f \circ T^i)_{i \in \mathbb{Z}}$ does not admit a martingale-coboundary decomposition in L^1 . But,

$$\sum_{k=1}^{\infty} \theta_k^2 \sqrt{N_k} \rho_k = \sum_{k=1}^{\infty} \frac{1}{k^2} < \infty$$

and Proposition 1.10 show that it satisfies the Dedecker-Rio criterion.

To verify that the process $(f \circ T^i)_{i \in \mathbb{Z}}$ satisfies the Maxwell-Woodroffe condition, by Proposition 1.11, we have to study the sums

$$\sum_{k=1}^{\infty} \theta_k^2 \min(n, N_k) \rho_k = \sum_{k=1}^{\lfloor \frac{\ln n}{2 \ln 4} \rfloor} \theta_k^2 N_k \rho_k + n \sum_{k=\lfloor \frac{\ln n}{2 \ln 4} \rfloor + 1}^{\infty} \theta_k^2 \rho_k. \quad (1.13)$$

The first term on the right can be estimated by

$$\sum_{k=1}^{\lfloor \frac{\ln n}{2 \ln 4} \rfloor} \theta_k^2 N_k \rho_k = \sum_{k=1}^{\lfloor \frac{\ln n}{2 \ln 4} \rfloor} \frac{4^k}{k^2} \leq \sum_{k=1}^{\lfloor \frac{\ln n}{2 \ln 4} \rfloor} 4^k = \mathcal{O}(\sqrt{n}).$$

For the second term,

$$\sum_{k=\lfloor \frac{\ln n}{2 \ln 4} \rfloor + 1}^{\infty} \theta_k^2 \rho_k = \sum_{k=\lfloor \frac{\ln n}{2 \ln 4} \rfloor + 1}^{\infty} \frac{1}{k^2 4^k} \leq \sum_{k=\lfloor \frac{\ln n}{2 \ln 4} \rfloor + 1}^{\infty} \frac{1}{4^k} = \mathcal{O}\left(\frac{1}{\sqrt{n}}\right).$$

From (1.13), we derive

$$\sum_{k=1}^{\infty} \theta_k^2 \min(n, N_k) \rho_k = \mathcal{O}(\sqrt{n})$$

and

$$\frac{(\sum_{k=1}^{\infty} \theta_k^2 \min(n, N_k) \rho_k)^{\frac{1}{2}}}{n^{\frac{3}{2}}} = \mathcal{O}\left(n^{-\frac{5}{4}}\right).$$

Therefore, by Proposition 1.11,

$$\sum_{n=1}^{\infty} \frac{\|E(S_n(f)|\mathcal{F}_0)\|_2}{n^{\frac{3}{2}}} < \infty.$$

Finally,

$$\sum_{k \geq 1} \theta_k \sqrt{\rho_k} = \sum_{k \geq 1} \frac{1}{k 2^k} < \infty.$$

Thus, by Proposition 1.12, the Hannan criterion holds and (viii) is proved.

Counterexample 2, proof of (vii)

Here, we give a process which satisfies the martingale-coboundary decomposition in L^1 , the Hannan criterion and the Maxwell-Woodroffe condition but fails to satisfy the Dedecker-Rio criterion. We consider the function f defined in (1.5), this time, by the sequences

$$\rho_k = \frac{1}{4^k}, \quad N_k = k^4 \text{ and } \theta_k = \frac{2^k}{k^{\frac{3}{2}}}, \quad k \geq 1.$$

We have

$$\sum_{k=1}^{\infty} \theta_k^2 \rho_k = \sum_{k=1}^{\infty} \frac{1}{k^3} < \infty,$$

$$\sum_{k=1}^{\infty} \theta_k \sqrt{N_k} \rho_k = \sum_{k=1}^{\infty} \frac{\sqrt{k}}{2^k} < \infty$$

and

$$\sum_{k=1}^{\infty} \theta_k^2 \sqrt{N_k} \rho_k = \sum_{k=1}^{\infty} \frac{1}{k} = \infty.$$

By Propositions 1.8, 1.9 and 1.10, the function f belongs to L^2 and admits a martingale-coboundary decomposition in L^1 but does not satisfy the Dedecker-Rio criterion.

The process $(f \circ T^i)_{i \in \mathbb{Z}}$ verifies the Maxwell-Woodroffe condition. Indeed,

$$\sum_{k=1}^{\infty} \theta_k^2 \min(n, N_k) \rho_k = \sum_{k=1}^{\lfloor n^{\frac{1}{4}} \rfloor} k + n \sum_{k=\lfloor n^{\frac{1}{4}} \rfloor + 1}^{\infty} \frac{1}{k^3} = \mathcal{O}(\sqrt{n})$$

and, as in counterexample 1, using Proposition 1.11, we deduce

$$\sum_{n=1}^{\infty} \frac{\|E(S_n(f)|\mathcal{F}_0)\|_2}{n^{\frac{3}{2}}} < \infty.$$

By Proposition 1.12, the Hannan criterion holds because

$$\sum_{k \geq 1} \theta_k \sqrt{\rho_k} = \sum_{k \geq 1} \frac{1}{k^{\frac{3}{2}}} < \infty.$$

This prove (vii).

Counterexample 3, proof of (v)

This counterexample shows that none of the conditions (M-C), (D-R) and (M-W) implies (HC). We consider the function f defined by the sequences

$$\rho_k = \frac{1}{4^k}, N_k = k \text{ and } \theta_k = \frac{2^k}{k}, k \geq 1.$$

Then $f \in L^2$ and using Propositions 1.9, 1.10, 1.11 and 1.12, we get:

$$\sum_{k \geq 1} \theta_k \rho_k \sqrt{N_k} = \sum_{k \geq 1} \frac{1}{2^k \sqrt{k}} < \infty$$

and then (M-C) is verified;

$$\sum_{k \geq 1} \theta_k^2 \rho_k \sqrt{N_k} = \sum_{k \geq 1} \frac{1}{k^{\frac{3}{2}}} < \infty$$

and then (D-R) is verified;

$$\sum_{k \geq 1} \theta_k^2 \min(n, N_k) \rho_k = \sum_{k \geq 1} \frac{\min(n, k)}{k^2}$$

but $\sum_{k=1}^n \frac{1}{k} \leq 1 + \ln(n)$ and $\sum_{k=n+1}^{\infty} \frac{n}{k^2} \leq 1$, then

$$\sum_{n \geq 1} n^{-\frac{3}{2}} \left(\sum_{k \geq 1} \theta_k^2 \min(n, N_k) \rho_k \right)^{\frac{1}{2}} \leq \sum_{n \geq 1} n^{-\frac{3}{2}} \sqrt{\ln(n) + 2} < \infty$$

and (M-W) is verified;

$$\sum_{k \geq 1} \theta_k \sqrt{\rho_k} = \sum_{k \geq 1} \frac{1}{k} = \infty$$

and then (HC) is not satisfied.

(v) is proved

Counterexample 4, proof of (iv)

We will show that (HC) does not imply any conditions (M-C), (D-R) or (M-W). We consider the function f defined by the sequences

$$\rho_k = \frac{1}{4^k}, N_k = 2^{4k} \text{ and } \theta_k = \frac{2^k}{k^{\frac{3}{2}}}, k \geq 1.$$

Then $f \in L^2$ and:

$$\sum_{k \geq 1} \theta_k \rho_k \sqrt{N_k} = \sum_{k \geq 1} \frac{2^k}{k^{\frac{3}{2}}} = \infty,$$

thus (M-C) does not hold;

$$\sum_{k \geq 1} \theta_k^2 \rho_k \sqrt{N_k} = \sum_{k \geq 1} \frac{2^{2k}}{k^3} = \infty,$$

thus (D-R) does not hold;

$$\sum_{k \geq 1} \theta_k^2 \min(n, N_k) \rho_k = \sum_{k \geq 1} \frac{\min(n, 2^{4k})}{k^3} \geq n \sum_{k \geq \lfloor \frac{\ln n}{4 \ln 2} \rfloor} \frac{1}{k^3} \geq 8 \ln^2 2 \frac{n}{\ln^2 n}$$

and then

$$\sum_{n \geq 1} n^{-\frac{3}{2}} \left(\sum_{k \geq 1} \theta_k^2 \min(n, N_k) \rho_k \right)^{\frac{1}{2}} \geq 8 \ln^2 2 \sum_{n \geq 1} \frac{1}{n \ln n} = \infty,$$

thus (M-W) does not hold.

On the other hand,

$$\sum_{k \geq 1} \theta_k \sqrt{\rho_k} = \sum_{k \geq 1} \frac{1}{k^{\frac{3}{2}}} < \infty$$

and then (HC) holds proving (iv).

Counterexample 5, proof of (iii)

This example shows (iii). We consider the function f defined by the sequences

$$\rho_k = \frac{C}{k^{\frac{3}{2}}}, N_k = k \text{ and } \theta_k = 1, k \geq 1,$$

where $0 < C < \left(2 \sum_{k=1}^{\infty} \frac{1}{k^{\frac{3}{2}}} \right)^{-1}$.

Then $f \in L^2$ and:

$$\sum_{k \geq 1} \theta_k \rho_k \sqrt{N_k} = C \sum_{k \geq 1} \frac{1}{k} = \infty,$$

thus (M-C) does not hold;

$$\sum_{k \geq 1} \theta_k^2 \rho_k \sqrt{N_k} = C \sum_{k \geq 1} \frac{1}{k} = \infty,$$

thus (D-R) does not hold;

$$\sum_{k \geq 1} \theta_k \sqrt{\rho_k} = C \sum_{k \geq 1} \frac{1}{k^{\frac{3}{4}}} = \infty$$

and then (HC) does not hold. But,

$$\begin{aligned} \sum_{k \geq 1} \theta_k^2 \min(n, N_k) \rho_k &= C \sum_{k \geq 1} \frac{\min(n, k)}{k^{\frac{3}{2}}} \\ &\leq C \sum_{k=1}^n \frac{1}{\sqrt{k}} + nC \sum_{k=n+1}^{\infty} \frac{1}{k^{\frac{3}{2}}} \\ &= O(\sqrt{n}) \end{aligned}$$

and then

$$\sum_{n \geq 1} n^{-\frac{3}{2}} \left(\sum_{k \geq 1} \theta_k^2 \min(n, N_k) \rho_k \right)^{\frac{1}{2}} < \infty,$$

thus (M-W) holds.

Counterexample 6, proof of (i)

In this paragraph our example will verify the martingale-coboundary decomposition in L^1 with the invariance principle but not the other criteria. We consider the function f defined by (1.5) with the sequences

$$\rho_k = \frac{1}{k^2 4^k}, \quad N_k = 4^k \text{ and } \theta_k = 2^k, k \geq 1.$$

We have

$$\sum_{k=1}^{\infty} \theta_k^2 \rho_k = \sum_{k=1}^{\infty} \frac{1}{k^2} < \infty \quad \text{and} \quad \sum_{k=1}^{\infty} \theta_k \sqrt{N_k} \rho_k = \sum_{k=1}^{\infty} \frac{1}{k^2} < \infty.$$

This implies that f belongs to L^2 and admits a martingale-coboundary decomposition in L^1 (Propositions 1.8 and 1.9).

Further,

$$\sum_{k=1}^{\infty} \theta_k^2 \sqrt{N_k} \rho_k = \sum_{k=1}^{\infty} \frac{2^k}{k^2} = \infty$$

and

$$\sum_{k=1}^{\infty} \theta_k \sqrt{\rho_k} = \sum_{k=1}^{\infty} \frac{1}{k} = \infty.$$

Hence the Dedecker-Rio criterion and the Hannan criterion do not hold.

For the Maxwell-Woodroffe condition,

$$\sum_{k=1}^{\infty} \theta_k^2 \min(n, N_k) \rho_k = \sum_{k=1}^{\lfloor \frac{\ln n}{\ln 4} \rfloor} \frac{4^k}{k^2} + n \sum_{k=\lfloor \frac{\ln n}{\ln 4} \rfloor + 1}^{\infty} \frac{1}{k^2} \geq C \frac{n}{\ln^2 n},$$

for some $C > 0$. Therefore,

$$\sum_{n=1}^{\infty} \frac{(\sum_{k=1}^{\infty} \theta_k^2 \min(n, N_k) \rho_k)^{\frac{1}{2}}}{n^{\frac{3}{2}}} \geq \sqrt{C} \sum_{n=1}^{\infty} \frac{1}{n \ln n} = \infty$$

and by Proposition 1.11, the Maxwell-Woodroffe condition does not hold.

To prove (i), it remains to show that the invariance principle holds. Actually, to do that, we will add hypotheses in the definition of the sets A_k . All the preceding results of this paragraph will remain valid.

We have shown that f admits a martingale-coboundary decomposition in L^1 . Thus, $f = m + g - g \circ T$, where $m, g \in L^1$ and $(m \circ T^i)_{i \in \mathbb{Z}}$ is a martingale difference sequence. Here, we assume that $\mu(T^{-(N_k+1)} A_k \triangle A_k) \leq \varepsilon_k$ for all k (in Lemma 1.7, take $N_k + 1$ instead of N_k). It is clear that this assumption does not change the previous results. Now, we can show

Proposition 1.13. *In the decomposition $f = m + g - g \circ T$, m belongs to L^2 .*

By the Billingsley and Ibragimov theorem for martingale difference sequences and by stochastic boundedness of partial sums of $g - g \circ T$, it follows

Corollary 1.14. *The process $(f \circ T^i)_{i \in \mathbb{Z}}$ verifies the CLT.*

Proof.

Actually, we shall prove that $g - g \circ T \in L^2$. In fact, see Volný [85], $g = \sum_{i=0}^{\infty} E(f \circ T^i | \mathcal{F}_0)$.

So,

$$\begin{aligned} g - g \circ T &= \sum_{i=0}^{\infty} E(f \circ T^i | \mathcal{F}_0) - \sum_{i=0}^{\infty} E(f \circ T^{i+1} | \mathcal{F}_1) \\ &= \sum_{k=1}^{\infty} \theta_k \left(\sum_{i=0}^{N_k} e_{-N_k+i} \mathbb{1}_{T^{-i} A_k} - \sum_{i=0}^{N_k} e_{-N_k+i+1} \mathbb{1}_{T^{-(i+1)} A_k} \right) \\ &= \sum_{k=1}^{\infty} \theta_k (e_{-N_k} - e_1) \mathbb{1}_{A_k} \\ &\quad + \sum_{k=1}^{\infty} \theta_k \sum_{i=0}^{N_k} e_{-N_k+i} (\mathbb{1}_{T^{-i} A_k \setminus A_k} - \mathbb{1}_{A_k \setminus T^{-i} A_k}) \\ &\quad - \sum_{k=1}^{\infty} \theta_k \sum_{i=0}^{N_k} e_{-N_k+i+1} (\mathbb{1}_{T^{-(i+1)} A_k \setminus A_k} - \mathbb{1}_{A_k \setminus T^{-(i+1)} A_k}). \end{aligned}$$

Now, by (1.4),

$$\begin{aligned} \left\| \sum_{k=1}^{\infty} \theta_k \sum_{i=0}^{N_k} e_{-N_k+i} (\mathbb{1}_{T^{-i}A_k \setminus A_k} - \mathbb{1}_{A_k \setminus T^{-i}A_k}) \right\|_2 &\leq \sum_{k=1}^{\infty} \theta_k \sum_{i=0}^{N_k} \left\| \mathbb{1}_{T^{-i}A_k \triangle A_k} \right\|_2 \\ &\leq \sum_{k=1}^{\infty} \theta_k N_k \sqrt{\varepsilon_k} < \infty. \end{aligned} \quad (1.14)$$

In the same way,

$$\left\| \sum_{k=1}^{\infty} \theta_k \sum_{i=0}^{N_k} e_{-N_k+i+1} (\mathbb{1}_{T^{-(i+1)}A_k \setminus A_k} - \mathbb{1}_{A_k \setminus T^{-(i+1)}A_k}) \right\|_2 \leq \sum_{k=1}^{\infty} \theta_k N_k \sqrt{\varepsilon_k} < \infty. \quad (1.15)$$

By disjointness of the sets A_k , by independence of the functions $e_{-N_k} - e_1$ and $\mathbb{1}_{A_k}$,

$$\begin{aligned} \left\| \sum_{k=1}^{\infty} \theta_k (e_{-N_k} - e_1) \mathbb{1}_{A_k} \right\|_2^2 &\leq \sum_{k=1}^{\infty} \theta_k^2 \|e_{-N_k} - e_1\|_2^2 \mu(A_k) \\ &\leq 4 \sum_{k=1}^{\infty} \theta_k^2 \rho_k < \infty. \end{aligned} \quad (1.16)$$

(1.14), (1.15) and (1.16) lead to the proposition. $\square$

Let $v_k = \theta_k \sum_{i=0}^{N_k} e_{-N_k} \circ T^i$. For each $R \in \mathbb{N}^*$, the quantity $\sum_{k=1}^R \|v_k\|_{\infty}$ is finite. Thus, there exists an (not necessarily strictly) increasing sequence $(R_n)_{n \in \mathbb{N}} \rightarrow \infty$ such that

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{R_n-1} \|v_k\|_{\infty} \xrightarrow{n \rightarrow \infty} 0. \quad (1.17)$$

The sequence $(R_n)_{n \in \mathbb{N}}$ being fixed, we construct the sets A_k in the following way. For all k , let n_k be the greatest integer such that $R_{n_k} \leq k$. To define the sets A_k , we apply Lemma 1.7 with $(\max(n_k, N_k + 1))_{k \in \mathbb{N}}$ instead of $(N_k)_{k \in \mathbb{N}}$. Again, it is easy to see that previous results remain valid. With this construction, we have the following property.

$$\forall k \geq R_n, \forall i, j \in \{0, \dots, n\}, \quad \mu(T^{-i}A_k \triangle T^{-j}A_k) \leq \varepsilon_k. \quad (1.18)$$

Proposition 1.15. *The process $(f \circ T^i)_{i \in \mathbb{Z}}$ satisfies the invariance principle.*

Proof.

Since $m \in L^2$, as recalled in introduction, it is enough to show that

$$\frac{1}{\sqrt{n}} \max_{i \leq n} |g \circ T^i| \xrightarrow{n \rightarrow \infty} 0 \text{ in probability.}$$

We have

$$\begin{aligned}
g &= \sum_{i=0}^{\infty} E(f \circ T^i | \mathcal{F}_0) \\
&= \sum_{k=1}^{\infty} \theta_k \left(\sum_{i=0}^{N_k} e_{-N_k+i} \right) \mathbb{1}_{A_k} + \sum_{k=1}^{\infty} \theta_k \left(\sum_{i=0}^{N_k} e_{-N_k+i} (\mathbb{1}_{T^{-i}A_k \setminus A_k} - \mathbb{1}_{A_k \setminus T^{-i}A_k}) \right) \\
&= g_1 + g_2.
\end{aligned}$$

By Markov inequality, for all $\lambda > 0$,

$$\begin{aligned}
\mu \left\{ \max_{i \leq n} |g_2 \circ T^i| \geq \lambda \sqrt{n} \right\} &\leq \frac{E(\max_{i \leq n} |g_2 \circ T^i|)}{\lambda \sqrt{n}} \\
&\leq \frac{1}{\lambda \sqrt{n}} \sum_{k=1}^{\infty} \theta_k (N_k + 1) \varepsilon_k \xrightarrow{n \rightarrow \infty} 0, \quad \text{by (1.4).}
\end{aligned}$$

So, $\frac{1}{\sqrt{n}} \max_{i \leq n} |g_2 \circ T^i|$ converges to 0 in probability.

It remains to prove the same thing for $g_1 = \sum_{k=1}^{\infty} v_k \mathbb{1}_{A_k}$. By (1.17),

$$\frac{1}{\sqrt{n}} \max_{i \leq n} \left| \sum_{k=1}^{R_n-1} v_k \circ T^i \mathbb{1}_{A_k} \circ T^i \right| \leq \frac{1}{\sqrt{n}} \sum_{k=1}^{R_n-1} \|v_k\|_{\infty} \xrightarrow{n \rightarrow \infty} 0.$$

Hence, it converges to zero in probability.

Now, for all $\lambda > 0$, $\mu \left\{ \max_{i \leq n} \left| \sum_{k=R_n}^{\infty} v_k \circ T^i \mathbb{1}_{A_k} \circ T^i \right| \geq 2\lambda \sqrt{n} \right\}$ is smaller than

$$\mu \left\{ \max_{i \leq n} \left| \sum_{k=R_n}^{\infty} v_k \circ T^i \mathbb{1}_{A_k} \right| \geq \lambda \sqrt{n} \right\} + \mu \left\{ \max_{i \leq n} \sum_{k=R_n}^{\infty} |v_k \circ T^i| \mathbb{1}_{T^{-i}A_k \triangle A_k} \geq \lambda \sqrt{n} \right\}.$$

For the first term, the Tchebychev inequality gives

$$\begin{aligned}
\mu \left\{ \max_{i \leq n} \left| \sum_{k=R_n}^{\infty} v_k \circ T^i \mathbb{1}_{A_k} \right| \geq \lambda \sqrt{n} \right\} &\leq \mu \left\{ \sum_{k=R_n}^{\infty} \max_{i \leq n} |v_k \circ T^i| \mathbb{1}_{A_k} \geq \lambda \sqrt{n} \right\} \\
&\leq \frac{E \left(\left(\sum_{k=R_n}^{\infty} \max_{i \leq n} |v_k \circ T^i| \mathbb{1}_{A_k} \right)^{\frac{1}{3}} \right)}{\lambda n^{\frac{1}{6}}} \\
&\leq \frac{\sum_{k=R_n}^{\infty} \theta_k^{\frac{1}{3}} (N_k + 1)^{\frac{1}{3}} \rho_k}{\lambda n^{\frac{1}{6}}} \xrightarrow{n \rightarrow \infty} 0,
\end{aligned}$$

because

$$\sum_{k=1}^{\infty} \theta_k^{\frac{1}{3}} N_k^{\frac{1}{3}} \rho_k = \sum_{k=1}^{\infty} \frac{1}{k^2 2^k} < \infty.$$

For the second term, Markov inequality, assumptions (1.18) and (1.4) show convergence to zero with n . Thus, $\frac{1}{\sqrt{n}} \max_{i \leq n} |g_1 \circ T^i| \xrightarrow[n \rightarrow \infty]{} 0$ in probability. $\square$

(i) is proved.

1.6 A slightly different model

To prove (ii) and (vi), we have to improve our general model. We assume

$$N_{k+1} \geq 2N_k, \forall k \geq 1. \quad (1.19)$$

We introduce a new model function h defined by

$$h = \sum_{k \geq 1} \theta_k \sum_{j=N_k+1}^{2N_k} e_{-j} \mathbb{1}_{A_k}$$

where the sets A_k are defined by applying Lemma 1.7 with $2N_k$ instead of N_k . Moreover, we assume that the ε_k 's are sufficiently small to have

$$\sum_{k=1}^{\infty} \theta_k N_k^2 \sqrt{\varepsilon_k} < \infty. \quad (1.20)$$

Proposition 1.16. *The function h belongs to L^2 if and only if $\sum_{k=1}^{\infty} \theta_k^2 N_k \rho_k < \infty$.*

Proof.

It suffices to see that

$$E \left| \sum_{k=1}^{\infty} \theta_k^2 \sum_{j=N_k+1}^{2N_k} e_{-j} \mathbb{1}_{A_k} \right|^2 = \sum_{k=1}^{\infty} \theta_k^2 E \left| \sum_{j=N_k+1}^{2N_k} e_{-j} \right|^2 \mu(A_k)$$

and to use the independence between the e_i 's. $\square$

From now on, we assume that $h \in L^2$. Remark that the stationary process $(h \circ T^i)_{i \in \mathbb{Z}}$ is also adapted to the filtration $(\mathcal{F}_i)_{i \in \mathbb{Z}}$.

Proposition 1.17. *The stationary process $(h \circ T^i)_{i \in \mathbb{Z}}$ admits a martingale-coboundary decomposition in $L^1(\Omega)$ if and only if $\sum_{k=1}^{\infty} \theta_k N_k^{\frac{3}{2}} \rho_k < \infty$.*

Proof.

Let $I_n = \sum_{i=1}^n E(h \circ T^i | \mathcal{F}_0)$. As in the proof of Proposition 1.9, using the properties of the sets A_k , the independence between the A_k 's and the e_i 's and (1.20) we can see that $\|I_n\|_1 < \infty$ if and only if

$$\sum_{k=1}^{\infty} \theta_k E \left| \sum_{i=1}^n \sum_{j=N_k+1}^{2N_k} E(e_{i-j} | \mathcal{F}_0) \right| \mu(A_k) < \infty.$$

Let us denote by B_n^k the function $\sum_{i=1}^n \sum_{j=N_k+1}^{2N_k} E(e_{i-j} | \mathcal{F}_0)$. Recall that $E(e_i | \mathcal{F}_0)$ is equal to e_i for $i \leq 0$ and is equal to 0 for $i > 0$. Then

for $n \geq 2N_k$,

$$B_n^k = N_k \sum_{j=0}^{N_k-1} e_{-j} + \sum_{j=0}^{N_k-1} (N_k - j) e_{-N_k-j};$$

for $N_k < n < 2N_k$,

$$B_n^k = \sum_{j=0}^{2N_k-n} (n - N_k + j - 1) e_{-j} + \sum_{j=2N_k-n+1}^{N_k-1} N_k e_{-j} + \sum_{j=0}^{N_k-1} (N_k - j) e_{-N_k-j};$$

and for $n \leq N_k$,

$$B_n^k = \sum_{j=0}^n (n - j) e_{-N_k+j} + \sum_{j=1}^{N_k} \min(n, N_k - j) e_{-N_k-j}.$$

In each case, by independence between the e_i , there exists $B > 0$ such that for all $n \geq 1$,

$$\|B_n^k\|_2 \leq B N_k^{\frac{3}{2}}.$$

Then

$$\begin{aligned} \sum_{k=1}^{\infty} \theta_k E \left| \sum_{i=1}^n \sum_{j=N_k+1}^{2N_k} E(e_{i-j} | \mathcal{F}_0) \right| \mu(A_k) &= \sum_{k=1}^{\infty} \theta_k E |B_n^k| \mu(A_k) \\ &\leq B \sum_{k=1}^{\infty} \theta_k N_k^{\frac{3}{2}} \mu(A_k). \end{aligned}$$

This implies that I_n belongs to the space L^1 for all n . In the same way, we can see that it also implies that $(I_n)_{n \in \mathbb{N}}$ is a Cauchy sequence in L^1 (the details are left to the reader).

Further, by the Marcinkiewicz-Zygmund inequality, there exists $A > 0$, such that for all $n \geq 2N_k$ we have

$$\|B_n^k\|_1 \geq AN_k^{\frac{3}{2}}.$$

Therefore

$$E \left| \sum_{i=1}^{\infty} E(h \circ T^i | \mathcal{F}_0) \right| \geq A \sum_{k=1}^{\infty} \theta_k N_k^{\frac{3}{2}} \mu(A_k).$$

□

Proposition 1.18. *If $\sum_{k=1}^{\infty} \theta_k^2 N_k^2 \rho_k < \infty$, then the stationary process $(h \circ T^i)_{i \in \mathbb{Z}}$ satisfies the Dedecker-Rio criterion.*

Proof.

Let $J_n = \sum_{i=1}^n hE(h \circ T^i | \mathcal{F}_0)$. As in the proof of Proposition 1.10, using the properties of the sets A_k and (1.20) we can see that $\|J_n\|_1 < \infty$ if and only if

$$E \left| \sum_{k=1}^{\infty} \theta_k h \left(\sum_{i=1}^n \sum_{j=N_k+1}^{2N_k} E(e_{i-j} | \mathcal{F}_0) \right) \mathbb{1}_{A_k} \right| < \infty.$$

Let us define B_n^k as in the previous proof. We have $\|B_n^k\|_2 \leq BN_k^{\frac{3}{2}}$, where B is a positive constant.

Thus, by the Cauchy-Schwarz inequality,

$$E \left| \sum_{k=1}^{\infty} \theta_k h B_n^k \mathbb{1}_{A_k} \right| \leq \sum_{k=1}^{\infty} \theta_k^2 \left\| \sum_{j=N_k+1}^{2N_k} e_{-j} \right\|_2 \|B_n^k\|_2 \rho_k \leq B \sum_{k=1}^{\infty} \theta_k^2 N_k^2 \rho_k.$$

Therefore, $\sum_{k=1}^{\infty} \theta_k^2 N_k^2 \rho_k < \infty$ implies that J_n belongs to L^1 for all n . In the same way, we can prove that it also implies that $(J_n)_{n \in \mathbb{N}}$ is a Cauchy sequence in L^1 (again, the details are left to the reader). □

Proposition 1.19. *For the previously defined function h ,*

$$\sum_{i \in \mathbb{Z}} \|P_0(h \circ T^i)\|_2 < \infty \text{ if and only if } \sum_{k \geq 1} \theta_k N_k \sqrt{\rho_k} < \infty.$$

Proof.

We have, for all $i \geq 1$,

$$\begin{aligned}
P_0(h \circ T^i) &= E(h \circ T^i | \mathcal{F}_0) - E(h \circ T^i | \mathcal{F}_{-1}) \\
&= \sum_{k \geq 1} \theta_k \left[\sum_{j=N_k+1}^{2N_k} E(e_{i-j} | \mathcal{F}_0) - E(e_{i-j} | \mathcal{F}_{-1}) \right] \mathbb{1}_{A_k} \circ T^i \\
&= \sum_{k \geq 1} \theta_k e_0 \sum_{j=N_k+1}^{2N_k} \mathbb{1}_{\{i=j\}} \mathbb{1}_{A_k} + \sum_{k \geq 1} \theta_k e_0 \sum_{j=N_k+1}^{2N_k} \mathbb{1}_{\{i=j\}} (\mathbb{1}_{T^{-i}A_k} - \mathbb{1}_{A_k}) \\
&= I_1(i) + I_2(i).
\end{aligned}$$

Hence, because of assumption (1.19), for all $i \geq 1$,

$$\begin{aligned}
\|I_1(i)\|_2 &= \left(\sum_{k \geq 1} \theta_k^2 \mu(A_k) \left(\sum_{j=N_k+1}^{2N_k} \mathbb{1}_{\{i=j\}} \right)^2 \right)^{\frac{1}{2}} \\
&= \sum_{k \geq 1} \theta_k \sqrt{\mu(A_k)} \mathbb{1}_{\{i \in \{N_k+1, \dots, 2N_k\}\}}.
\end{aligned}$$

Finally, for each $k \geq 1$ there exist N_k integers i such that $i \in \{N_k + 1, \dots, 2N_k\}$, then

$$\sum_{i \geq 1} \|I_1(i)\|_2 = \sum_{k \geq 1} \theta_k N_k \sqrt{\mu(A_k)}.$$

Further, by (1.20),

$$\sum_{i \geq 1} \|I_2(i)\|_2 < \infty,$$

which concludes the proof. $\square$

1.7 Other counterexamples

Counterexample 7, proof of (vi)

We choose the sequences for the definition of the function h taking

$$\rho_k = \frac{1}{4^k}, \quad N_k = 2^k \text{ and } \theta_k = \frac{1}{k^{\frac{3}{2}}}, \quad k \geq 1.$$

Then,

$$\sum_{k=1}^{\infty} \theta_k^2 N_k \rho_k = \sum_{k=1}^{\infty} \frac{1}{2^k k^3} < \infty$$

and by Proposition 1.16, h belongs to L^2 . We also have

$$\begin{aligned}\sum_{k=1}^{\infty} \theta_k N_k^{\frac{3}{2}} \rho_k &= \sum_{k=1}^{\infty} \frac{1}{2^{\frac{k}{2}} k^{\frac{3}{2}}} < \infty, \\ \sum_{k=1}^{\infty} \theta_k^2 N_k^2 \rho_k &= \sum_{k=1}^{\infty} \frac{1}{k^3} < \infty, \\ \sum_{k \geq 1} \theta_k N_k \sqrt{\rho_k} &= \sum_{k \geq 1} \frac{1}{k^{\frac{3}{2}}} < \infty.\end{aligned}$$

Propositions 1.17, 1.18 and 1.19 show that the function h satisfies (M-C), (D-R) and (HC).

Using hypothesis (1.20) and the same observations as in the proof of Proposition 1.11, we see that the convergence of $\sum_{n=1}^{\infty} n^{-\frac{3}{2}} \|E(S_n(f)|\mathcal{F}_0)\|_2$ is equivalent to the convergence of

$$\sum_{n=1}^{\infty} n^{-\frac{3}{2}} \left(\sum_{k=1}^{\infty} \theta_k^2 E \left| \sum_{i=1}^n \sum_{j=N_k+1}^{2N_k} E(e_{-j+i}|\mathcal{F}_0) \right|^2 \rho_k \right)^{\frac{1}{2}}.$$

For all $n \geq 2N_k$,

$$\sum_{i=1}^n \sum_{j=N_k+1}^{2N_k} E(e_{-j+i}|\mathcal{F}_0) = N_k \sum_{j=0}^{N_k-1} e_{-j} + \sum_{j=0}^{N_k-1} (N_k - j) e_{-N_k-j},$$

and so,

$$E \left| \sum_{i=1}^n \sum_{j=N_k+1}^{2N_k} E(e_{-j+i}|\mathcal{F}_0) \right|^2 \geq N_k^3. \quad (1.21)$$

Here, $N_k = 2^k$, so

$$\begin{aligned}\sum_{k=1}^{\infty} \theta_k^2 E \left| \sum_{i=1}^n \sum_{j=N_k+1}^{2N_k} E(e_{-j+i}|\mathcal{F}_0) \right|^2 \rho_k &\geq \sum_{k=1}^{\lfloor \frac{\ln n}{\ln 2} \rfloor - 1} \theta_k^2 N_k^3 \rho_k \\ &= \sum_{k=1}^{\lfloor \frac{\ln n}{\ln 2} \rfloor - 1} \frac{2^k}{k^3} \\ &\geq C \frac{n}{\ln^2 n},\end{aligned}$$

where C is a positive constant. We derive that

$$\sum_{n=1}^{\infty} \frac{\|E(S_n(f)|\mathcal{F}_0)\|_2}{n^{\frac{3}{2}}} = \infty,$$

i.e. the Maxwell-Woodroffe condition does not hold and (vi) is proved.

Counterexample 8, proof of (ii)

Here we choose

$$\rho_k = \frac{1}{4^k}, \quad N_k = 4^k \quad \text{and} \quad \theta_k = \frac{1}{k2^k}, \quad k \geq 1.$$

Then, the function $h \in L^2$ and satisfies the Dedecker-Rio criterion. Indeed

$$\sum_{k=1}^{\infty} \theta_k^2 N_k \rho_k = \sum_{k=1}^{\infty} \frac{1}{4^k k^2} < \infty$$

and

$$\sum_{k=1}^{\infty} \theta_k^2 N_k^2 \rho_k = \sum_{k=1}^{\infty} \frac{1}{k^2} < \infty.$$

But,

$$\sum_{k=1}^{\infty} \theta_k N_k^{\frac{3}{2}} \rho_k = \sum_{k=1}^{\infty} \frac{1}{k} = \infty$$

and

$$\sum_{k=1}^{\infty} \theta_k N_k \sqrt{\rho_k} = \sum_{k=1}^{\infty} \frac{1}{k} = \infty.$$

Then, h does not admit a martingale-coboundary decomposition in L^1 and fails to the Hannan criterion. Further, here, $N_k = 4^k$, so by (1.21),

$$\begin{aligned} \sum_{k=1}^{\infty} \theta_k^2 E \left| \sum_{i=1}^n \sum_{j=N_k+1}^{2N_k} E(e_{-j+i} | \mathcal{F}_0) \right|^2 \rho_k &\geq \sum_{k=1}^{\lfloor \frac{\ln n}{\ln 4} \rfloor - 1} \theta_k^2 N_k^3 \rho_k \\ &= \sum_{k=1}^{\lfloor \frac{\ln n}{\ln 4} \rfloor - 1} \frac{4^k}{k^2} \\ &\geq C \frac{n}{\ln^2 n}, \end{aligned}$$

As in counterexample 1.7 we deduce that the Maxwell-Woodroffe condition does not hold and (ii) is proved.

Comments

A large part of this chapter comes from the papers by Durieu and Volný [40] and Durieu [37]. Lemma 1.7 as well as Counterexamples 5, 6 and 8 do not appear in these papers. The idea to include the Hannan criterion in the comparison was suggested by Jérôme Dedecker.

Chapitre 2

Empirical Invariance Principle for Ergodic Torus Automorphisms ; Genericity

Abstract

In this chapter, we consider the dynamical system given by an algebraic ergodic automorphism T on a torus. We study a Central Limit Theorem for the empirical process associated to the stationary process $(f \circ T^i)_{i \in \mathbb{N}}$, where f is a given $\mathbb{R}$ -valued function. We give a sufficient condition on f for this Central Limit Theorem to hold. In a second part, we prove that the distribution function of a Morse function is continuously differentiable if the dimension of the manifold is at least 3 and Hölder continuous if the dimension is 1 or 2. As a consequence, the Morse functions satisfy the empirical invariance principle, which is therefore generically verified.

Keywords: Empirical process; Partially hyperbolic dynamical system; Functional central limit theorem; Multiple mixing; Distribution function; Morse function; Genericity.

AMS Classification: 60G30; 60F17; 28D05; 37D30; 57R45.

2.1 Introduction

We are interested in the study of the convergence of an empirical process defined on a dynamical system. This kind of work already appears in the book of Billingsley [6] where a limit theorem for the empirical process defined by the transformation $x \mapsto 2x \bmod 1$ on $[0, 1]$ is established. In 2004, the case of expanding maps of the interval was studied by Collet, Martinez and Schmitt [13], using properties of the transfer operator. The reader can also see (for more examples) the paper by Dedecker and Priour [18]. Here, we deal with the empirical invariance principle (Central Limit Theorem for an empirical process) in the case of a linear transformation of the torus.

Let us begin with some definitions. Let $(X, \mathcal{A}, \mu)$ be a probability space and $T : X \longrightarrow X$ a measurable μ -invariant transformation of X (*i.e.* $\mu(T^{-1}A) = \mu(A)$, $\forall A \in \mathcal{A}$). If f is a measurable function from X to $[0, 1]$, the distribution function associated to f is $F : [0, 1] \longrightarrow [0, 1]$ defined by

$$F(t) = \mu\{f \leq t\}, \quad t \in [0, 1].$$

The empirical distribution function of order n associated to f is the function on X defined for each t by,

$$F_n(t) = \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{1}_{[0,t]}(f \circ T^i).$$

We define the empirical process $\{Y_n(t), t \in [0, 1]\}_{n \in \mathbb{N}}$ by

$$Y_n(t) = \sqrt{n} (F_n(t) - F(t)) = \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} (\mathbb{1}_{[0,t]}(f \circ T^i) - F(t)).$$

We say that the empirical invariance principle holds if $\{Y_n(t), t \in [0, 1]\}_{n \in \mathbb{N}}$ converges in distribution to a Gaussian process in the space $D([0, 1])$ of cadlag functions, provided with the Skorohod topology.

Herein X is a torus and T an ergodic algebraic automorphism. Ergodicity means that for all $A \in \mathcal{A}$, $T^{-1}(A) = A$ implies $\mu(A) = 0$ or 1 . For this kind of transformation there are two cases: T can be hyperbolic or quasi-hyperbolic (see next Section).

Actually, for hyperbolic torus automorphism, an empirical invariance principle can be deduced from known results about functionals of absolutely regular (or β -mixing) processes using a well adapted Markov partition which is weak Bernoulli.

In the quasi-hyperbolic case, according to Lind [66], no regular partition of the torus is weak Bernoulli. So the Borovkova, Burton and Dehling theorem ([8]) cannot apply. The question is the following: does the empirical invariance principle hold in this case?

In Section 2.2, we first prove that the answer is positive under a rather technical condition on f , see Theorem 2.2. The proof uses a multiple mixing property of ergodic torus automorphisms and works as well in the hyperbolic case.

In fact, the technical condition on the function f holds as soon as f and its distribution function are Hölder continuous, see Theorem 2.3.

In Section 2.3, we show that the distribution functions of Morse functions defined on a compact Riemannian manifold are at least Hölder continuous, in fact $\mathcal{C}^1$ as soon as the dimension of the manifold is at least 3, see Theorem 2.16.

This result is used in Section 2.2 in order to prove that the set of functions for which the empirical invariance principle holds contains an open and dense subset of $\mathcal{C}^r(\mathbb{T}^d)$ for $r \geq 2$, see Theorem 2.5.

2.2 Empirical Invariance Principle

Let $\mathbb{T}^d = \mathbb{R}^d / \mathbb{Z}^d$ be the d -dimensional torus and let $T : \mathbb{T}^d \rightarrow \mathbb{T}^d$ be an algebraic ergodic automorphism. The transformation T can be represented by a square matrix with integer coefficients and determinant equal to ± 1 . Here, ergodicity is equivalent to the fact that no eigenvalue of T is a root of unity. In that case, it is known that the modulus of at least one eigenvalue is strictly bigger than 1 (and the modulus of another one is strictly smaller than 1). The automorphism T is said to be *hyperbolic* if no eigenvalue has modulus 1 and *quasi-hyperbolic* (or partially hyperbolic) otherwise.

For example of an hyperbolic torus automorphism, we can consider, on $\mathbb{T}^2$, the transformation $T : x \mapsto Mx \pmod{1}$, where

$$M = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$$

which has two eigenvalues $\lambda = \frac{3}{2} - \frac{\sqrt{5}}{2}$ and λ^{-1} .

For example of quasi-hyperbolic torus automorphism, we need to consider at least the torus of dimension 4. The following example is given in Le Borgne [61]. Consider T on $\mathbb{T}^4$, defined by $T : x \mapsto Mx \pmod{1}$, where

$$M = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 \end{pmatrix}.$$

We denote by μ the Lebesgue measure on $\mathbb{T}^d$ (product measure). The aim is to get a limit theorem for the empirical process associated to a function f defined on the torus.

2.2.1 Hyperbolic case

Assume that the automorphism T is hyperbolic. One can show that the empirical invariance principle holds for Hölder continuous function f having a Lipschitz continuous distribution function. This is a consequence of Borovkova, Burton and Dehling [8], Theorem 5, which is recalled here.

We say that a process $(X_k)_{k \in \mathbb{Z}}$ is a *functional* of a process $(Z_k)_{k \in \mathbb{Z}}$ if there exists a function $g : \mathbb{R}^{\mathbb{Z}} \rightarrow \mathbb{R}$ such that for any $k \in \mathbb{Z}$, $X_k = g(Z_{n+k})_{n \in \mathbb{Z}}$. By definition, a stationary process $(X_k)_{k \in \mathbb{Z}}$ is a *1-approximating functional* of $(Z_k)_{k \in \mathbb{Z}}$ if there exist nonnegative constants $(a_k)_{k \geq 0}$ with $a_k \rightarrow 0$ as $k \rightarrow \infty$ and

$$E |X_0 - E(X_0 | Z_{-k}, \dots, Z_k)| \leq a_k$$

holds for any $k \geq 0$.

Recall also, that the sequence of random variables $(Z_n)_{n \in \mathbb{Z}}$ is *β -mixing* if $\beta_k \rightarrow 0$,

$$\beta_k := \sup_{n \in \mathbb{N}} \sup_{\substack{\mathcal{A} \in \mathcal{P}(\mathcal{A}_1^n) \\ \mathcal{B} \in \mathcal{P}(\mathcal{A}_{n+k}^\infty)}} \sum_{A \in \mathcal{A}} \sum_{B \in \mathcal{B}} |\mu(A \cap B) - \mu(A)\mu(B)|,$$

where $\mathcal{P}(\mathcal{A}_k^l) = \{\mathcal{A}_k^l\text{-measurable partitions}\}$.

Theorem (Borovkova, Burton and Dehling, 2001). *Let $(X_k)_{k \in \mathbb{Z}}$ and $(Z_k)_{k \in \mathbb{Z}}$ be two stationary processes.*

If, for some $0 < \delta < 1$,

$$(i) \ (Z_n)_{n \in \mathbb{Z}} \text{ is } \beta\text{-mixing with } \sum_{k \geq 0} k^2 \beta_k^{\frac{\delta}{2+\delta}} < \infty,$$

$$(ii) \ (X_n)_{n \in \mathbb{Z}} \text{ is a 1-approximation of } (Z_n)_{n \in \mathbb{Z}}, \text{ with } \sum_{k \geq 0} k^2 a_k^{\frac{\delta}{2+\delta}} < \infty,$$

$$(iii) \ \text{the distribution function of } X_0 : \Omega \rightarrow [0, 1] \text{ is Lipschitz continuous,}$$

then the empirical invariance principle holds for $(X_n)_{n \in \mathbb{Z}}$.

To apply this theorem, the strategy is to encode the dynamical system into a stationary process $(Z_i)_{i \in \mathbb{Z}}$ having a β -mixing property. Then the stationary process $(f \circ T^i)_{i \in \mathbb{Z}}$ will be viewed as a functional of $(Z_i)_{i \in \mathbb{Z}}$ and one can show that it is a 1-approximation with coefficients decreasing to zero exponentially fast. This can be done using a well adapted Markov partition of the torus. Since the transformation T is hyperbolic, there exist two stable sub-spaces E^s and E^u associated to the eigenvalues with modulus smaller than 1

(resp. eigenvalues with modulus greater than 1) such that $R^d = E^s \oplus E^u$. A collection $\mathcal{P} = \{R_1, \dots, R_n\}$ of compact rectangles along these two directions is a *Markov partition* if

- (i) $\bigcup_{i=1}^n R_i = \mathbb{T}^d$;
- (ii) for all $i \neq j$, $\text{int}(R_i \cap R_j) = \emptyset$;
- (iii) for all i , there exists j such that $\partial^s T R_i \subset \partial^s R_j$;
- (iv) for all i , there exists j such that $\partial^u T R_i \supset \partial^u R_j$;

where for any subset A , the notation ∂A stands for the boundary of A and if $R = A \times B$ is a rectangle along the directions E^s and E^u , $\partial^s R = \partial A \times B$ and $\partial^u R = A \times \partial B$.

According to Bowen [9], every hyperbolic torus automorphism admits a Markov partition. In particular, if $\mathcal{P} = \{R_1, \dots, R_n\}$ is a Markov partition and if $\mathcal{P}_k^l = \bigvee_{j=k}^l T^{-j} \mathcal{P}$, then $\sigma(\mathcal{P}_{-\infty}^\infty) = \mathcal{B}$ and $\text{diam}(\mathcal{P}_{-l}^l) \leq C\lambda^l$ for $\lambda < 1$. We say that the partition is *regular*. Further, if

$$Z_0 = \begin{cases} \mathbb{T}^d & \longrightarrow \{1, \dots, n\} \\ x & \longmapsto i \text{ if } x \in R_i \end{cases}$$

and $Z_i = Z_0 \circ T^i$, again by Bowen [9], $(Z_i)_{i \in \mathbb{Z}}$ is a β -mixing stationary process with exponential rate (we say that the partition $\mathcal{P}$ is weak Bernoulli).

Now, if $\alpha > 0$ and $C > 0$ are the Hölder constants of a Hölder continuous function f which is defined on the torus,

$$E |f - E(f|\mathcal{P}_{-l}^l)| \leq C \text{diam}(\mathcal{P}_{-l}^l)^\alpha \leq C\lambda^{\alpha l}.$$

Thus the process $(f \circ T^i)_{i \in \mathbb{Z}}$ is a 1-approximation of $(Z_i)_{i \in \mathbb{Z}}$ with exponential rate. Therefore, we proved the following corollary.

Corollary 2.1. *Let T be an hyperbolic torus automorphism, $f : \mathbb{T}^d \longrightarrow [0, 1]$ a Hölder-continuous function such that its distribution function F is Lipschitz. Then the empirical invariance principle holds for the process $(f \circ T^i)_{i \in \mathbb{Z}}$.*

Note that this can be done, in general, for the class of Anosov diffeomorphisms.

2.2.2 Quasi-hyperbolic case

Let T be a quasi-hyperbolic torus automorphism. According to Lind [66], Theorem 4, we cannot find a regular partition of the torus which is weak Bernoulli. So, we cannot use the same techniques. Nevertheless, we are able to prove that the empirical invariance principle holds as well in the quasi-hyperbolic case.

As hyperbolicity is a particular case of quasi-hyperbolicity, the following statements and proofs work in both cases.

For any subset A of the torus, the notation $\partial_\varepsilon A$ stands the " ε -boundary" of A :

$$\partial_\varepsilon A := \{x/d(x, \partial A) \leq \varepsilon\}$$

where d is the Euclidian metric on the torus.

Our main result is the following.

Theorem 2.2. *Let $f : \mathbb{T}^d \longrightarrow \mathbb{R}$ be a bounded measurable function satisfying the condition*

$$\exists C, \xi > 0 \text{ such that } \forall t \in \mathbb{R} \text{ and } \forall \varepsilon > 0, \quad \mu(\partial_\varepsilon \{f \leq t\}) \leq C\varepsilon^\xi. \quad (*)$$

Moreover, its distribution function F is assumed to be continuous.

Then the process $(f \circ T^i)_{i \in \mathbb{N}}$ verifies the empirical invariance principle:

let $[a, b]$ be a compact interval of $\mathbb{R}$ such that $f(\mathbb{T}^d) \subset [a, b]$,

$$\{Y_n(t) = \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} (\mathbb{1}_{[a,t]}(f \circ T^i) - F(t)) , t \in [a, b]\}_{n \in \mathbb{N}}$$

converges in distribution to a Gaussian process Y in $D([a, b])$. Further, Y is μ -almost surely continuous.

This theorem is proved in Section 2.2.3.

The following theorem shows that even in the hyperbolic case, our result is slightly more general than the one deduced from the theorem of Borovkova et al.

Theorem 2.3. *If $f : \mathbb{T}^d \longrightarrow \mathbb{R}$ is a Hölder continuous function having a Hölder continuous distribution function, then the process $(f \circ T^i)_{i \in \mathbb{N}}$ satisfies the empirical invariance principle.*

Proof.

Let $f : \mathbb{T}^d \longrightarrow \mathbb{R}$ be a function satisfying the assumptions of Theorem 2.3 and F its distribution function. We will show that Condition $(*)$ holds.

For $\varepsilon > 0$ and $t \in \mathbb{R}$, the set $\partial_\varepsilon \{f \leq t\}$ is the union of the two following disjoint sets:

$$\partial_\varepsilon \{f \leq t\}^+ := \partial_\varepsilon \{f \leq t\} \cap \{f > t\},$$

$$\partial_\varepsilon\{f \leq t\}^- := \partial_\varepsilon\{f \leq t\} \cap \{f \leq t\}.$$

First of all, the continuity of the function f implies that $\partial\{f \leq t\} \subset \{f = t\}$.

Now, for $x \in \partial_\varepsilon\{f \leq t\}^+$, $f(x) \leq t + K\varepsilon^\xi$, where K and ξ denote the Hölder constants of f . We have

$$\begin{aligned} \mu(\partial_\varepsilon\{f \leq t\}^+) &\leq \mu(f \leq t + K\varepsilon^\xi) - \mu(f \leq t) \\ &= F(t + K\varepsilon^\xi) - F(t) \\ &\leq L(K\varepsilon^\xi)^\zeta \end{aligned}$$

where L and ζ are the Hölder constants of F .

We obtain the same result for $\partial_\varepsilon\{f \leq t\}^-$ and thus

$$\mu(\partial_\varepsilon\{f \leq t\}) \leq 2LK^\zeta \varepsilon^{\xi\zeta}.$$

□

Remark 2.4. *The fact that the function is Hölder continuous is not enough to get Condition (*), even if its distribution function is continuous. A counterexample is given in Section 2.2.4.*

Now Theorem 2.3 together with the upcoming result of Section 2.3 (Theorem 2.16) leads to the following genericity result.

Theorem 2.5. *The set of functions f for which $(f \circ T^i)_{i \in \mathbb{N}}$ satisfies the empirical invariance principle contains an open and dense subset of $C^r(\mathbb{T}^d)$ for $r \geq 2$.*

2.2.3 Proof of Theorem 2.2

We will deduce Theorem 2.2 from the following:

Theorem 2.6. *Let $f : \mathbb{T}^d \longrightarrow [0, 1]$ be a uniformly distributed function such that Condition (*) is satisfied.*

Then $\{Y_n(t) = \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} (\mathbb{1}_{[0,t]}(f \circ T^i) - t), t \in [0, 1]\}_{n \in \mathbb{N}}$ converges in distribution to a μ -almost surely continuous Gaussian process in $D([0, 1])$.

Theorem 2.6 implies Theorem 2.2.

Let $f : \mathbb{T}^d \longrightarrow \mathbb{R}$ be a bounded function verifying Condition (*) with a continuous distribution function F . We can assume $f(\mathbb{T}^d) \subset [0, 1]$ without loss of generality.

Let us defined as usual the pseudo-inverse function of the distribution function by $F^{-1}(t) = \sup\{s \in [0, 1] / F(s) \leq t\}$, $t \in [0, 1]$. Then, thanks to the continuity of F , $F \circ F^{-1} = \text{id}$. Let us consider the function $g := F \circ f$. We have

$$\{g \leq t\} = \{F \circ f \leq t\} = \{f \leq F^{-1}(t)\}.$$

thanks again to the continuity of F . From this equality we deduce that g is uniformly distributed and satisfies Condition $(*)$ (with the same constants as f).

Let us denote by Y_n and Z_n the empirical processes respectively defined by the functions f and g . For each $t \in [0, 1]$, the equality $Y_n(t) = Z_n \circ F(t)$ holds μ -almost surely, because

$$\mu \{ \mathbb{1}_{[0,t]}(f) = \mathbb{1}_{[0,F(t)]}(g) \} = 1.$$

Moreover the equality $\mathbb{1}_{[0,t]}(f) = \mathbb{1}_{[0,F(t)]}(g)$ holds as soon as F^{-1} does not have a jump at t . Therefore the process equality $Y_n = Z_n \circ F$ holds μ -almost surely.

By Theorem 2.6, Z_n converges in distribution to an almost surely continuous Gaussian process Z . We recall the following result (see Theorem 5.1 in Billingsley [6]):

Let $\psi : D([0, 1]) \longrightarrow D([0, 1])$ and D_ψ the set of discontinuities of ψ . If a sequence of random variables h_n converges in distribution to a variable h such that $\mu(h \in D_\psi) = 0$, then $\psi(h_n)$ converges in distribution to $\psi(h)$.

Let us consider the mapping $\psi : h \longmapsto h \circ F$ from $D([0, 1])$ to $D([0, 1])$. It is continuous on $C([0, 1])$ for the induced topology which is the topology of the uniform convergence (see Billingsley [6] p.112). The theorem of Billingsley applied to ψ gives the result for Y_n and Theorem 2.2 is proved. $\square$

Proof of Theorem 2.6.

We consider a uniformly distributed function $f : \mathbb{T}^d \longrightarrow [0, 1]$ (i.e. $F \equiv \text{Id}_{[0,1]}$).

Since we cannot find a regular partition of the torus which is weak Bernoulli, our proof will not use the theorem of Borovkova et al. To get the invariance principle, the proof consists in two steps. First we get the finite-dimensional Central Limit Theorem and then we show the tightness of the process using a multiple mixing inequality to prove a suitable 4-th moment bound.

The following proposition was proved by Le Borgne [61].

Proposition 2.7. *Let $T : \mathbb{T}^d \longrightarrow \mathbb{T}^d$ be an ergodic torus automorphism. If a measurable set A verifies:*

$$\exists C, \xi > 0, \forall \varepsilon > 0, \mu(\partial_\varepsilon A) \leq C\varepsilon^\xi,$$

then the Central Limit Theorem holds for the function $\mathbb{1}_A - \mu(A)$.

For each $t \in [0, 1]$, we denote by φ_t the function

$$\varphi_t = \mathbb{1}_{\{f \leq t\}} - t.$$

Applying Proposition 2.7, we get the following one.

Proposition 2.8. *If a function f satisfies Condition (*), then for all $t \in [0, 1]$, φ_t satisfies the Central Limit Theorem. i.e.*

$$\frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} \varphi_t \circ T^i \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, \sigma).$$

It follows easily that the same is true for the finite dimensional case.

Proposition 2.9. *If a function f satisfies Condition (*), then for all $k \in \mathbb{N}$ and for all $(t_1, \dots, t_k) \in [0, 1]^k$, $(Y_n(t_1), \dots, Y_n(t_k))$ converges in distribution to a k -dimensional Gaussian vector.*

Now it remains to show that the process $\{Y_n(t), t \in [0, 1]\}_{n \in \mathbb{N}}$ is tight. Following Billingsley [6], it is sufficient to prove:

$$\forall \varepsilon > 0, \forall \eta > 0, \exists \zeta \in]0, 1[, \exists n_0 \in \mathbb{N} \text{ such that } \forall n \geq n_0,$$

$$\mu \left(\sup_{|t-s| \leq \zeta} |Y_n(t) - Y_n(s)| \geq \varepsilon \right) \leq \eta. \quad (2.1)$$

The first step is to establish the following lemma.

Lemma 2.10. *There exist $C > 0$ and $\delta \in (0, 1)$ such that for all $s < t \in [0, 1]$, if $\varphi = \varphi_t - \varphi_s$,*

$$E \left(\left(\sum_{i=1}^n \varphi \circ T^i \right)^4 \right) \leq C \left(n^2 (t-s)^{\frac{2}{1+\delta}} + n (t-s)^{\frac{1}{1+\delta}} \right).$$

To prove Lemma 2.10, we will use a multiple mixing property of ergodic torus automorphisms.

A version of the following proposition (in the case of diagonal flows and for Hölder observables) can be found in the paper by Le Borgne [63] (see also Le Borgne and Pène [64]). Here, it is adapted to the case of ergodic torus automorphisms and for $\mathcal{C}^1$ functions. For a proof of the next proposition, the reader can see the appendix.

Proposition 2.11. $\exists C > 0, \exists \alpha > 0, \forall m, p \in \mathbb{N}^*, \forall \phi_1, \dots, \phi_m, \psi_1, \dots, \psi_p \in \mathcal{C}^1(\mathbb{T}^d), \forall k_1 \leq \dots \leq k_m \leq 0 \leq l_1 \leq \dots \leq l_p, \forall n \in \mathbb{N},$

$$\left| \text{Cov} \left(\prod_{j=1}^m \phi_j \circ T^{k_j}, \prod_{j=1}^p \psi_j \circ T^{l_j+n} \right) \right| \leq C \left(\sum_{j=1}^m \prod_{i \neq j} \|\phi_i\|_\infty \|\phi_j\|_{\mathcal{C}^1} |k_j|^r \right) \left(\sum_{j=1}^p \prod_{i \neq j} \|\psi_i\|_\infty \|\psi_j\|_{\mathcal{C}^1} \right) e^{-\alpha n}$$

where r is the size of the biggest Jordan's block of T restricted to its neutral subspace.

In Lemma 2.10, φ is a discontinuous function. We will use the fact that the boundary of the set $\{s < f \leq t\}$ is enough regular (Condition $(*)$) to approximate φ by $\mathcal{C}^1$ functions. In the sequel, C always denotes a constant, but its value may change.

Lemma 2.12. *Under Condition $(*)$, there exists $C > 0$, for all $\beta > 0$, for all $s, t \in [0, 1]$, there exists a sequence $(\varphi_k)_{k \in \mathbb{N}}$ of $\mathcal{C}^1$ functions such that*

$$\|\varphi_k\|_{\mathcal{C}^1} \leq C e^{\beta k} \quad \text{and} \quad \|\varphi - \varphi_k\|_{L^p} \leq C e^{-\frac{\gamma k}{p}}, \quad \forall p \geq 1$$

where $\gamma = \frac{\xi\beta}{d+1}$ and $\varphi = \varphi_t - \varphi_s$.

Proof.

The considered balls are defined with respect to the Euclidian norm of $\mathbb{R}^d$ which is inducted on the torus by identifying $\mathbb{T}^d$ to $[0, 1]^d$. Let $\rho : \mathbb{T}^d \longrightarrow [0, +\infty)$ be a $\mathcal{C}^1$ function such that $E(\rho) = 1$ and ρ equals 0 outside $B(0, \frac{1}{2})$.

Write

$$\rho_k(x) = \begin{cases} e^{\frac{\beta k d}{d+1}} \rho(e^{\frac{\beta k}{d+1}} x) & \text{if } x \in B(0, \frac{1}{2} e^{-\frac{\beta k}{d+1}}) \\ 0 & \text{else} \end{cases}.$$

Then ρ_k is $\mathcal{C}^1$, $\rho_k^{-1}((0, +\infty)) \subset B(0, \frac{1}{2} e^{-\frac{\beta k}{d+1}})$, $E(\rho_k) = 1$ and for $i = 1, \dots, d$,

$$\left\| \frac{\partial \rho_k}{\partial x_i} \right\|_\infty \leq e^{\frac{\beta k d}{d+1}} e^{\frac{\beta k}{d+1}} \left\| \frac{\partial \rho}{\partial x_i} \right\|_\infty \leq C e^{\beta k}.$$

Write $\varphi_k = \varphi * \rho_k$. Then

$$\|\varphi_k\|_{\mathcal{C}^1} = \|\varphi_k\|_\infty + \max_{i=1}^d \left\| \varphi * \frac{\partial \rho_k}{\partial x_i} \right\|_\infty \leq 1 + \max_{i=1}^d \left\| \frac{\partial \rho_k}{\partial x_i} \right\|_\infty \leq C e^{\beta k}$$

and

$$\begin{aligned}
\|\varphi_k - \varphi\|_p^p &= E(|\varphi_k - \varphi|^p) \leq E|\varphi_k - \varphi| \\
&\leq \|\varphi_k - \varphi\|_\infty \mu\left(\partial_{\frac{1}{2}e^{-\frac{\beta k}{d+1}}} \{f \in (s, t]\}\right) \\
&\leq Ce^{-\frac{\beta k}{d+1}\xi} \\
&= Ce^{\gamma k}.
\end{aligned}$$

□

The function φ is always bounded by 1. From this, we deduce the following remark which is useful in the proof of Lemma 2.10.

Remark 2.13. For all $p \geq 1$, $E(|\varphi|^p) \leq 2|t - s|$.

Proof of Lemma 2.10.

Developing the term $E\left(\left(\sum_{i=1}^n \varphi \circ T^i\right)^4\right)$, we will have to study terms like

$$E(\varphi \varphi \circ T^i \varphi \circ T^{i+j} \varphi \circ T^{i+j+k})$$

with $i + j + k \leq n$. We fix three integers i, j, k like this. Note that,

$$E(\varphi \varphi \circ T^i \varphi \circ T^{i+j} \varphi \circ T^{i+j+k}) = \text{Cov}(\varphi, (\varphi \varphi \circ T^j \varphi \circ T^{j+k}) \circ T^i).$$

We wish to apply Proposition 2.11. To do that, we consider the sequence of $\mathcal{C}^1$ functions $(\varphi_l)_{l \in \mathbb{N}}$ defined by Lemma 2.12. We have

$$\text{Cov}(\varphi, (\varphi \varphi \circ T^j \varphi \circ T^{j+k}) \circ T^i) = E(\varphi_i (\varphi_i \varphi_i \circ T^j \varphi_i \circ T^{j+k}) \circ T^i) \quad (2.2)$$

$$+ E((\varphi - \varphi_i) (\varphi \varphi \circ T^j \varphi \circ T^{j+k}) \circ T^i) \quad (2.3)$$

$$+ E(\varphi_i ((\varphi - \varphi_i) \varphi \circ T^j \varphi \circ T^{j+k}) \circ T^i) \quad (2.4)$$

$$+ E(\varphi_i (\varphi_i (\varphi - \varphi_i) \circ T^j \varphi \circ T^{j+k}) \circ T^i) \quad (2.5)$$

$$+ E(\varphi_i (\varphi_i \varphi_i \circ T^j (\varphi - \varphi_i) \circ T^{j+k}) \circ T^i). \quad (2.6)$$

We will distinguish two cases.

First case : $|t - s| \geq e^{-\gamma i}$.

Notice that γ is the constant equal to $\frac{\xi\beta}{d+1}$ that appears in Lemma 2.12. For the seek of clarity it will be fixed farther.

Let $p, q \geq 1$ such that $\frac{1}{p} + \frac{3}{q} = 1$. By Hölder inequality and Lemma 2.12,

$$\begin{aligned} |(2.3)| &\leq \|\varphi - \varphi_i\|_p \|\varphi\|_q^3 \\ &\leq Ce^{-\frac{\gamma_i}{p}} \|\varphi\|_q^3 \\ &\leq Ce^{-\frac{\gamma_i}{p}} 2^{\frac{3}{q}} |t - s|^{\frac{3}{q}}. \end{aligned}$$

On the other hand,

$$|(2.6)| \leq \|\varphi - \varphi_i\|_p \|\varphi_i\|_q^3$$

and

$$\|\varphi_i\|_q \leq \|\varphi - \varphi_i\|_q + \|\varphi\|_q \leq Ce^{-\frac{\gamma_i}{q}} + \|\varphi\|_q.$$

Thus by assumption,

$$\begin{aligned} |(2.6)| &\leq Ce^{-\frac{\gamma_i}{p}} \left[Ce^{-\frac{\gamma_i}{q}} + \|\varphi\|_q \right]^3 \\ &\leq Ce^{-\frac{\gamma_i}{p}} \left[C|t - s|^{\frac{1}{q}} + 2^{\frac{1}{q}} |t - s|^{\frac{1}{q}} \right]^3 \\ &\leq Ce^{-\frac{\gamma_i}{p}} |t - s|^{\frac{3}{q}}. \end{aligned}$$

In the same way,

$$|(2.4)| \leq \|\varphi - \varphi_i\|_p \|\varphi_i\|_q \|\varphi\|_q^2 \leq Ce^{-\frac{\gamma_i}{p}} |t - s|^{\frac{3}{q}}.$$

The same thing is true for (2.5). We get

$$|(2.3) + (2.4) + (2.5) + (2.6)| \leq Ce^{-\frac{\gamma_i}{p}} |t - s|^{\frac{3}{q}}. \quad (2.7)$$

Now $(2.2) = \text{Cov}(\varphi_i, (\varphi_i \varphi_i \circ T^j \varphi_i \circ T^{j+k}) \circ T^i)$, because $E(\varphi_i) = 0$. So we can apply the multiple mixing inequality. There exist constants $C > 0$ and $\alpha > 0$ which depend only on T , such that

$$\begin{aligned} |(2.2)| &\leq C \|\varphi_i\|_{C^1} \|\varphi_i\|_{\infty}^2 \|\varphi_i\|_{C^1} e^{-\alpha i} \\ &\leq Ce^{2\beta i} e^{-\alpha i} \end{aligned}$$

because $\|\varphi_i\|_{\infty} \leq 1$. As $|t - s| \geq e^{-\gamma_i}$,

$$|(2.2)| \leq Ce^{-\eta i} |t - s|^{\frac{3}{q}} \quad (2.8)$$

where $\eta = \alpha - 2\beta - \frac{3\gamma}{q}$.

The constant β being arbitrary, we can choose $\beta > 0$ such that $\frac{\xi + 2(d+1)}{d+1} \beta < \frac{\alpha}{2}$. Then, for all $q \geq 3$, $\eta > \frac{\alpha}{2} > 0$. Now, we can fix $\delta \in (0, 1)$ and choose $p = \frac{1+\delta}{\delta}$ (so,

$q = 3 + 3\delta$) in order to have $\frac{\gamma}{p} < \eta$. This is possible because $p \rightarrow \infty$ when $\delta \rightarrow 0$. We get $e^{-\eta} \leq e^{-\frac{\gamma}{p}}$ and by (2.7) and (2.8),

$$\begin{aligned} |E(\varphi(\varphi \varphi \circ T^j \varphi \circ T^{j+k}) \circ T^i)| &\leq C|t-s|^{\frac{3}{q}} e^{-\frac{\gamma i}{p}} \\ &= C|t-s|^{\frac{1}{1+\delta}} e^{-\frac{\gamma \delta i}{1+\delta}}. \end{aligned}$$

Second case : $|t-s| < e^{-\gamma i}$.

By Hölder inequality with p and q previously fixed,

$$\begin{aligned} |E(\varphi(\varphi \varphi \circ T^j \varphi \circ T^{j+k}) \circ T^i)| &\leq \|\varphi\|_p \|\varphi\|_q^3 \\ &\leq 2^{\frac{1}{p}} |t-s|^{\frac{1}{p}} 2^{\frac{3}{q}} |t-s|^{\frac{3}{q}} \\ &\leq C e^{-\frac{\gamma i}{p}} |t-s|^{\frac{3}{q}} \\ &= C|t-s|^{\frac{1}{1+\delta}} e^{-\frac{\gamma \delta i}{1+\delta}}. \end{aligned}$$

In each case, we have

$$|E(\varphi \varphi \circ T^i \varphi \circ T^{i+j} \varphi \circ T^{i+j+k})| \leq C|t-s|^{\frac{1}{1+\delta}} e^{-\frac{\gamma \delta}{1+\delta} i}. \quad (2.9)$$

In the same way, playing with k instead of i , we get

$$\begin{aligned} |E(\varphi \varphi \circ T^i \varphi \circ T^{i+j} \varphi \circ T^{i+j+k})| &= |\text{Cov}((\varphi \varphi \circ T^i \varphi \circ T^{i+j}), \varphi \circ T^{i+j} \circ T^k)| \\ &\leq C|t-s|^{\frac{1}{1+\delta}} (1+(i+j)^r) e^{-\frac{\gamma \delta}{1+\delta} k} \end{aligned} \quad (2.10)$$

where the term $(i+j)^r$ is related to the neutral subspace of T and appears in application of Proposition 2.11.

Now, with j , there is a remainder term:

$$\begin{aligned} &|E(\varphi \varphi \circ T^i \varphi \circ T^{i+j} \varphi \circ T^{i+j+k})| \\ &= |\text{Cov}((\varphi \varphi \circ T^i), (\varphi \circ T^i \varphi \circ T^{i+k}) \circ T^j) + E(\varphi \varphi \circ T^i) E(\varphi \varphi \circ T^k)| \\ &\leq C|t-s|^{\frac{1}{1+\delta}} (1+i^r) e^{-\frac{\gamma \delta}{1+\delta} j} + |E(\varphi \varphi \circ T^i) E(\varphi \varphi \circ T^k)| \end{aligned}$$

Again the term i^r appears in application of Proposition 2.11.

So, we study

$$\begin{aligned} E(\varphi \varphi \circ T^i) &= E(\varphi_i \varphi_i \circ T^i) \\ &\quad + E((\varphi - \varphi_i) \varphi \circ T^i) \\ &\quad + E(\varphi_i (\varphi - \varphi_i) \circ T^i). \end{aligned}$$

By Proposition 2.11 (or by the exponential mixing inequality (see Lind [66])),

$$|E(\varphi_i \varphi_i \circ T^i)| \leq C \|\varphi_i\|_{\mathcal{C}^1}^2 e^{-\alpha i} \leq C e^{2\beta i - \alpha i}.$$

As above, considering two cases and keeping the same notations, we have

$$|E(\varphi \varphi \circ T^i)| \leq C |t - s|^{\frac{1}{1+\delta}} e^{-\frac{\gamma\delta}{1+\delta} i}$$

and the same inequality holds for k . Thus we get

$$\begin{aligned} |E(\varphi \varphi \circ T^i \varphi \circ T^{i+j} \varphi \circ T^{i+j+k})| &\leq C |t - s|^{\frac{1}{1+\delta}} (1 + (i+k)^r) e^{-\frac{\gamma\delta}{1+\delta} j} \\ &\quad + C |t - s|^{\frac{2}{1+\delta}} e^{-\frac{\gamma\delta}{1+\delta} (i+k)}. \end{aligned} \quad (2.11)$$

Now we can obtain the desired majoration.

$$\begin{aligned} E \left(\left(\sum_{i=1}^n \varphi \circ T^i \right)^4 \right) &\leq 4!n \sum_{\{i,j,k:i+j+k \leq n\}} |E(\varphi \varphi \circ T^i \varphi \circ T^{i+j} \varphi \circ T^{i+j+k})| \\ &\leq 4!n \left[\sum_{i=1}^n \sum_{j,k \leq i} |E(\varphi (\varphi \varphi \circ T^j \varphi \circ T^{j+k}) \circ T^i)| \right. \\ &\quad + \sum_{j=1}^n \sum_{i,k \leq j} |E((\varphi \varphi \circ T^i)(\varphi \circ T^i \varphi \circ T^{i+k}) \circ T^j)| \\ &\quad \left. + \sum_{k=1}^n \sum_{i,j \leq k} |E((\varphi \varphi \circ T^i \varphi \circ T^{i+j}) \varphi \circ T^{i+j} \circ T^k)| \right] \\ &= 4!n [\quad I_1 \quad + \quad I_2 \quad + \quad I_3 \quad]. \end{aligned}$$

By (2.9),

$$\begin{aligned} I_1 &\leq C \left[\sum_{i=1}^n \sum_{j,k \leq i} e^{-\frac{\gamma\delta}{1+\delta} i} \right] |t - s|^{\frac{1}{1+\delta}} \\ &\leq C \left[\sum_{i=1}^n i^2 e^{-\frac{\gamma\delta}{1+\delta} i} \right] |t - s|^{\frac{1}{1+\delta}} \\ &\leq C |t - s|^{\frac{1}{1+\delta}} \end{aligned}$$

because the series converges.

By (2.10),

$$\begin{aligned}
I_3 &\leq C \left[\sum_{k=1}^n \sum_{i,j \leq k} (1 + (i+j)^r) e^{-\frac{\gamma\delta}{1+\delta}k} \right] |t-s|^{\frac{1}{1+\delta}} \\
&\leq C \left[\sum_{k=1}^n k^2 (1 + (2k)^r) e^{-\frac{\gamma\delta}{1+\delta}k} \right] |t-s|^{\frac{1}{1+\delta}} \\
&\leq C |t-s|^{\frac{1}{1+\delta}}.
\end{aligned}$$

Finally, by (2.11),

$$\begin{aligned}
I_2 &\leq C \left[\sum_{j=1}^n \sum_{i,k \leq j} (1 + i^r) e^{-\frac{\gamma\delta}{1+\delta}j} \right] |t-s|^{\frac{1}{1+\delta}} + C \left[\sum_{j=1}^n \sum_{i,k \leq j} e^{-\frac{\gamma\delta}{1+\delta}(i+k)} \right] |t-s|^{\frac{2}{1+\delta}} \\
&\leq C \left[\sum_{j=1}^n j^2 (1 + j^r) e^{-\frac{\gamma\delta}{1+\delta}j} \right] |t-s|^{\frac{1}{1+\delta}} + C \left[n \left(\sum_{i=1}^n e^{-\frac{\gamma\delta}{1+\delta}i} \right) \left(\sum_{k=1}^n e^{-\frac{\gamma\delta}{1+\delta}k} \right) \right] |t-s|^{\frac{2}{1+\delta}} \\
&\leq C |t-s|^{\frac{1}{1+\delta}} + Cn |t-s|^{\frac{2}{1+\delta}}
\end{aligned}$$

because, again, the series converge.

In conclusion, there exists $\delta > 0$ such that for all $s, t \in [0, 1]$,

$$E \left(\left(\sum_{i=1}^n \varphi \circ T^i \right)^4 \right) \leq C \left(n^2 |t-s|^{\frac{2}{1+\delta}} + n |t-s|^{\frac{1}{1+\delta}} \right).$$

□

Lemma 2.10 is the key inequality for the sequel. Now, the method leading to the tightness of the process $(Y_n)_{n \in \mathbb{N}}$ is the classical chaining one. The reader can see, for more details, the paper by Dehling and Philipp [21]. For convenience, we expose here the main steps of the method.

To check Condition (2.1), we study $|Y_n(t) - Y_n(s)|$ for $s \leq t \leq s + \zeta$ where ζ is to be determined. The idea is to introduce a subdivision of step h of the interval, where h is also to be determined. Next Lemma is proved in Billingsley [6].

Lemma 2.14. $\forall h \in [0, 1]$ and $\forall s \leq t \leq s + h$,

$$|Y_n(t) - Y_n(s)| \leq |Y_n(s+h) - Y_n(s)| + h\sqrt{n}.$$

We deduce

Lemma 2.15. $\forall s, h \in [0, 1)$ and $\forall m \in \mathbb{N}$ such that $s + mh \leq 1$,

$$\sup_{s \leq t \leq s+mh} |Y_n(t) - Y_n(s)| \leq 3 \max_{i \leq m} |Y_n(s + ih) - Y_n(s)| + h\sqrt{n}.$$

The proof is left to the reader.

Now, assume that $h < \frac{\varepsilon}{\sqrt{n}}$. Then Lemma 2.15 gives

$$\begin{aligned} \mu \left(\sup_{s \leq t \leq s+mh} |Y_n(t) - Y_n(s)| \geq 4\varepsilon \right) &\leq \mu \left(3 \max_{i \leq m} |Y_n(s + ih) - Y_n(s)| + h\sqrt{n} \geq 4\varepsilon \right) \\ &\leq \mu \left(\max_{i \leq m} |Y_n(s + ih) - Y_n(s)| \geq \varepsilon \right). \end{aligned} \quad (2.12)$$

Lemma 2.10 gives: $\forall s, t \in [0, 1]$,

$$E \left((Y_n(t) - Y_n(s))^4 \right) \leq C \left(|t - s|^{\frac{2}{1+\delta}} + \frac{1}{n} |t - s|^{\frac{1}{1+\delta}} \right).$$

Applying this with $s + ih$ and $s + (i + k)h$, we get

$$E \left((Y_n(s + (i + k)h) - Y_n(s + ih))^4 \right) \leq C \left((kh)^{\frac{2}{1+\delta}} + \frac{1}{n} (kh)^{\frac{1}{1+\delta}} \right).$$

We can choose $h \geq \frac{\varepsilon}{n^{1+\delta}}$ (for n large enough). Thus, we get $h^{\frac{1}{1+\delta}} \geq \frac{\varepsilon^{\frac{1}{1+\delta}}}{n} \geq \frac{\varepsilon}{n}$ ($\varepsilon < 1$) and

$$E \left((Y_n(s + (i + k)h) - Y_n(s + ih))^4 \right) \leq C \left((kh)^{\frac{2}{1+\delta}} + \frac{(kh)^{\frac{2}{1+\delta}}}{\varepsilon} \right) \leq \frac{2C}{\varepsilon} (kh)^{\frac{2}{1+\delta}}.$$

According to Billingsley [6], Theorem 12.2,

$$\mu \left(\max_{i \leq m} |Y_n(s + ih) - Y_n(s)| \geq \varepsilon \right) \leq \frac{C(mh)^{\frac{2}{1+\delta}}}{\varepsilon^5}. \quad (2.13)$$

From (2.12) and (2.13), we derive

$$\mu \left(\sup_{s \leq t \leq s+mh} |Y_n(t) - Y_n(s)| \geq 4\varepsilon \right) \leq \frac{C(mh)^{\frac{2}{1+\delta}}}{\varepsilon^5}. \quad (2.14)$$

In conclusion, let ε, η be fixed, δ defined by Lemma 2.10, and ζ belonging to $[0, 1]$ such that $\zeta < \left(\frac{\eta \varepsilon^5}{C} \right)^{\frac{1+\delta}{2}}$. Pick n_0 large enough to verify

$$\frac{\zeta n_0^{1+\delta}}{\varepsilon} - \frac{\zeta \sqrt{n_0}}{\varepsilon} \geq 1.$$

Then for all $n \geq n_0$, there exist $m \in \mathbb{N}$ and $h \in (0, 1)$ such that

$$\frac{\varepsilon}{n^{1+\delta}} \leq h < \frac{\varepsilon}{\sqrt{n}} \quad \text{and} \quad \zeta = mh.$$

Finally, (2.14) implies

$$\mu \left(\sup_{s \leq t \leq s+\zeta} |Y_n(t) - Y_n(s)| \geq 4\varepsilon \right) \leq \frac{C}{\varepsilon^5} \zeta^{\frac{2}{1+\delta}} \leq \eta.$$

The tightness condition (2.1) is verified and the result is proved. $\square$

2.2.4 A counterexample

We have shown in Theorem 2.3 that Condition (*) holds as soon as the function f and its distribution function F are Hölder continuous. In this section, we prove that the fact that f is Hölder continuous function is not enough to get Condition (*), even if F is continuous.

We define :

$\forall k \geq 1$,

$$s_k = 1 - \frac{2}{2^k} \quad \text{and} \quad S_k = 1 - \frac{1}{2^k}$$

and the intervals $I_k = [s_k, S_k]$, $k \geq 1$. Further, for each $k \geq 2$, we write

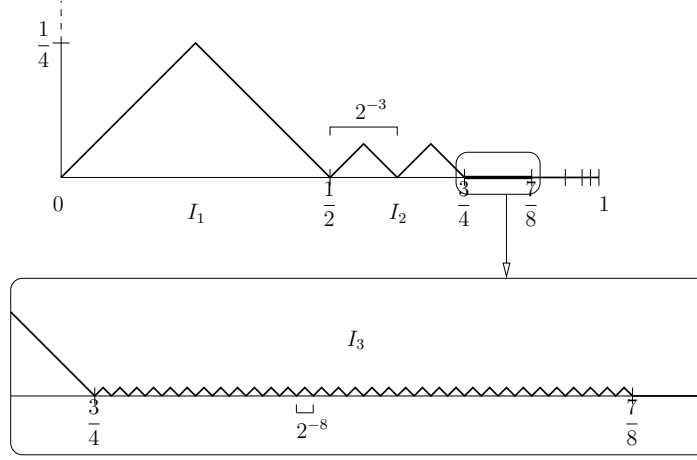
$$s_{(k)}^l = s_k + \frac{l}{2^{k^2}} = \frac{2^{k^2} - 2^{k^2-k+1} + l}{2^{k^2}}, \quad 0 \leq l \leq 2^{k^2-k}.$$

So, we have, $s_k = s_{(k)}^0$ and $S_k = s_{(k)}^{k^2-k}$.

Let $f : [0, 1] \longrightarrow [0, 1]$ be defined as follows:

$$\left\{ \begin{array}{l} \forall x \in \left[0, \frac{1}{4}\right], \quad f(x) = x; \\ \forall x \in \left[\frac{1}{4}, \frac{1}{2}\right], \quad f(x) = \frac{1}{2} - x; \\ \forall k \geq 2, \quad \forall 0 \leq l \leq 2^{k^2-k}, \quad f(s_{(k)}^l) = \begin{cases} 0 & \text{if } l \text{ is even} \\ \frac{1}{2^{k^2}} & \text{if } l \text{ is odd} \end{cases} \\ \text{and } f \text{ affine function on each interval } [s_{(k)}^l, s_{(k)}^{l+1}]. \end{array} \right.$$

Finally, we obtain a continuous function with $f(1) = 0$. Further, f is continuous on the one dimensional torus.



Hence, f is clearly Lipschitz continuous with a Lipschitz constant equal to 1. Remark that its distribution function is continuous (the preimages by f of each point are at most countable).

We will show that f does not satisfy Condition (*).

Let $\varepsilon_k = \frac{1}{2^{k^2}}$, $k \geq 2$. Then, by definition of f , for all $k \geq 2$,

$$I_k \subset \partial_{\varepsilon_k} \{f = 0\}.$$

Thus,

$$\mu(\partial_{\varepsilon_k} \{f = 0\}) \geq \frac{1}{2^k} = 2^k \varepsilon_k^{\frac{2}{k}}.$$

Now, for all $C > 0$ and $\xi > 0$, there exists an integer $k_0 \geq 2$ such that $2^{k_0} > C$ and $\frac{2}{k_0} < \xi$. So, there exist $\varepsilon = \varepsilon_{k_0}$ and $t = 0$ such that

$$\mu(\partial_{\varepsilon} \{f \leq t\}) > C \varepsilon^{\xi}.$$

To get a counterexample in dimension $d > 1$, it is enough to take the function

$$\begin{aligned} g : [0, 1]^d &\longrightarrow [0, 1] \\ (x_1, \dots, x_d) &\longmapsto \frac{1}{d}(f(x_1) + \dots + f(x_d)). \end{aligned}$$

2.3 Regularity of the distribution functions of Morse functions

In Theorem 2.3 the empirical invariance principle is stated under the assumption that the distribution function of f is Hölder continuous. For this reason it is of interest to determine

a class as large as possible of functions that present some characteristics allowing the study of the regularity properties of their distribution functions.

The class of Morse functions seems to be a good candidate. Let us recall that a $\mathcal{C}^r$ function f , with $r \geq 2$, is a Morse function if

- (i) its critical points, or singularities, that is the points where the differential vanishes, are isolated;
- (ii) at each critical point the Hessian, which is well defined (see Hirsch [57] or Milnor [72]), is a non degenerate quadratic form.

According to the well known Morse Lemma, it is possible to find local coordinates in a neighbourhood of a critical point such that f can be written

$$f(x_1, \dots, x_d) = x_1^2 + x_2^2 + \dots + x_p^2 - x_{p+1}^2 - x_{p+2}^2 - \dots - x_d^2$$

with $d = \dim M$ since the critical point is not degenerate.

On the one hand these characteristics are sufficient to study the behaviour of the distribution function of f . On the other hand the set of Morse functions is open and dense in the set of $\mathcal{C}^r$ functions, with $r \geq 2$ or $r = +\infty$, for the $\mathcal{C}^r$ convergence topology.

In the statement of next Theorem we consider a compact Riemannian manifold M instead of a torus because this generalization does not change anything in the proof.

The measure on the compact Riemannian manifold is of course the natural Lebesgue one defined by its metric (it exists even if M is not orientable, see Berger-Gostiaux [3]). We will assume that the volume of M is equal to 1 but this does not matter. Notice that the torus, viewed as $\mathbb{R}^d/\mathbb{Z}^d$, is a compact Riemannian manifold, the measure of which is the Lebesgue measure on $[0, 1]^d$.

2.3.1 Main result

Theorem 2.16. *Let M be a compact, d -dimensional Riemannian manifold. The natural measure is denoted by μ and we assume $\mu(M) = 1$.*

Let f be a Morse function on M and F its distribution function: $F(a) = \mu\{f \leq a\}$.

1. $\dim M = 1$. *The distribution function F is $\frac{1}{2}$ -Hölder continuous and is $\mathcal{C}^1$ outside the singular values of f .*
2. $\dim M = 2$. *If f has some hyperbolic singularities, then F is $\frac{1}{2}$ -Hölder continuous. Otherwise, F is piecewise $\mathcal{C}^1$ hence Lipschitz. In any case it is $\mathcal{C}^1$ outside the singular values of f .*

3. $\dim M \geq 3$. The distribution function F is $\mathcal{C}^1$.

This theorem will be proved in next Section. The following corollary is straightforward.

Corollary 2.17. *The set of $\mathcal{C}^r$ functions whose distribution functions are $\mathcal{C}^1$ (resp. $\frac{1}{2}$ -Hölder continuous) contains an open and dense subset of $\mathcal{C}^r(M; \mathbb{R})$ for $\dim M \geq 3$ (resp. $\dim M = 1, 2$).*

The dimension 2 is particular. On the sphere S^2 it is easy to find functions without hyperbolic singularities, hence functions whose all critical points are extrema: consider for example the "height" function. According to Theorem 2.16 their distribution functions are piecewise $\mathcal{C}^1$ and therefore Lipschitz. Moreover the set of these functions is open (but not dense!) in $\mathcal{C}^r(M)$ because the non degenerate singularities are stable.

On the opposite any Morse function on the 2-dimensional torus has at least two hyperbolic singularities. This intuitive fact can be proved by considering the index of the gradient of f , see for instance Hirsch [57]. Therefore the distribution function of a Morse function on the torus $\mathbb{T}^2$ is never Lipschitz and we can state:

The set of $\mathcal{C}^r$ functions on the 2-dimensional torus $\mathbb{T}^2$ whose distribution functions are Lipschitz is contained in a closed subset with empty interior of $\mathcal{C}^r(\mathbb{T}^2; \mathbb{R})$ for $r \geq 2$.

2.3.2 Proof of Theorem 2.16

Let us begin with some notations. A point of M is usually denoted by x . With a clear abuse of notation the local coordinates for the same point will be $x = (x_1, x_2, \dots, x_d)$ or even $(x, y) = (x_1, \dots, x_p, y_1, \dots, y_q)$ when we will identify $\mathbb{R}^d$ with the product $\mathbb{R}^p \times \mathbb{R}^q$.

Around the hyperbolic singularities we will consider the domain $D_{p,q}(r)$ of $\mathbb{R}^p \times \mathbb{R}^q$ defined as the intersection of

$$\{\|y\| \leq r, \quad \|x\|^2 \leq \|y\|^2 + r^2\} \cup \{\|x\| \leq r, \quad \|y\|^2 \leq \|x\|^2 + r^2\}$$

and

$$\{\|x\| \|y\| \leq r^2\}.$$

There are two reasons to define these rather complicated sets. The first one is that $D_{p,q}(r)$ has to be symmetric with respect to x and y . Secondly we need the boundary of $D_{p,q}(r)$ to be parallel to the gradient of f near the set $\{\|x\| = \|y\|\}$. In these local coordinates f is given by

$$f(x, y) = x_1^2 + x_2^2 + \dots + x_p^2 - y_1^2 - y_2^2 - \dots - y_q^2 + a = \|x\|^2 - \|y\|^2 + a.$$

and one can verify that $\text{grad } f$ is parallel to the hypersurface $\{\|x\| \|y\| = r^2\}$.

To finish the following change of coordinates is used several times:

$$\begin{aligned} \mathbb{R}_+^* \times S^{d-1} &\longmapsto \mathbb{R}^d \\ (s, z) &\longmapsto sz \end{aligned}$$

where S^{d-1} stands for the $(d-1)$ -dimensional sphere. Let us denote by ϕ this diffeomorphism, by dx the canonical volume of $\mathbb{R}^d$, and by σ the induced volume on S^{d-1} . Then we have

$$\phi^* dx = s^{d-1} ds \wedge \sigma$$

or, in other words, we replace dx by $s^{d-1} ds \wedge \sigma$ in the integrals (see Berger-Gostiaux [3]). In the proof of Lemma 2.19 we will do two changes of variable of this kind and will write σ_{d-1} instead of σ to avoid confusion.

In order to prove Theorem 2.16 we will first consider the regular values of f and then the singular ones. If a is a singular value we can assume without loss of generality that the set $\{f = a\}$ contains an unique singularity x_0 , according to the fact that the singularities are isolated. We will distinguish the case where x_0 is an extremum, and the one where x_0 is a hyperbolic singularity.

First case: a is a regular value of f .

We proceed as in [72]. Let $r > 0$ such that the set $\{a - 2r \leq f \leq a + 2r\}$ does not contain any singularity. Let X be a $\mathcal{C}^1$ vector field on M that verifies

$$\forall x \in \{a - r \leq f \leq a + r\} \quad X(x) = \frac{\text{grad } f(x)}{\|\text{grad } f(x)\|^2}$$

and vanishes outside the set $\{a - 2r \leq f \leq a + 2r\}$. The vector field X is complete, and we denote its flow by Φ_t . For all $x \in \{a - r \leq f \leq a + r\}$ we have

$$\frac{d}{dt} f(\Phi_t(x)) = \langle \text{grad } f, \frac{\text{grad } f}{\|\text{grad } f\|^2} \rangle_{\Phi_t(x)} = 1,$$

and therefore $f(\Phi_t(x)) = f(x) + t$, as long as $\Phi_t(x) \in \{a - r \leq f \leq a + r\}$. Consequently we have

$$\{f \leq a + t\} = \Phi_t\{f \leq a\}$$

for $|t| \leq r$. Then

$$F(a + t) = \mu(\{f \leq a + t\}) = \mu(\Phi_t\{f \leq a\}) = \int_{\{f \leq a\}} |J\Phi_t| d\mu.$$

where $J\Phi_t(x)$ stands for the Jacobian of Φ_t evaluated between orthonormal basis at the points x and $\Phi_t(x)$.

Now f is $\mathcal{C}^2$, the vector field X and its flow Φ_t are $\mathcal{C}^1$. Thus $J\Phi_t$ is continuously derivable with respect to t and does not vanish. As a conclusion the distribution function F is $\mathcal{C}^1$ on $]a - r, a + r[$ and

$$F'(a + t) = \int_{\{f \leq a\}} \frac{\partial}{\partial t} |J\Phi_t| d\mu.$$

Second case: $f(x_0) = a$ is a local extremum.

We can assume without loss of generality that x_0 is a minimum.

Let (U, φ) be a local card centered at x_0 such that $\varphi(U) = B(0, 2r)$, where $B(0, 2r)$ is the ball of radius $2r > 0$ in $\mathbb{R}^d$, and f can be written

$$f(x) = x_1^2 + x_2^2 + \cdots + x_d^2 + a = \|x\|^2 + a.$$

We can assume that x_0 is the unique singularity of the set $\{a - 2r \leq f \leq a + 2r\}$. Let us consider the manifold with boundary $N = M \setminus U$. If the set $\{a - 2r \leq f \leq a + 2r\} \cap N$ is not empty we can build a vector field X on N as in the previous case which verifies

$$\forall x \in \{a - r \leq f \leq a + r\} \cap N \quad X(x) = \frac{\text{grad } f(x)}{\|\text{grad } f(x)\|^2}$$

and is zero outside of $\{a - 2r \leq f \leq a + 2r\}$. The previous proof shows that

$$F_1(a + t) = \mu(\{f \leq a + t\} \cap N)$$

is $\mathcal{C}^1$ for $t \in]-r, r[$. Now let us consider the function

$$F_2(a + t) = \mu(\{f \leq a + t\} \cap U) = \mu(\varphi^{-1}(B(0, \sqrt{t}))).$$

We have

$$F_2(a + t) = \int_{B(0, \sqrt{t})} |J\varphi^{-1}(x)| dx$$

where $J\varphi^{-1}(x)$ stands for the Jacobian of φ^{-1} evaluated between orthonormal basis at the points x and $\varphi^{-1}(x)$. According to Lemma 2.18, with $\theta = |J\varphi^{-1}|$, this function is $\mathcal{C}^1$ (resp. piecewise $\mathcal{C}^1$) (resp. $\frac{1}{2}$ -Hölder continuous) for $-r < t < r$ if $d \geq 3$ (resp. $d = 2$) (resp. $d = 1$) and so is $F = F_1 + F_2$.

Third case: $f(x_0) = a$ is a hyperbolic singularity.

Let (U, φ) be a local card centered at x_0 such that $\varphi(U) = D_{p,q}(2r)$, and such that in these coordinates f is given by

$$f(x, y) = x_1^2 + x_2^2 + \cdots + x_p^2 - y_1^2 - y_2^2 - \cdots - y_q^2 + a = \|x\|^2 - \|y\|^2 + a.$$

We can assume that x_0 is the unique singularity of the set $\{a - 2r \leq f \leq a + 2r\}$. Let us consider the manifold with corners $N = M \setminus U$.

The vector field $\text{grad } f$ is tangent to the boundary of N at any point (x, y) of this boundary such that $\|x\| \|y\| = 4r^2$. Therefore we can build a vector field X on N that verifies

$$\forall x \in \{a - r \leq f \leq a + r\} \cap N \quad X(x) = \frac{\text{grad } f(x)}{\|\text{grad } f(x)\|^2}$$

that is zero outside of $\{a - 2r \leq f \leq a + 2r\}$, and that is tangent to the boundary of N . The function

$$F_1(a + t) = \mu(\{f \leq a + t\} \cap N)$$

is as previously $\mathcal{C}^1$ for $t \in]-r, r[$. Now the function

$$F_2(a + t) = \mu(\{f \leq a + t\} \cap U)$$

is equal to

$$F_2(a + t) = \int_{D_{p,q}(r)} |J\varphi^{-1}(x, y)| \, dx dy.$$

According to Lemma 2.19 this function is $\mathcal{C}^1$ (resp. $\frac{1}{2}$ -Hölder continuous) for $-r < t < r$ if $d \geq 3$ (resp. $d = 2$) and so is $F = F_1 + F_2$. $\square$

Lemma 2.18. *Let θ be a $\mathcal{C}^1$ positive map on $B(0, r) \subset \mathbb{R}^d$, the ball of radius $r > 0$, and*

$$V(t) = \int_{B(0, \sqrt{t})} \theta(x) dx.$$

If $d \geq 2$ then V is a $\mathcal{C}^1$ function on $[0, r^2]$ with

$$\begin{aligned} V'(0) &= 0 & \text{if } d \geq 3 \\ V'(0) &= \pi\theta(0) & \text{if } d = 2 \end{aligned}.$$

If $d = 1$ then V is not derivable at 0 but the function

$$t \longmapsto V(t^2)$$

is $\mathcal{C}^1$ on $[0, r]$.

Proof.

Let us do the change of variable

$$\begin{aligned} \mathbb{R}_+^* \times S^{d-1} &\longmapsto \mathbb{R}^d \\ (s, z) &\longmapsto sz \end{aligned}.$$

We have

$$V(t) = \int_{[0, \sqrt{t}] \times S^{d-1}} \theta(sz) s^{d-1} ds \wedge \sigma = \int_0^{\sqrt{t}} s^{d-1} ds \int_{S^{d-1}} \theta(sz) \sigma = \int_0^{\sqrt{t}} s^{d-1} \omega(s) ds$$

where $\omega(s)$ is the $\mathcal{C}^1$ function defined by

$$\omega(s) = \int_{S^{d-1}} \theta(sz) \sigma.$$

Whenever $d \geq 2$ we set $u = s^2$ and we obtain

$$V(t) = \frac{1}{2} \int_0^t u^{\frac{d-2}{2}} \omega(\sqrt{u}) du.$$

As $d \geq 2$, the function $u \mapsto u^{\frac{d-2}{2}} \omega(\sqrt{u})$ is continuous and V is $\mathcal{C}^1$. Moreover $V'(0) = 0$ if $d \geq 3$ and $V'(0) = \frac{1}{2} \omega(0) = \pi \theta(0)$ if $d = 2$.

In the case $d = 1$, the function

$$V(t) = \int_{-\sqrt{t}}^{\sqrt{t}} \theta(x) dx$$

is not derivable at 0 because θ is positive. But clearly

$$t \mapsto V(t^2) = \int_{-t}^t \theta(x) dx$$

is $\mathcal{C}^1$. □

Lemma 2.19. *Let θ be a $\mathcal{C}^1$ positive map on $D_{p,q}(r) \subset \mathbb{R}^p \times \mathbb{R}^q$, and*

$$V(t) = \int_{\{\|x\|^2 - \|y\|^2 \leq t\} \cap D_{p,q}(r)} \theta(x, y) dx dy.$$

If $d = p + q \geq 3$ then V is $\mathcal{C}^1$ on $] -r^2, r^2[$.

If $d = p + q = 2$ then V is not derivable at 0 but the function

$$t \mapsto V(t^2)$$

is $\mathcal{C}^1$ on $] -r, r[$.

Proof.

Because of the symmetry of the domain $D_{p,q}(r)$ it is enough to compute $V(t)$ for $t \geq 0$ and to verify that its derivative at 0 is symmetric with respect to p and q , in order to ensure that V is $\mathcal{C}^1$, not only piecewise $\mathcal{C}^1$.

For this purpose we first compute the integral of θ on the domain

$$\{\|y\| \leq r, \quad \|y\|^2 \leq \|x\|^2 \leq \|y\|^2 + t\}$$

and we denote this integral by $V_1(t)$. We use the two changes of variable

$$\begin{aligned} \mathbb{R}_+^* \times S^{p-1} &\longmapsto \mathbb{R}^p & \text{and} & \quad \mathbb{R}_+^* \times S^{q-1} \longmapsto \mathbb{R}^q \\ (s, z) &\longmapsto sz & & \quad (\tau, \zeta) \longmapsto \tau\zeta \end{aligned}$$

$$\begin{aligned} V_1(t) &= \int_{\{\|y\| \leq r\}} \int_{\{\|y\| \leq \|x\| \leq \sqrt{\|y\|^2 + t}\}} \theta(x, y) dx \\ &= \int_{\{\|y\| \leq r\}} \int_{\|y\|}^{\sqrt{\|y\|^2 + t}} s^{p-1} ds \int_{S^{p-1}} \theta(sz, y) \sigma_{p-1}(z) \\ &= \int_0^r \tau^{q-1} d\tau \int_{S^{q-1}} \sigma_{q-1}(\zeta) \int_{\tau}^{\sqrt{\tau^2 + t}} s^{p-1} ds \int_{S^{p-1}} \theta(sz, \tau\zeta) \sigma_{p-1}(z) \\ &= \int_0^r \tau^{q-1} d\tau \int_{\tau}^{\sqrt{\tau^2 + t}} s^{p-1} \omega(s, \tau) ds \end{aligned}$$

where

$$\omega(s, \tau) = \int_{S^{p-1} \times S^{q-1}} \theta(sz, \tau\zeta) \sigma_{p-1}(z) \wedge \sigma_{q-1}(\zeta)$$

is a $\mathcal{C}^1$ map. We have now to distinguish three cases.

First case: $p \geq 2$

We can set $u = s^2$ and obtain

$$V_1(t) = \frac{1}{2} \int_0^r \tau^{q-1} d\tau \int_{\tau^2}^{\tau^2 + t} u^{\frac{p-2}{2}} \omega(\sqrt{u}, \tau) du.$$

Clearly V_1 is $\mathcal{C}^1$ on $[0, r^2]$ and

$$V_1'(t) = \frac{1}{2} \int_0^r \tau^{q-1} (\tau^2 + t)^{\frac{p-2}{2}} \omega(\sqrt{\tau^2 + t}, \tau) d\tau.$$

In particular at $t = 0$

$$V_1'(0) = \frac{1}{2} \int_0^r \tau^{d-3} \omega(\tau, \tau) d\tau.$$

Second case: $p = 1, q \geq 2$

We set

$$g(u, \tau) = \int_{\tau}^u \omega(s, \tau) ds \quad \text{and} \quad h(t, \tau) = \tau^{q-1} g(\sqrt{\tau^2 + t}, \tau).$$

For $0 < \tau \leq r$, and $0 \leq t \leq r^2$, we have

$$\begin{aligned} \frac{\partial}{\partial t} h(t, \tau) &= \tau^{q-1} \frac{\partial}{\partial u} g(\sqrt{\tau^2 + t}, \tau) \frac{1}{2\sqrt{\tau^2 + t}} \\ &\leq \frac{1}{2} \tau^{q-2} \frac{\partial}{\partial u} g(\sqrt{\tau^2 + t}, \tau) \end{aligned}$$

and

$$\lim_{t \rightarrow 0} \frac{\partial}{\partial t} h(t, \tau) = \frac{\partial}{\partial t} h(0, \tau) = \frac{1}{2} \tau^{q-2} \frac{\partial}{\partial u} g(\tau, \tau).$$

Therefore V_1 is $\mathcal{C}^1$ and

$$V_1'(0) = \int_0^r \frac{1}{2} \tau^{q-2} \frac{\partial}{\partial u} g(\tau, \tau) d\tau = \frac{1}{2} \int_0^r \tau^{d-3} \omega(\tau, \tau) d\tau.$$

Third case: $p = q = 1$

We have

$$V_1(t) = \int_0^r d\tau \int_{\tau}^{\sqrt{\tau^2 + t}} \omega(s, \tau) ds.$$

In the particular case where $\omega = 1$, it is clear that V_1 is not derivable at $t = 0$ because the slope at 0 is infinite (the computation is left to the reader). In the general case there exists $m > 0$ such that $\omega \geq m$, and V_1 is no more derivable at $t = 0$.

However let us consider

$$V_1(t^2) = \int_0^r d\tau \int_{\tau}^{\sqrt{\tau^2 + t^2}} \omega(s, \tau) ds.$$

We have

$$\frac{\partial}{\partial t} \int_{\tau}^{\sqrt{\tau^2 + t^2}} \omega(s, \tau) ds = \frac{t}{\sqrt{\tau^2 + t^2}} \omega(\sqrt{\tau^2 + t^2}, \tau) \leq \omega(\sqrt{\tau^2 + t^2}, \tau)$$

and

$$\frac{\partial}{\partial t} \int_{\tau}^{\sqrt{\tau^2 + t^2}} \omega(s, \tau) ds \xrightarrow[t \rightarrow 0]{} 0.$$

Therefore $t \mapsto V_1(t^2)$ is $\mathcal{C}^1$ and its derivative vanishes at $t = 0$.

It remains to compute the integral of θ on the compact

$$\{\|y\| \leq r, \quad \|y\|^2 \leq \|x\|^2 \leq \|y\|^2 + t\} \cap \{\|x\| \|y\| \leq r^2\}.$$

Let us denote this integral by $W(t)$. We have

$$W(t) = \int_{\{m(t) \leq \|y\| \leq r\}} dy \int_{\frac{r^2}{\|y\|}}^{\sqrt{\|y\|^2 + t}} \theta(x, y) dx$$

where $m(t)$ is the minimum value of $\|y\|$ for which $\frac{r^2}{\|y\|} \leq \sqrt{\|y\|^2 + t}$. Notice that

$$m(t) = \sqrt{\frac{\sqrt{t^2 + 4r^4} - t}{2}}$$

is $\mathcal{C}^1$ and has a positive lower bound as long as $t \leq r^2$. Therefore

$$\psi(t, y) = \int_{\frac{r^2}{\|y\|}}^{\sqrt{\|y\|^2 + t}} \theta(x, y) dx$$

is $\mathcal{C}^1$, and, using a now classical change of variable,

$$W(t) = \int_{m(t)}^r \tau^{q-1} d\tau \int_{S^{q-1}} \psi(t, \tau\zeta) \sigma_{q-1}(\zeta)$$

is also $\mathcal{C}^1$. Let us show that $W'(0) = 0$. First at all we have $\psi(0, r\zeta) = 0$ because for $t = 0$, $m(t) = r$ and the bounds of the integral are r and r . Hence

$$\begin{aligned} W'(0) &= \int_{m(0)}^r \tau^{q-1} \frac{\partial}{\partial t} \left(\int_{S^{q-1}} \psi(t, \tau\zeta) \sigma_{q-1}(\zeta) \right)_{t=0} d\tau \\ &\quad - m'(0) \left(\tau^{q-1} \int_{S^{q-1}} \psi(0, \tau\zeta) \sigma_{q-1}(\zeta) \right)_{\tau=m(0)} = 0. \end{aligned}$$

□

Comments

Almost all this chapter correspond to the paper by Durieu and Jouan [38]. This work took advantage of discussions with Stéphane Le Borgne.

Chapitre 3

New Techniques for Empirical Processes of Dependent Data

Abstract

In this chapter, we present a new technique for proving empirical process invariance principle for stationary processes $(X_n)_{n \geq 0}$. The main novelty of our approach lies in the fact that we only require the central limit theorem and a moment bound for a restricted class of functions $(f(X_n))_{n \geq 0}$, not containing the indicator functions. Our approach can be applied to Markov chains and dynamical systems, using spectral properties of the transfer operator. Our proof consists of a novel application of chaining techniques.

Keywords: Stationary processes; Moment inequalities; Strongly ergodic Markov chains; Dynamical system; Empirical distribution; Invariance principle.

AMS Classification: 60G10; 60J10; 60F17; 28D05; 62G20.

3.1 Introduction

Let $(X_n)_{n \geq 0}$ be a stationary process of $\mathbb{R}$ -valued random variables with marginal distribution function $F(t) = P(X_0 \leq t)$. Define the empirical distribution function $(F_n(t))_{t \in \mathbb{R}}$ and the empirical process $(U_n(t))_{t \in \mathbb{R}}$ by

$$\begin{aligned} F_n(t) &:= \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(-\infty, t]}(X_i), \quad t \in \mathbb{R}, \\ U_n(t) &:= \sqrt{n}(F_n(t) - F(t)), \quad t \in \mathbb{R}. \end{aligned}$$

The empirical process plays a prominent role in non-parametric statistical inference about the distribution function F . In many statistical applications, information about the distribution of the empirical process is needed.

In the case of i.i.d. observations, Donsker [29] proved in 1952 that the empirical process converges in distribution to a Brownian bridge process, thus confirming an earlier conjecture of Doob [30]. In 1968, Billingsley [6] extended Donsker's theorem to some weakly dependent processes, specifically to functionals of ϕ -mixing processes. One of the applications of Billingsley's theorem is to the empirical process of data generated by the continued fraction dynamical system $T : [0, 1] \rightarrow [0, 1]$, where $T(x)$ is the fractional part of $\frac{1}{x}$. Since 1968, many authors have studied the empirical process of weakly dependent data. Invariance principles for empirical distribution of strong mixing random variables were proved in 1977 by Berkes and Philipp [4] and in 1980 for the multivariate case by Philipp and Pinzur [78]. Later, absolutely regular processes were studied by Doukhan et al. [31] and Borovkova et al. [8]. Many other weak dependence conditions have been studied, see Doukhan and Louichi [32], Prieur [80], Dedecker and Prieur [18], Wu and Shao [91], Wu [90]. From the point of view of dynamical systems, an empirical process invariance principle for expanding maps of the interval was proved by Collet et al [13]. Another one for ergodic torus automorphisms was proved by Durieu and Jouan [38] (see Chapter 2).

Proofs of empirical process invariance principles usually consist of two parts, establishing finite-dimensional convergence and tightness of the empirical process.

Finite-dimensional convergence, *i.e.* convergence in distribution of the sequence of vectors $(U_n(t_1), \dots, U_n(t_k))_{n \geq 1}$, is an immediate consequence of the multivariate CLT for partial sums of the process

$$(\mathbb{1}_{(-\infty, t_1]}(X_n), \dots, \mathbb{1}_{(-\infty, t_k]}(X_n))_{n \geq 1}.$$

Tightness is far more difficult to establish. One ingredient is usually a probability bound on the increments of the empirical process

$$U_n(t) - U_n(s) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \{\mathbb{1}_{(s, t]}(X_i) - (F(t) - F(s))\},$$

for a fixed pair $s < t$. Such bounds can in the simplest approach be obtained from bounds on the 4-th moments of $U_n(t) - U_n(s)$. Other results require higher order moment bounds or even exponential bounds.

The traditional approach to empirical process invariance principles, as outlined above, works well in situations when the sequence of indicator variables $(\mathbb{1}_{(s,t]}(X_n))_{n \geq 0}$ inherits good properties from the original process $(X_n)_{n \geq 0}$. This holds, for example, when $(X_n)_{n \geq 0}$ is strong (uniform, beta) mixing, because then $(\mathbb{1}_{(s,t]}(X_n))_{n \geq 0}$ has the same property (for an overview of this theory, see Dehling and Philipp [21] and references therein). There are, however, situations where this is not the case or at least not easy to establish. For some types of Markov processes and dynamical systems, see e.g. Hennion and Hervé [54], one has good control over the properties of $(f(X_n))_{n \geq 0}$ when f is a Lipschitz function, but not for indicator functions. For example, Gouëzel [47] gave a uniformly expanding map of the interval which has a spectral gap on the space of Lipschitz functions but not on the space of bounded variation functions (see Section 3.6). In this chapter, we develop an approach that is strictly based on properties of Lipschitz functions $f(X_i)$ of the original data (we could also work with another class of function approximating well the indicators). We make two basic assumptions, namely that the partial sums of Lipschitz functions satisfy the CLT and that a suitable 4-th moment bound is satisfied.

For our proof we develop a variant of the classical chaining technique that uses only Lipschitz functions at all stages of the chaining argument. We replace the usual finite-dimensional convergence plus tightness approach by a method of approximation by a sequence of finite-dimensional processes, which are different from the coordinate projections $(U_n(t_1), \dots, U_n(t_k))$. We show convergence in distribution of the finite-dimensional processes and prove that the finite-dimensional process approximates the empirical process. In the final step, we use an improved version of a Theorem of Billingsley [6], see our Theorem 3.3 below, to establish convergence in distribution of the empirical process.

3.2 Main results

In the present section, we make two assumptions concerning the process $(X_i)_{i \geq 0}$,

1. For any Lipschitz function f , the CLT holds, *i.e.*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \{f(X_i) - Ef(X_i)\} \xrightarrow{\mathcal{D}} N(0, \sigma^2), \quad (3.1)$$

where $N(0, \sigma^2)$ denotes a normal law with mean zero and variance

$$\sigma^2 = E(f(X_0) - Ef(X_0))^2 + 2 \sum_{i=1}^{\infty} \text{Cov}(f(X_0), f(X_i)).$$

2. A bound on the 4-th central moments of partial sums of $(f(X_i))_{i \geq 0}$, f bounded Lipschitz with $E(f(X_0)) = 0$, of the type

$$E \left\{ \sum_{i=1}^n f(X_i) \right\}^4 \leq C m_f^3 \left(n \|f(X_0)\|_1 \log^\alpha (1 + \|f\|) + n^2 \|f(X_0)\|_1^2 \log^\beta (1 + \|f\|) \right), \quad (3.2)$$

where C is some universal constant, α and β are some nonnegative integers,

$$\|f\| = \sup_x |f(x)| + \sup_{x \neq y} \frac{|f(x) - f(y)|}{|x - y|}$$

and

$$m_f = \max\{1, \sup_x |f(x)|\}.$$

In Section 3.5, some conditions under which such an upper bound holds are given.

We shall assume some regularity for the distribution function of X_0 . We define the modulus of continuity of a function $f : \mathbb{R} \rightarrow \mathbb{R}$ by

$$\omega_f(\delta) = \sup \{|f(s) - f(t)| : s, t \in \mathbb{R}, |s - t| < \delta\}.$$

We can now state our main result.

Theorem 3.1. *Let $(X_i)_{i \geq 0}$ be a $\mathbb{R}$ -valued stationary random process such that the conditions (3.1) and (3.2) hold. Assume that X_0 has a distribution function F satisfying the following condition,*

$$\omega_F(\delta) \leq D |\log(\delta)|^{-\gamma} \text{ for some } D > 0 \text{ and } \gamma > \max\left\{\frac{\alpha}{2}, \beta\right\}, \quad (3.3)$$

then

$$(U_n(t))_{t \in \mathbb{R}} \xrightarrow{\mathcal{D}} (W(t))_{t \in \mathbb{R}},$$

where $W(t)$ is a mean-zero Gaussian process with covariances

$$\begin{aligned} EW(s) \cdot W(t) &= \text{Cov}(\mathbb{1}_{(-\infty, s]}(X_0), \mathbb{1}_{(-\infty, t]}(X_0)) \\ &\quad + \sum_{k=1}^{\infty} \text{Cov}(\mathbb{1}_{(-\infty, s]}(X_0), \mathbb{1}_{(-\infty, t]}(X_k)) \\ &\quad + \sum_{k=1}^{\infty} \text{Cov}(\mathbb{1}_{(-\infty, s]}(X_k), \mathbb{1}_{(-\infty, t]}(X_0)). \end{aligned}$$

Further, almost surely, $(W(t))_{t \in \mathbb{R}}$ has continuous sample paths.

Remark 3.2. *In particular, if the random variable X_0 has a Hölder-continuous distribution function then (3.3) holds.*

In order to prove Theorem 3.1, we apply the following theorem, which is a stronger version of Theorem 4.2 of Billingsley [6] in the complete case. We do not need to assume a priori that $X^{(m)}$ has a limit in distribution.

Theorem 3.3. *Let (S, ρ) be a complete separable metric space and let $X_n, X_n^{(m)}$ and $X^{(m)}$, $n, m \geq 1$ be S -valued random variables satisfying*

$$X_n^{(m)} \xrightarrow{\mathcal{D}} X^{(m)} \text{ as } n \rightarrow \infty, \forall m \quad (3.4)$$

$$\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} P(\rho(X_n, X_n^{(m)}) \geq \varepsilon) = 0, \forall \varepsilon > 0. \quad (3.5)$$

Then there exists an S -valued random variable X such that

$$X_n \xrightarrow{\mathcal{D}} X \text{ as } n \rightarrow \infty.$$

Moreover $X^{(m)} \xrightarrow{\mathcal{D}} X$ as $m \rightarrow \infty$.

Both theorems are proved in Section 3.3 and Section 3.4.

3.3 Proof of Theorem 3.1

3.3.1 The bounded case

We first prove the result for bounded variables. Let $(X_i)_{i \geq 0}$ be a $[0, 1]$ -valued stationary random process such that (3.1), (3.2) and (3.3) hold.

In our approach we work with Lipschitz continuous approximations to the indicator functions $\mathbb{1}_{(-\infty, t]}(x)$. Given a partition

$$0 = t'_0 < \dots < t'_m = 1$$

we define

$$t_j = F^{-1}(t'_j)$$

where F^{-1} is given by

$$F^{-1}(t) = \sup\{s \in [0, 1] : F(s) \leq t\}.$$

Thus, by continuity of F , we have a partition

$$0 \leq t_0 < \dots < t_m = 1.$$

We introduce the functions $\varphi_j : [0, 1] \rightarrow \mathbb{R}$ by

$$\varphi_j(x) = \varphi\left(\frac{x - t_{j-1}}{t_{j-1} - t_{j-2}}\right), \quad \text{for } j = 2, \dots, m$$

where

$$\varphi(x) = \mathbb{1}_{(-\infty, -1]}(x) - x\mathbb{1}_{(-1, 0]}(x) \quad (3.6)$$

and $\varphi_1 \equiv 0$.

The function φ_j will serve as a Lipschitz-continuous approximation to the indicator function $\mathbb{1}_{(-\infty, t_{j-1}]}(x)$. Note that $\varphi_j(x)$ depends on the partition, not only on the point t_{j-1} . We now define the process

$$\begin{aligned} F_n^{(m)}(t) &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m \mathbb{1}_{[t_{j-1}, t_j)}(t) \varphi_j(X_i) \\ &= \sum_{j=1}^m \left(\frac{1}{n} \sum_{i=1}^n \varphi_j(X_i) \right) \mathbb{1}_{[t_{j-1}, t_j)}(t). \end{aligned}$$

Note that $F_n^{(m)}(t)$ is a piecewise constant approximation to the empirical distribution function $F_n(t)$. For $t \in [t_{j-1}, t_j]$, we have the inequality

$$F_n(t_{j-2}) \leq F_n^{(m)}(t) \leq F_n(t_{j-1}).$$

We define further

$$F^{(m)}(t) = E(F_n^{(m)}(t)) = \sum_{j=1}^m E(\varphi_j(X_0)) \mathbb{1}_{[t_{j-1}, t_j)}(t),$$

and finally the centered and normalized process

$$U_n^{(m)}(t) = \sqrt{n} (F_n^{(m)}(t) - F^{(m)}(t)). \quad (3.7)$$

Our proof of Theorem 3.1 now consists of two parts, each of which will be formulated separately as a proposition below. The theorem will follow by application of Theorem 3.3, where (S, ρ) is the space of cadlag functions $D[0, 1]$ provided with the Skorohod metric.

Proposition 3.4. *For any partition $0 = t'_0 < \dots < t'_m = 1$, there exists a piecewise constant Gaussian process $(W^{(m)}(t))_{0 \leq t \leq 1}$ such that*

$$(U_n^{(m)}(t))_{0 \leq t \leq 1} \xrightarrow{\mathcal{D}} (W^{(m)}(t))_{0 \leq t \leq 1}.$$

The sample paths of the processes $(W^{(m)}(t))_{0 \leq t \leq 1}$ are constant on each of the intervals $[t_{j-1}, t_j)$, $1 \leq j \leq m$, and $W^{(m)}(0) = 0$. The vector $(W^{(m)}(t_1), \dots, W^{(m)}(t_m))$ has a multivariate normal distribution with mean zero and covariances

$$\begin{aligned} \text{Cov}(W^{(m)}(t_{i-1}), W^{(m)}(t_{j-1})) &= \text{Cov}(\varphi_i(X_0), \varphi_j(X_0)) \\ &\quad + \sum_{k=1}^{\infty} \text{Cov}(\varphi_i(X_0), \varphi_j(X_k)) \\ &\quad + \sum_{k=1}^{\infty} \text{Cov}(\varphi_i(X_k), \varphi_j(X_0)) \end{aligned}$$

Proof.

Using (3.1) and the Cramér-Wold device, we can show that for any Lipschitz functions $f_1, \dots, f_k$, the multivariate CLT holds, *i.e.*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \{(f_1(X_i), \dots, f_k(X_i)) - E(f_1(X_0), \dots, f_k(X_0))\} \xrightarrow{\mathcal{D}} N(0, \Sigma_{f_1, \dots, f_k}),$$

where $N(0, \Sigma_{f_1, \dots, f_k})$ denotes a multivariate normal law with mean zero and covariance matrix

$$\Sigma_{f_1, \dots, f_k} = (\sigma_{f_i, f_j})_{1 \leq i, j \leq k}$$

where for any Lipschitz functions f, g we define

$$\begin{aligned} \sigma_{f, g} = \text{Cov}(f(X_0), g(X_0)) &+ \sum_{k=1}^{\infty} \text{Cov}(f(X_0), g(X_k)) \\ &+ \sum_{k=1}^{\infty} \text{Cov}(f(X_k), g(X_0)). \end{aligned}$$

This result proves the proposition. □

Proposition 3.5. *For any $\varepsilon, \eta > 0$ there exists a partition $0 = t'_0 < \dots < t'_m = 1$ such that*

$$\limsup_{n \rightarrow \infty} P \left(\sup_{0 \leq t \leq 1} |U_n(t) - U_n^{(m)}(t)| > \varepsilon \right) \leq \eta.$$

Proof.

By a variant of the well known chaining technique we will control

$$P \left(\sup_{0 \leq t \leq 1} |U_n(t) - U_n^{(m)}(t)| \geq \varepsilon \right),$$

and then show that this probability can be made arbitrarily small by choosing a partition $0 = t'_0 < \dots < t'_m = 1$ that is fine enough. From here on we assume that the partition $0 = t'_0 < \dots < t'_m = 1$ is regularly distributed. Let $h = \frac{1}{m} = t'_j - t'_{j-1}$, for $j = 1, \dots, m$.

On the interval $[t'_{j-1}, t'_j]$ we introduce a sequence of refining partitions

$$t'_{j-1} = s_0^{(k)} < s_1^{(k)} < \dots < s_{2^k}^{(k)} = t'_j$$

by

$$s_l^{(k)} = t'_{j-1} + l \cdot \frac{h}{2^k}, \quad 0 \leq l \leq 2^k.$$

Let us define

$$s_l^{(k)} = F^{-1}(s_l'^{(k)}) \quad , \quad 0 \leq l \leq 2^k.$$

We now have partitions of $[t_{j-1}, t_j]$,

$$t_{j-1} = s_0^{(k)} < s_1^{(k)} < \dots < s_{2^k}^{(k)} = t_j.$$

For convenience, we also consider the points

$$s_{-1}^{(k)} = F^{-1}\left(t'_{j-1} - \frac{h}{2^k}\right)$$

and the points

$$s_{2^k+1}^{(k)} = F^{-1}\left(t'_{j-1} + (2^k + 1)\frac{h}{2^k}\right).$$

For any $t \in [t_{j-1}, t_j)$ and $k \geq 0$ we define the index

$$l(k, t) = \max \left\{ l : s_l^{(k)} \leq t \right\}.$$

In this way we obtain a chain

$$t_{j-1} = s_{l(0,t)}^{(0)} \leq s_{l(1,t)}^{(1)} \leq \dots \leq s_{l(k,t)}^{(k)} \leq t \leq s_{l(k,t)+1}^{(k)},$$

linking the left endpoint t_{j-1} to t . Note that for $t \in [t_{j-1}, t_j)$ we have by definition $U_n^{(m)}(t) = U_n^{(m)}(t_{j-1})$. We define the functions $\psi_l^{(k)}$, $k \geq 0$, $0 \leq l \leq 2^k$, by

$$\psi_l^{(k)}(x) = \varphi\left(\frac{x}{s_l^{(k)} - s_{l-1}^{(k)}}\right),$$

where φ is defined as in (3.6). Note that $\psi_{l(0,t)}^{(0)}(x - s_{l(0,t)}^{(0)}) = \varphi_j(x)$. To be consistent, in the case $j = 1$, we have to fix $\psi_0^{(k)} \equiv 0$, for all $k \geq 0$. We build a chain bridging the gap between

$$F_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(-\infty, t]}(X_i)$$

and

$$F_n^{(m)}(t) = \frac{1}{n} \sum_{i=1}^n \varphi_j(X_i)$$

by the functions

$$\begin{aligned} \varphi_j(x) &= \psi_{l(0,t)}^{(0)}(x - s_{l(0,t)}^{(0)}) \\ &\leq \psi_{l(1,t)}^{(1)}(x - s_{l(1,t)}^{(1)}) \\ &\leq \dots \\ &\leq \psi_{l(K,t)}^{(K)}(x - s_{l(K,t)}^{(K)}) \\ &\leq \mathbb{1}_{(-\infty, t]}(x) \\ &\leq \psi_{l(K,t)+2}^{(K)}(x - s_{l(K,t)+2}^{(K)}), \end{aligned}$$

where K is some integer to be chosen later. In this way we get

$$\begin{aligned} F_n(t) - F_n^{(m)}(t) &= \sum_{k=1}^K \frac{1}{n} \sum_{i=1}^n \left(\psi_{l(k,t)}^{(k)}(X_i - s_{l(k,t)}^{(k)}) - \psi_{l(k-1,t)}^{(k-1)}(X_i - s_{l(k-1,t)}^{(k-1)}) \right) \\ &\quad + \frac{1}{n} \sum_{i=1}^n \left(\mathbb{1}_{(-\infty, t]}(X_i) - \psi_{l(K,t)}^{(K)}(X_i - s_{l(K,t)}^{(K)}) \right). \end{aligned} \quad (3.8)$$

Observe that by definition of $s_{l(k,t)}^{(k)}$ and of $\psi^{(K)}$,

$$\begin{aligned} 0 &\leq \mathbb{1}_{(-\infty, t]}(X_i) - \psi_{l(K,t)}^{(K)}(X_i - s_{l(K,t)}^{(K)}) \\ &\leq \psi_{l(K,t)+2}^{(K)}(X_i - s_{l(K,t)+2}^{(K)}) - \psi_{l(K,t)}^{(K)}(X_i - s_{l(K,t)}^{(K)}). \end{aligned}$$

From (3.8) we get by centering and normalization

$$\begin{aligned} U_n(t) - U_n^{(m)}(t) &= \sum_{k=1}^K \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \left(\psi_{l(k,t)}^{(k)}(X_i - s_{l(k,t)}^{(k)}) - E\psi_{l(k,t)}^{(k)}(X_i - s_{l(k,t)}^{(k)}) \right) \right. \\ &\quad \left. - \left(\psi_{l(k-1,t)}^{(k-1)}(X_i - s_{l(k-1,t)}^{(k-1)}) - E\psi_{l(k-1,t)}^{(k-1)}(X_i - s_{l(k-1,t)}^{(k-1)}) \right) \right\} \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \left(\mathbb{1}_{(-\infty, t]}(X_i) - F(t) \right) \right. \\ &\quad \left. - \left(\psi_{l(K,t)}^{(K)}(X_i - s_{l(K,t)}^{(K)}) - E\psi_{l(K,t)}^{(K)}(X_i - s_{l(K,t)}^{(K)}) \right) \right\}. \end{aligned}$$

For the last term on the r.h.s. we have the following upper and lower bounds,

$$\begin{aligned} &\frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \left(\mathbb{1}_{(-\infty, t]}(X_i) - F(t) \right) - \left(\psi_{l(K,t)}^{(K)}(X_i - s_{l(K,t)}^{(K)}) - E\psi_{l(K,t)}^{(K)}(X_i - s_{l(K,t)}^{(K)}) \right) \right\} \\ &\leq \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \left(\psi_{l(K,t)+2}^{(K)}(X_i - s_{l(K,t)+2}^{(K)}) - E\psi_{l(K,t)+2}^{(K)}(X_i - s_{l(K,t)+2}^{(K)}) \right) \right. \\ &\quad \left. - \left(\psi_{l(K,t)}^{(K)}(X_i - s_{l(K,t)}^{(K)}) - E\psi_{l(K,t)}^{(K)}(X_i - s_{l(K,t)}^{(K)}) \right) \right\} \\ &\quad + \sqrt{n} \left(E\psi_{l(K,t)+2}^{(K)}(X_i - s_{l(K,t)+2}^{(K)}) - F(t) \right) \end{aligned}$$

and

$$\begin{aligned} &\frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \left(\mathbb{1}_{(-\infty, t]}(X_i) - F(t) \right) - \left(\psi_{l(K,t)}^{(K)}(X_i - s_{l(K,t)}^{(K)}) - E\psi_{l(K,t)}^{(K)}(X_i - s_{l(K,t)}^{(K)}) \right) \right\} \\ &\geq -\sqrt{n} \left(F(t) - E\psi_{l(K,t)}^{(K)}(X_i - s_{l(K,t)}^{(K)}) \right). \end{aligned}$$

Now choose $K = 4 + \left\lfloor \log \left(\frac{\sqrt{nh}}{\varepsilon} \right) \log^{-1}(2) \right\rfloor$ and note that

$$\frac{\varepsilon}{2^4} \leq \sqrt{n} \frac{h}{2^K} \leq \frac{\varepsilon}{2^3}$$

and thus

$$\begin{aligned} & \sqrt{n} \left| E\psi_{l(K,t)+2}^{(K)}(X_i - s_{l(K,t)+2}^{(K)}) - E\psi_{l(K,t)}^{(K)}(X_i - s_{l(K,t)}^{(K)}) \right| \\ & \leq \sqrt{n} \left| F(s_{l(K,t)+2}^{(K)}) - F(s_{l(K,t)-1}^{(K)}) \right| \\ & \leq \frac{\varepsilon}{2}. \end{aligned}$$

Thus we get for all $t \in [t_{j-1}, t_j]$,

$$\begin{aligned} & |U_n(t) - U_n^{(m)}(t)| \\ & \leq \sum_{k=1}^K \frac{1}{\sqrt{n}} \left| \sum_{i=1}^n \left\{ \left(\psi_{l(k,t)}^{(k)}(X_i - s_{l(k,t)}^{(k)}) - E\psi_{l(k,t)}^{(k)}(X_i - s_{l(k,t)}^{(k)}) \right) \right. \right. \\ & \quad \left. \left. - \left(\psi_{l(k-1,t)}^{(k-1)}(X_i - s_{l(k-1,t)}^{(k-1)}) - E\psi_{l(k-1,t)}^{(k-1)}(X_i - s_{l(k-1,t)}^{(k-1)}) \right) \right\} \right| \\ & \quad + \frac{1}{\sqrt{n}} \left| \sum_{i=1}^n \left\{ \left(\psi_{l(K,t)+2}^{(K)}(X_i - s_{l(K,t)+2}^{(K)}) - E\psi_{l(K,t)+2}^{(K)}(X_i - s_{l(K,t)+2}^{(K)}) \right) \right. \right. \\ & \quad \left. \left. - \left(\psi_{l(K,t)}^{(K)}(X_i - s_{l(K,t)}^{(K)}) - E\psi_{l(K,t)}^{(K)}(X_i - s_{l(K,t)}^{(K)}) \right) \right\} \right| \\ & \quad + \frac{\varepsilon}{2}. \end{aligned}$$

Note that by definition of $l(k, t)$ and of $s_l^{(k)}$, we have $s_{l(k-1,t)}^{(k-1)} \in \{s_{l(k,t)}^{(k)}, s_{l(k,t)-1}^{(k)}\}$ and thus

$$l(k-1, t) = \left\lfloor \frac{l(k, t)}{2} \right\rfloor.$$

Therefore

$$\begin{aligned} & \sup_{t_{j-1} \leq t \leq t_j} |U_n(t) - U_n^{(m)}(t)| \\ & \leq \sum_{k=1}^K \frac{1}{\sqrt{n}} \max_{0 \leq l \leq 2^{k-1}} \left| \sum_{i=1}^n \left(\left(\psi_l^{(k)}(X_i - s_l^{(k)}) - E\psi_l^{(k)}(X_i - s_l^{(k)}) \right) \right. \right. \\ & \quad \left. \left. - \left(\psi_{\lfloor \frac{l}{2} \rfloor}^{(k-1)}(X_i - s_{\lfloor \frac{l}{2} \rfloor}^{(k-1)}) - E\psi_{\lfloor \frac{l}{2} \rfloor}^{(k-1)}(X_i - s_{\lfloor \frac{l}{2} \rfloor}^{(k-1)}) \right) \right) \right| \\ & \quad + \frac{1}{\sqrt{n}} \max_{0 \leq l \leq 2^{K-1}} \left| \sum_{i=1}^n \left(\left(\psi_{l+2}^{(K)}(X_i - s_{l+2}^{(K)}) - E\psi_{l+2}^{(K)}(X_i - s_{l+2}^{(K)}) \right) \right. \right. \\ & \quad \left. \left. - \left(\psi_l^{(K)}(X_i - s_l^{(K)}) - E\psi_l^{(K)}(X_i - s_l^{(K)}) \right) \right) \right| \\ & \quad + \frac{\varepsilon}{2}. \end{aligned}$$

Now take $\varepsilon_k := \frac{\varepsilon}{4k(k+1)}$ and note that $\sum_{k=1}^K \varepsilon_k \leq \frac{\varepsilon}{4}$. Then we obtain

$$\begin{aligned}
& P \left(\sup_{t_{j-1} \leq t \leq t_j} |U_n(t) - U_n^{(m)}(t)| \geq \varepsilon \right) \\
& \leq \sum_{k=1}^K \sum_{l=0}^{2^k-1} P \left(\frac{1}{\sqrt{n}} \left| \sum_{i=1}^n \left\{ \left(\psi_l^{(k)}(X_i - s_l^{(k)}) - E\psi_l^{(k)}(X_i - s_l^{(k)}) \right) \right. \right. \right. \\
& \quad \left. \left. \left. - \left(\psi_{\lfloor \frac{l}{2} \rfloor}^{(k-1)}(X_i - s_{\lfloor \frac{l}{2} \rfloor}^{(k-1)}) - E\psi_{\lfloor \frac{l}{2} \rfloor}^{(k-1)}(X_i - s_{\lfloor \frac{l}{2} \rfloor}^{(k-1)}) \right) \right\} \right| \geq \varepsilon_k \right) \\
& \quad + \sum_{l=0}^{2^K-1} P \left(\frac{1}{\sqrt{n}} \left| \sum_{i=1}^n \left\{ \left(\psi_{l+2}^{(K)}(X_i - s_{l+2}^{(K)}) - E\psi_{l+2}^{(K)}(X_i - s_{l+2}^{(K)}) \right) \right. \right. \right. \\
& \quad \left. \left. \left. - \left(\psi_l^{(K)}(X_i - s_l^{(K)}) - E\psi_l^{(K)}(X_i - s_l^{(K)}) \right) \right\} \right| \geq \frac{\varepsilon}{4} \right).
\end{aligned}$$

At this point we use Markov's inequality together with the 4-th moment bound (3.2).

$$\begin{aligned}
& P \left(\sup_{t_{j-1} \leq t \leq t_j} |U_n(t) - U_n^{(m)}(t)| \geq \varepsilon \right) \\
& \leq C \sum_{k=1}^K \sum_{l=0}^{2^k-1} \left\{ \frac{1}{n\varepsilon_k^4} \left\| \psi_l^{(k)}(X_0 - s_l^{(k)}) - \psi_{\lfloor \frac{l}{2} \rfloor}^{(k-1)}(X_0 - s_{\lfloor \frac{l}{2} \rfloor}^{(k-1)}) \right\|_1 \right. \\
& \quad \cdot \log^\alpha \left(1 + \left\| \psi_l^{(k)} - \psi_{\lfloor \frac{l}{2} \rfloor}^{(k-1)} \right\| \right) \\
& \quad + \frac{1}{\varepsilon_k^4} \left\| \psi_l^{(k)}(X_0 - s_l^{(k)}) - \psi_{\lfloor \frac{l}{2} \rfloor}^{(k-1)}(X_0 - s_{\lfloor \frac{l}{2} \rfloor}^{(k-1)}) \right\|_1^2 \\
& \quad \cdot \log^\beta \left(1 + \left\| \psi_l^{(k)} - \psi_{\lfloor \frac{l}{2} \rfloor}^{(k-1)} \right\| \right) \Big\} \\
& \quad + C \sum_{l=0}^{2^K-1} \left\{ \frac{4^4}{n\varepsilon^4} \left\| \psi_{l+2}^{(K)}(X_0 - s_{l+2}^{(K)}) - \psi_l^{(K)}(X_0 - s_l^{(K)}) \right\|_1 \right. \\
& \quad \cdot \log^\alpha \left(1 + \left\| \psi_{l+2}^{(K)} - \psi_l^{(K)} \right\| \right) \\
& \quad + \frac{4^4}{\varepsilon^4} \left\| \psi_{l+2}^{(K)}(X_0 - s_{l+2}^{(K)}) - \psi_l^{(K)}(X_0 - s_l^{(K)}) \right\|_1^2 \\
& \quad \cdot \log^\beta \left(1 + \left\| \psi_{l+2}^{(K)} - \psi_l^{(K)} \right\| \right) \Big\}.
\end{aligned}$$

Note that

$$\begin{aligned}
\left\| \psi_l^{(k)}(X_0 - s_l^{(k)}) - \psi_{\lfloor \frac{l}{2} \rfloor}^{(k-1)}(X_0 - s_{\lfloor \frac{l}{2} \rfloor}^{(k-1)}) \right\|_1 & \leq \left| F(s_l^{(k)}) - F(s_{\lfloor \frac{l}{2} \rfloor - 1}^{(k-1)}) \right| \\
& \leq \left| F(s_l^{(k)}) - F(s_{l-3}^{(k)}) \right| \\
& = \frac{3h}{2^k}
\end{aligned}$$

and

$$\begin{aligned} \left\| \psi_{l+2}^{(K)}(X_0 - s_{l+2}^{(K)}) - \psi_l^{(K)}(X_0 - s_l^{(K)}) \right\|_1 &\leq \left| F(s_{l+2}^{(K)}) - F(s_{l-1}^{(K)}) \right| \\ &= \frac{3h}{2^K}. \end{aligned}$$

If (3.3) is satisfied,

$$\begin{aligned} \left\| \psi_l^{(k)} \right\| &\leq 1 + \left[\inf \left\{ s > 0 : \forall t, F(t+s) - F(t) \geq \frac{h}{2^k} \right\} \right]^{-1} \\ &\leq 1 + \left[\inf \left\{ s > 0 : D |\log(s)|^{-\gamma} \geq \frac{h}{2^k} \right\} \right]^{-1} \\ &= 1 + \exp \left(\left(\frac{D2^k}{h} \right)^{\frac{1}{\gamma}} \right). \end{aligned}$$

Thus we have

$$\begin{aligned} P \left(\sup_{t_{j-1} \leq t \leq t_j} |U_n(t) - U_n(t_j)| \geq \varepsilon \right) &\leq 4^4 C \sum_{k=1}^K 2^k \frac{(k(k+1))^4}{\varepsilon^4} \frac{1}{n} \frac{3h}{2^k} \log^\alpha \left(2 + \exp \left(\left(\frac{D2^k}{h} \right)^{\frac{1}{\gamma}} \right) \right) \\ &\quad + 4^4 C \sum_{k=1}^K 2^k \frac{(k(k+1))^4}{\varepsilon^4} \frac{(3h)^2}{2^{2k}} \log^\beta \left(2 + \exp \left(\left(\frac{D2^k}{h} \right)^{\frac{1}{\gamma}} \right) \right) \\ &\quad + 4^4 C 2^K \frac{1}{\varepsilon^4} \frac{1}{n} \frac{3h}{2^K} \log^\alpha \left(2 + \exp \left(\left(\frac{D2^K}{h} \right)^{\frac{1}{\gamma}} \right) \right) \\ &\quad + 4^4 C 2^K \frac{1}{\varepsilon^4} \frac{(3h)^2}{2^{2K}} \log^\beta \left(2 + \exp \left(\left(\frac{D2^K}{h} \right)^{\frac{1}{\gamma}} \right) \right) \\ &\leq \frac{1}{n} \frac{C'}{\varepsilon^4} \sum_{k=1}^K k^8 h \left(\frac{D2^k}{h} \right)^{\frac{\alpha}{\gamma}} + \frac{C'}{\varepsilon^4} \sum_{k=1}^K \frac{k^8}{2^k} h^2 \left(\frac{D2^k}{h} \right)^{\frac{\beta}{\gamma}} \\ &\leq D^{\frac{\alpha}{\gamma}} \frac{1}{n} \frac{C'}{\varepsilon^4} h \left(\frac{2^K}{h} \right)^{\frac{\alpha}{\gamma}} \sum_{k=1}^K k^8 + D^{\frac{\beta}{\gamma}} \frac{C'}{\varepsilon^4} h^{2-\frac{\beta}{\gamma}} \sum_{k=1}^{\infty} k^8 2^{k(\frac{\beta}{\gamma}-1)} \\ &\leq \frac{h}{n} \frac{C'''}{\varepsilon^4} \left(\frac{\sqrt{n}}{\varepsilon} \right)^{\frac{\alpha}{\gamma}} K^9 + \frac{C'''}{\varepsilon^4} h^{2-\frac{\beta}{\gamma}} \end{aligned}$$

where C' and C'' are some constants and we used convergence of the series $\sum_{k=1}^{\infty} k^8 2^{k(\frac{\beta}{\gamma}-1)}$.

Finally, using $mh = 1$,

$$\begin{aligned}
& P \left(\sup_{0 \leq t \leq 1} |U_n(t) - U_n^{(m)}(t)| \geq \varepsilon \right) \\
& \leq \sum_{j=1}^m P \left(\sup_{t_{j-1} \leq t \leq t_j} |U_n(t) - U_n^{(m)}(t)| \geq \varepsilon \right) \\
& \leq m h n^{\frac{\alpha}{2\gamma}-1} \frac{C''}{\varepsilon^{4+\frac{\alpha}{\gamma}}} K^9 + m \frac{C''}{\varepsilon^4} h^{2-\frac{\beta}{\gamma}} \\
& \leq n^{\frac{\alpha}{2\gamma}-1} \frac{C''}{\varepsilon^{4+\frac{\alpha}{\gamma}}} \left(4 + \log \frac{\sqrt{n}h}{\varepsilon} \right)^9 + \frac{C''}{\varepsilon^4} h^{1-\frac{\beta}{\gamma}}
\end{aligned}$$

Now, the first of the two final summands converges to zero as $n \rightarrow \infty$. The second can be made arbitrarily small by choosing a partition that is fine enough (*i.e.* h small). $\square$

We used a different technique than the usual finite dimensional convergence plus tightness. Of course, since the weak convergence implies the finite dimensional convergence and the tightness, these two properties are satisfied. Nevertheless, we can also deduce a thightness criterion implying that, almost surely, the limit process has continuous sample paths (see Billingsley [6], Theorem 15.5).

Proposition 3.6. *For all $\varepsilon, \eta > 0$, there exist $\delta > 0$ and $N \geq 0$ such that for all $n \geq N$,*

$$P \left(\sup_{|t-s| < \delta} |U_n(t) - U_n(s)| \geq \varepsilon \right) \leq \eta.$$

In particular, $P(W \in C(\mathbb{R})) = 1$.

Proof.

Let $\varepsilon > 0$ and $\eta > 0$.

Let m be an integer such that

$$\frac{C}{\varepsilon^4} \frac{D_{\gamma}^{\frac{\beta}{\gamma}}}{m^{1+\frac{\beta}{\gamma}}} < \frac{\eta}{4}. \quad (3.9)$$

By Proposition 3.5, there exists $N \geq 0$ such that for all $n \geq N$,

$$P \left(\sup_{0 \leq t \leq 1} |U_n(t) - U_n^{(m)}(t)| \geq \frac{\varepsilon}{3} \right) \leq \frac{\eta}{4}.$$

Let $\delta > 0$ such that $\delta < \min\{t'_j - t'_{j-1} : j = 1 \dots m\}$. Then, for all $n \geq N$,

$$\begin{aligned} & P \left(\sup_{|t-s|<\delta} |U_n(t) - U_n(s)| \geq \varepsilon \right) \\ & \leq 2P \left(\sup_{0 \leq t \leq 1} |U_n(t) - U_n^m(t)| \geq \frac{\varepsilon}{3} \right) + P \left(\sup_{|t-s|<\delta} |U_n^m(t) - U_n^m(s)| \geq \frac{\varepsilon}{3} \right) \\ & \leq \frac{\eta}{2} + P \left(\sup_{|t-s|<\delta} |U_n^m(t) - U_n^m(s)| \geq \frac{\varepsilon}{3} \right). \end{aligned}$$

We recall that

$$\begin{aligned} \|\varphi_j(X_0) - \varphi_{j+1}(X_0)\|_1 & \leq P(t_{j-2} \leq X_0 \leq t_j) \leq \frac{2}{m} \\ \|\varphi_j\| & \leq 1 + \exp \left(\left(\frac{D}{m} \right)^{\frac{1}{\gamma}} \right). \end{aligned}$$

Thus, by the 4-th moment bound (3.2),

$$P \left(\sup_{|t-s|<\delta} |U_n^m(t) - U_n^m(s)| \geq \frac{\varepsilon}{3} \right) \leq \frac{C}{n\varepsilon^4} \left(\frac{D}{m} \right)^{\frac{\alpha}{\gamma}} + \frac{C}{\varepsilon^4} \frac{D^{\frac{\beta}{\gamma}}}{m^{1+\frac{\beta}{\gamma}}}.$$

Now there exists $N' \geq N$ such that

$$\frac{C}{n\varepsilon^4} \left(\frac{D}{m} \right)^{\frac{\alpha}{\gamma}} \leq \frac{\eta}{4}.$$

Finally, by (3.9),

$$P \left(\sup_{|t-s|<\delta} |U_n(t) - U_n(s)| \geq \varepsilon \right) \leq \eta.$$

□

3.3.2 The unbounded case

Let $(X_i)_{i \geq 0}$ be a $\mathbb{R}$ -valued stationary random process such that (3.1), (3.2) and (3.3) hold. We will show that it can be reduced to the case of bounded variables.

For all $x < y \in \mathbb{R}$, we say that the closed interval $[x, y]$ is a 'bad' interval (for the function F) if

$$F(y) - F(x) \geq y - x.$$

We say that $[x, y]$ is a maximal 'bad' interval (for the function F) if for all 'bad' intervals $[a, b]$, we have $[a, b] \subset [x, y]$ or $[a, b] \cap [x, y] = \emptyset$.

We denote by I^{max} the set of all maximal 'bad' intervals.

Lemma 3.7.

(i) The Lebesgue measure of

$$I := \bigcup_{[x,y] \in I^{max}} [x, y]$$

is smaller than 1.

(ii) For all $[x, y] \in I^{max}$, we have

$$F(y) - F(x) = y - x.$$

Proof.

Because the distribution function F is non-decreasing and takes values in $[0, 1]$, the first assertion is clear.

If for $x < y$, $F(y) - F(x) > y - x$, then there exists $\varepsilon > 0$ such that

$$F(y) - F(x) > y - x + \varepsilon.$$

Thus, for all $z > y$ such that $z - y \leq \varepsilon$, by monotonicity of F , we have

$$\begin{aligned} F(z) - F(x) &\geq F(y) - F(x) \\ &> y - x + \varepsilon \\ &\geq z - x \end{aligned}$$

and then $[x, y]$ is not maximal. □

We define the function g from $\mathbb{R}$ to $]0, 1[$ by

$$\text{for all } [x, y] \in I^{max}, \text{ for all } t \in [x, y], g(t) := F(x) + t - x$$

and

$$\text{for all } t \notin I, g(t) := F(t).$$

Then g is a 1-Lipschitz function.

We define the $[0, 1]$ -valued stationary random process $(Y_i)_{i \geq 0}$ by

$$Y_i = g(X_i), \quad i \geq 0.$$

Since g is Lipschitz, $(Y_i)_{i \geq 0}$ satisfies (3.1) and (3.2).

We also have

$$G(t) := P(Y_0 \leq t) = F \circ g^{-1}(t)$$

where

$$g^{-1}(t) = \sup\{s \in \mathbb{R} : F(s) \leq t\}.$$

Clearly, G is the identity on $g(\mathbb{R} \setminus I)$. Further, for all $[x, y] \in I^{max}$, the graph of G on $g([x, y])$ is the graph of F on $[x, y]$ and the Lebesgue measure of $g([x, y])$ is equal to the Lebesgue measure of $[x, y]$. Then

$$\omega_G(\delta) \leq \max\{\omega_F(\delta), \delta\}$$

and (3.3) holds.

We define the associated distribution functions and empirical processes

$$\begin{aligned} F_n(t) &:= \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(-\infty, t]}(X_i), \quad t \in \mathbb{R}, \\ U_n(t) &:= \sqrt{n}(F_n(t) - F(t)), \quad t \in \mathbb{R}, \\ G_n(t) &:= \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{[0, t]}(Y_i), \quad 0 \leq t \leq 1, \\ V_n(t) &:= \sqrt{n}(G_n(t) - G(t)), \quad 0 \leq t \leq 1. \end{aligned}$$

We have

$$U_n(t) = V_n(g(t)), \quad t \in \mathbb{R}.$$

By the theorem for bounded variables (Section 3.3.1),

$$(V_n(t))_{0 \leq t \leq 1} \xrightarrow{\mathcal{D}} (V(t))_{0 \leq t \leq 1},$$

where $V(t)$ is a mean-zero Gaussian process such that $P(V \in C[0, 1]) = 1$.

Applying Theorem 5.1 of Billingsley [6] with

$$\begin{aligned} h : D[0, 1] &\longrightarrow D(\mathbb{R}) \\ x &\longmapsto x \circ g, \end{aligned}$$

we get the weak convergence of $(U_n(t))_{t \in \mathbb{R}}$ to a Gaussian process

$$(W(t))_{t \in \mathbb{R}} = (V \circ g(t))_{t \in \mathbb{R}}$$

such that $P(W \in C(\mathbb{R})) = 1$.

This concludes the proof of Theorem 3.1.

3.4 Proof of Theorem 3.3

Theorem 3.3 is an improvement of Theorem 4.2 of Billingsley [6] for complete separable metric space. Note that, in Billingsley, the not complete case is considered and the fact that $X^{(m)} \xrightarrow{\mathcal{D}} X$ as $m \rightarrow \infty$ is required as an assumption. The proof is based on the following lemma.

Lemma 3.8. *Let (X, d) be a complete metric space and let $x_n, x_n^{(m)}, x^m \in X$, $n \geq 1, m \geq 1$ be given with the properties*

$$\lim_{n \rightarrow \infty} d(x_n^{(m)}, x^{(m)}) = 0 \quad \forall m \quad (3.10)$$

$$\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} d(x_n, x_n^{(m)}) = 0. \quad (3.11)$$

Then $x := \lim_{m \rightarrow \infty} x^{(m)}$ exists and

$$\lim_{n \rightarrow \infty} d(x_n, x) = 0.$$

Proof. We will first show that $x^{(m)}$ is a Cauchy sequence. Given $\epsilon > 0$, choose M so big that $\forall m \geq M$

$$\limsup_{n \rightarrow \infty} d(x_n, x_n^{(m)}) < \frac{\epsilon}{4}.$$

Now take $m_1, m_2 \geq M$. For all n sufficiently large, we have then

$$\begin{aligned} d(x_n^{(m_1)}, x^{(m_1)}) &< \frac{\epsilon}{4} \\ d(x_n^{(m_2)}, x^{(m_2)}) &< \frac{\epsilon}{4} \\ d(x_n, x_n^{(m_1)}) &< \frac{\epsilon}{4} \\ d(x_n, x_n^{(m_2)}) &< \frac{\epsilon}{4}, \end{aligned}$$

and hence, by the triangle inequality $d(x^{(m_1)}, x^{(m_2)}) < \epsilon$. Thus $(x^{(m)})_{m \geq 1}$ is a Cauchy sequence and hence $x := \lim_{m \rightarrow \infty} x^{(m)}$ exists.

It remains to show that $\lim_{n \rightarrow \infty} x_n = x$. Given $\epsilon > 0$, choose m_0 so that

$$\limsup_{n \rightarrow \infty} d(x_n, x_n^{(m_0)}) < \frac{\epsilon}{4}$$

and $d(x^{(m_0)}, x) < \frac{\epsilon}{4}$. Then choose N such that for all $n \geq N$

$$\begin{aligned} d(x_n, x_n^{(m_0)}) &< \frac{\epsilon}{4} \\ d(x_n^{(m_0)}, x^{(m_0)}) &< \frac{\epsilon}{4}. \end{aligned}$$

Using the triangle inequality, we get

$$d(x_n, x) < \varepsilon$$

for all $n \geq N$. □

Proof of Theorem 3.3. Let $\mu_n, \mu_n^{(m)}$ and $\mu^{(m)}$ denote the distributions of the random variables $X_n, X_n^{(m)}$ and $X^{(m)}$ respectively. These are elements of $M_1(S)$, the space of probability measures on S . We consider the Prohorov metric d on $M_1(S)$, defined by

$$d(\mu, \nu) = \inf \{ \varepsilon > 0 : \mu(A) \leq \nu(A^\varepsilon) + \varepsilon \quad \forall A \subset S \text{ measurable} \}.$$

Note that $(M_1(S), d)$ is a complete metric space. If Y, Z are two S -valued random variables with distributions P_Y, P_Z , satisfying

$$P(\rho(Y, Z) \geq \varepsilon) \leq \varepsilon,$$

then $d(P_Y, P_Z) \leq \varepsilon$. Moreover d metrizes the topology of weak convergence, i. e. $\mu_n \rightarrow \mu$ if and only if $d(\mu_n, \mu) \rightarrow 0$. We now apply Lemma 3.1 to $\mu_n, \mu_n^{(m)}, \mu^{(m)}$. Note that (3.10) is a direct consequence of (3.4). Given $\varepsilon > 0$, by (3.5) we can find m_0 such that for all $m \geq m_0$,

$$\limsup_{n \rightarrow \infty} P(\rho(X_n, X_n^{(m)}) \geq \varepsilon) < \varepsilon.$$

Fix such an m ; then we can find n_0 such that $\forall n \geq n_0$

$$P(\rho(X_n, X_n^{(m)}) \geq \varepsilon) \leq \varepsilon$$

and thus $d(\mu_n, \mu_n^{(m)}) \leq \varepsilon$. Hence

$$\limsup_{n \rightarrow \infty} d(\mu_n, \mu_n^{(m)}) \leq \varepsilon$$

for all $m \geq m_0$, showing that (3.11) holds. Thus by Lemma 3.1, there exists a probability distribution μ on S such that

$$\begin{aligned} \lim_{n \rightarrow \infty} d(\mu^{(m)}, \mu) &= 0 \\ \lim_{m \rightarrow \infty} d(\mu_n, \mu) &= 0. \end{aligned}$$

Finally, let X be an S -valued random variable with distribution μ . Then $X^{(m)} \xrightarrow{\mathcal{D}} X$ as $m \rightarrow \infty$ and $X_n \rightarrow X$ as $n \rightarrow \infty$. □

3.5 A Fourth Moment Inequality for Functionals of Stationary Processes

Fourth moment bound for partial sums of stationary processes is a key tool in the study of functional limit theorems. In particular, it is the case in the investigation of empirical process invariance principle. As recalled in introduction, since the work of Donsker [29] and the work of Billingsley [6], the chaining technique seems to be a suitable way to get the tightness of the process $(Y_n)_{n \geq 0}$. Fourth moment inequality is a central point in this technique. In many cases (such as the i.i.d. case), if an inequality of the type

$$E \left[\sum_{i=0}^{n-1} \{ \mathbb{1}_{(s,t]}(X_i) - (F(t) - F(s)) \} \right]^4 \leq C(n(t-s) + n^2(t-s)^2), \quad (3.12)$$

for all $s < t$ is established then the tightness of the process follows (as in Chapter 2).

Again, the difficulty is to deal with the sequence of indicator variables $(\mathbb{1}_{(s,t]}(X_i))_{i \geq 0}$. In 2004, Collet, Martinez and Schmitt [13] proved a fourth moment inequality for expanding maps of the interval (see also Dedecker and Priour [18]). They use spectral properties of the transfer operator associated on the space of bounded variation functions and the fact that indicators belong to this space. In other cases, such as some type of Markov chains and dynamical system (see Hennion and Hervé [54]), one can have good properties of the Markov operator or the transfer operator on other spaces of functions which do not contain the indicator functions (for example, on the space of Lipschitz function, see the example by Gouëzel [47]).

In the previous sections, a new technique for proving empirical process invariance principle is developed using approximations of indicators by regular functions. The fourth moment bound that we develop in Corollary 3.11 is well adapted to this situation. The main point to note is that the Banach norm only appears through its logarithm. In Subsection 3.5.1 and Subsection 3.5.2 the fourth moment inequality is stated and proved for strongly ergodic Markov chain. In Subsection 3.5.3, we state the same moment bound for a class of dynamical systems.

3.5.1 Fourth moment inequality for Markov chains

Let $(X_n)_{n \geq 0}$ be a homogeneous Markov chain with a stationary measure ν . Denote by P the associated Markov operator and E the state space. Consider a Banach space $(\mathcal{B}, \|\cdot\|)$ of ν -measurable functions from E to $\mathbb{R}$, which contains the function $\mathbf{1} = \mathbb{1}_E$. We will assume that the chain is $\mathcal{B}$ -geometrically ergodic.

Definition 3.9. The Markov chain $(X_n)_{n \geq 0}$ is $\mathcal{B}$ -geometrically ergodic or strongly ergodic (with respect to $\mathcal{B}$) if

a) there exist $\kappa > 0$ and $0 < \theta < 1$ such that for all $f \in \mathcal{B}$,

$$\|P^n f - \Pi f\| \leq \kappa \theta^n \|f\| \quad (3.13)$$

where $\Pi f = E_\nu(f) \mathbf{1}$.

b) there exists $p \geq 1$ such that $(\mathcal{B}, \|\cdot\|)$ is continuously included in $(L^p(\nu), \|\cdot\|_p)$.
i.e. $\exists C > 0$ such that $\forall f \in \mathcal{B}$,

$$\|f\|_p \leq C \|f\| \quad (3.14)$$

where $\|f\|_p = \left(\int |f|^p d\nu \right)^{\frac{1}{p}}$.

Further, we assume that there exists a constant $M > 0$ such that for all $f \in \mathcal{B}$ and for all $n \in \mathbb{N}$,

$$\|f P^n f\| \leq M \|f\| \|P^n f\|. \quad (3.15)$$

In particular, this is the case if $(\mathcal{B}, \|\cdot\|)$ is a Banach algebra. In the sequel, with no loss of generality, we assume $M = 1$.

Strong ergodicity covers a large class of examples (discussed in Section 3.6). It corresponds to the fact that the Markov transition operator acting on $\mathcal{B}$ has 1 as simple eigenvalue and the rest of the spectre is included in a closed ball of radius strictly smaller than 1.

For a function $\varphi : E \longrightarrow \mathbb{R}$, we consider the partial sum

$$S_n(\varphi) = \sum_{i=1}^n \varphi(X_i).$$

The aim is to get a fourth moment inequality for this partial sum when the function φ belongs to the space $\mathcal{B}$. Our main results are the following.

Theorem 3.10. If $(X_n)_{n \geq 0}$ is a $\mathcal{B}$ -geometrically ergodic Markov chain with stationary measure ν , then for all $\varphi \in \mathcal{B}$ such that $E_\nu(\varphi) = 0$ and $\varphi \in L^4(\nu) \cap L^{3q}(\nu)$,

$$\begin{aligned} & E_\nu [S_n(\varphi)^4] \\ & \leq K \left[n \|\varphi(X_0)\|_1^4 \log^3(\|\varphi\| + 1) \right. \\ & \quad + n (\|\varphi(X_0)\|_q^3 + \|\varphi(X_0)\|_q^2 + \|\varphi(X_0)\|_q + \|\varphi(X_0)\|_q^2) \log^2(\|\varphi\| + 1) \\ & \quad \left. + n^2 (\|\varphi(X_0)\|_1 \log(\|\varphi\| + 1) + \|\varphi(X_0)\|_q)^2 \right] \end{aligned}$$

where $\frac{1}{p} + \frac{1}{q} = 1$ and K is a constant.

As consequence, the following corollary gives a simpler inequality in the case in which the function φ is bounded.

Corollary 3.11. *If $(X_n)_{n \geq 0}$ is a $\mathcal{B}$ -geometrically ergodic Markov chain with stationary measure ν , then for all $\varphi \in \mathcal{B}$ such that $E_\nu(\varphi) = 0$ and φ is bounded,*

$$E_\nu [S_n(\varphi)^4] \leq K m_\varphi^3 [n \|\varphi(X_0)\|_q \log^3(\|\varphi\| + 1) + n^2 \|\varphi(X_0)\|_q^2 \log^2(\|\varphi\| + 1)],$$

where $m_\varphi = \max\{1, \sup_x |\varphi(x)|\}$.

Assume that one can prove a multivariate central limit theorem for functions in $\mathcal{B}$, then by the technique of the preceding sections, if the space $\mathcal{B}$ contains a class of functions approximating the indicators, an empirical process invariance principle follows.

To be complete, we state the following result, which is a corollary of Gordin's Theorem [44] and from which we can deduce a multivariate central limit theorem.

Proposition 3.12. *If $(X_n)_{n \geq 0}$ is ergodic and $\mathcal{B}$ -geometrically ergodic with $p \geq 2$ in (3.14), then for all $\varphi \in \mathcal{B}$, $\frac{S_n(\varphi - E_\nu \varphi)}{\sqrt{n}}$ converges in distribution to a centred normal law.*

See Hennion and Hervé [54], Thm.IX.2, for sufficient conditions on $\mathcal{B}$ to have the property that strong ergodicity implies ergodicity.

3.5.2 Proof of Theorem 3.10

Let us consider the assumptions of Theorem 3.10 hold. In the sequel, all the expectations are considered with respect to the measure ν and $\mathcal{F}_i$ denotes the σ -algebra generated by X_i .

Let $\varphi \in \mathcal{B}$ with $E(\varphi) = 0$. As the Markov chain is a stationary process, we have the following bound:

$$E[S_n(\varphi)^4] \leq 4!n \sum_{\substack{i,j,k \geq 0: \\ i+j+k \leq n}} E(\varphi(X_0) \varphi(X_i) \varphi(X_{i+j}) \varphi(X_{i+j+k}))$$

which can be decomposed into three sums.

$$E[S_n(\varphi)^4] \leq 4!n \left[\sum_{i=1}^n \sum_{j,k \leq i} E(\varphi(X_0) \varphi(X_i) \varphi(X_{i+j}) \varphi(X_{i+j+k})) \right] \quad (3.16)$$

$$+ \sum_{j=1}^n \sum_{i,k \leq j} E(\varphi(X_0) \varphi(X_i) \varphi(X_{i+j}) \varphi(X_{i+j+k})) \quad (3.17)$$

$$+ \sum_{k=1}^n \sum_{i,j \leq k} E(\varphi(X_0) \varphi(X_i) \varphi(X_{i+j}) \varphi(X_{i+j+k})) \Big]. \quad (3.18)$$

So to study the terms $E(\varphi(X_0)\varphi(X_i)\varphi(X_{i+j})\varphi(X_{i+j+k}))$, we will consider three cases according to the greatest integer between i, j and k .

Before, we can see that in all cases, by Hölder inequality, we have

$$|E(\varphi(X_0)\varphi(X_i)\varphi(X_{i+j})\varphi(X_{i+j+k}))| \leq \|\varphi(X_0)\|_4^4 \leq \|\varphi(X_0)^4\|_1. \quad (3.19)$$

Further, let n_0 be a positive integer such that

$$-\frac{\log(\|\varphi\| + 1)}{\log \theta} < n_0 \leq -\frac{\log(\|\varphi\| + 1)}{\log \theta} + 1.$$

Note that $\theta^{n_0}\|\varphi\| \leq 1$.

First case: $i, j \leq k$.

Here, we use the properties of the Markov operator P on the space $\mathcal{B}$ to get another majoration. Applying successively Hölder inequality, properties (3.14) and (3.13), we get

$$\begin{aligned} & |E(\varphi(X_0)\varphi(X_i)\varphi(X_{i+j})\varphi(X_{i+j+k}))| \\ &= |E(\varphi(X_0)\varphi(X_i)\varphi(X_{i+j})(E(\varphi(X_{i+j+k})|\mathcal{F}_{i+j}) - E(\varphi(X_{i+j+k}))))| \\ &\leq \|\varphi(X_0)\varphi(X_i)\varphi(X_{i+j})\|_q \|P^k\varphi(X_0) - \Pi\varphi(X_0)\|_p \\ &\leq \|\varphi(X_0)\|_{3q}^3 C \|P^k\varphi - \Pi\varphi\| \\ &\leq \|\varphi(X_0)\|_q^3 C \kappa \theta^k \|\varphi\|. \end{aligned} \quad (3.20)$$

Now, for sum (3.18), using (3.19) for the $n_0 - 1$ first terms and (3.20) for the others, we obtain

$$\begin{aligned} & \sum_{k=1}^n \sum_{i,j \leq k} E(\varphi(X_0)\varphi(X_i)\varphi(X_{i+j})\varphi(X_{i+j+k})) \\ &\leq \sum_{k=1}^{n_0-1} k^2 \|\varphi(X_0)^4\|_1 + \sum_{k=n_0}^n k^2 C \kappa \|\varphi(X_0)\|_q^3 \theta^k \|\varphi\| \\ &\leq (n_0 - 1)^3 \|\varphi(X_0)^4\|_1 + C \kappa \|\varphi(X_0)\|_q^3 \sum_{k=n_0}^n k^2 \theta^{k-n_0}. \end{aligned}$$

There exists a constant C_1 which only depends on θ , such that

$$\begin{aligned} \sum_{k=n_0}^n k^2 \theta^{k-n_0} &\leq \sum_{k \geq 2} (k + n_0 - 2)^2 \theta^{k-2} \\ &\leq \sum_{k \geq 2} k^2 \theta^{k-2} + 2(n_0 - 2) \sum_{k \geq 2} k \theta^{k-2} + (n_0 - 2)^2 \sum_{k \geq 2} \theta^{k-2} k \\ &\leq C_1 (n_0 - 1)^2, \end{aligned}$$

because the three series converge.

Thus, writing $C_2 = -\frac{1}{\log \theta}$, we get $n_0 - 1 \leq C_2 \log(\|\varphi\| + 1)$ and

$$\begin{aligned} & \sum_{k=1}^n \sum_{i,j \leq k} E(\varphi(X_0) \varphi(X_i) \varphi(X_{i+j}) \varphi(X_{i+j+k})) \\ & \leq C_2^3 \|\varphi(X_0)^4\|_1 \log^3(\|\varphi\| + 1) + C_3 \|\varphi(X_0)^3\|_q \log^2(\|\varphi\| + 1), \end{aligned} \quad (3.21)$$

where $C_3 = C\kappa C_1 C_2^2$.

Second case: $i, k \leq j$.

We can decompose the expectation as follows,

$$\begin{aligned} & |E(\varphi(X_0) \varphi(X_i) \varphi(X_{i+j}) \varphi(X_{i+j+k}))| \\ & \leq |E(\varphi(X_0) \varphi(X_i) (E(\varphi(X_{i+j}) E(\varphi(X_{i+j+k}) | \mathcal{F}_{i+j}) | \mathcal{F}_i) - E(\varphi(X_0) \varphi(X_k))))| \\ & \quad + |E(\varphi(X_0) \varphi(X_i)) E(\varphi(X_0) \varphi(X_k))|. \end{aligned} \quad (3.22)$$

In the right hand side, let us call (I) the first term and (II) the second one.

Since $E(\varphi P^k \varphi(X_0)) = E(\varphi(X_0) E(\varphi(X_k) | \mathcal{F}_0)) = E(\varphi(X_0) \varphi(X_k))$, we have

$$\begin{aligned} (I) & \leq \|\varphi(X_0) \varphi(X_i)\|_q \|P^j(\varphi P^k \varphi)(X_0) - \Pi(\varphi P^k \varphi)(X_0)\|_p \\ & \leq \|\varphi(X_0)\|_{2q}^2 C \|P^j(\varphi P^k \varphi) - \Pi(\varphi P^k \varphi)\| \\ & \leq C \|\varphi(X_0)^2\|_q \kappa \theta^j \|\varphi P^k \varphi\| \end{aligned}$$

and by assumption (3.15),

$$\|\varphi P^k \varphi\| \leq \|\varphi\| \|P^k \varphi\| \leq \kappa \theta^k \|\varphi\|^2.$$

Therefore,

$$(I) \leq C \kappa^2 \|\varphi(X_0)^2\|_q \theta^{j+k} \|\varphi\|^2. \quad (3.23)$$

Now, thanks to the decomposition (3.22) (using also inequality (3.19)), for n big enough,

$$\begin{aligned} & \sum_{j=1}^n \sum_{i,k \leq j} E(\varphi(X_0) \varphi(X_i) \varphi(X_{i+j}) \varphi(X_{i+j+k})) \\ & \leq \sum_{j=1}^{2n_0-2} j^2 \|\varphi(X_0)^4\|_1 + \sum_{j=2n_0-1}^n \sum_{i,k \leq j} ((I) + (II)) \\ & \leq 8(n_0 - 1)^3 \|\varphi(X_0)^4\|_1 + \sum_{j=2n_0-1}^n j^2 (I) + \sum_{j=1}^n \sum_{i,k \leq j} (II), \end{aligned}$$

where n_0 has been defined previously.

Inequality (3.23) and $\theta^{n_0}\|\varphi\| \leq 1$ imply

$$\begin{aligned} \sum_{j=2n_0-1}^n j^2(I) &\leq \sum_{j=2n_0-1}^n j^2 C\kappa^2 \|\varphi(X_0)^2\|_q \theta^j \|\varphi\|^2 \\ &\leq C\kappa^2 \|\varphi(X_0)^2\|_q \sum_{j=2n_0-1}^n j^2 \theta^{j-2n_0}. \end{aligned}$$

As before, there exists a constant C_4 depending on θ such that,

$$\sum_{j=2n_0-1}^n j^2 \theta^{j-2n_0} \leq C_4(n_0 - 1)^2.$$

So,

$$\sum_{j=2n_0-1}^n j^2(I) \leq C\kappa^2 \|\varphi(X_0)^2\|_q C_4(n_0 - 1)^2.$$

For the third term, we have

$$\sum_{j=1}^n \sum_{i,k \leq j} (II) \leq n \left(\sum_{i=1}^n |E(\varphi(X_0)\varphi(X_i))| \right) \left(\sum_{k=1}^n |E(\varphi(X_0)\varphi(X_k))| \right).$$

We can see that

$$\begin{aligned} |E(\varphi(X_0)\varphi(X_i))| &\leq \|\varphi(X_0)\|_q \|P^i\varphi(X_0) - \Pi\varphi(X_0)\|_p \\ &\leq C\|\varphi(X_0)\|_q \|P^i\varphi - \Pi\varphi\| \\ &\leq C\kappa\|\varphi(X_0)\|_q \theta^i \|\varphi\| \end{aligned} \tag{3.24}$$

and, in the same way,

$$|E(\varphi(X_0)\varphi(X_k))| \leq C\kappa\|\varphi(X_0)\|_q \theta^k \|\varphi\|. \tag{3.25}$$

Alternatively, by Hölder inequality,

$$|E(\varphi(X_0)\varphi(X_i))| \leq \|\varphi(X_0)^2\|_1 \quad \text{and} \quad |E(\varphi(X_0)\varphi(X_k))| \leq \|\varphi(X_0)^2\|_1. \tag{3.26}$$

Thus, by (3.24), (3.25) and (3.26),

$$\begin{aligned} \sum_{i=1}^n |E(\varphi(X_0)\varphi(X_i))| &\leq \sum_{i=1}^{n_0-1} \|\varphi(X_0)^2\|_1 + \sum_{i=n_0}^n C\kappa\|\varphi(X_0)\|_q \theta^i \|\varphi\| \\ &\leq (n_0 - 1)\|\varphi(X_0)^2\|_1 + C\kappa\|\varphi(X_0)\|_q \sum_{i=n_0}^n \theta^{i-n_0} \\ &\leq (n_0 - 1)\|\varphi(X_0)^2\|_1 + C_5\|\varphi(X_0)\|_q, \end{aligned}$$

where $C_5 = C\kappa \sum_{i \geq 0} \theta^i < \infty$ and

$$\sum_{k=1}^n |E(\varphi(X_0) \varphi(X_k))| \leq (n_0 - 1) \|\varphi(X_0)^2\|_1 + C_5 \|\varphi(X_0)\|_q.$$

Finally,

$$\begin{aligned} & \sum_{j=1}^n \sum_{i,k \leq j} E(\varphi(X_0) \varphi(X_i) \varphi(X_{i+j}) \varphi(X_{i+j+k})) \\ & \leq 8C_2^3 \|\varphi(X_0)^4\|_1 \log^3(\|\varphi\| + 1) + C_6 \|\varphi(X_0)^2\|_q \log^2(\|\varphi\| + 1) \\ & \quad + n (C_2 \|\varphi(X_0)^2\|_1 \log(\|\varphi\| + 1) + C_5 \|\varphi(X_0)\|_q)^2, \end{aligned} \quad (3.27)$$

where $C_6 = C\kappa^2 C_4 C_2^2$.

Third case: $j, k \leq i$.

Three uses of the operator properties give

$$\begin{aligned} & |E(\varphi(X_0) \varphi(X_i) \varphi(X_{i+j}) \varphi(X_{i+j+k}))| \\ & = |E(\varphi(X_0) E(\varphi(X_i) E(\varphi(X_{i+j}) E(\varphi(X_{i+j+k}) | \mathcal{F}_{i+j}) | \mathcal{F}_i) | \mathcal{F}_0))| \\ & = |E[\varphi(X_0) [E(\varphi(X_i) E(\varphi(X_{i+j}) E(\varphi(X_{i+j+k}) | \mathcal{F}_{i+j}) | \mathcal{F}_i) | \mathcal{F}_0) \\ & \quad - E(\varphi(X_i) \varphi(X_{i+j}) \varphi(X_{i+j+k}))]]| \\ & \leq \|\varphi(X_0)\|_q \|P^i(\varphi P^j(\varphi P^k \varphi))(X_0) - \Pi(\varphi P^j(\varphi P^k \varphi))(X_0)\|_p \\ & \leq \|\varphi(X_0)\|_q C\kappa \theta^i \|\varphi P^j(\varphi P^k \varphi)\| \end{aligned} \quad (3.28)$$

and

$$\begin{aligned} \|\varphi P^j(\varphi P^k \varphi)\| & \leq \|\varphi\| \|P^j(\varphi P^k \varphi)\| \\ & \leq \|\varphi\| (\|P^j(\varphi P^k \varphi) - \Pi \varphi P^k \varphi\| + |E(\varphi(X_0) \varphi(X_k))|) \\ & \leq \|\varphi\| (\kappa \theta^j \|\varphi P^k \varphi\| + |E(\varphi(X_0) \varphi(X_k))|) \\ & \leq \kappa^2 \theta^{j+k} \|\varphi\|^3 + C\kappa \|\varphi(X_0)\|_q \theta^k \|\varphi\|^2, \end{aligned} \quad (3.29)$$

where we used inequality (3.25) at the last line.

From (3.28) and (3.29), we derive

$$\begin{aligned} & |E(\varphi(X_0) \varphi(X_i) \varphi(X_{i+j}) \varphi(X_{i+j+k}))| \\ & \leq C\kappa^3 \|\varphi(X_0)\|_q \theta^{i+j+k} \|\varphi\|^3 + C^2 \kappa^2 \|\varphi(X_0)\|_q^2 \theta^{i+k} \|\varphi\|^2 \\ & \leq C_7 \theta^i \|\varphi(X_0)\|_q \|\varphi\|^2 (\|\varphi\| + \|\varphi(X_0)\|_q) \end{aligned}$$

where $C_7 = \max\{C\kappa^3, C^2\kappa^2\}$.

With this last inequality and (3.19), the sum (3.16) can be bounded in the same way as before. We use the integer n_0 to get

$$\begin{aligned}
& \sum_{i=1}^n \sum_{j,k \leq i} E(\varphi(X_0) \varphi(X_i) \varphi(X_{i+j}) \varphi(X_{i+j+k})) \\
& \leq \sum_{i=1}^{3n_0-3} i^2 \|\varphi(X_0)\|_1^4 + C_7 \sum_{i=3n_0-2}^n i^2 \theta^i \|\varphi(X_0)\|_q \|\varphi\|^2 (\|\varphi\| + \|\varphi(X_0)\|_q) \\
& \leq 27(n_0 - 1)^3 \|\varphi(X_0)\|_1^4 + C_7 (\|\varphi(X_0)\|_q + \|\varphi(X_0)\|_q^2) \sum_{i=3n_0-2}^n i^2 \theta^{i-3n_0}.
\end{aligned}$$

The sum is bounded by the corresponding serie which is finite (majoration by $C_8(n_0 - 1)^2$, where C_8 depends only on θ). So, we can conclude the study of the third case by

$$\begin{aligned}
& \sum_{i=1}^n \sum_{j,k \leq i} E(\varphi(X_0) \varphi(X_i) \varphi(X_{i+j}) \varphi(X_{i+j+k})) \\
& \leq 27C_2^3 \|\varphi(X_0)\|_1^4 \log^3(\|\varphi\| + 1) \\
& \quad + C_7 C_8 C_2^2 (\|\varphi(X_0)\|_q + \|\varphi(X_0)\|_q^2) \log^2(\|\varphi\| + 1).
\end{aligned} \tag{3.30}$$

To conclude, let K be the maximum of all the constants appearing in (3.21), (3.27) and (3.30):

$$\begin{aligned}
& E[S_n(\varphi)^4] \\
& \leq 4!K [n \|\varphi(X_0)\|_1^4 \log^3(\|\varphi\| + 1) \\
& \quad + n (\|\varphi(X_0)\|_q^3 + \|\varphi(X_0)\|_q^2 \|\varphi(X_0)\|_q + \|\varphi(X_0)\|_q + \|\varphi(X_0)\|_q^2) \log^2(\|\varphi\| + 1) \\
& \quad + n^2 (\|\varphi(X_0)\|_1^2 \log(\|\varphi\| + 1) + \|\varphi(X_0)\|_q^2)] .
\end{aligned}$$

□

3.5.3 Fourth moment inequality for dynamical systems

In Section 3.5.1, we dealt with homogeneous Markov chains through its operator. In the same way, the techniques can be applied to dynamical systems, using transfer operator. Here, we state the result for dynamical systems but the proof (which is essentially the same as in Section 3) is left to the reader.

Let $(\Omega, \mathcal{A}, \mu)$ be a probability space and T a measurable measure preserving transformation (*i.e.* $\forall A \in \mathcal{A}, \mu(T^{-1}A) = \mu(A)$). Let us consider the Perron-Frobenius operator (or the transfer operator) of T , $P : L^1(\mu) \longrightarrow L^1(\mu)$ defined by

$$\int_{\Omega} Pf(x)g(x)d\mu(x) = \int_{\Omega} f(x)g \circ T(x)d\mu(x)$$

for all $f \in L^1(\mu)$ and $g \in L^\infty(\mu)$.

As in the Markov case, we assume there exists a Banach space $(\mathcal{B}, \|\cdot\|)$ of μ -measurable functions from Ω to $\mathbb{R}$ which contains $\mathbf{1}$ and satisfies (3.15) and that P verifies same assumptions:

(i) there exist $\kappa > 0$ and $0 < \theta < 1$ such that for all $f \in \mathcal{B}$,

$$\|P^n f - \Pi f\| \leq \kappa \theta^n \|f\| \quad (3.31)$$

where $\Pi f = E_\mu(f) \mathbf{1}$.

(ii) there exists $p \geq 1$ such that $(\mathcal{B}, \|\cdot\|)$ is continuously included in $(L^p(\mu), \|\cdot\|_p)$.

Again, assertion (i) follows from some quasi-compactness of the Perron-Frobenius operator, see Hennion and Hervé [54], Baladi [2].

Under these assumptions, we have same fourth moment bound:

Theorem 3.13. *For all $f \in \mathcal{B}$ such that $E_\mu(f) = 0$ and f is bounded,*

$$E_\mu \left[\left(\sum_{i=1}^n f \circ T^i \right)^4 \right] \leq K m_f^3 [n \|f\|_q \log^3(\|f\| + 1) + n^2 \|f\|_q^2 \log^2(\|f\| + 1)] .$$

where $\frac{1}{p} + \frac{1}{q} = 1$ and $m_f = \max\{1, \sup_x |f(x)|\}$.

3.6 Applications

We give some examples where the fourth moment inequality applies and then leads to some empirical process invariance principles.

3.6.1 Uniform ergodicity

Let $(X_n)_{n \geq 0}$ be a Markov chain on the state space E . Denote by $(\mathcal{B}^\infty, \|\cdot\|_\infty)$ the space of bounded measurable functions from E to $\mathbb{R}$ provided with the uniform norm. One says that the Markov chain $(X_n)_{n \geq 0}$ is *uniformly ergodic* if it is $\mathcal{B}^\infty$ -geometrically ergodic. This condition is equivalent to the fact that the process satisfies the so-called Doeblin's condition (see Meyn and Tweedie [71]).

In this situation, if X_0 has a distribution function which is enough regular, our fourth moment bound (Corollary 3.11) implies inequality (3.12). Then tightness follows and empirical process invariance principle will follow from the multivariate central limit theorem. Note that this result is already known since Billingsley [6].

See many examples of uniformly ergodic Markov chains in Meyn and Tweedie [71], like T-chains on compact spaces. Another example is given by the Knudsen Gas model (see e.g. Pène [77]).

3.6.2 Expanding maps

For a large class of expanding maps of the interval (the class of Lasota-Yorke transformations), empirical process invariance principles have already been established in Collet, Martinez and Schmitt [13] and Dedecker and Priour [18]. Our result also covers these ones. Many of such transformation are studied in Broise [10], as continuous fraction expansions, β -transformations, Gauss maps. For these examples, one can show that the Perron-Frobenius operator admits a spectral gap on the space BV of bounded variation functions. Since the indicator functions belong to BV, Theorem 3.13 with condition on the initial distribution prove inequality (3.12).

But according to Gouëzel [47], there exist some uniformly expanding maps of the interval for which the transfer operator does not act continuously on the space BV, but admits a spectral gap on the space of Lipschitz functions. The example given by Gouëzel is a transformation of the interval $[0, 1)$. Let $(a_n)_{n \geq 1}$ be a sequence of positive numbers with $\sum a_n < \frac{1}{4}$ and let $N > 0$ be an integer. Denote by I_n the subintervals $[4 \sum_{i=1}^{n-1} a_i, 4 \sum_{i=1}^n a_i)$. We decompose I_n into two subintervals of length $2a_n$ denoted by $I_n^{(1)}$ and $I_n^{(2)}$. We can find a map v_n (resp. w_n) on $[0, 1)$ with image $I_n^{(1)}$ (resp. $I_n^{(2)}$) such that the derivative at a point x is equal to $a_n(1 + 2\cos^2(2\pi n^4 x))$ (resp. $a_n(1 + 2\sin^2(2\pi n^4 x))$). The map T is defined on I_n in such a way that v_n and w_n are two inverse branches of it. It remains the interval $[4 \sum_{i=1}^{\infty} a_i, 1)$ that we subdivide into N subintervals of equal length. T is defined as an affine transformation on each of these subintervals onto $[0, 1)$.

Theorem. (Gouëzel [47]) If $a_n = \frac{1}{100n^3}$ and $N = 4$, then T is a Lebesgue measure preserving transformation and its associated transfer operator has a spectral gap on the space of Lipschitz functions with a simple eigenvalue at 1 and no other eigenvalue of modulus 1. Further, the transfer operator does not act continuously on BV .

For this example, we cannot apply the result by Collet et al. [13]. But, the spectral gap property gives the condition (3.31) on the space of Lipschitz functions. By Theorem 3.13, the 4-th moment bound (3.2) can be deduced and then, Theorem 3.1 can be used to get an invariance principle for the associated empirical process.

3.6.3 Subshifts

Let E be a finite set and $\mathcal{E} = E^{\mathbb{N}}$. The metric d defined on $\mathcal{E}$ is

$$d(x, y) = 2^{-\inf\{k \geq 0 : x_k \neq y_k\}}.$$

Let $A = (a(i, j))_{i, j \in E}$ be a matrix with coefficients in $\{0, 1\}$ and

$$\Omega = \{x \in \mathcal{E} : a(x_k, x_{k+1}) = 1, \forall k \geq 0\}.$$

Write T the shift operator on Ω , i.e. $(Tx)_k = x_{k+1}$, $\forall k \geq 0$. Denote by $\mathcal{B}$ the space of complex valued functions on Ω , which are Lipschitz continuous with respect to the metric d . The norm on $\mathcal{B}$ is $\|\cdot\| = \|\cdot\|_{\infty} + m(\cdot)$, where

$$m(f) = \sup \left\{ \frac{|f(x) - f(y)|}{d(x, y)}, x \neq y \right\}. \quad (3.32)$$

Note that, since (Ω, d) is compact, $\mathcal{B} \subset L^{\infty}$. The Ruelle-Perron-Frobenius Theorem shows that the transfer operator P has some quasi-compact properties on $\mathcal{B}$. See Parry and Pollicott [74] or Hennion and Hervé [54]. If 1 is the only eigenvalue of modulus 1 and is simple, then conditions (i) and (ii) holds and Theorem 3.13 is satisfied. If f is a Lipschitz continuous function on Ω , then by Theorem 3.1, an empirical process invariance principle is satisfied for the process $(f \circ T^i)_{i \geq 0}$.

3.6.4 Linear processes

Let $(A, \|\cdot\|_A)$ be a measurable Banach space and $(a_i)_{i \geq 0}$ a sequence of linear forms on A such that there exist $C > 0$ and $0 < \theta < 1$ such that

$$|a_i| \leq C\theta^i, \quad (3.33)$$

where $|a_i| = \sup_{\|x\|_A \leq 1} |a_i(x)|$. Let $(e_i)_{i \in \mathbb{Z}}$ be an i.i.d. bounded random sequence with values in a compact subset $B \subset A$ and marginal distribution μ . We define the real-valued linear process $(X_k)_{k \geq 0}$ by

$$X_k = \sum_{i \geq 0} a_i(e_{k-i}), \quad k \geq 0.$$

Several results have already been established for empirical processes of linear processes (see Doukhan and Surgailis [33], Wu [90], Dedecker and Prieur [18]). Here, assumption on the $(a_i)_{i \geq 0}$ is stronger than in the mentioned papers, but there will be no assumption on the distribution of the e_i 's and assumption on the distribution function of X_0 will be weaker. Note that $(X_k)_{k \geq 0}$ can be viewed as a functional of a Markov chain.

Let $Y_k = (e_k, e_{k-1}, \dots)$, then $(Y_k)_{k \geq 0}$ is a stationary Markov chain on $B^{\mathbb{N}}$ (with stationary measure $\mu^{\otimes \mathbb{N}}$) and $X_k = \Phi(Y_k)$ where

$$\Phi : B^{\mathbb{N}} \longrightarrow \mathbb{R}, \quad \Phi(x_0, x_1, \dots) = \sum_{i \geq 0} a_i(x_i).$$

Let Q be the Markov transition operator of the chain. On $B^{\mathbb{N}}$, we define a metric d by

$$d(x, y) = \sum_{i \geq 0} \theta^i \|x_i - y_i\|_A$$

where $x = (x_i)_{i \geq 0}$ and $y = (y_i)_{i \geq 0}$. As B is compact, then $(B^{\mathbb{N}}, d)$ is also compact. Let us denote by $\mathcal{L}$ the space of all Lipschitz functions from $B^{\mathbb{N}}$ to $\mathbb{R}$ provided with the norm $\|\cdot\|$ defined by

$$\|f\| = \sup_{x \in B^{\mathbb{N}}} |f(x)| + \sup_{x \neq y} \frac{|f(x) - f(y)|}{d(x, y)}.$$

For all $f \in \mathcal{L}$ and for all $x = (x_i)_{i \geq 0}$ and $y = (y_i)_{i \geq 0} \in B^{\mathbb{N}}$, we have

$$\begin{aligned} |Q^k f(x) - Q^k f(y)| &= |E(f(Y_k)|Y_0 = x) - E(f(Y_k)|Y_0 = y)| \\ &= |E(f(e_k, \dots, e_1, x_0, \dots)) - E(f(e_k, \dots, e_1, y_0, \dots))| \\ &\leq \|f\| E\{d((e_k, \dots, e_1, x_0, \dots), (e_k, \dots, e_1, y_0, \dots))\} \\ &\leq C\theta^k \|f\| d(x, y), \end{aligned}$$

and

$$\begin{aligned} |Q^k f(x) - Ef(Y_0)| &= |E(f(Y_k)|Y_0 = x) - Ef(Y_k)| \\ &\leq E|f(e_k, e_{k-1}, \dots, e_1, x_0, \dots) - f(e_k, e_{k-1}, \dots)| \\ &\leq C\theta^k \|f\| E\{d(x, Y_0)\}. \end{aligned}$$

Then, we have for all $f \in \mathcal{L}$,

$$\|Q^k f - Ef(Y_0)\| \leq C\theta^k \|f\|.$$

Since $(\mathcal{L}, \|\cdot\|) \subset (L^\infty(\mu^{\otimes \mathbb{N}}), \|\cdot\|_\infty)$, by Corollary 3.11, $(Y_k)_{k \geq 0}$ satisfies the 4-th moment bound (3.2) with $\alpha = 3$ and $\beta = 2$ for all Lipschitz functions. Further, for all $f \in \mathcal{L}$ the sequence $\sum_{i=0}^n Q^i f(Y_0)$ converges in $L^2(\mu^{\otimes \mathbb{N}})$ and so by Gordin's theorem (see Gordin [44]), the CLT (3.1) is satisfied. Clearly, the function Φ is a Lipschitz continuous function on $B^\mathbb{N}$, and for all Lipschitz function $g : \mathbb{R} \rightarrow \mathbb{R}$, $g \circ \Phi$ is also a Lipschitz continuous function on $B^\mathbb{N}$. Thus conditions (3.1) and (3.2) hold for the process $(X_k)_{k \geq 0}$, for all Lipschitz function on $\mathbb{R}$. Then Theorem 3.1 applies and we have

Corollary 3.14. *Let $(X_k)_{k \geq 0}$ be a real linear process defined by a sequence of linear forms $(a_i)_{i \geq 0}$ and a sequence of i.i.d. bounded random variables $(e_i)_{i \in \mathbb{Z}}$, both on a measurable Banach space A . Assume (a_i) satisfies (3.33) and the distribution function F of X_0 satisfies*

$$\omega_F(\delta) \leq D |\log(\delta)|^{-\gamma} \text{ for some } D > 0 \text{ and } \gamma > 2.$$

Then $(U_n(t))_{t \in \mathbb{R}}$ converges in distribution to a mean-zero Gaussian process.

In the paper by Dedecker and Priour [18], Corollary 1, X_0 has a bounded density. Here, the existence of a density is not needed. Our result is comparable to a result of Wu and Shao [91].

For a concrete example, consider $A = \{0, 1\}$, $a_i = \frac{2}{3^i}$, $i \geq 0$ and $e_k = 0$ or 1 with probability $\frac{1}{2}$, $k \in \mathbb{Z}$. Then

$$X_k = 2 \sum_{i \geq 0} \frac{e_{k-i}}{3^i}, \quad k \geq 0$$

is a stationary process with values in $[0, 1]$ and the common distribution function of all the X_k is the Cantor function, which is not absolutely continuous but $\frac{\log 2}{\log 3}$ -Hölder continuous (see Dovgoshey et al. [34]).

3.6.5 Random iterative Lipschitz models

Let G be a semi-group of Lipschitz transformations of a metric space (E, d) and let $\mathcal{G}$ be a σ -algebra on G . We assume that the action of G on E is measurable.

Let $(g_n)_{n \geq 1}$ be an i.i.d. sequence of G -valued random variables with distribution η . Let X_0 be a E -valued random variable independent of $(g_n)_{n \geq 1}$. We consider the Markov chain $(X_n)_{n \geq 0}$ defined by

$$X_n = g_n(X_{n-1})$$

with transition operator P defined by

$$Pf(x) = \int_G f(g(x)) d\eta(g).$$

One says η is contracting if

$$\limsup_n \left\{ \int_G \frac{d(g(x), g(y))}{d(x, y)} d\eta^{*n}(g) : x, y \in E, x \neq y \right\}^{\frac{1}{n}} < 1$$

where η^{*n} denotes the distribution of $g_n \circ \dots \circ g_1$.

Assume (E, d) is compact. Let $\mathcal{B}_0$ be the space of $\mathbb{C}$ -valued Lipschitz continuous functions on E provided with the norm $\|\cdot\|_0 = \|\cdot\|_\infty + m(\cdot)$, where m is defined as in (3.32). It is shown in Hennion and Hervé [54], Thm.X.3, that if η is contracting, then there exists a unique P -invariant measure on E and $(X_n)_{n \geq 0}$ is $\mathcal{B}_0$ -geometrically ergodic with respect to this measure. Then fourth moment bound holds and thanks to Theorem 3.1, an empirical invariance principle follows. One example of application is given by products of invertible random matrices (see Hennion and Hervé [54] X.4).

In the case where (E, d) is not compact but every closed ball in E is compact, one can have a similar result but with another Banach space (see Hennion and Hervé [54] Thm.X.4). Here, the Banach space is the space $\mathcal{B}_1$ of locally Lipschitz functions with weight. These are the $\mathbb{C}$ -valued functions f such that

$$m_1(f) = \sup \left\{ \frac{|f(x) - f(y)|}{d(x, y)p(x)p(y)} : x \neq y \right\} < \infty$$

where $p(x) = 1 + d(x, x_0)$ for a fixed $x_0 \in E$. The norm is

$$\|f\|_1 = \sup \left\{ \frac{|f(x)|}{p(x)^2} : x \in E \right\} + m_1(f).$$

As an example, we can mention a large class of autoregressive models.

3.6.6 Autoregressive models

The autoregressive (AR) processes are used in statistical time series analysis. The process $(X_n)_{n \geq 0} \subset \mathbb{R}^d$ is called autoregressive with initial value $X_0 \in \mathbb{R}^d$ if it satisfies, for all $n \geq 1$,

$$X_n = AX_{n-1} + Y_n$$

where $A \in \mathcal{M}(\mathbb{R}^d)$ and $(Y_n)_{n \geq 1} \subset \mathbb{R}^d$ is an i.i.d. sequence of random variables, independent of X_0 . See Hennion and Hervé [54] Thm.X.16 for conditions under which $(X_n)_{n \in \mathbb{N}}$ is $\mathcal{B}_1$ -geometrically ergodic.

Comments

The fourth moment inequality comes from the paper by Durieu [36] while the remainder of the chapter corresponds to the paper by Dehling, Durieu and Volný [20]. We are grateful to Loïc Hervé for several lectures introducing us to the spectral gap method, and to Jérôme Dedecker for his critical comments on an earlier version of the second paper.

Chapitre 4

On Sums of Indicator Functions in Dynamical Systems

Abstract

In this chapter, we are interesting in the limit theorem question for sums of indicator functions. We show that in every aperiodic dynamical system, for every increasing sequence $(a_n)_{n \in \mathbb{N}} \subset \mathbb{R}_+$ such that $a_n \nearrow \infty$ and $\frac{a_n}{n} \rightarrow 0$ as $n \rightarrow \infty$, there exist a measurable set A such that the sequence of the distributions of the partial sums $\frac{1}{a_n} \sum_{i=0}^{n-1} (\mathbb{1}_A - \mu(A)) \circ T^i$ is dense in the set of the probability measures on $\mathbb{R}$. Further, in the ergodic case, we prove that there exists a dense G_δ of such sets.

Keywords: Dynamical system; Ergodic theorem; Sums of random variables; Limit theorem; Genericity.

AMS Classification: 28D05; 37A50; 60F05; 60G10.

4.1 Introduction

In [11], Burton and Denker found that in every aperiodic dynamical system there exists a process $(f \circ T^i)$ for which the CLT holds and they posed the question how big is the subset of $f \in L^2$ with this property. Clearly, we have to study the space L_0^2 of f with $E(f|\mathcal{I}) = 0$. As already observed by Burton and Denker, because the coboundaries are dense in L_0^2 , this set is dense. In Volný [84], it has been found that for any sequence $a_n \rightarrow \infty$, $\frac{a_n}{n} \rightarrow 0$, there exists a dense G_δ part G of L_0^2 such that for any $f \in G$ and any probability law ν there exists a sequence $n_k \rightarrow \infty$ such that $\frac{1}{a_{n_k}} S_{n_k}(f)$ converge in law to ν . The same result takes place for all spaces L^p , $1 \leq p \leq \infty$. In Liardet and Volný [65] a similar result was found for the space of continuous functions and for irrational rotations of the circle for spaces of smooth functions. As a corollary we get that generically, the rate of convergence in the ergodic theorems (of Birkhoff and of von Neumann) is arbitrarily slow. This gave a new proof of a result of del Junco and Rosenblatt [22]. In the paper of del Junco and Rosenblatt a similar result on the rate of convergence in the ergodic theorems was found for functions $\mathbb{1}_A - \mu(A)$, the genericity was studied in the space of $A \in \mathcal{A}$ equipped with the (pseudo)metric of the measure of symmetric difference.

Here we shall study the distributional convergence for the functions $\mathbb{1}_A - \mu(A)$. The research was motivated by the study of the invariance principle of the empirical process of strictly stationary sequences $(X_i)_{i \in \mathbb{N}}$ in Dehling, Durieu and Volný [20].

4.2 Result

Let $(\Omega, \mathcal{A}, \mu)$ be a probability space. We say that a measurable transformation T , from Ω to Ω , is measure preserving if for all $A \in \mathcal{A}$, $\mu(T^{-1}A) = \mu(A)$. If it is the case, $(\Omega, \mathcal{A}, \mu, T)$ is called a dynamical system. Further, the dynamical system is aperiodic if

$$\mu\{x \in \Omega / \exists n \geq 1, T^n x = x\} = 0.$$

It is ergodic if for any $A \in \mathcal{A}$, if $T^{-1}A = A$ then $\mu(A) = 0$ or 1 .

On $\mathcal{A}$ we consider the pseudo-metric Θ defined by

$$\Theta(A, B) = \mu(A \Delta B), \quad A, B \in \mathcal{A}.$$

Our main result is the following.

Theorem 4.1. *Let $(\Omega, \mathcal{A}, \mu, T)$ be an aperiodic ergodic dynamical system, $(a_n)_{n \in \mathbb{N}} \subset \mathbb{R}_+$ be an increasing sequence such that $a_n \nearrow \infty$ and $\frac{a_n}{n} \rightarrow 0$ as $n \rightarrow \infty$.*

There exists a Θ -dense G_δ of sets $A \in \mathcal{A}$ having the property that for all probability ν on $\mathbb{R}$, there exists a subsequence $(n_k)_{k \in \mathbb{N}}$ satisfying

$$\frac{1}{a_{n_k}} S_{n_k}(\mathbb{1}_A - \mu(A)) \xrightarrow[k \rightarrow \infty]{\mathcal{D}} \nu.$$

4.3 Some preliminary results

Let $\mathcal{M}$ be the set of all probability measures on $\mathbb{R}$ and $\mathcal{M}_0$ be the set of all probability measures on $\mathbb{R}$ which have zero-mean. Recall that $\mathcal{M}_0$ is dense in $\mathcal{M}$ for the topology of the weak convergence. We denote by d the Lévy metric on $\mathcal{M}$. For all μ and ν in $\mathcal{M}$ with distribution functions F and G ,

$$d(\mu, \nu) = \inf\{\varepsilon > 0 / G(t - \varepsilon) - \varepsilon \leq F(t) \leq G(t + \varepsilon) + \varepsilon\}.$$

The space $(\mathcal{M}, d)$ is a complete separable metric space (see Dudley [35]).

If $X : \Omega \longrightarrow \mathbb{R}$ is a random variable, we denote by $\mathcal{L}_\Omega(X)$ the distribution of X on $\mathbb{R}$.

Lemma 4.2. *Let $(\Omega, \mathcal{A}, \mu)$ be a non-atomic probability space and ν be a probability on $\mathbb{R}$. Then, there exists a measurable random variable $X : \Omega \longrightarrow \mathbb{R}$, such that*

$$\mathcal{L}_\Omega(X) = \nu.$$

Proof.

It is well known that $(\Omega, \mathcal{A}, \mu)$ is isomorphic to $([0, 1], \mathcal{B}[0, 1], \lambda)$, where λ is the Lebesgue measure on $[0, 1]$ (see, Parthasarathy [75]). Using the pseudo-inverse of the distribution function of ν , we can find a random variable X such that $\mathcal{L}_\Omega(X) = \nu$. $\square$

Let ν be a probability on $\mathbb{R}$. For $B \in \mathcal{B}(\mathbb{R})$, ν_B denotes the probability on $\mathbb{R}$ defined by

$$\nu_B(A) = \nu(B)^{-1} \nu(A \cap B).$$

For $x \in \mathbb{R}$, ν_x denotes the probability on $\mathbb{R}$ defined by

$$\nu_x(B) = \nu(xB)$$

where $xB = \{xb / b \in B\}$.

Lemma 4.3.

(i) *For each probability ν on $\mathbb{R}$, for all Borelian sets B ,*

$$d(\nu_B, \nu) \leq \nu(\mathbb{R} \setminus B).$$

(ii) For all probabilities ν and η on $\mathbb{R}$, for all $x \geq 1$,

$$d(\nu_x, \eta_x) \leq d(\nu, \eta).$$

Proof.

(i) For all $A, B \in \mathcal{B}(\mathbb{R})$, we have

$$\begin{aligned} \nu_B(A) &= \nu(B)^{-1} \nu(A \cap B) \\ &= \nu(A \cap B) + \frac{\nu(\mathbb{R} \setminus B)}{\nu(B)} \nu(A \cap B) \\ &\leq \nu(A) + \nu(\mathbb{R} \setminus B) \end{aligned}$$

and

$$\nu(A) = \nu(A \cap B) + \nu(A \cap (\mathbb{R} \setminus B)) \leq \nu_B(A) + \nu(\mathbb{R} \setminus B).$$

By definition of the Lévy metric, this gives (i).

(ii) Let $x \geq 1$. For all $t \in \mathbb{R}$, for any $\varepsilon > d(\eta, \nu)$,

$$\begin{aligned} \eta_x((-\infty, t]) &= \eta((-\infty, tx]) \\ &\leq \nu((-\infty, tx + \varepsilon]) + \varepsilon \\ &\leq \nu_x((-\infty, t + \varepsilon]) + \varepsilon \end{aligned}$$

and

$$\begin{aligned} \eta_x((-\infty, t - \varepsilon]) - \varepsilon &= \eta((-\infty, tx - \varepsilon x]) - \varepsilon \\ &\leq \nu((-\infty, tx]) \\ &= \nu_x((-\infty, t]). \end{aligned}$$

So, $d(\nu_x, \eta_x) \leq d(\nu, \eta)$. □

Lemma 4.4. *Let $(a_n)_{n \in \mathbb{N}} \subset \mathbb{R}$ be an increasing sequence such that $a_n \nearrow \infty$ as $n \rightarrow \infty$. For each probability $\nu \in \mathcal{M}_0$ and $\varepsilon > 0$, there exist $C_0 \geq 1$ and $n_0 \in \mathbb{N}$, for all $C \geq C_0$ and $n \geq n_0$, there exists a probability η on $\mathbb{Z}$ with support $S \subset [-a_n C, a_n C]$ such that*

$$d(\eta_{a_n}, \nu) \leq \varepsilon$$

and

$$E(\eta) = 0.$$

Proof.

Let $\nu \in \mathcal{M}_0$ and $\varepsilon > 0$ be fixed and choose $\alpha > 0$ such that $4\alpha \leq \varepsilon$ and $\alpha < 1$. There exists $C_0 \geq 1$ such that

$$\int_{\mathbb{R} \setminus [-C_0, C_0]} |x| d\nu(x) \leq \alpha.$$

In particular, $\nu(\mathbb{R} \setminus [-C_0, C_0]) \leq \alpha$. Let $C \geq C_0$ and define $\tau = \nu_{[-C, C]}$. By Lemma 4.3,

$$d(\tau, \nu) \leq \nu(\mathbb{R} \setminus [-C, C]) \leq \alpha$$

and we have

$$|E(\tau)| = \nu([-C, C])^{-1} |E(\nu) - \int_{\mathbb{R} \setminus [-C, C]} x d\nu(x)| \leq \frac{\alpha}{1 - \alpha}.$$

Now, choose $n_0 \in \mathbb{N}$ (independent of C) such that $\frac{1}{a_{n_0}} < \alpha$ and fix $n \geq n_0$. Then we define the probability η' on $\mathbb{Z}$ by $\eta'(\{k\}) := \tau\left(\left[\frac{k}{a_n}, \frac{k+1}{a_n}\right)\right)$, $k \in \mathbb{Z}$.

We have, for all $t \in \mathbb{R}$,

$$\begin{aligned} \eta'_{a_n}((-\infty, t]) &= \eta'((-\infty, \lfloor ta_n \rfloor]) \\ &= \tau\left(\left(-\infty, \frac{\lfloor ta_n \rfloor + 1}{a_n}\right)\right) \\ &\leq \tau\left(\left(-\infty, t + \frac{1}{a_n}\right]\right) \end{aligned}$$

and

$$\tau((-\infty, t]) \leq \eta'_{a_n}((-\infty, t]).$$

Thus $d(\eta'_{a_n}, \tau) \leq \frac{1}{a_n} \leq \alpha$ and $d(\eta'_{a_n}, \nu) \leq 2\alpha$.

Further,

$$a_n E(\tau) - 1 \leq E(\eta') \leq a_n E(\tau).$$

Thus

$$|E(\eta')| \leq a_n |E(\tau)| + 1 \leq a_n \frac{\alpha}{1 - \alpha} + 1 \leq 2a_n \alpha.$$

Now we denote by η the probability on $\{-\lfloor a_n C \rfloor, \dots, \lfloor a_n C \rfloor\}$ defined by

$$\eta(\{i\}) = \begin{cases} \eta'(\{i\}) - \frac{E(\eta')}{2\lfloor a_n C \rfloor i} & \text{if } i \neq 0 \\ \eta'(0) & \text{if } i = 0 \end{cases}$$

Then $E(\eta) = 0$ and

$$d(\eta'_{a_n}, \eta_{a_n}) \leq d(\eta', \eta) \leq \left| \frac{E(\eta')}{2[a_n C]} \right| \leq 2\alpha.$$

Therefore $d(\eta_{a_n}, \nu) \leq \varepsilon$. $\square$

Proposition 4.5. *Let $(\Omega, \mathcal{A}, \mu, T)$ be an aperiodic ergodic dynamical system, $(a_n)_{n \in \mathbb{N}} \subset \mathbb{R}_+$ be an increasing sequence such that $a_n \nearrow \infty$ and $\frac{a_n}{n} \rightarrow 0$ as $n \rightarrow \infty$ and $\varepsilon > 0$. Let $A \in \mathcal{A}$ be a set such that $\mu(A) < 1$ and ν be a probability in $\mathcal{M}_0$. There exists $N \in \mathbb{N}$ such that for any $n \geq N$, there exists a set $B \in \mathcal{A}$ such that $\mu(B) \leq \varepsilon$, $A \cap B = \emptyset$ and*

$$d(\mathcal{L}_\Omega(\frac{1}{a_n} S_n(\mathbb{1}_B - \mu(B))), \nu) \leq \varepsilon$$

Proof.

Fix $\varepsilon > 0$ and $A \in \mathcal{A}$ such that $\mu(A) < 1$. Let α be a positive constant such that $\alpha \leq \frac{\varepsilon}{5}$ and $\mu(A) + 2\alpha < 1$.

By Lemma 4.4 applied to ν and α we get the constants $C_0 \geq 1$ and $n_0 \geq 1$ for which the conclusion of the lemma holds. Fix $C \geq C_0$ and set $\gamma := \frac{\alpha}{C+1}$. Let n_1 be an integer such that, for all $n \geq n_1$,

$$2 \frac{a_n C + 1}{n} \leq \alpha. \quad (4.1)$$

By Birkoff's ergodic theorem, there exist a set $E \in \mathcal{A}$ of measure greater than $1 - \frac{\gamma}{2}$ and an integer $n_2 \geq 1$ such that for all $n \geq n_2$, for all $x \in E$,

$$\left| \frac{1}{n} S_n(\mathbb{1}_A - \mu(A))(x) \right| \leq \alpha. \quad (4.2)$$

We denote by $\bar{n}$ the maximum of n_0 , n_1 and n_2 and we choose $N \in \mathbb{N}$ such that

$$\frac{\bar{n}}{N} \leq \alpha.$$

For any $n \geq N$, there exists a Rokhlin tower of height n with base $F \subset E$ and junk set of measure smaller than γ .

Indeed, let G be the base of a Rokhlin tower of height n and of measure greater than $1 - \frac{\gamma}{2}$. Because $\mu(\Omega \setminus E) \leq \frac{\gamma}{2}$, there exists an integer $i_0 \in \{0, \dots, n-1\}$ such that

$$\mu((T^{i_0} G) \cap E) \geq \frac{1}{n} (1 - \frac{\gamma}{2} - \frac{\gamma}{2}) = \frac{1 - \gamma}{n}.$$

If $F = T^{i_0} G \cap E$, then $F \subset E$ and the sets $F, TF, \dots, T^{n-1}F$ are disjoint. So, F is the base of a Rokhlin tower of height n with a junk set of measure smaller than γ .

From now on, n is fixed. By Lemma 4.4, there exists a centered probability η on $\mathbb{Z}$ with support $S \subset [-a_n C, a_n C]$ such that

$$d(\eta_{a_n}, \nu) \leq \alpha.$$

By Lemma 4.2, there exists a function $h : F \rightarrow \mathbb{Z}$ such that $\mathcal{L}_F(h) = \eta$. In particular, $E(h) = 0$. We set $d = \lfloor a_n C \rfloor + 1$ and

$$\begin{aligned} g : F &\rightarrow \mathbb{Z} \\ x &\mapsto h(x) + d. \end{aligned}$$

Note that $1 \leq g \leq 2d$ almost surely and $E(g) = d$. We now set

$$F_i := g^{-1}(\{i\}), \quad i = 1, \dots, 2d.$$

Note that the F_i 's depend on ν , α , C and n . Further $\{F_1, F_2, \dots, F_{2d}\}$ is a partition of the set F .

By (4.2), for each $x \in F$, the sub-orbit $\{x, Tx, \dots, T^{\bar{n}-1}x\}$ hits A at most $\bar{n}(\mu(A) + \alpha)$ times. Since by (4.1), $\frac{2d}{n} \leq \alpha$ and $\mu(A) + 2\alpha < 1$, we can find $2d$ points in this sub-orbit which are not in A . So, for each $i = 1, \dots, 2d$, we can find a measurable set $B_i \subset \bigcup_{j=0}^{\bar{n}-1} T^j F_i$, such that $B_i \cap A = \emptyset$ and for any $x \in F_i$, $S_{\bar{n}}(\mathbb{1}_{B_i})(x) = i$.

Finally, we set

$$B = \bigcup_{i=1}^{2d} B_i.$$

We have $B \in \mathcal{A}$ and $A \cap B = \emptyset$.

From the definition of B and (4.1),

$$\mu(B) = E(g_1)\mu(F) = d\mu(F) \leq \frac{a_n C + 1}{n} \leq \alpha \leq \varepsilon.$$

We define

$$\Omega_k = \bigcup_{i=0}^{n-\bar{n}-1} T^{-i} F_k, \quad k = 1, \dots, 2d$$

and

$$\bar{\Omega} = \bigcup_{k=1}^{2d} \Omega_k.$$

Since the $T^{-i} F_k$ are disjoint, using (4.1) and the fact that $\gamma \leq \alpha$, we have

$$\mu(\bar{\Omega}) = (n - \bar{n})\mu(F) \geq 1 - \gamma - \frac{\bar{n}}{n} \geq 1 - 3\alpha. \quad (4.3)$$

Further, by definition, $S_n(\mathbb{1}_B)(x) = k$ if and only if $x \in \Omega_k$. Therefore, for all $k = 1, \dots, 2d$,

$$\mu_{\overline{\Omega}}(S_n(\mathbb{1}_B) = k) = \mu_{\overline{\Omega}}(\Omega_k) = \mu_F(F_k).$$

Thus $\mathcal{L}_{\overline{\Omega}}(S_n(\mathbb{1}_B)) = \mathcal{L}_F(g)$ and by (4.3) and Lemma 4.3,

$$d(\mathcal{L}_{\Omega}(S_n(\mathbb{1}_B)), \mathcal{L}_F(g)) \leq 3\alpha.$$

Again, by Lemma 4.3,

$$d(\mathcal{L}_{\Omega}(\frac{1}{a_n}S_n(\mathbb{1}_B)), \mathcal{L}_F(\frac{g}{a_n})) \leq 3\alpha$$

and

$$d(\mathcal{L}_{\Omega}(\frac{1}{a_n}(S_n(\mathbb{1}_B) - d)), \mathcal{L}_F(\frac{g-d}{a_n})) \leq 3\alpha. \quad (4.4)$$

Now, remark that

$$d(\mathcal{L}_{\Omega}(\frac{1}{a_n}(S_n(\mathbb{1}_B) - d)), \mathcal{L}_{\Omega}(\frac{1}{a_n}S_n(\mathbb{1}_B - \mu(B)))) \leq \alpha. \quad (4.5)$$

Indeed, since $\mu(B) = d\mu(F)$, we have

$$-\gamma d \leq n\mu(B) - d \leq 0$$

and then

$$\left| \frac{1}{a_n}S_n(\mathbb{1}_B - \mu(B)) - \frac{1}{a_n}(S_n(\mathbb{1}_B) - d) \right| \leq \frac{\gamma d}{a_n} \leq \alpha.$$

To conclude, using (4.4), (4.5) and the fact that $d(\mathcal{L}_F(\frac{h}{a_n}), \nu) \leq \alpha$, we get

$$d(\mathcal{L}_{\Omega}(\frac{1}{a_n}S_n(\mathbb{1}_B - \mu(B))), \nu) \leq 5\alpha \leq \varepsilon.$$

□

Proposition 4.6. *Let $(\Omega, \mathcal{A}, \mu, T)$ be an aperiodic ergodic dynamical system, $(a_n)_{n \in \mathbb{N}} \subset \mathbb{R}_+$ be an increasing sequence such that $a_n \nearrow \infty$ and $\frac{a_n}{n} \rightarrow 0$ as $n \rightarrow \infty$ and $\varepsilon > 0$. For any set $A \in \mathcal{A}$ such that $\mu(A) < 1$, there exists a set $B \in \mathcal{A}$ such that*

$$(i) \quad \mu(A \triangle B) \leq \varepsilon,$$

$$(ii) \quad \text{there exists a sequence } (n_k)_{k \geq 1} \text{ such that for all } k \geq 1,$$

$$d(\mathcal{L}_{\Omega}(\frac{1}{a_{n_k}}S_{n_k}(\mathbb{1}_B - \mu(B))), \delta_0) \leq \varepsilon.$$

where δ_0 is the centered Dirac measure.

This proposition will be proved as a corollary of the following lemma.

Lemma 4.7. *Let $(\Omega, \mathcal{A}, \mu, T)$ be an aperiodic ergodic dynamical system, $(a_n)_{n \in \mathbb{N}} \subset \mathbb{R}_+$ be an increasing sequence such that $a_n \nearrow \infty$ and $\frac{a_n}{n} \rightarrow 0$ as $n \rightarrow \infty$ and $\varepsilon > 0$. For any set $A \in \mathcal{A}$ such that $\mu(A) < 1$, for any $N \in \mathbb{N}$, there exist $n \geq N$ and a set $C \in \mathcal{A}$ such that $\mu(A \Delta C) \leq \varepsilon$ and*

$$d(\mathcal{L}_\Omega(\frac{1}{a_n}S_n(\mathbb{1}_C - \mu(C))), \delta_0) \leq \varepsilon.$$

Proof.

Let $\varepsilon > 0$, $A \in \mathcal{A}$ such that $\mu(A) < 1$ and $N \in \mathbb{N}$ be fixed and let α be a positive constant such that $6\alpha \leq \varepsilon$ and $\mu(A) + 2\alpha < 1$.

By Birkoff's ergodic theorem there exist a set $G \in \mathcal{A}$ with $\mu(G) > 1 - \alpha$ and an integer M such that for all $k \geq M$ and for all $x \in G$,

$$\left| \frac{1}{k} S_k(\mathbb{1}_A - \mu(A)) \right| \leq \alpha. \quad (4.6)$$

There exists an integer $n \geq N$, such that

$$\frac{M}{a_n} \leq \alpha. \quad (4.7)$$

We choose an integer $p \geq \frac{1 - 2\alpha}{M\alpha}$ and we set $N' = Mnp$.

Let F be the base of a Rokhlin tower of height N' with a junk set of measure smaller than α . For $k = 0, \dots, M - 1$, we define the sets

$$F_k = \bigcup_{i=0}^{np-1} T^{iM+k} F.$$

Since $\mu(G) > 1 - \alpha$, there exists $k_0 \in \{0, \dots, M - 1\}$ such that

$$\mu(F_{k_0} \setminus G) \leq \frac{2\alpha}{M}.$$

Further, $H = T^{k_0} F$ is the base of a Rokhlin tower of height N' with a junk set J such that $\mu(J) \leq 2\alpha$.

For $x \in \Omega$, we denote by $s_l(x) = \{x, Tx, \dots, T^{l-1}x\}$ the sub-orbit of length l which begins at x . The set A will be modified in a way that for each sub-orbit $s_M(x)$, $x \in F_{k_0}$, the average of visits of the set along the sub-orbit be close to the measure $\mu(A)$.

For $x \in H$, we define the following transformation:

• If $x \in G$, by (4.6), along the sub-orbit $s_M(x)$, the number of visits of the set A belongs to $[M(\mu(A) - \alpha), M(\mu(A) + \alpha)]$. Further, $M(\mu(A) + \alpha) \leq (1 - \alpha)M$. So by adding or removing at most αM points to A along the sub-orbit $s_M(x)$, we can transform the set A to get a set $A_0(x)$ (which differs from A only on $s_M(x)$) such that along the sub-orbit $s_M(x)$, the number of visits of the set $A_0(x)$ belongs to $[M\mu(A) - 1, M\mu(A) + 1]$.

• If $x \notin G$, we can also transform A in order to have that the number of visits of the set $A_0(x)$ along the segment $s_M(x)$ belongs to $[M\mu(A) - 1, M\mu(A) + 1]$. Here, to do that we possibly need to add or remove M points to A along $s_M(x)$.

Doing this transformation for all points of H , we can get a set $A_0 \in \mathcal{A}$ (taking into account all the changes done to A along $\{H, TH, \dots, T^{M-1}H\}$) having the property that for all $x \in H$,

$$|S_M(\mathbb{1}_{A_0} - \mu(A))(x)| \leq 1 \quad (4.8)$$

(the problem of measurability for A_0 is not discussed, but we can see that the transformations can be done in a measurable way).

Now we do the same transformations for all points of $T^M H, T^{2M} H, \dots$, and $T^{(pn-1)M} H$. Finally, we can get a measurable set B deduced from A and having the property (4.8) on each $T^{kM} H$. Further, doing these transformations level per level, we can have B in such a way that for all $x \in H$ and for all $k \in \{1, \dots, pn\}$,

$$|S_{kM}(\mathbb{1}_B - \mu(A))(x)| \leq 1.$$

Note that we did not change the set A on the junk set J , so $A \cap J = B \cap J$. Recall that we did at most αM changes for points of $F_{k_0} \cap G$ and at most M for points of $F_{k_0} \setminus G$, so

$$\begin{aligned} \mu(A \Delta B) &\leq \alpha M \mu(F_{k_0} \cap G) + M \mu(F_{k_0} \setminus G) \\ &\leq 3\alpha. \end{aligned} \quad (4.9)$$

We also have

$$\begin{aligned} |\mu(B) - \mu(A)| &\leq \left| \int_H S_{N'}(\mathbb{1}_B)(x) d\mu(x) + \mu(B \cap J) - \mu(A) N' \mu(H) - \mu(A) \mu(J) \right| \\ &\leq \int_H |S_{N'}(\mathbb{1}_B - \mu(A))(x)| d\mu(x) + |\mu(B \cap J) - \mu(A) \mu(J)| \\ &\leq \mu(H) + |\mu(B \cap J) - \mu(A) \mu(J)|. \end{aligned}$$

Then, by changing the set B on only one level of the tower and on the junk set, we can obtain a new set $C \in \mathcal{A}$ such that $\mu(A) = \mu(C)$ and $\mu(B \Delta C) \leq \mu(H) + \mu(J) \leq 3\alpha$. Thus, by (4.9), $\mu(A \Delta C) \leq 6\alpha \leq \varepsilon$ and we have the following property: for all $x \in H$, for all $k, k' \in \{1, \dots, pn\}$,

$$|S_{kM}(\mathbb{1}_C - \mu(C))(x) - S_{k'M}(\mathbb{1}_C - \mu(C))(x)| \leq 3. \quad (4.10)$$

Let $\tilde{\Omega} = \bigcup_{i=0}^{N'-n} T^i H$. We have

$$\mu(\Omega \setminus \tilde{\Omega}) \leq 3\alpha$$

and, by (4.10) and (4.7), for all $x \in \tilde{\Omega}$,

$$\frac{1}{a_n} |S_n(\mathbb{1}_C - \mu(C))(x)| \leq \frac{3 + 2M}{a_n} \leq 3\alpha$$

From theses two inequalities, we deduce that

$$d(\mathcal{L}_\Omega(\frac{1}{a_n} S_n(\mathbb{1}_C - \mu(C))), \delta_0) \leq 6\alpha \leq \varepsilon.$$

□

Proof of Proposition 4.6.

Let $\varepsilon > 0$ and $A \in \mathcal{A}$ such that $\mu(A) < 1$ be fixed and choose $\varepsilon_1 \leq \frac{\varepsilon}{2}$.

By Lemma 4.7, there exist $n_1 \in \mathbb{N}$ and a set $C_1 \in \mathcal{A}$ such that $\mu(A \Delta C_1) \leq \varepsilon_1$ and

$$d(\mathcal{L}_\Omega(\frac{1}{a_{n_1}} S_{n_1}(\mathbb{1}_{C_1} - \mu(C_1))), \delta_0) \leq \varepsilon_1.$$

Let $\varepsilon_2 \leq \frac{\varepsilon_1}{2n_1}$. By Lemma 4.7, there exist an integer $n_2 \geq n_1$ and a set $C_2 \in \mathcal{A}$ such that $\mu(C_1 \Delta C_2) \leq \varepsilon_2$ and

$$d(\mathcal{L}_\Omega(\frac{1}{a_{n_2}} S_{n_2}(\mathbb{1}_{C_2} - \mu(C_2))), \delta_0) \leq \varepsilon_2.$$

We will procede by induction. After step $k-1$, we choose $\varepsilon_k \leq \frac{\varepsilon_{k-1}}{2n_{k-1}}$. By application of Lemma 4.7, there exist an integer $n_k \geq n_{k-1}$ and a set $C_k \in \mathcal{A}$ such that $\mu(C_{k-1} \Delta C_k) \leq \varepsilon_k$ and

$$d(\mathcal{L}_\Omega(\frac{1}{a_{n_k}} S_{n_k}(\mathbb{1}_{C_k} - \mu(C_k))), \delta_0) \leq \varepsilon_k.$$

Finally, we set $B = \bigcup_{n \geq 1} \bigcap_{k \geq n} C_k$. We have

$$\mu(A \Delta B) \leq \sum_{k \geq 1} \varepsilon_k \leq \varepsilon$$

and for all $k \geq 1$,

$$\begin{aligned}
& d(\mathcal{L}_\Omega(\frac{1}{a_{n_k}} S_{n_k}(\mathbb{1}_B - \mu(B))), \delta_0) \\
& \leq d(\mathcal{L}_\Omega(\frac{1}{a_{n_k}} S_{n_k}(\mathbb{1}_{C_k} - \mu(C_k))), \delta_0) + d(\mathcal{L}_\Omega(\frac{1}{a_{n_k}} S_{n_k}(\mathbb{1}_{B \Delta C_k} - \mu(B \Delta C_k))), \delta_0) \\
& \leq \varepsilon_k + n_k \mu(B \Delta C_k) \\
& \leq \varepsilon_k + n_k \sum_{i \geq k+1} \mu(C_{i-1} \Delta C_i) \\
& \leq \varepsilon_k + n_k \sum_{i \geq k+1} \varepsilon_i \\
& \leq \varepsilon.
\end{aligned}$$

□

4.4 Proof of Theorem 4.1

Let $(\Omega, \mathcal{A}, \mu, T)$ be an aperiodic ergodic dynamical system. Let $(a_n)_{n \in \mathbb{N}} \subset \mathbb{R}_+$ be an increasing sequence such that $a_n \nearrow \infty$ and $\frac{a_n}{n} \rightarrow 0$ as $n \rightarrow \infty$.

Let $(\varepsilon_k)_{k \geq 1}$ be a decreasing sequence of positive reals such that ε_k goes to 0 as k goes to ∞ .

For each $\nu \in \mathcal{M}_0$ and for each $k \geq 1$, we define

$$H_k^\nu = \{A \in \mathcal{A} / \exists n \geq k \text{ such that } d(\mathcal{L}_\Omega(\frac{1}{a_n} S_n(\mathbb{1}_A - \mu(A))), \nu) < \varepsilon_k\}.$$

For each $\nu \in \mathcal{M}_0$ and for each $k \geq 1$, it is clear that H_k^ν is an open set in $\mathcal{A}$. We now prove that it is dense.

Assume that ν and k are fixed and let $\varepsilon > 0$ and $A \in \mathcal{A}$. By Proposition 4.6, there exists a set $B \in \mathcal{A}$ such that $\mu(A \Delta B) < \frac{\varepsilon}{2}$ and there exists a sequence $(n_i)_{i \geq 1}$ such that for all $i \geq 1$,

$$d(\mathcal{L}_\Omega(\frac{1}{a_{n_i}} S_{n_i}(\mathbb{1}_B - \mu(B))), \delta_0) \leq \frac{\varepsilon_k}{2}. \quad (4.11)$$

By Proposition 4.5, there exists an integer i_0 such that, for the integer $n = n_{i_0} \geq k$, there exists a set $C \in \mathcal{A}$ satisfying $\mu(C) < \frac{\varepsilon}{2}$, $C \cap B = \emptyset$ and

$$d(\mathcal{L}_\Omega(\frac{1}{a_n} S_n(\mathbb{1}_C - \mu(C))), \nu) < \frac{\varepsilon_k}{2}. \quad (4.12)$$

Hence, $\mu((B \cup C) \triangle A) < \varepsilon$ and by (4.11) and (4.12),

$$\begin{aligned} d(\mathcal{L}_\Omega(\frac{1}{a_n} S_n(\mathbb{1}_{B \cup C} - \mu(B \cup C))), \nu) \\ \leq d(\mathcal{L}_\Omega(\frac{1}{a_n} S_n(\mathbb{1}_C - \mu(C))), \nu) + d(\mathcal{L}_\Omega(\frac{1}{a_n} S_n(\mathbb{1}_B - \mu(B))), \delta_0) \\ < \varepsilon_k, \end{aligned}$$

i.e. $B \cup C$ belongs to H_k^ν . Therefore H_k^ν is dense in $\mathcal{A}$ for the pseudo-metric Θ .

Let M be a countable subset of $\mathcal{M}_0$ which is dense in $\mathcal{M}$ and set

$$H = \bigcap_{\nu \in M} \bigcap_{k=1}^{\infty} H_k^\nu.$$

By Baire's theorem, H is a dense G_δ (for the metric of the measure of the symmetric difference).

Further, for each $A \in H$, the sequence $(\mathcal{L}_\Omega(\frac{1}{a_n} S_n(\mathbb{1}_A - \mu(A))))_{n \geq 1}$ is dense in $\mathcal{M}$ for the Lévy metric d .

Indeed, let $A \in H$, $\eta \in \mathcal{M}$ and $\varepsilon > 0$. By density of M , there exist $\nu \in M$ such that

$$d(\nu, \eta) < \frac{\varepsilon}{2}.$$

But $A \in H_k^\nu$ for all $k \geq 1$, then there exists an increasing sequence $(n_k)_{k \geq 1}$ such that

$$d(\mathcal{L}_\Omega(\frac{1}{a_{n_k}} S_{n_k}(\mathbb{1}_A - \mu(A))), \nu) \leq \varepsilon_k.$$

Thus, there exists $K \in \mathbb{N}$ such that

$$d(\mathcal{L}_\Omega(\frac{1}{a_{n_K}} S_{n_K}(\mathbb{1}_A - \mu(A))), \nu) \leq \frac{\varepsilon}{2}$$

and

$$d(\mathcal{L}_\Omega(\frac{1}{a_{n_K}} S_{n_K}(\mathbb{1}_A - \mu(A))), \eta) \leq \varepsilon.$$

□

4.5 A remark on the non-ergodic case

Assume that $(\Omega, \mathcal{A}, \mu, T)$ is an aperiodic dynamical system which is not ergodic. Denote by $\mathcal{I}$ the σ -algebra of the invariant sets and $(\mu^x)_{x \in \chi}$ the ergodic components of the measure

μ . If there exist a set $A \in \mathcal{A}$, a probability measure ν on $\mathbb{R}$ and a sequence $(n_k)_{k \in \mathbb{N}}$ such that

$$\frac{1}{a_{n_k}} S_{n_k}(\mathbb{1}_A - \mu(A)) \xrightarrow[k \rightarrow \infty]{\mathcal{D}} \nu,$$

then $E(\mathbb{1}_A | \mathcal{I}) = \mu(A)$ almost surely.

Indeed, if there exists $x \in \chi$ such that $\mu^x(A) - \mu(A) = c > 0$, then by Birkoff's ergodic theorem,

$$\frac{1}{n} S_n(\mathbb{1}_A - \mu(A)) \xrightarrow[n \rightarrow \infty]{} c$$

μ^x -almost surely. Therefore

$$\frac{1}{a_{n_k}} S_{n_k}(\mathbb{1}_A - \mu(A)) \xrightarrow[k \rightarrow \infty]{} +\infty$$

and we have a contradiction.

So, to find a set which satisfies the conclusion of Theorem 4.1, we have to consider the sets A such that $E(\mathbb{1}_A | \mathcal{I})$ is almost surely constant. The class of such sets is not, in general, dense in $\mathcal{A}$. So, in the non-ergodic case, we cannot expect the result of genericity.

Nevertheless, in the forthcoming section, we prove the existence of such sets by an explicit construction.

4.6 A constructive proof

We fix, for all the section, an aperiodic dynamical system $(\Omega, \mathcal{A}, \mu, T)$ and an increasing sequence $(a_n)_{n \in \mathbb{N}} \subset \mathbb{R}_+$ such that $a_n \nearrow \infty$ and $\frac{a_n}{n} \rightarrow 0$ as $n \rightarrow \infty$.

In this section, we show that one can prove the existence of an arbitrarily small set $A \in \mathcal{A}$ such that the sequence of the distributions of $\left(\frac{1}{a_n} S_n(\mathbb{1}_A - \mu(A)) \right)_{n \in \mathbb{N}}$ is dense in $\mathcal{M}$ without the help of the ergodic theorem. In the following proposition, ergodicity is not needed.

Proposition 4.8. *For all $\varepsilon > 0$, there exists a set $A \in \mathcal{A}$, with $\mu(A) < \varepsilon$ such that for all probability measure ν on $\mathbb{R}$, there exists a subsequence $(n_k)_{k \in \mathbb{N}}$ such that*

$$\frac{1}{a_{n_k}} S_{n_k}(\mathbb{1}_A - \mu(A)) \xrightarrow[k \rightarrow \infty]{\mathcal{D}} \nu.$$

Using the separability of the set $\mathcal{M}_0$ which is dense in $\mathcal{M}$, we can prove that the next proposition is equivalent to Proposition 4.8.

Proposition 4.9. *For all $\varepsilon > 0$ and for all sequences $(\nu_k)_{k \in \mathbb{N}}$ in $\mathcal{M}_0$, there exist a set $A \in \mathcal{A}$, with $\mu(A) < \varepsilon$, and a subsequence $(n_k)_{k \in \mathbb{N}}$ such that*

$$d(\mathcal{L}_\Omega(\frac{1}{a_{n_k}} S_{n_k}(\mathbb{1}_A - \mu(A))), \nu_k) \xrightarrow{k \rightarrow \infty} 0.$$

Proposition 4.9 implies Proposition 4.8. Indeed, if M is a countable and dense subset of $\mathcal{M}_0$, we can find a sequence $(\nu_k)_{k \in \mathbb{N}}$ such that for all $\eta \in M$, there exists an infinite set $K_\eta \in \mathbb{N}$ verifying that for all $k \in K_\eta$, $\nu_k = \eta$. Let A be the set corresponding to the sequence $(\nu_k)_{k \in \mathbb{N}}$. Let $\tau \in \mathcal{M}_0$, by density of M , there exist a probability $\eta \in M$ such that $d(\tau, \eta)$ is small enough.

There also exists a subsequence $(k_j)_{j \in \mathbb{N}}$ such that $\nu_{k_j} = \eta$ for all $j \in \mathbb{N}$ and by Proposition 4.9, there exists a sequence $(n_{k_j})_{j \in \mathbb{N}}$ such that

$$d(\mathcal{L}_\Omega(\frac{1}{a_{n_{k_j}}} S_{n_{k_j}}(\mathbb{1}_A - \mu(A))), \eta) \xrightarrow{j \rightarrow \infty} 0.$$

This prove one implication and the converse is clear.

Now to prove Proposition 4.9, we will construct explicitly the set A . To do that, we will use the two following lemmas.

Lemma 4.10. *For all probability ν on $\mathbb{R}$, for all $\varepsilon > 0$, there exists $C_0 \geq 1$ and $n_0 \in \mathbb{N}$, for all $C \geq C_0$ and $n \geq n_0$, there exists a probability η on $\mathbb{Z}$ with support $S \subset [-a_n C, a_n C]$ such that for all $i \in \mathbb{Z}$, $\eta(\{i\}) \in \mathbb{Q}$,*

$$d(\eta_{a_n}, \nu) \leq \varepsilon$$

and

$$E(\eta) = 0.$$

Proof.

This Lemma is a consequence of Lemma 4.4 (which has a constructive proof) and of the fact that for all probability measure on $\mathbb{Z}$ with finite support, we can find a probability on $\mathbb{Z}$ with same support which is arbitrarily close to the first one and takes values in $\mathbb{Q}$. $\square$

Lemma 4.11. *For all $n \geq 1$ and for all $\varepsilon > 0$, there exists a set $F \in \mathcal{A}$ such that $\{F, \dots, T^{n-1}F\}$ is a Rokhlin tower of measure greater than $1 - \varepsilon$ and the stay time in the junk set $J = \Omega \setminus (\cup_{i=0}^{n-1} T^i F)$ is almost surely 1, i.e. for a.e. $x \in J$, $Tx \in F$.*

Proof.

This can be view as a consequence of the Alpern's theorem [1], by constructing a castle with two towers of height n and $n + 1$ and the base of the second tower of measure less than ε . $\square$

Proof of Proposition 4.9

Let $(\nu_k)_{k \geq 1}$ be a fixed sequence in $\mathcal{M}_0$ and $\varepsilon > 0$. Let $(\varepsilon_k)_{k \geq 1}$ be a decreasing sequence of nonnegative real such that $\sum_{k \geq 1} \varepsilon_k < \varepsilon$ and $\sum_{k \geq 1} k\varepsilon_k < \infty$.

We will give a sequence of pairwise disjoint sets $A_k \in \mathcal{A}$ with $\mu(A_k)$ decreasing sufficiently fast and a sequence $(n_k)_{k \geq 1}$ such that for all $j = 1, \dots, k$,

$$d(\mathcal{L}_\Omega(\frac{1}{a_{n_j}} S_{n_j}(\mathbb{1}_{\cup_{i=1}^k A_i} - \mu(\cup_{i=1}^k A_i))), \nu_j) \leq \sum_{i=j}^k \varepsilon_i. \quad (4.13)$$

Thus, if $A = \cup_{j=1}^\infty A_j$, then $A \in \mathcal{A}$ and satisfies the conclusion of Proposition 4.9.

Some definitions

By induction, we will define some sequences that help us to find a good set A . We first consider the probability ν_1 and the constant ε_1 . We set $\alpha_1 = \frac{\varepsilon_1}{8}$. Applying Lemma 4.10 to ν_1 and $\frac{\varepsilon_1}{2}$, we get two constants $C(\nu_1, \varepsilon_1)$ and $n(\nu_1, \varepsilon_1)$ and we choose $C_1 = C(\nu_1, \varepsilon_1)$ and $n_1 \geq n(\nu_1, \varepsilon_1)$ such that

$$\frac{a_{n_1} C_1 + 1}{n_1} \leq \alpha_1.$$

Let $d_1 = \lfloor a_{n_1} C_1 \rfloor + 1$. We get a corresponding probability η_1 on $\{-d_1 + 1, \dots, d_1 - 1\}$ such that $d(\eta_1, \nu_1) \leq \alpha_1$. Since for all i , $\eta_1(\{i\}) \in \mathbb{Q}$, there exist $q_1 \in \mathbb{N}$ and $q_1^{(i)} \in \mathbb{N}$, $i = 1 \dots, 2d_1$, such that

$$\eta_1(\{i - d_1\}) = \frac{q_1^{(i)}}{q_1}, \text{ for all } i = 1, \dots, 2d_1.$$

We set $N_1 = q_1 n_1$.

By induction, we define the sequences $(\alpha_k)_{k \geq 2}$, $(C_k)_{k \geq 2}$, $(n_k)_{k \geq 2}$, $(q_k)_{k \geq 2}$ as follows. We consider the probability ν_k and ε_k . Applying Lemma 4.10 to ν_k and $\frac{\varepsilon_k}{2}$, we get two constants $C(\nu_k, \varepsilon_k)$ and $n(\nu_k, \varepsilon_k)$. We define $\alpha_k = \frac{a_{n_{k-1}}}{2n_{k-1}} \varepsilon_k$. We choose $C_k = \max\{C(\nu_k, \alpha_k), C_{k-1}\}$ and $n_k \geq n(\nu_k, \alpha_k)$ a multiple of N_{k-1} such that

$$\frac{a_{n_k} C_k + 1}{n_k} \leq \alpha_k \quad (4.14)$$

and

$$\frac{N_{k-1}}{a_{n_k}} \leq \alpha_k \quad (4.15)$$

We set $d_k = \lfloor a_{n_k} C_k \rfloor + 1$. We get a corresponding probability η_k on $\{-d_k + 1, \dots, d_k - 1\}$ such that $d(\eta_k, \nu_k) \leq \alpha_k$. There exist $q_k \in \mathbb{N}$ and $q_k^{(i)} \in \mathbb{N}$, $i = 1 \dots, 2d_k$, such that

$$\eta_k(\{i - d_k\}) = \frac{q_k^{(i)}}{q_k}. \text{ Set } N_k = q_k n_k.$$

For $k \geq 1$, we also set $p_k = \frac{n_{k+1}}{n_k}$ and we consider a sequence $(\beta_k)_{k \geq 1}$ such that for all $k \geq 1$,

$$\sum_{j \geq k} \beta_j \leq \alpha_k. \quad (4.16)$$

We define the sequence $(\gamma_k)_{k \geq 1}$ such that

$$\gamma_k \leq \frac{\beta_{k+1}}{2p_{k+1}}. \quad (4.17)$$

Note, we will also use the fact that in particular,

$$\gamma_k \leq \frac{an_{k+1}}{n_{k+1}} \alpha_{k+1}. \quad (4.18)$$

and even that $\gamma_k \leq \frac{a_{n_k}}{n_k}$. Further, for all $k \geq 1$, we apply Lemma 4.11. We get a set $F_k \in \mathcal{A}$ such that $\{F_k, TF_k, \dots, T^{n_{k+1}-1}F_k\}$ is a Rokhlin tower of height n_{k+1} and the junk set $J_k = \Omega \setminus \bigcup_{i=0}^{n_{k+1}-1} T^i F_k$ is a set with stay time 1 and $\mu(J_k) \leq \gamma_k$.

Step 1: the set A_1

As we saw in Proposition 4.5, we can find a set A_1 and an integer n_1 such that

$$d(\mathcal{L}_\Omega(\frac{1}{a_{n_1}} S_{n_1}(\mathbb{1}_{A_1} - \mu(A_1))), \nu_1) \leq \varepsilon_1.$$

But, we want that the set A_1 becomes negligible for larger partial sums. In this way we will be able to find a set A_2 to approach ν_2 with a partial sum of length n_2 and the set A_1 will not interfere. We proceed as follows.

The set $A_{1,1}$

Consider the set $F_1 \in \mathcal{A}$ which is defined before. Then $\{F_1, TF_1, \dots, T^{n_1-1}F_1\}$ is a Rokhlin tower given by Lemma 4.11 of measure greater than $1 - \gamma_1$. Write $F_1^i = T^{in_1}F_1$, $i = 0, \dots, p_1 - 1$. We thus have p_1 towers $\{F_1^i, \dots, T^{in_1-1}F_1^i\}$ of height n_1 .

By Lemma 4.2, let h_1 be a measurable function from F_1 to $\mathbb{Z}$ such that $\mathcal{L}_G(h_1) = \eta_1$. We denote by g_1 the positive function equal to $h_1 + d_1$. Let

$$A_{F_1, i} = A_{F_1, i}^{(g_1)} := g_1^{-1}(\{i\}), \quad i = 1, \dots, 2d_1.$$

Now, for all $i = 1, \dots, 2d_1$, let $\{A_{F_1, i, 1}, \dots, A_{F_1, i, q_1^{(i)}}\}$ be a partition of the set $A_{F_1, i}$ into sets of measure $\frac{1}{q_1}$. We thus have a partition of F_1 into

$$\{A_{F_1, 1, 1}, \dots, A_{F_1, 1, q_1^{(1)}}, A_{F_1, 2, 1}, \dots, A_{F_1, 2, q_1^{(2)}}, \dots, A_{F_1, 2d_1, 1}, \dots, A_{F_1, 2d_1, q_1^{(2d_1)}}\}.$$

By induction, we deduce partitions of F_1^l for $l = 1, \dots, p_1 - 1$ setting

$$A_{F_1^l, i, j} = \begin{cases} T^{n_1} A_{F_1^{l-1}, i, j+1} & \text{if } 1 \leq j \leq q_1^{(i-1)} \\ T^{n_1} A_{F_1^{l-1}, i+1, 1} & \text{if } j = q_1^{(i)} \text{ and } i < 2d_1 \\ T^{n_1} A_{F_1^{l-1}, 1, 1} & \text{if } j = q_1^{(2d_1)} \text{ and } i = 2d_1 \end{cases}.$$

For all $l = 0, \dots, p_1 - 1$ and for all $i = 1, \dots, 2d_1$, we set

$$A_{F_1^l, i} = \bigcup_{j=1}^{q_1^{(i)}} A_{F_1^l, i, j}$$

and

$$A_{F_1^l} = \bigcup_{i=1}^{2d_1} \{T^k A_{F_1^l, i} / k = 0, \dots, i-1\}.$$

Now, we can define a "first version" of the set A_1 , called $A_{1,1}$, as follows,

$$A_{1,1} = \bigcup_{l=0}^{p_1-1} A_{F_1^l}.$$

Lemma 4.12.

(i) $\mu(A_{1,1}) \leq \alpha_1;$

(ii)

$$A_{1,1} \subset \bigcup_{j=0}^{p_1-1} \{T^{i+jn_1} F_1 / i = 0, \dots, 2d_1 - 1\};$$

(iii)

$$d(\mathcal{L}_\Omega(\frac{1}{a_{n_1}} S_{n_1}(\mathbb{1}_{A_{1,1}} - \mu(A_{1,1}))), \nu_1) \leq \varepsilon_1;$$

(iv)

$$d(\mathcal{L}_\Omega(\frac{1}{a_{n_2}} S_{n_2}(\mathbb{1}_{A_{1,1}} - \mu(A_{1,1}))), \delta_0) \leq \alpha_2,$$

where δ_0 is the zero-mean Dirac measure.

Proof.

For each $l = 0, \dots, p_1 - 1$, $\mu(A_{F_1^l}) = E(g_1)\mu(F_1) = d_1\mu(F_1)$. Therefore, by (4.14),

$$\mu(A_{1,1}) = \sum_{l=0}^{p_1-1} \mu(A_{F_1^l}) = p_1 d_1 \mu(F_1) \leq \frac{p_1}{n_2} d_1 = \frac{d_1}{n_1} \leq \alpha_1.$$

(ii) is clear.

Let $\Omega_1 = \bigcup_{l=0}^{p_1-1} \bigcup_{i=0}^{n_1-2d_1-1} T^{-i} F_1^l$. Then, we have

$$\mathcal{L}_{\Omega_1}\left(\frac{1}{a_{n_1}} S_{n_1}(\mathbb{1}_{A_{1,1}})\right) = \mathcal{L}_{F_1}(g_1)$$

and since $\gamma_1 \leq \alpha_1$,

$$\mu(\Omega_1) = p_1(n_1 - 2d_1)\mu(F_1) \geq (n_2 - 2p_1d_1)\frac{(1 - \gamma_1)}{n_2} = 1 - \gamma_1 - \frac{2d_1}{n_1} \geq 1 - 3\alpha_1.$$

Thus

$$d(\mathcal{L}_{\Omega}\left(\frac{1}{a_{n_1}} S_{n_1}(\mathbb{1}_{A_{1,1}})\right), \mathcal{L}_{F_1}(g_1)) \leq 3\alpha_1.$$

We can conclude as in the proof of Proposition 4.5 that

$$d(\mathcal{L}_{\Omega}\left(\frac{1}{a_{n_1}} S_{n_1}(\mathbb{1}_{A_{1,1}} - \mu(A_{1,1}))\right), \mathcal{L}_{F_1}(g_1)) \leq \varepsilon_1$$

and (iii) follows.

For (iv), by definition of $A_{1,1}$, we have

$$(p_1 - q_1)d_1 \leq S_{n_2}(\mathbb{1}_{A_{1,1}}) \leq (p_1 + q_1)d_1$$

and

$$|S_{n_2}(\mathbb{1}_{A_{1,1}} - \mu(A_{1,1}))| \leq p_1d_1|1 - n_2\mu(F_1)| + q_1d_1 \leq p_1d_1\gamma_1 + q_1d_1.$$

Thus, using $\gamma_1 \leq \frac{a_{n_2}}{n_2}\alpha_2$, we have

$$\frac{1}{a_{n_2}}|S_{n_2}(\mathbb{1}_{A_{1,1}} - \mu(A_{1,1}))| \leq \frac{d_1}{n_1}\alpha_2 + \frac{d_1}{n_1}\frac{N_1}{a_{n_2}} \leq 2\alpha_1\alpha_2 \leq \alpha_2.$$

□

Unfortunately, the set $A_{1,1}$ is not enough well defined to be negligible for higher partial sums. So, we will need to change a small part of $A_{1,1}$ at each step. The set $A_{1,1}$ can be considered as a first version of the set A_1 . In the sequel we procede by induction.

The sets $A_{1,k}$, $k \geq 2$

We give here the general algorithm to deduce the set $A_{1,k}$ from $A_{1,k-1}$.

Consider $F_k \in \mathcal{A}$. Then $\{F_k, TF_k, \dots, T^{m_{k+1}-1}F_k\}$ is a Rokhlin tower given by Lemma 4.11 of measure greater than $1 - \gamma_k$.

We are going to adapt the set $A_{1,k-1}$ in order to make it negligible for partial sum of size n_{k+1} . To do that, we will change this set on a part of measure smaller than β_k to get $A_{1,k}$.

For all x in F_k , we define the first time in F_{k-1} by $n(x) = \inf\{n \geq 0 / T^n x \in F_{k-1}\}$ and

$$F'_k = \bigcup_{x \in F_k} T^{n(x)} x = \left(\bigcup_{i=0}^{n_k} T^i F_k \right) \cap F_{k-1}.$$

For $x \in F'_k$ and $i = 0, \dots, p_k - 1$, let

$$\rho_i(x) = \sum_{j=in_k}^{(i+1)n_k-1} \mathbb{1}_{A_{1,k-1}} \circ T^j(x).$$

For all $x \in F'_k$, $\rho_0(x) = S_{n_k}(\mathbb{1}_{A_{1,k-1}})(x) = \frac{n_k}{n_1} d_1$ but for $i > 0$ it can be different. The differences appear when the orbit of the point x meets the junk set J_{k-1} . But by definition of the Rokhlin tower (see Lemma 4.11), it can meet J_{k-1} only one time in every n_k consecutive iterates of T .

We have for $i > 0$, $(p_1 - q_1)d_1 \leq \rho_i(x) \leq (p_1 + q_1)d_1$.

For $\rho_i(x) = p_1 d_1 - j$, $j > 0$, let $B_i(x)$ be a set composed by j elements of the set $\{T^{in_k} x, \dots, T^{(i+1)n_k-1} x\} \setminus A_{1,k-1}$, in such a way that every n_1 -consecutive points in $\{T^{in_k} x, \dots, T^{(i+1)n_k-1} x\}$ meet $A_{1,k-1} \cup B_i(x)$ at most $2d_1$ times.

If $\rho_i(x) = p_1 d_1 + j$, $j > 0$, let $C_i(x)$ be the set composed by the j first elements of $\{T^{in_k} x, \dots\} \cap A_{1,k-1}$.

Let

$$B = \bigcup_{x \in F'_k} \bigcup_{i=0}^{p_k-1} B_i(x) \quad , \quad C = \bigcup_{x \in F'_k} \bigcup_{i=0}^{p_k-1} C_i(x)$$

and

$$A_{1,k} = (A_{1,k-1} \setminus C) \cup B.$$

Lemma 4.13.

(i)

$$\mu(A_{1,k-1} \triangle A_{1,k}) \leq \beta_k;$$

(ii)

$$d(\mathcal{L}_\Omega(\frac{1}{a_{n_{k+1}}} S_{n_{k+1}}(\mathbb{1}_{A_{1,k}} - \mu(A_{1,k})), \delta_0) \leq \alpha_{k+1}.$$

Proof.

First, since the orbit of a point x can only meet J_{k-1} one time every n_k and using (4.17), we have

$$\mu(A_{1,k-1} \triangle A_{1,k}) \leq 2p_k \mu(J_{k-1}) \leq 2p_k \gamma_{k-1} \leq \beta_k.$$

On the other hand, for all $x \in F'_k$ and $i = 0, \dots, p_k - 1$, we have

$$\sum_{j=in_k}^{(i+1)n_k-1} \mathbb{1}_{A_{1,k}} \circ T^j(x) = p_{k-1} d_{k-1}$$

and thus

$$S_{n_{k+1}}(\mathbb{1}_{A_{1,k}})(x) = p_k p_{k-1} d_{k-1} = \frac{n_{k+1}}{n_{k-1}} d_{k-1}.$$

Further $\mu(A_{1,k}) = \frac{n_{k+1}}{n_{k-1}} d_{k-1} \mu(F_k)$ and

$$(p_k - 1) p_{k-1} d_{k-1} \leq S_{n_{k+1}}(\mathbb{1}_{A_{1,k}}) \leq (p_k + 1) p_{k-1} d_{k-1}.$$

Then, by (4.18) and (4.15),

$$\frac{1}{a_{n_{k+1}}} |S_{n_{k+1}}(\mathbb{1}_{A_{1,k}} - \mu(A_{1,k}))| \leq \frac{1}{a_{n_{k+1}}} \left(\frac{n_{k+1} d_{k-1}}{n_{k-1}} \gamma_k + p_{k-1} d_{k-1} \right) \leq \alpha_{k-1} \alpha_{k+1} + \alpha_{k-1} \alpha_{k+1}.$$

(ii) follows. □

The set A_1

We can now define the set we want as

$$A_1 = \lim_{k \rightarrow \infty} A_{1,k}$$

which is well defined because the sequence $(\mu(A_{1,k} \triangle A_{1,k+1}))_{k \geq 1}$ is summable.

Lemma 4.14.

(i) $\mu(A_1) \leq 2\alpha_1$.

(ii)

$$d(\mathcal{L}_\Omega(\frac{1}{a_{n_1}} S_{n_1}(\mathbb{1}_{A_1} - \mu(A_1))), \nu_1) \leq 2\varepsilon_1.$$

(iii) For all $k \geq 2$,

$$d(\mathcal{L}_\Omega(\frac{1}{a_{n_k}} S_{n_k}(\mathbb{1}_{A_1} - \mu(A_1))), \delta_0) \leq \varepsilon_k.$$

Proof.

For all $k \geq 1$, we have

$$\mu(A_1 \triangle A_{1,k}) \leq \sum_{j=k+1}^{\infty} \mu(A_{1,j-1} \triangle A_{1,j}) \leq \sum_{j=k+1}^{\infty} \beta_j \leq \alpha_{k+1} \quad (4.19)$$

and then $\mu(A_1) \leq \mu(A_{1,1}) + \mu(A_1 \triangle A_{1,1}) \leq 2\alpha_1$. Further (4.19) implies, for all n ,

$$d(\mathcal{L}_\Omega(\frac{1}{a_n}(S_n(\mathbb{1}_{A_{1,k}} - \mu(A_{1,k})) - S_n(\mathbb{1}_{A_1} - \mu(A_1)))), \delta_0) \leq \frac{n}{a_n} \alpha_{k+1}.$$

(ii) and (iii) follows. $\square$

Step 2: The set A_2

The set $A_{2,2}$

We will consider $F_2 \in \mathcal{A}$ and we repeat what we did to find the set $A_{1,1}$, working with n_2, q_2, p_2, d_2 instead of n_1, q_1, p_1, d_1 . The difference comes to the fact that we want $A_1 \cap A_2 = \emptyset$. Recall that η_2 is the probability measure on $\mathbb{Z}$ given by Lemma 4.10 applied to ν_2 and α_2 and with constants C_2 and n_2 . Let h_2 be a function from F_2 to $\mathbb{Z}$ given by Lemma 4.2 such that $\mathcal{L}_G(h_2) = \eta_2$ and we called g_2 the positive function equal to $h_2 + d_2$. Let

$$A_{F_2,i} = A_{F_2,i}^{(g_2)} := g_2^{-1}(\{i\}), \quad i = 1, \dots, 2d_2$$

and for all $i = 1, \dots, 2d_2$, let $\{A_{F_2,i,1}, \dots, A_{F_2,i,q_2^{(i)}}\}$ be a partition of the set $A_{F_2,i}$ into sets of measure $\frac{1}{q_2}$. We thus have a partition of F_2 into

$$\{A_{F_2,1,1}, \dots, A_{F_2,1,q_2^{(1)}}, A_{F_2,2,1}, \dots, A_{F_2,2,q_2^{(2)}}, \dots, A_{F_2,2d_1,1}, \dots, A_{F_2,2d_1,q_2^{(2d_1)}}\}.$$

By induction, we deduce partitions of $F_2^l = T^{ln_2} F_2$ for $l = 1, \dots, p_2 - 1$. We set

$$A_{F_2^l,i,j} = \begin{cases} T^{n_2} A_{F_2^{l-1},i,j+1} & \text{if } 1 \leq j \leq q_2^{(i-1)} \\ T^{n_2} A_{F_2^{l-1},i+1,1} & \text{if } j = q_2^{(i)} \text{ and } i < 2d_2 \\ T^{n_2} A_{F_2^{l-1},1,1} & \text{if } j = q_2^{(2d_1)} \text{ and } i = 2d_2 \end{cases}.$$

For all $l = 0, \dots, p_2 - 1$ and for all $i = 1, \dots, 2d_2$, we set

$$A_{F_2^l,i} = \bigcup_{j=1}^{q_2^{(i)}} A_{F_2^l,i,j}.$$

Because we want disjointness, we cannot define $A_{F_2^l}$ as $\bigcup_{i=1}^{2d_2} \{T^k A_{F_2^l, i} / k = 0, \dots, i-1\}$. So, for each $l \in \{0, \dots, p_2 - 1\}$, for each $x \in \mathcal{F}_2^l$, if $x \in A_{F_2^l, i}$, we denote by $D_l(x)$ the set composed by the i first elements of $\{x, Tx, \dots\} \setminus A_1$ and we set

$$D_l = \bigcup_{x \in F_2^l} D_l(x).$$

Since A_1 contains at most $2d_1$ points in each orbit of size n_1 and since $d_2 \geq d_1$,

$$D_l \subset \bigcup_{i=0}^{4d_2-1} T^i F_2^l.$$

We define $A_{2,2}$ as

$$A_{2,2} = \bigcup_{l=0}^{p_2-1} D_l.$$

Lemma 4.15.

(i) $\mu(A_{2,2}) \leq \alpha_2$;

(ii)

$$d(\mathcal{L}_\Omega(\frac{1}{a_{n_2}} S_{n_2}(\mathbb{1}_{A_{2,2}} - \mu(A_{2,2}))), \nu_2) \leq \varepsilon_2;$$

(iii)

$$d(\mathcal{L}_\Omega(\frac{1}{a_{n_3}} S_{n_3}(\mathbb{1}_{A_{2,2}} - \mu(A_{2,2}))), \delta_0) \leq \alpha_3.$$

Proof.

The proof follows the one of Lemma 4.12 and is left to the reader. $\square$

The sets $A_{2,k}$, $k \geq 3$

Now we define a sequence $A_{2,k}$, $k \geq 3$ such that

$$\mu(A_{2,k-1} \triangle A_{2,k}) \leq \beta_k$$

and for all $k \geq 3$,

$$d(\mathcal{L}_\Omega(\frac{1}{a_{n_{k+1}}} S_{n_{k+1}}(\mathbb{1}_{A_{2,k}} - \mu(A_{2,k}))), \delta_0) \leq \alpha_{k+1}.$$

To do that we use the same techniques than before, preserving the fact that for all $k \geq 1$, $A_{1,k}$ contains at most $2d_2$ points in each orbit of length n_2 .

The set A_2

The set

$$A_2 = \lim_{k \rightarrow \infty} A_{2,k}$$

is well defined, disjoint of A_1 and satisfies the following lemmas.

Lemma 4.16.

(i) $\mu(A_2) \leq 2\alpha_2$;

(ii)

$$d(\mathcal{L}_\Omega(\frac{1}{a_{n_2}} S_{n_2}(\mathbb{1}_{A_2} - \mu(A_2))), \nu_2) \leq 2\varepsilon_2;$$

(iii) For all $k \geq 3$,

$$d(\mathcal{L}_\Omega(\frac{1}{a_{n_k}} S_{n_k}(\mathbb{1}_{A_2} - \mu(A_2))), \delta_0) \leq \varepsilon_k.$$

Proof.

The proof follows the one of Lemma 4.14. □

Further

Lemma 4.17.

(i)

$$d(\mathcal{L}_\Omega(\frac{1}{a_{n_1}} S_{n_1}(\mathbb{1}_{A_1 \cup A_2} - \mu(A_1 \cup A_2))), \nu_1) \leq 2\varepsilon_1 + 2\varepsilon_2;$$

(ii)

$$d(\mathcal{L}_\Omega(\frac{1}{a_{n_2}} S_{n_2}(\mathbb{1}_{A_1 \cup A_2} - \mu(A_1 \cup A_2))), \nu_2) \leq 3\varepsilon_2.$$

Proof.

By disjointness, for all $n \geq 1$,

$$S_n(\mathbb{1}_{A_1 \cup A_2} - \mu(A_1 \cup A_2)) = S_n(\mathbb{1}_{A_1} - \mu(A_1)) + S_{n_1}(\mathbb{1}_{A_2} - \mu(A_2))$$

and

$$\begin{aligned} d(\mathcal{L}_\Omega(\frac{1}{a_{n_1}} S_{n_1}(\mathbb{1}_{A_1 \cup A_2} - \mu(A_1 \cup A_2))), \nu_1) &\leq d(\mathcal{L}_\Omega(\frac{1}{a_{n_1}} S_{n_1}(\mathbb{1}_{A_1} - \mu(A_1))), \nu_1) \\ &\quad + d(\mathcal{L}_\Omega(\frac{1}{a_{n_1}} S_{n_1}(\mathbb{1}_{A_2} - \mu(A_2))), \delta_0). \end{aligned}$$

From Lemma 4.14, the first term is smaller than $2\varepsilon_1$ and since $\mu(A_2) \leq 2\alpha_2$, by definition of α_2 ,

$$d(\mathcal{L}_\Omega(\frac{1}{a_{n_1}} S_{n_1}(\mathbb{1}_{A_2} - \mu(A_2))), \delta_0) \leq 2 \frac{n_1}{a_{n_1}} 2\alpha_2 \leq 2\varepsilon_2.$$

This give (i). In the same way, using Lemma 4.16 and Lemma 4.14, we get (ii). □

Step k: the set A_k , $k \geq 3$

By induction, we can find a set A_k disjoint of the sets A_i , $i < k$, satisfying the following lemma.

Lemma 4.18.

(i) $\mu(A_k) \leq 2\alpha_k;$

(ii)

$$d(\mathcal{L}_\Omega(\frac{1}{a_{n_k}} S_{n_k}(\mathbb{1}_{A_k} - \mu(A_k))), \nu_k) \leq 2\varepsilon_k;$$

(iii) For all $j \geq k + 1$,

$$d(\mathcal{L}_\Omega(\frac{1}{a_{n_j}} S_{n_j}(\mathbb{1}_{A_k} - \mu(A_k))), \delta_0) \leq \varepsilon_j.$$

Final step: the set A

Set

$$A = \bigcup_{k \geq 1} A_k.$$

Then $\mu(A) = \sum_{k \geq 1} \mu(A_k) \leq \sum_{k \geq 1} \varepsilon_k \leq \varepsilon$ and for all $j \geq 1$,

$$d(\mathcal{L}_\Omega(\frac{1}{a_{n_j}} S_{n_j}(\mathbb{1}_A - \mu(A))), \nu_j) \leq 2 \sum_{i \geq j} i \varepsilon_i$$

which goes to 0 when j goes to ∞ .

Thus Proposition 4.9 is proved. □

Comments

The content of this chapter comes from the paper by Durieu and Volný [39]. Thanks to Jean-Pierre Conze who found an error in an earlier version of this paper and thanks to Sébastien Gouëzel who suggested us a simplification in the proof which is done in Section 4.4.

Annexe A

Multiple Mixing for Ergodic Torus Automorphisms

We prove here the multiple mixing inequality for ergodic torus automorphisms which is used in Chapter 2.

A version of this inequality in the case of diagonal flows and for Hölder observables can be found in the paper by Le Borgne [63] (see also Le Borgne and Pène [64]). Here, it is adapted to the case of ergodic torus automorphisms and for $\mathcal{C}^1$ functions (because it is enough in our situation). Hence the proof is not new and follows the one of [64]. Because the setting is different, for convenience of the reader, we give here the main ideas.

Let T be a general (quasi-hyperbolic) ergodic automorphism of the torus of dimension d and μ the Lebesgue measure on $\mathbb{T}^d$.

Proposition. *There exist $C > 0$, $\alpha > 0$, for all $m, p \in \mathbb{N}^*$, for all $\phi_1, \dots, \phi_m, \psi_1, \dots, \psi_p \in \mathcal{C}^1(\mathbb{T}^d)$, for all $k_1 \leq \dots \leq k_m \leq 0 \leq l_1 \leq \dots \leq l_p$, for all $n \in \mathbb{N}$,*

$$\left| \text{Cov} \left(\prod_{j=1}^m \phi_j \circ T^{k_j}, \prod_{j=1}^p \psi_j \circ T^{l_j+n} \right) \right| \leq C \left(\sum_{j=1}^m \prod_{i \neq j} \|\phi_i\|_\infty \|\phi_j\|_{\mathcal{C}^1} |k_j|^r \right) \left(\sum_{j=1}^p \prod_{i \neq j} \|\psi_i\|_\infty \|\psi_j\|_{\mathcal{C}^1} \right) e^{-\alpha n}$$

where r is the size of the biggest Jordan's block of T restricted to its neutral subspace.

Proof.

Denote by E^s , E^u and E^c the T -stable subspaces of $\mathbb{R}^d$ corresponding respectively to the stable, the unstable and the central directions of T (where T is identified to its representative matrix). We have

$$\mathbb{R}^d = E^s \oplus E^u \oplus E^c$$

and there exists $\lambda > 1$ such that for all $n \geq 1$,

$$|T^n v| \leq \lambda^{-n} |v| \text{ for all } v \in E^s,$$

$$|T^n v| \geq \lambda^n |v| \text{ for all } v \in E^u,$$

$$|T^n v| \leq n^r |v| \text{ for all } v \in E^c,$$

where $|\cdot|$ denotes the maximum norm on $\mathbb{R}^d$ and r is the size of the greatest Jordan block of T restricted to the space E^c . Further μ can be written as the product measure of μ_s , μ_u and μ_c . Set

$$B_i(0, \rho) = \{x \in E^i / |x| \leq \rho\}, \rho > 0, i = s, u, c.$$

It is well known, since the work of Lind [66], that we have an exponential mixing inequality of the type

$$\exists C > 0, \exists \alpha > 0, \forall \phi, \psi \in \mathcal{C}^1(\mathbb{T}^d, \mathbb{R}), |\text{Cov}(\phi, \psi \circ T^n)| \leq C \|\phi\|_{\mathcal{C}^1} \|\psi\|_{\mathcal{C}^1} e^{-\alpha n}, \quad (\text{A.1})$$

where $\|g\|_{\mathcal{C}^1} = \|g\|_{\infty} + \sup_{x \in \mathbb{T}^d} \|Dg(x)\|_{\infty}$.

We cannot deduce directly the proposition because the $\mathcal{C}^1$ -norm behaves badly with the composition by T . But from this, we can deduce a property of good distribution of the stable leaves. Denote by $\|\cdot\|_{\mathcal{C}^1, s}$ (resp. $\|\cdot\|_{\mathcal{C}^1, u, c}$) the $\mathcal{C}^1$ -norm is the stable direction (resp. unstable-central direction).

Lemma. (good distribution of stable leaves) *There exist θ such that for all $\phi \in \mathcal{C}^1$, $x \in \mathbb{T}^d$ and $\rho > 0$,*

$$\frac{1}{\mu_s(T^{-n}B_s(0, \rho))} \left| \int_{T^{-n}B(0, \rho)} \phi(T^{-n}x + s) ds \right| \leq C \|\phi\|_{\mathcal{C}^1, u, c} e^{-\theta n}.$$

Let $\mathcal{A}_0$ be a sub- σ -algebra of the Borelian one for which the atoms are pieces of stable leaves and set $\mathcal{A}_n = T^{-n}\mathcal{A}_0$. Let ϕ and ψ be two $\mathcal{C}^1$ -function with zero mean.

$$\begin{aligned} \text{Cov}(\phi, \psi \circ T^n) &= \text{Cov}(\phi - E(\phi|\mathcal{A}_{-\lfloor \frac{n}{2} \rfloor}), \psi \circ T^n) + \text{Cov}(E(\phi|\mathcal{A}_{-\lfloor \frac{n}{2} \rfloor}), \psi \circ T^n) \\ &\leq \|\psi\|_{\infty} \|\phi - E(\phi|\mathcal{A}_{-\lfloor \frac{n}{2} \rfloor})\|_{\infty} + \|\phi\|_{\infty} \|E(\psi|\mathcal{A}_{\lceil \frac{n}{2} \rceil})\|_{\infty}. \end{aligned}$$

But, since the diameter of the atoms of $\mathcal{A}_{-n}$ decreases exponentially fast,

$$\|\phi - E(\phi|\mathcal{A}_k)\|_{\infty} \leq C \|\phi\|_{\mathcal{C}^1, s} e^{-\frac{\alpha n}{2}}$$

and, by the lemma,

$$\|E(\psi|\mathcal{A}_{\lceil \frac{n}{2} \rceil})\|_{\infty} \leq C \|\psi\|_{\mathcal{C}^1, u, c} e^{-\frac{\theta n}{2}}.$$

Thus, for $\gamma = \max\{\alpha, \theta\}$, we get

$$|\text{Cov}(\phi, \psi \circ T^n)| \leq C \|\phi\|_{\mathcal{C}^1, s} \|\psi\|_{\mathcal{C}^1, u, c} e^{-\gamma n}. \quad (\text{A.2})$$

Further, for all $n \geq 0$ and for all $\phi \in \mathcal{C}^1$, we have

$$\|\phi \circ T^n\|_{\mathcal{C}^1, s} \leq C \|\phi\|_{\mathcal{C}^1, s}. \quad (\text{A.3})$$

Indeed, if $v \in E^s$ and $h > 0$, by linearity of the map T ,

$$\begin{aligned} \frac{|\phi \circ T^n(x) - \phi \circ T^n(x + hv)|}{h} &= \frac{|\phi \circ T^n(x) - \phi(T^n(x) + hT^n(v))|}{h} \\ &\leq \|\phi\|_{\mathcal{C}^1, s} |T^n v| \\ &\leq \|\phi\|_{\mathcal{C}^1, s} \lambda^{-n} |v|. \end{aligned}$$

And, in the same way, we get

$$\|\phi \circ T^{-n}\|_{\mathcal{C}^1, u, c} \leq C n^r \|\phi\|_{\mathcal{C}^1, u, c} \quad (\text{A.4})$$

Now, to prove the proposition, we apply what precedes to

$$\phi = \prod_{j=1}^m \phi_j \circ T^{k_j} \text{ and } \psi = \prod_{j=1}^p \psi_j \circ T^{l_j}.$$

Using (A.3) and (A.4), the computation shows that

$$\|\phi\|_{\mathcal{C}^1, u, c} \leq \sum_{j=1}^m \prod_{i \neq j} \|\phi_i\|_{\infty} \|\phi_j\|_{\mathcal{C}^1, u, c} |k_j|^r$$

and

$$\|\psi\|_{\mathcal{C}^1, s} \leq \sum_{j=1}^p \prod_{i \neq j} \|\psi_i\|_{\infty} \|\psi_j\|_{\mathcal{C}^1, s}.$$

Then, by (A.2), the proposition is proved. $\square$

Proof of the lemma.

Let $K = x + B_s(0, \rho_1) + B_{u, c}(0, \rho_2)$, where $x \in \mathbb{T}^d$ and $\rho_1, \rho_2 > 0$ and set $\psi = \mathbb{1}_K$. Let $\phi \in \mathcal{C}^1$ such that $E\phi = 0$ and fix an integer k . We have

$$\begin{aligned} \langle \phi, \psi \circ T^k \rangle &= \int_{T^{-k}K} \phi d\mu \\ &= \int_{T^{-k}B_s(0, \rho_1) \times T^{-k}B_{u, c}(0, \rho_2)} \phi(T^{-k}x + t + s) d\mu_s(t) d\mu_{u, c}(s) \\ &= \int_{T^{-k}B_s(0, \rho_1) \times T^{-k}B_{u, c}(0, \rho_2)} [\phi(T^{-k}x + t + s) - \phi(T^{-k}x + t)] d\mu_s(t) d\mu_{u, c}(s) \\ &\quad + \mu_{u, c}(T^{-k}B_{u, c}(0, \rho_2)) \int_{T^{-k}B_s(0, \rho_1)} \phi(T^{-k}x + t) d\mu_s(t) \end{aligned}$$

By the mean-value theorem, the first term is less than

$$\mu(K)\|\phi\|_{\mathcal{C}^1, u, c} \rho_2.$$

The rectangle K being regular, there exist a $\mathcal{C}^1$ function $\tilde{\psi}$ and a constant $C > 0$ such that

$$\|\tilde{\psi}\|_{\mathcal{C}^1} \leq C e^{\frac{\alpha}{2}k} \text{ and } \|\psi - \tilde{\psi}\|_1 \leq C e^{\frac{\alpha}{2(d+1)}k},$$

see Lemma 2.12. Then, by (A.1), we get

$$| \langle \phi, \tilde{\psi} \circ T^k \rangle | \leq C \|\phi\|_{\mathcal{C}^1} e^{-\frac{\alpha}{2}k}$$

and

$$| \langle \phi, \psi \circ T^k - \tilde{\psi} \circ T^k \rangle | \leq C \|\phi\|_{\infty} \|\psi - \tilde{\psi}\|_1 e^{-\frac{\alpha}{2(d+1)}k}.$$

we deduce that there exists $\theta > 0$ such that

$$\frac{1}{\mu_s(T^{-k}B_s(0, \rho_1))} \left| \int_{T^{-k}B(0, \rho_1)} \phi(T^{-k}x + s) ds \right| \leq C \|\phi\|_{\mathcal{C}^1} e^{-\theta k}.$$

Now, the goal is to get the norm $\|\cdot\|_{\mathcal{C}^1, u, c}^1$ instead of the $\mathcal{C}^1$ -norm in the previous inequality. Assume that ϕ has a support in a small rectangle $L = x_0 + B_s(0, \rho_1) + B_{u, c}(0, \rho_2)$ (otherwise, we can decompose the function ϕ as a sum of functions with supports in small rectangles). We can show there exists a $\mathcal{C}^1$ function χ defined on L , such that $\int_L \chi = \int_L \phi$, $\|\chi\|_{\mathcal{C}^1} \leq C \|\phi\|_{\mathcal{C}^1, u, c}$ and $\int_{T^{-k}B_s(0, \rho_1)} \chi(x_0 + s) d\mu_s(s)$ is close to $\int_{T^{-k}B_s(0, \rho_1)} \phi(x_0 + s) d\mu_s(s)$ when k is large enough.

To do that we consider a $\mathcal{C}^\infty$ function Γ defined on $B_s(0, \rho_1)$ which has integral equal to 1 and we define χ by

$$\chi(x_0 + s + t) = \int_{B_s(0, \rho_1)} \phi(x_0 + u + t) d\mu_s(u) \Gamma(s).$$

We can check that the wandering properties hold and we deduce the lemma. $\square$

Bibliographie

- [1] Steve Alpern. Generic properties of measure preserving homeomorphisms. In *Ergodic theory (Proc. Conf., Math. Forschungsinst., Oberwolfach, 1978)*, volume 729 of *Lecture Notes in Math.*, pages 16–27. Springer, Berlin, 1979.
- [2] Viviane Baladi. *Positive transfer operators and decay of correlations*, volume 16 of *Advanced Series in Nonlinear Dynamics*. World Scientific Publishing Co. Inc., River Edge, NJ, 2000.
- [3] Marcel Berger and Bernard Gostiaux. *Géométrie différentielle: variétés, courbes et surfaces*. Mathématiques. [Mathematics]. Presses Universitaires de France, Paris, second edition, 1992.
- [4] István Berkes and Walter Philipp. An almost sure invariance principle for the empirical distribution function of mixing random variables. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 41(2):115–137, 1977/78.
- [5] Patrick Billingsley. The Lindeberg-Lévy theorem for martingales. *Proc. Amer. Math. Soc.*, 12:788–792, 1961.
- [6] Patrick Billingsley. *Convergence of probability measures*. John Wiley & Sons Inc., New York, 1968.
- [7] G. D. Birkhoff. Proof of the ergodic theorem. *Proceedings of the National Academy of Sciences*, 17:656–660, 1931.
- [8] Svetlana Borovkova, Robert Burton, and Herold Dehling. Limit theorems for functionals of mixing processes with applications to U -statistics and dimension estimation. *Trans. Amer. Math. Soc.*, 353(11):4261–4318 (electronic), 2001.
- [9] Rufus Bowen. *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*. Springer-Verlag, Berlin, 1975. Lecture Notes in Mathematics, Vol. 470.
- [10] Anne Broise. Transformations dilatantes de l'intervalle et théorèmes limites. *Astérisque*, (238):1–109, 1996. Études spectrales d'opérateurs de transfert et applications.
- [11] Robert Burton and Manfred Denker. On the central limit theorem for dynamical systems. *Trans. Amer. Math. Soc.*, 302(2):715–726, 1987.

- [12] Yuan Shih Chow and Henry Teicher. *Probability theory*. Springer Texts in Statistics. Springer-Verlag, New York, third edition, 1997. Independence, interchangeability, martingales.
- [13] P. Collet, S. Martinez, and B. Schmitt. Asymptotic distribution of tests for expanding maps of the interval. *Ergodic Theory Dynam. Systems*, 24(3):707–722, 2004.
- [14] I. P. Cornfeld, S. V. Fomin, and Ya. G. Sinaï. *Ergodic theory*, volume 245 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, New York, 1982. Translated from the Russian by A. B. Sosinskiĭ.
- [15] Jérôme Dedecker. A central limit theorem for stationary random fields. *Probab. Theory Related Fields*, 110(3):397–426, 1998.
- [16] Jérôme Dedecker and Florence Merlevède. The conditional central limit theorem in Hilbert spaces. *Stochastic Process. Appl.*, 108(2):229–262, 2003.
- [17] Jérôme Dedecker, Florence Merlevède, and Dalibor Volný. On the weak invariance principle for non-adapted sequences under projective criteria. *J. Theoret. Probab.*, 20(4):971–1004, 2007.
- [18] Jérôme Dedecker and Clémentine Prieur. An empirical central limit theorem for dependent sequences. *Stochastic Process. Appl.*, 117(1):121–142, 2007.
- [19] Jérôme Dedecker and Emmanuel Rio. On the functional central limit theorem for stationary processes. *Ann. Inst. H. Poincaré Probab. Statist.*, 36(1):1–34, 2000.
- [20] Herold Dehling, Olivier Durieu, and Dalibor Volný. New techniques for empirical processes of dependent data. *submitted*, ArXiv: 0806.2941.
- [21] Herold Dehling and Walter Philipp. Empirical process techniques for dependent data. In *Empirical process techniques for dependent data*, pages 3–113. Birkhäuser Boston, Boston, MA, 2002.
- [22] Andrés del Junco and Joseph Rosenblatt. Counterexamples in ergodic theory and number theory. *Math. Ann.*, 245(3):185–197, 1979.
- [23] Yves Derriennic. Some aspects of recent works on limit theorems in ergodic theory with special emphasis on the “central limit theorem”. *Discrete Contin. Dyn. Syst.*, 15(1):143–158, 2006.
- [24] Yves Derriennic and Michael Lin. The central limit theorem for Markov chains with normal transition operators, started at a point. *Probab. Theory Related Fields*, 119(4):508–528, 2001.
- [25] Yves Derriennic and Michael Lin. Fractional Poisson equations and ergodic theorems for fractional coboundaries. *Israel J. Math.*, 123:93–130, 2001.

- [26] Yves Derriennic and Michael Lin. The central limit theorem for Markov chains started at a point. *Probab. Theory Related Fields*, 125(1):73–76, 2003.
- [27] Dmitry Dolgopyat. Limit theorems for partially hyperbolic systems. *Trans. Amer. Math. Soc.*, 356(4):1637–1689 (electronic), 2004.
- [28] Monroe D. Donsker. An invariance principle for certain probability limit theorems. *Mem. Amer. Math. Soc.*, 1951(6):12, 1951.
- [29] Monroe D. Donsker. Justification and extension of Doob’s heuristic approach to the Komogorov-Smirnov theorems. *Ann. Math. Statistics*, 23:277–281, 1952.
- [30] J. L. Doob. Heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Statistics*, 20:393–403, 1949.
- [31] P. Doukhan, P. Massart, and E. Rio. Invariance principles for absolutely regular empirical processes. *Ann. Inst. H. Poincaré Probab. Statist.*, 31(2):393–427, 1995.
- [32] Paul Doukhan and Sana Louhichi. A new weak dependence condition and applications to moment inequalities. *Stochastic Process. Appl.*, 84(2):313–342, 1999.
- [33] Paul Doukhan and Donatas Surgailis. Functional central limit theorem for the empirical process of short memory linear processes. *C. R. Acad. Sci. Paris Sér. I Math.*, 326(1):87–92, 1998.
- [34] O. Dovgoshey, O. Martio, V. Ryazanov, and M. Vuorinen. The Cantor function. *Expo. Math.*, 24(1):1–37, 2006.
- [35] R. M. Dudley. *Real analysis and probability*, volume 74 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2002. Revised reprint of the 1989 original.
- [36] Olivier Durieu. A fourth moment inequality for functionals of stationary processes. to appear in *J. Appl. Probability*, ArXiv: 0806.2980.
- [37] Olivier Durieu. Independence of four projective criteria for the weak invariance principle. to appear in *Aléa Lat. Am. J. Probab. Math. Stat.*
- [38] Olivier Durieu and Philippe Jouan. Empirical invariance principle for ergodic torus automorphisms; genericity. *Stoch. Dyn.*, 8(2):173–195, 2008.
- [39] Olivier Durieu and Dalibor Volný. On sums of indicator functions in dynamical systems. *submitted*, ArXiv: 0810.2917.
- [40] Olivier Durieu and Dalibor Volný. Comparison between criteria leading to the weak invariance principle. *Ann. Inst. H. Poincaré Probab. Statist.*, 44(2):324–340, 2008.
- [41] S. J. Eigen and V. S. Prasad. Multiple Rokhlin tower theorem: a simple proof. *New York J. Math.*, 3A(Proceedings of the New York Journal of Mathematics Conference, June 9–13, 1997):11–14 (electronic), 1997/98.

- [42] P. Erdős and M. Kac. On certain limit theorems of the theory of probability. *Bull. Amer. Math. Soc.*, 52:292–302, 1946.
- [43] Carl-Gustav Esseen and Svante Janson. On moment conditions for normed sums of independent variables and martingale differences. *Stochastic Process. Appl.*, 19(1):173–182, 1985.
- [44] M. I. Gordin. The central limit theorem for stationary processes. *Dokl. Akad. Nauk SSSR*, 188:739–741, 1969.
- [45] M. I. Gordin. Abstracts of communication, t.1: A-k. International conference on probability theory, Vilnius. 1973.
- [46] M. I. Gordin and B. A. Lifšic. Central limit theorem for stationary Markov processes. *Dokl. Akad. Nauk SSSR*, 239(4):766–767, 1978.
- [47] Sébastien Gouëzel. An interval map with a spectral gap on lipschitz functions, but not on bounded variation functions. ArXiv:0809.0658v1.
- [48] Y. Guivarc’h and J. Hardy. Théorèmes limites pour une classe de chaînes de Markov et applications aux difféomorphismes d’Anosov. *Ann. Inst. H. Poincaré Probab. Statist.*, 24(1):73–98, 1988.
- [49] Allan Gut. *Probability: a graduate course*. Springer Texts in Statistics. Springer, New York, 2005.
- [50] P. Hall and C. C. Heyde. *Martingale limit theory and its application*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1980. Probability and Mathematical Statistics.
- [51] E. J. Hannan. Central limit theorems for time series regression. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 26:157–170, 1973.
- [52] E. J. Hannan. The central limit theorem for time series regression. *Stochastic Process. Appl.*, 9(3):281–289, 1979.
- [53] Boris Hasselblatt and Yakov Pesin. Partially hyperbolic dynamical systems. In *Handbook of dynamical systems. Vol. 1B*, pages 1–55. Elsevier B. V., Amsterdam, 2006.
- [54] Hubert Hennion and Loïc Hervé. *Limit theorems for Markov chains and stochastic properties of dynamical systems by quasi-compactness*, volume 1766 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2001.
- [55] C. C. Heyde. On the central limit theorem for stationary processes. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 30:315–320, 1974.
- [56] C. C. Heyde. On the central limit theorem and iterated logarithm law for stationary processes. *Bull. Austral. Math. Soc.*, 12:1–8, 1975.

- [57] Morris W. Hirsch. *Differential topology*. Springer-Verlag, New York, 1976. Graduate Texts in Mathematics, No. 33.
- [58] I. A. Ibragimov. A central limit theorem for a class of dependent random variables. *Teor. Veroyatnost. i Primenen.*, 8:89–94, 1963.
- [59] Yitzhak Katznelson. Ergodic automorphisms of T^n are Bernoulli shifts. *Israel J. Math.*, 10:186–195, 1971.
- [60] C. Kipnis and S. R. S. Varadhan. Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Comm. Math. Phys.*, 104(1):1–19, 1986.
- [61] Stéphane Le Borgne. *Dynamique symbolique et propriétés stochastiques des automorphismes du tore: cas hyperbolique et quasi hyperbolique*. PhD thesis, Université de Rennes 1, 1997.
- [62] Stéphane Le Borgne. Limit theorems for non-hyperbolic automorphisms of the torus. *Israel J. Math.*, 109:61–73, 1999.
- [63] Stéphane Le Borgne. Exemples de systèmes dynamiques quasi-hyperboliques à décorrelations lentes. <http://perso.univ-rennes1.fr/stephane.leborgne/>, 2006.
- [64] Stéphane Le Borgne and Françoise Pène. Vitesse dans le théorème limite central pour certains systèmes dynamiques quasi-hyperboliques. *Bull. Soc. Math. France*, 133(3):395–417, 2005.
- [65] Pierre Liardet and Dalibor Volný. Sums of continuous and differentiable functions in dynamical systems. *Israel J. Math.*, 98:29–60, 1997.
- [66] D. A. Lind. Dynamical properties of quasihyperbolic toral automorphisms. *Ergodic Theory Dynamical Systems*, 2(1):49–68, 1982.
- [67] J. W. Lindeberg. Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Math. Z.*, 15(1):211–225, 1922.
- [68] Carlangelo Liverani. Central limit theorem for deterministic systems. In *International Conference on Dynamical Systems (Montevideo, 1995)*, volume 362 of *Pitman Res. Notes Math. Ser.*, pages 56–75. Longman, Harlow, 1996.
- [69] Michael Maxwell and Michael Woodroffe. Central limit theorems for additive functionals of Markov chains. *Ann. Probab.*, 28(2):713–724, 2000.
- [70] Florence Merlevède, Magda Peligrad, and Sergey Utev. Recent advances in invariance principles for stationary sequences. *Probab. Surv.*, 3:1–36 (electronic), 2006.
- [71] S. P. Meyn and R. L. Tweedie. *Markov chains and stochastic stability*. Communications and Control Engineering Series. Springer-Verlag London Ltd., London, 1993.
- [72] J. Milnor. *Morse theory*. Based on lecture notes by M. Spivak and R. Wells. Annals of Mathematics Studies, No. 51. Princeton University Press, Princeton, N.J., 1963.

- [73] Donald Ornstein. Bernoulli shifts with the same entropy are isomorphic. *Advances in Math.*, 4:337–352 (1970), 1970.
- [74] William Parry and Mark Pollicott. Zeta functions and the periodic orbit structure of hyperbolic dynamics. *Astérisque*, (187-188):268, 1990.
- [75] K. R. Parthasarathy. *Probability measures on metric spaces*. Probability and Mathematical Statistics, No. 3. Academic Press Inc., New York, 1967.
- [76] Magda Peligrad and Sergey Utev. A new maximal inequality and invariance principle for stationary sequences. *Ann. Probab.*, 33(2):798–815, 2005.
- [77] Françoise Pène. Rate of convergence in the multidimensional central limit theorem for stationary processes. Application to the Knudsen gas and to the Sinai billiard. *Ann. Appl. Probab.*, 15(4):2331–2392, 2005.
- [78] Walter Philipp and Laurence Pinzur. Almost sure approximation theorems for the multivariate empirical process. *Z. Wahrsch. Verw. Gebiete*, 54(1):1–13, 1980.
- [79] David Pollard. *Empirical processes: theory and applications*. NSF-CBMS Regional Conference Series in Probability and Statistics, 2. Institute of Mathematical Statistics, Hayward, CA, 1990.
- [80] Clémentine Prieur. An empirical functional central limit theorem for weakly dependent sequences. *Probab. Math. Statist.*, 22(2, Acta Univ. Wratislav. No. 2470):259–287, 2002.
- [81] J. Rousseau-Egele. Un théorème de la limite locale pour une classe de transformations dilatantes et monotones par morceaux. *Ann. Probab.*, 11(3):772–788, 1983.
- [82] Ja. G. Sinaï. A weak isomorphism of transformations with invariant measure. *Dokl. Akad. Nauk SSSR*, 147:797–800, 1962.
- [83] Aad W. van der Vaart and Jon A. Wellner. *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York, 1996. With applications to statistics.
- [84] Dalibor Volný. On limit theorems and category for dynamical systems. *Yokohama Math. J.*, 38(1):29–35, 1990.
- [85] Dalibor Volný. Approximating martingales and the central limit theorem for strictly stationary processes. *Stochastic Process. Appl.*, 44(1):41–74, 1993.
- [86] Dalibor Volný. Martingale approximation of non adapted stochastic processes with nonlinear growth of variance. In *Dependence in probability and statistics*, volume 187 of *Lecture Notes in Statist.*, pages 141–156. Springer, New York, 2006.
- [87] Dalibor Volný. A nonadapted version of the invariance principle of Peligrad and Utev. *C. R. Math. Acad. Sci. Paris*, 345(3):167–169, 2007.

- [88] Dalibor Volný and Pavel Samek. On the invariance principle and the law of iterated logarithm for stationary processes. In *Mathematical physics and stochastic analysis (Lisbon, 1998)*, pages 424–438. World Sci. Publishing, River Edge, NJ, 2000.
- [89] Michael Woodroffe. A central limit theorem for functions of a Markov chain with applications to shifts. *Stochastic Process. Appl.*, 41(1):33–44, 1992.
- [90] Wei Biao Wu. Empirical processes of stationary sequences. *Statist. Sinica*, 18(1):313–333, 2008.
- [91] Wei Biao Wu and Xiaofeng Shao. Limit theorems for iterated random functions. *J. Appl. Probab.*, 41(2):425–436, 2004.
- [92] Wei Biao Wu and Michael Woodroffe. Martingale approximations for sums of stationary processes. *Ann. Probab.*, 32(2):1674–1690, 2004.