



HAL
open science

Filtrations à temps discret

Gael Ceillier

► **To cite this version:**

Gael Ceillier. Filtrations à temps discret. Mathématiques [math]. Université Joseph-Fourier - Grenoble I, 2010. Français. NNT: . tel-00561467

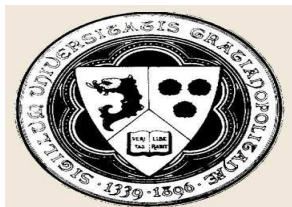
HAL Id: tel-00561467

<https://theses.hal.science/tel-00561467>

Submitted on 1 Feb 2011

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.



UNIVERSITÉ DE GRENOBLE

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité **MSTII / Mathématiques**

Arrêté ministériel : 7 août 2006

Présentée et soutenue publiquement par

Gaël CEILLIER

le **01/12/2010**

« Filtrations à temps discret »

Thèse dirigée par **Christophe Leuridan**

JURY

Soutenance à Grenoble le 1^{er} décembre 2010 devant le jury composé de :

Rapporteurs	Michel ÉMERY	Directeur de recherche CNRS, IRMA
	Anatoly VERSHIK	Professeur, Steklov Institute of Mathematics
Examineurs	Xavier BRESSAUD	Professeur, Université Paul Sabatier
	Didier PIAU	Professeur, Université de Grenoble
	Marc YOR	Professeur, Université Paris VI
	Christophe LEURIDAN	Maître de Conférences, Université de Grenoble (Directeur de thèse)

Thèse préparée au sein de l'Institut Fourier, Laboratoire de mathématiques dans l' Ecole Doctorale de Grenoble, MSTII.

À Manon, pour qui tout commence au moment où j'écris ces mots.

Remerciements,

Je tiens d'abord à exprimer mes remerciements à Christophe Leuridan, pour avoir accepté de devenir mon directeur de thèse et pour l'aide précieuse qu'il a su m'apporter. Je voudrais qu'il sache combien sa confiance, son soutien et sa compréhension, m'ont été précieux durant ces trois années.

Je souhaiterais, ensuite remercier Michel Émery, pour avoir accepté d'être l'un des rapporteurs de ce manuscrit, ainsi que pour la qualité des conseils qu'il m'a prodigué durant ces dernières années. J'aimerais également le remercier pour m'avoir donné l'opportunité de lui présenter mes travaux et pour la sympathie qu'il a semblé me porter.

It is quite an honor for me that Anatoly Vershik accepted to be one of my referees, and I would like to express my most sincere thanks for his careful reading and helpful comments about my work.

Je tiens aussi à exprimer ma gratitude aux examinateurs de mon jury de thèse, Xavier Bressaud, Didier Piau et Marc Yor.

Merci à Didier Piau, pour m'avoir aidé par ses questions, ses remarques et son humour à mieux exposer mes idées. Je voudrais qu'il sache combien son soutien tout au long de ma thèse a joué dans la pertinence de mes travaux.

Merci à Marc Yor de m'avoir fait l'honneur d'accepter de devenir un des membres de mon jury.

Enfin, merci à Xavier Bressaud, pour l'intérêt qu'il porte à mes travaux et pour avoir accepté de faire partie de mon jury de thèse.

Je remercie également Stéphane Laurent, dont les travaux ont initié les miens et pour les nombreux échanges que nous avons eus.

Enfin, j'aimerais remercier Frédéric Simon pour son aide dans mes calculs et son amitié. Merci aussi aux autres habitants de la "C13", en particulier à Ivan pour avoir supporté mes affres pendant ma première année de thèse, à Maxime pour être resté ce qu'il est, et aussi à Alex, Jean, Antoine et Pol pour tous les bons moments passé ensemble.

Un grand merci à Damien qui est parti en éclaireur à Grenoble, et qui fut non seulement un parfait colocataire, mais aussi un guide culinaire.

Merci à Bastien et à Lucas pour le ski, la coinche et pour tous ces moments de détente.

Merci à Romain pour tout l'Espoir qu'il m'a donné pendant ces longues soirées de travail, ainsi qu'à Magalie et leur petite Marine pour leur accueil.

Merci surtout à Samira, pour tout ce qu'elle a chamboulé dans ma vie et pour tous les projets que nous avons ensemble.

Merci à ma famille, à mon docteur de grand frère qui m'a doublé de quelques mois, à ma soeur Léa qui même loin est toujours aussi proche et à Morgan dont l'intelligence et le talent forcent mon admiration.

Je tiens à remercier, tout particulièrement mes parents dont le soutien sans failles m'a mené là où je suis maintenant.

Contents

1	Introduction	9
1.1	Tribus et filtrations	9
1.2	Filtrations de type produit	11
1.3	Filtrations possédant une paramétrisation génératrice	14
1.4	Filtrations standard	16
1.5	Critères de standardité	18
1.5.1	Critère de Vershik de premier niveau	18
1.5.2	Critère de Vershik de second niveau	19
1.5.3	I-confort	20
1.6	Processus des mots découpés	21
1.7	Processus stationnaires à valeurs dans un ensemble fini	24
1.7.1	Une condition sur les écarts moyens	25
1.7.2	Chaînes à mémoire variable	25
1.7.3	Marche aléatoire en scène aléatoire	26
1.7.4	Une variante de la marche aléatoire en scène aléatoire	27
1.8	Pistes de recherche	27
2	Introduction for experts	29
2.1	Sigma-fields and filtrations	29
2.2	Product type filtrations	30
2.3	Filtrations with a generating parametrization	31
2.4	Standard filtrations	32
2.5	Criterion of standardness	33
2.5.1	Vershik's first level criterion	33

2.5.2	Vershik's second level criterion	34
2.5.3	I-cosiness	35
2.6	Split-word processes	36
2.7	Stationary processes with values in a finite set	38
3	Filtrations of split-word processes	41
3.1	Introduction	41
3.1.1	Results	43
3.1.2	Examples	45
3.2	Laurent's method and tools	47
3.2.1	Change of innovations	47
3.2.2	Canonical word and coupling	48
3.2.3	Adapting Laurent's proof	52
3.3	Improving on condition ∇	53
3.3.1	Statement of the main result	53
3.3.2	Construction of the new innovations	54
3.3.3	Proof of the main result	56
3.3.4	Proof of some auxiliary facts	59
3.4	Proof of non standardness under Δ	61
3.4.1	Preliminary notions	62
3.4.2	Proof of the inequality $\mathbb{P}[X'_{n_0} \neq X''_{n_0}] \geq \mathbb{E}[e_n(X'_n, X''_n)]$ for $n \leq n_0$	65
3.4.3	Bounding below $\mathbb{E}[e_n(X'_n, X''_n)]$ when X'_n and X''_n are independent	66
3.4.4	Negation of the I-cosiness criterion	67
4	Sufficient conditions of standardness for filtrations of stationary processes taking values in a finite space	69
4.1	Introduction	69
4.1.1	Setting	69
4.1.2	Statement of the results	71
4.2	Pointwise influence	74
4.2.1	Construction of a governing sequence	74

4.2.2	Upper bound of the error	78
4.2.3	Proof of the result with maximum influences	82
4.3	Average influences	84
4.3.1	Priming lemma	84
4.3.2	Approximation until a given time	87
4.3.3	Successive approximations and end of the proof of the result with average influences	88
4.4	Examples	92
4.4.1	Parity of the number of ones in a row	92
4.4.2	Random walk in random scenery	93
4.4.3	Random walk in random scenery with misreadings	94
4.4.4	Random walk in renewed random scenery	95
4.5	Proofs of auxiliary facts	98
4.5.1	Inequalities involving $\alpha_n, \delta_n, \gamma_n$ and η_n	98
4.5.2	Proof of lemma 4.1.1	99
4.5.3	Proof of corollary 4.1.2	100
4.5.4	Proof that $\mathcal{H}(\gamma)$ provides a positive lower bound for $p(a x)$	100
5	Annexe : preuve de quelques faits simples	101
5.0.5	Preuve de la proposition 1.3.1	101
5.0.6	Preuve de la proposition 1.3.3	103
5.0.7	Preuve du corollaire 1.4.1	103

Chapter 1

Introduction

1.1 Tribus et filtrations

Kolmogorov a proposé en 1933 une théorie des probabilités dont le cadre est un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$.

Dans nombre de situations, on est amené à travailler avec une sous-tribu de \mathcal{A} . Une sous-tribu représente une certaine information. Par exemple, si ξ est une variable aléatoire à valeurs dans un espace mesuré (E, \mathcal{E}) , la tribu

$$\sigma(\xi) := \{\xi^{-1}(A) ; A \in \mathcal{E}\}$$

représente l'information apportée par la variable ξ . Quand on connaît la valeur prise par ξ en un point $\omega \in \Omega$, on sait quels sont les évènements de $\sigma(\xi)$ qui sont réalisés en ω .

Une tribu est dite triviale (ou dégénérée) si elle ne contient pas d'information, c'est-à-dire si elle ne contient que des évènements de probabilité 0 ou 1.

Dans toute cette thèse, nous travaillons avec des tribus complètes, c'est-à-dire des tribus \mathcal{B} vérifiant la propriété suivante : si \mathcal{B} contient un évènement A de probabilité nulle, alors \mathcal{B} contient toutes les parties de A .

Une sous-tribu de $(\Omega, \mathcal{A}, \mathbb{P})$ est dite essentiellement séparable si l'on peut l'engendrer (modulo les évènements négligeables) par une suite d'évènements, ou de façon équivalente si elle est de la forme $\sigma(\xi)$ pour une certaine variable aléatoire réelle ξ . La séparabilité essentielle d'une tribu \mathcal{B} équivaut à la séparabilité de l'espace de Hilbert $L^2(\Omega, \mathcal{B}, \mathbb{P})$. Par conséquent toute sous-tribu d'une tribu essentiellement séparable est essentiellement séparable.

Les probabilités modernes (martingales, chaînes de Markov, équations différentielles stochastiques...) font intervenir systématiquement (bien que ce soit souvent de manière implicite) des familles croissantes de tribus appelées filtrations. In-

tuitivement, les filtrations représentent une certaine information disponible qui augmente en fonction d'un paramètre que nous appellerons le temps.

Les objets auxquels nous nous intéressons dans cette thèse sont les filtrations à temps discret, et plus précisément les suites croissantes $\mathcal{F} = (\mathcal{F}_n)_n$ de tribus indexées par les entiers naturels, les entiers négatifs ou par les entiers relatifs. En fait, on étudie les filtrations en lien avec la probabilité \mathbb{P} . Bien que ce soit sous-entendu, le véritable objet de l'étude est donc un espace probabilisé filtré

$$(\Omega, \mathcal{A}, \mathbb{P}, \mathcal{F}).$$

Les exemples de filtrations que nous considérons sont tous les filtrations naturelles de certains processus. Si on se donne un processus $X = (X_n)_n$, la filtration naturelle de X est la famille de tribus \mathcal{F}^X définie par

$$\mathcal{F}_n^X = \sigma(\{X_k ; k \leq n\}).$$

En d'autres termes, la tribu \mathcal{F}_n^X représente l'information contenue par la trajectoire de X jusqu'au temps n .

Bien que définir la notion de filtration soit simple, les classer à isomorphisme près est pour l'instant complètement hors de portée.

Une définition naïve d'isomorphisme d'un espace probabilisé filtré $(\Omega, \mathcal{A}, \mathbb{P}, \mathcal{F})$ dans un espace probabilisé filtré $(\Omega', \mathcal{A}', \mathbb{P}', \mathcal{F}')$ est une application bimesurable $\Phi : \Omega \rightarrow \Omega'$ telle que $\mathcal{F}_n = \Phi^{-1}(\mathcal{F}'_n)$ pour tout n , et $\mathbb{P}'(A') = \mathbb{P}(\Phi^{-1}(A'))$ pour tout $A' \in \mathcal{A}'$.

Cette définition est en fait trop restrictive, pour au moins deux raisons :

- il faut travailler modulo les évènements négligeables,
- les tribus \mathcal{A} et \mathcal{A}' peuvent ne pas contenir les singletons.

Émery et Schachermayer utilisent dans [5] une définition plus souple : un isomorphisme d'un espace probabilisé filtré $(\Omega, \mathcal{A}, \mathbb{P}, (\mathcal{F}_n)_{n \in \mathbb{Z}})$ dans un espace probabilisé filtré $(\Omega', \mathcal{A}', \mathbb{P}', (\mathcal{F}'_n)_{n \in \mathbb{Z}})$ est une application linéaire de l'espace vectoriel $\mathbf{L}^0(\Omega, \mathcal{F}_\infty, \mathbb{P})$ des variables aléatoires réelles sur $(\Omega, \mathcal{F}_\infty, \mathbb{P})$ dans $\mathbf{L}^0(\Omega', \mathcal{F}'_\infty, \mathbb{P}')$ qui préserve les lois des variables aléatoires, qui commute avec les opérations boréliennes et qui envoie \mathcal{F} sur \mathcal{F}' .

Exemple 1 (Espaces probabilisés filtrés isomorphes) *Les espaces suivants sont isomorphes :*

- $\{0, 1\}^\infty$ muni de la tribu produit, de la probabilité $\bigotimes_{n \geq 1} \frac{1}{2}(\delta_0 + \delta_1)$ et de la filtration engendrée par les applications coordonnées $p_n : x \in \{0, 1\}^\infty \mapsto x_n$.
- $[0, 1]$ muni de la tribu borélienne, de la mesure de Lebesgue et de la filtration

dyadique : pour $n \geq 0$, \mathcal{F}_n est engendrée par les intervalles $[k2^{-n}, (k+1)2^{-n}[$ pour $0 \leq k \leq 2^n - 1$.

- \mathbb{R} muni de la tribu des boréliens 1-périodiques, de la loi uniforme sur $[0, 1[$ vue comme probabilité sur \mathbb{R} et de la filtration engendrée par les applications $(f_n)_{n \geq 1}$ de \mathbb{R} dans $\{0, 1\}$ définies par $f_n(x) = 1$ si la partie entière de $2^n x$ est paire, et $f_n(x) = 0$ sinon.
- \mathbb{Z}^∞ muni de la tribu produit, de la loi de la marche aléatoire simple symétrique sur \mathbb{Z} et de la filtration naturelle de cette marche.

À partir de maintenant, nous nous intéressons essentiellement aux filtrations indexées par les entiers négatifs ou par les entiers relatifs. Nous allons définir trois grandes familles de filtrations :

- les filtrations de type produit,
- les filtrations admettant une paramétrisation génératrice,
- les filtrations standard.

Ces familles sont intéressantes pour des filtrations indexées par les entiers relatifs ou les entiers négatifs puisque les difficultés apparaissent du côté de $-\infty$. Un des objectifs de cette thèse est de déterminer à quelle(s) famille(s) appartiennent certains exemples de filtrations.

1.2 Filtrations de type produit

Les filtrations les plus simples sont les filtrations de type produit.

Définition 1 (Filtration de type produit) Une filtration $\mathcal{F} = (\mathcal{F}_n)_n$ est de type produit s'il existe une famille $I = (I_n)_n$ de variables aléatoires indépendantes telles que $\mathcal{F} = \mathcal{F}^I$.

On peut montrer facilement que si \mathcal{F} est de type produit alors toute filtration isomorphe à \mathcal{F} est aussi de type produit, cette notion est donc un invariant pour les filtrations. La notion de filtration de type produit est à première vue très simple, elle réserve cependant des surprises. Pour une filtration, une propriété plus faible que d'être de type produit est de posséder une suite d'innovations.

Définition 2 (Innovations) Une suite de variables $I = (I_n)_n$ sera appelée suite d'innovations (ou système d'innovations) pour \mathcal{F} si pour tout n , la variable I_n est indépendante de \mathcal{F}_{n-1} et $\mathcal{F}_n = \mathcal{F}_{n-1} \vee \sigma(I_n)$.

La définition ci-dessus impose que les variables aléatoires I_n sont indépendantes entre elles.

Définition 3 (Filtration à accroissements indépendants) *Une filtration est dite à accroissement indépendants si elle admet un système d'innovations.*

Cette notion a été introduite en ces termes par Émery et Schachermayer [5], elle est aussi utilisée par Laurent, mais il appelle ces filtrations des « filtrations de type produit local ».

Parmi les filtrations à accroissements indépendants, les filtrations (r_n) -adiques constituent une classe intéressante.

Définition 4 (Filtrations (r_n) -adiques) *Soit (r_n) une suite d'entiers strictement positifs, une filtration $\mathcal{F} = (\mathcal{F}_n)_n$ est (r_n) -adique si elle admet un système d'innovations $(I_n)_n$ tel que pour tout n , la loi de I_n soit uniforme sur $\{1, \dots, r_n\}$. Une filtration sera dite dyadique si elle est (r_n) -adique avec $r_n = 2$ pour tout n .*

La proposition suivante fait le lien entre les filtrations de type produit et les innovations.

Proposition 1.2.1 *Si \mathcal{F} est une filtration de type produit engendrée par une suite de variables aléatoires indépendantes (I_n) , alors :*

- les I_n forment un système d'innovations pour \mathcal{F} ,
- la tribu $\mathcal{F}_{-\infty} = \bigcap_n \mathcal{F}_n$ est triviale.

On pourrait être tenté de croire que la réciproque est vraie, c'est-à-dire que si $\mathcal{F}_{-\infty}$ est triviale et si \mathcal{F} admet une suite d'innovations $(I_n)_n$, alors $\mathcal{F} = \mathcal{F}^I$. Mais cette implication, utilisée à tort dans [11] par Kallianpur et Wiener est fautive. (Cette erreur de Kallianpur et Wiener est fameuse et l'article n'est pas publié mais il en existe une description dans [18].)

Voici un contre-exemple classique à cette affirmation : soit $X = (X_n)_{n \in \mathbb{Z}}$ une suite de variables aléatoires indépendantes uniformes sur $\{-1, 1\}$. Posons pour tout $n \in \mathbb{Z}$, $I_n = X_n X_{n-1}$. Alors, $\mathcal{F}_{-\infty}^X$ est triviale et $\mathcal{F}_n^X = \mathcal{F}_{n-1}^X \vee \sigma(I_n)$. Cependant $\mathcal{F}^X \neq \mathcal{F}^I$ puisque X_0 est indépendant de \mathcal{F}_0^I .

Voici un exemple plus étonnant de construction d'innovations. Soit μ une probabilité sur le cercle \mathbb{R}/\mathbb{Z} symétrique par rapport à $\dot{0}$. Soit $(X_n)_{n \in \mathbb{Z}}$ une marche aléatoire sur le cercle à pas de loi μ , c'est-à-dire une chaîne de Markov stationnaire telle que pour tout $n \in \mathbb{Z}$,

- la position X_n suit la loi uniforme sur $[0, 1[= \mathbb{R}/\mathbb{Z}$,
- le pas $\xi_n = X_n - X_{n-1}$ est indépendant de \mathcal{F}_{n-1}^X et de loi μ .

Un exemple intéressant est le cas où $\mu = \frac{1}{2}(\delta_a + \delta_{-a})$ avec $a \in \mathbb{R}/\mathbb{Z}$ irrationnel.

On montre facilement que les pas $(\xi_n)_{n \in \mathbb{Z}}$ forment un système d'innovations pour la chaîne X . Cependant ces innovations ne permettent pas de retrouver toute la suite, puisqu'on montre facilement que X_0 est indépendant de \mathcal{F}^ξ . On peut se demander si d'autres innovations permettent de retrouver le processus X .

Vershik [26] mentionne cet exemple (page 744, remarques sur les exemples 2 et 3), et affirme que la filtration \mathcal{F}^X est de type produit si μ ne charge que deux irrationnels. Toutefois, Vershik ne construit pas de suite de variables aléatoires indépendantes qui engendrent \mathcal{F}^X .

Dans [17], Leuridan construit explicitement une telle suite lorsque le support de μ n'est pas contenu dans un sous-groupe fini de \mathbb{R}/\mathbb{Z} . Dans le cadre d'une marche aléatoire à pas de loi symétrique par rapport à 0, la construction de Leuridan donne une bonne idée de ce que l'on peut choisir comme innovations. Les innovations $(\eta_n)_{n \in \mathbb{Z}}$ qu'il utilise pour engendrer la filtration \mathcal{F}^X sont de la forme $\eta_n = H_n \xi_n$ où H_n est une variable aléatoire mesurable pour \mathcal{F}_{n-1}^X et à valeurs dans $\{-1, 1\}$.

Cet exemple illustre une méthode générale donnée entre autres par Laurent dans sa thèse [12, proposition 4.3.7] pour changer d'innovations.

Proposition 1.2.2 *Soient \mathcal{B}, \mathcal{C} deux sous-tribus essentiellement séparables de $(\Omega, \mathcal{A}, \mathbb{P})$. Supposons qu'il existe une variable aléatoire réelle V indépendante de \mathcal{C} telle que $\mathcal{B} = \mathcal{C} \vee \sigma(V)$. Alors les variables aléatoires V' indépendantes de \mathcal{C} telles que $\mathcal{B} = \mathcal{C} \vee \sigma(V')$ sont les variables de la forme*

$$V' = \phi_Y(V) \text{ p.s.}$$

où Y est une variable aléatoire réelle engendrant \mathcal{C} et $(\phi_y)_{y \in \mathbb{R}}$ est une famille de transformations inversibles de $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ dans lui-même telle que pour tout $y \in \mathbb{R}$ $\phi_y(\mathbb{P}_V) = \mathbb{P}_{V'}$.

Les filtrations considérées dans les exemples précédents sont des filtrations de type produit. On peut se demander de manière générale si toute filtration \mathcal{F} admettant une suite d'innovations est de type produit dès que $\mathcal{F}_{-\infty}$ est triviale. La réponse est négative, un contre-exemple simple a été donné par Vinokourov :

Exemple 2 (Vinokourov [27]) *Soit $p \in]0, 1[$, on note $X = (X_n)_{n \in \mathbb{Z}}$ la chaîne de Markov stationnaire à valeurs dans $\{-1, 1\}$ de probabilité de transition :*

$$\begin{aligned} p_{1,1} &= p_{-1,-1} = p \\ p_{1,-1} &= p_{-1,1} = 1 - p. \end{aligned}$$

On vérifie facilement que $\mathcal{F}_{-\infty}^X$ est triviale, et que X admet le système d'innovations $(I_n)_{n \in \mathbb{Z}}$ où $I_n = X_n X_{n-1}$. Mais $\mathcal{F}^X \neq \mathcal{F}^I$ puisque X_0 est indépendante de \mathcal{F}_0^I . Si $p \neq \frac{1}{2}$, on peut montrer à l'aide de la proposition 1.2.2 que les seules innovations de \mathcal{F}^X sont les fonctions injectives des variables I_n , la filtration \mathcal{F}^X n'est donc pas de type produit.

1.3 Filtrations possédant une paramétrisation génératrice

On a vu que dans l'exemple de Vinokourov, la filtration naturelle de $(X_n)_{n \in \mathbb{Z}}$ n'est pas de type produit si $p \neq \frac{1}{2}$, cependant elle vérifie une propriété plus faible que nous allons voir.

Supposons par exemple que $p < \frac{1}{2}$. Alors quitte à grossir l'espace probabilisé, on peut construire une suite $(U_n)_{n \in \mathbb{Z}}$ de variables aléatoires indépendantes à valeurs dans $\{-1, 0, 1\}$, de loi $p\delta_{-1} + p\delta_1 + (1 - 2p)\delta_0$, telles que la suite $(X_n)_{n \in \mathbb{Z}}$ vérifie la relation de récurrence suivante : pour tout n dans \mathbb{Z} ,

- U_n est indépendante de $\mathcal{F}_{n-1}^{X,U}$,
- si $U_n = 0$ alors $X_n = X_{n-1}$,
- si $U_n \neq 0$ alors $X_n = U_n$.

La relation de récurrence montre que $\mathcal{F}_n^X \subset \mathcal{F}_{n-1}^X \vee \sigma(U_n)$.

À la différence des innovations $(I_n)_{n \in \mathbb{Z}}$ considérées précédemment, les variables aléatoires $(U_n)_{n \in \mathbb{Z}}$ permettent de reconstituer la chaîne $(X_n)_{n \in \mathbb{Z}}$. En effet, pour tout $n \in \mathbb{Z}$, il existe presque sûrement un entier $m \leq n$ telle que $U_m \neq 0$. La relation de récurrence montre que $X_n = U_m$ où m est le dernier entier avant n tel que $U_m \neq 0$. Cet exemple est une bonne illustration de la notion de paramétrisation génératrice.

Définition 5 (Paramétrisation) Soit $\mathcal{F} = (\mathcal{F}_n)_n$ une filtration de $(\Omega, \mathcal{A}, \mathbb{P})$. Soit $U = (U_n)_n$ une suite de variables aléatoires indépendantes, nous dirons que U est une paramétrisation de \mathcal{F} si pour tout n ,

- U_n est indépendante de $\mathcal{F}_{n-1} \vee \mathcal{F}_{n-1}^U$,
- $\mathcal{F}_n \subset \mathcal{F}_{n-1} \vee \sigma(U_n)$.

En particulier toute suite d'innovations de \mathcal{F} est une paramétrisation de \mathcal{F} . La définition de paramétrisation que nous utilisons s'inspire de celle utilisée par Schachermayer [20] à un retournement du temps près (Schachermayer demande

en plus que les variables aléatoires U_n suivent la loi uniforme sur $[0, 1]$, mais cette restriction a peu d'importance).

Proposition 1.3.1 *Toute filtration $(\mathcal{F}_n)_{n \leq 0}$ telle que \mathcal{F}_0 est essentiellement séparable possède une paramétrisation.*

La démonstration de cette proposition est faite à la fin de cette thèse (section 5.0.5).

Définition 6 (Inclusion) *Soient $\mathcal{F} = (\mathcal{F}_n)_n$ et $\mathcal{G} = (\mathcal{G}_n)_n$ deux filtrations. On dit que \mathcal{F} est incluse dans \mathcal{G} (ce qu'on note $\mathcal{F} \subset \mathcal{G}$) si $\mathcal{F}_n \subset \mathcal{G}_n$ pour tout n .*

Notation 1 *Soient $\mathcal{F} = (\mathcal{F}_n)_n$ et $\mathcal{G} = (\mathcal{G}_n)_n$ deux filtrations. On note $\mathcal{F} \vee \mathcal{G}$ la filtration définie par $(\mathcal{F} \vee \mathcal{G})_n = \mathcal{F}_n \vee \mathcal{G}_n = \sigma(\mathcal{F}_n \cup \mathcal{G}_n)$ pour tout n .*

Définition 7 (Paramétrisation génératrice) *Soient $\mathcal{F} = (\mathcal{F}_n)_n$ une filtration de $(\Omega, \mathcal{A}, \mathbb{P})$ et $U = (U_n)_n$ une paramétrisation de \mathcal{F} . Si de plus $\mathcal{F} \subset \mathcal{F}^U$, on dit que U est une paramétrisation génératrice de \mathcal{F} .*

Autrement dit, une paramétrisation $(U_n)_n$ est génératrice si et seulement si à tout instant n , l'information contenue dans \mathcal{F}_n peut être reconstituée à partir de $(U_k)_{k \leq n}$. Dans le cas où \mathcal{F} est la filtration naturelle d'un processus $(X_n)_n$, cela revient à trouver des fonctions g_n telles que $X_n = g_n((U_k)_{k \leq n})$.

Lorsque U est une paramétrisation génératrice de \mathcal{F} , les filtrations \mathcal{F} et \mathcal{F}^U sont non seulement emboîtées, mais elles vérifient une propriété plus forte que l'inclusion appelée immersion (voir proposition 1.3.3 plus loin).

Définition 8 (Immersion) *La filtration \mathcal{F} est immergée dans la filtration \mathcal{G} , ce que nous notons désormais $\mathcal{F} \hookrightarrow \mathcal{G}$, si*

- $\mathcal{F} \subset \mathcal{G}$,
- pour tout n , \mathcal{F}_n est indépendante de \mathcal{G}_{n-1} conditionnellement à \mathcal{F}_{n-1} .

Intuitivement, la deuxième partie de l'énoncé signifie que \mathcal{G}_{n-1} ne donne pas plus d'information sur \mathcal{F}_n que \mathcal{F}_{n-1} . En particulier une filtration $(\mathcal{F}_n)_{n \in \mathbb{Z}}$ n'est pas immergée dans la filtration $(\mathcal{F}_{n+1})_{n \in \mathbb{Z}}$ à moins d'être constante. De façon équivalente, \mathcal{F} est immergée dans \mathcal{G} si et seulement si toute \mathcal{F} -martingale est une \mathcal{G} -martingale.

L'exemple le plus simple d'immersion non triviale est obtenu par grossissement indépendant.

Définition 9 On dit que \mathcal{G} est un grossissement indépendant de \mathcal{F} s'il existe une filtration \mathcal{H} indépendante de \mathcal{F} et telle que $\mathcal{G} = \mathcal{F} \vee \mathcal{H}$.

Proposition 1.3.2 Si \mathcal{G} est un grossissement indépendant de \mathcal{F} , alors \mathcal{F} est immergée dans \mathcal{G} .

La propriété suivante, démontrée dans la section 5.0.6, donne d'autres exemples d'immersions.

Proposition 1.3.3 Si \mathcal{F} admet une paramétrisation génératrice U , alors \mathcal{F} est immergée dans \mathcal{F}^U .

Reciproquement toute filtration immergée dans une filtration de type produit admet-elle une paramétrisation génératrice ? Cette question reste ouverte même pour les filtrations essentiellement séparables.

D'après la proposition 1.3.1, si $\mathcal{F} = (\mathcal{F}_n)_n$ est immergée dans la filtration \mathcal{F}^V avec $(V_n)_n$ des variables aléatoires réelles indépendantes, on peut construire une paramétrisation $(U_n)_n$ de la filtration \mathcal{F} . On a alors pour tout n ,

- $\mathcal{F}_n \subset \mathcal{F}_n^V$,
- $\mathcal{F}_n \subset \mathcal{F}_{n-1} \vee \sigma(U_n)$.

Cependant, nous ne savons pas construire en général de suite (U_n) de variables aléatoires réelles indépendantes vérifiant les deux inclusions à la fois $\mathcal{F}_n \subset \mathcal{F}_n^U$ et $\mathcal{F}_n \subset \mathcal{F}_{n-1} \vee \sigma(U_n)$.

1.4 Filtrations standard

Puisque nous nous intéressons à des propriétés invariantes par isomorphisme, il nous faut assouplir légèrement la notion d'immersion.

Définition 10 Une filtration \mathcal{F} est immersible dans une filtration \mathcal{G} s'il existe une filtration \mathcal{F}' isomorphe à \mathcal{F} telle que \mathcal{F}' soit immergée dans \mathcal{G} .

La notion d'immersibilité permet de définir les filtrations standard.

Définition 11 Une filtration est standard si elle est immersible dans une filtration de type produit.

Il est évident d'après la définition de l'immersibilité que la standardité est une propriété invariante par isomorphisme.

La notion de standardité a d'abord été introduite par Vershik [26] pour des suites décroissantes de partitions mesurables, c'est-à-dire pour des temps discrets. Cette notion a ensuite été adaptée et étudiée par Dubins, Feldman, Smorodinsky et Tsirelson [4, 23] dans le cadre des filtrations à temps continu. L'article [5] de Émery et Schachermayer donne un panorama pour les filtrations à temps discret, avec un point de vue probabiliste.

La proposition 1.3.1 possède un corollaire intéressant dans le domaine de la standardité :

Corollaire 1.4.1 *Soit $\mathcal{F} = (\mathcal{F}_n)_{n \leq 0}$ une filtration telle que \mathcal{F}_0 est essentiellement séparable. Pour tout $n_0 \leq 0$, la filtration $(\mathcal{F}_n)_{n \leq n_0}$ est standard si et seulement si toute la filtration \mathcal{F} est standard.*

Ce corollaire, démontré à la section 5.0.7, montre que le caractère standard d'une filtration est une propriété asymptotique, qui ne dépend que du comportement vers $-\infty$ de la filtration.

On dispose donc de trois familles emboîtées de filtrations :

$$\begin{aligned} & \{ \text{Filtrations de type produit} \} \\ & \quad \subset \{ \text{Filtrations admettant une paramétrisation génératrice} \} \\ & \quad \subset \{ \text{Filtrations standard} \} \end{aligned}$$

De nombreux articles étudient des exemples de filtrations, pour déterminer si elles appartiennent ou non à ces familles de filtrations.

Dans certains cas ces notions coïncident. Le critère de Vershik [26] (théorème 3.2, équivalence entre les points 2 et 3) ou le théorème 15.2.4 dans [12], montrent que pour les filtrations (r_n) -adiques le fait d'être de type produit est équivalent au fait d'être standard. Ce fait reste vrai plus généralement pour les filtrations homogènes (c'est-à-dire possédant un système d'innovations de lois uniformes sur $[0, 1]$ ou sur des ensembles finis).

On a envie de dire que les filtrations standard sont relativement simples puisqu'on peut les immerger dans des filtrations de type produit que l'on comprend bien. Mais la limite entre filtrations standard et filtrations non standard reste subtile. Un résultat important de la théorie des filtrations standard est le théorème d'isomorphisme lacunaire, dû à Vershik. Dans sa formulation d'origine ce théorème s'applique aux suites décroissantes de partitions mesurables, la formulation suivante, qui utilise le vocabulaire des filtrations, est proposée par Émery et Schachermayer [5].

Théorème 1 (Isomorphisme lacunaire (Vershik, [26])) *Soit $(\mathcal{F}_n)_{n \leq 0}$ une filtration telle que \mathcal{F}_0 est essentiellement séparable. On suppose de plus que la tribu asymptotique $\mathcal{F}_{-\infty}$ est triviale. Alors il existe une application $\varphi : \mathbb{Z}^- \rightarrow \mathbb{Z}^-$ strictement croissante telle que la filtration extraite $(\mathcal{F}_{\varphi(n)})_{n \leq 0}$ est standard.*

Le théorème précédent donne l'existence d'une telle extractrice φ , cependant il ne donne pas de moyen de la construire. Quand on a une filtration $(\mathcal{F}_n)_{n \leq 0}$ non standard on peut donc se demander pour quels choix de φ la filtration $(\mathcal{F}_{\varphi(n)})_{n \leq 0}$ est standard. Dans [26], Vershik donne un exemple de filtration $(\mathcal{F}_n)_{n \leq 0}$ non standard, telle que $(\mathcal{F}_{2n})_{n \leq 0}$ est standard. Gorbulsky dans [8] donne un autre exemple de filtration qui vérifie cette propriété. Nous verrons dans la première partie de cette thèse que le processus des mots découpés donne un exemple de filtration non standard à la limite de la standardité.

1.5 Critères de standardité

Définir la notion de standardité est relativement simple, mais savoir si une filtration donnée est standard ou non est souvent complexe. En particulier prouver directement qu'une filtration n'est pas standard est très difficile. C'est pourquoi des conditions nécessaires et suffisantes de standardité ont été établies. Nous allons rappeler quelques-uns de ces critères, en commençant par les critères de Vershik, puis en donnant quelques critères de confort.

1.5.1 Critère de Vershik de premier niveau

Tout d'abord nous allons traiter du critère de Vershik «de premier niveau» (suivant la terminologie de Émery et Schachermayer [5]). Dans sa formulation initiale ([26] théorème 3.2, équivalence entre les points 1 et 2), ce critère est énoncé pour des suites décroissantes (r_n) -adiques de partitions mesurables. Ce critère a été étendu au cas séparable par Feldman et Smorodinsky [6]. Nous énonçons ici une formulation plus générale écrite dans un langage probabiliste et proposée par Laurent [12, définition 3.1.3] :

Définition 12 *Sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, soit $\mathcal{F} = (\mathcal{F}_n)_{n \leq 0}$ une filtration à accroissements indépendants telle que \mathcal{F}_0 soit essentiellement séparable. On dit que \mathcal{F} satisfait le critère de Vershik de premier niveau si pour toute variable aléatoire ξ mesurable pour \mathcal{F}_0 et à valeurs dans un ensemble fini F et tout réel $\delta > 0$, il existe un entier $n_0 < 0$, des innovations V_{n_0}, \dots, V_0 de \mathcal{F} aux instants $n_0, \dots, 0$ et une variable aléatoire ζ mesurable pour $\sigma(V_{n_0}, \dots, V_0)$ et à valeurs dans F tels que $\mathbb{P}[\xi \neq \zeta] < \delta$.*

Proposition 1.5.1 *Soit $\mathcal{F} = (\mathcal{F}_n)_{n \leq 0}$ une filtration à accroissements indépendants. Alors \mathcal{F} est de type produit si et seulement si \mathcal{F} vérifie le critère de Vershik de premier niveau.*

1.5.2 Critère de Vershik de second niveau

Nous venons de voir que le critère de Vershik de premier niveau donne une condition nécessaire et suffisante pour qu'une filtration soit de type produit. Le critère de Vershik « de second niveau » donne une condition nécessaire et suffisante de standardité. Encore une fois, la formulation initiale ([26] théorème 3.2, équivalence entre les points 1 et 3) de ce critère est énoncée pour des suites décroissantes (r_n) -adiques de partitions mesurables. Pour de telles suites être de type produit équivaut à être standard, pour cette raison il y a équivalence entre les deux critères dans la théorie de Vershik. Ces critères sont regroupés sous le nom de critère de standardité dans [26]. Nous allons énoncer plus loin la formulation de ce critère proposée par Émery et Schachermayer [5] et écrite dans un langage probabiliste. Le second critère est plus élaboré que le premier et nécessite d'introduire quelques notations préalables.

Définition 13 (Dispersion d'une variable aléatoire) *Soit ξ une variable aléatoire à valeurs dans un espace métrique séparable (E, ρ) . On appelle dispersion de ξ la distance moyenne entre deux copies indépendantes ξ' et ξ'' de ξ :*

$$\text{disp}(\xi) = \mathbb{E}[\rho(\xi', \xi'')].$$

Définition 14 (Espaces de Kantorovich-Rubinstein itérés) *Pour tout espace métrique compact (K, ρ) , on note K' l'ensemble des probabilités sur K et ρ' la distance de Kantorovich-Rubinstein sur K' , définie par*

$$\rho'(\mu, \nu) = \inf \int \rho(x, y) d\Lambda(x, y)$$

où la borne inférieure est prise sur les probabilités Λ sur $K \times K$ de marginales μ et ν . Cette distance ρ' munit K' d'une structure d'espace métrique compact. On définit alors par récurrence les espaces métriques compacts $(K_n, \rho_n)_{n \leq 0}$ par $(K_0, \rho_0) = (K, \rho)$ et par la relation de récurrence $(K_{n-1}, \rho_{n-1}) = (K'_n, \rho'_n)$, pour tout $n \leq 0$.

Définition 15 (Tours de mesures associées à une filtration) *Soit $(\mathcal{F}_n)_{n \leq 0}$ une filtration de $(\Omega, \mathcal{A}, \mathbb{P})$. Soit ξ une variable aléatoire définie sur $(\Omega, \mathcal{A}, \mathbb{P})$, mesurable pour \mathcal{F}_0 et à valeurs dans l'espace métrique compact (K, ρ) . On définit par récurrence les variables $\pi_n^{\mathcal{F}} \xi$ à valeurs dans K_n en posant $\pi_0^{\mathcal{F}} \xi = \xi$ et pour tout $n \leq 0$,*

$$\pi_{n-1}^{\mathcal{F}} \xi = \mathcal{L}(\pi_n^{\mathcal{F}} \xi \mid \mathcal{F}_{n-1}).$$

La variable aléatoire $\pi_{-1}^{\mathcal{F}}\xi$ est donc la loi conditionnelle de ξ sachant \mathcal{F}_{-1} vue comme variable aléatoire à valeurs dans (K_{-1}, ρ_{-1}) , la variable aléatoire $\pi_{-2}^{\mathcal{F}}\xi$ est la loi conditionnelle de $\pi_{-1}^{\mathcal{F}}\xi$ sachant \mathcal{F}_{-2} , et ainsi de suite.

Nous pouvons maintenant donner la version probabiliste du critère de Vershik de second niveau, suivant la terminologie de Émery et Schachermayer [5].

Définition 16 (Critère de Vershik de second niveau) *Si $\mathcal{F} = (\mathcal{F}_n)_{n \leq 0}$ est une filtration sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, on dit que \mathcal{F} vérifie le second critère de Vershik si toute variable \mathcal{F}_0 -mesurable ξ à valeurs dans $[0, 1]$ vérifie*

$$\text{disp}(\pi_n^{\mathcal{F}}\xi) \rightarrow 0 \quad \text{quand } n \rightarrow -\infty.$$

Théorème 2 (Vershik [26]) *Soit $\mathcal{F} = (\mathcal{F}_n)_{n \leq 0}$ une filtration sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. Alors \mathcal{F} est standard si et seulement si \mathcal{F} satisfait le critère de second niveau de Vershik.*

1.5.3 I-confort

Le critère de Vershik de second niveau a été adapté par Tsirelson [23] en un critère de confort pour des filtrations à temps continu. La notion de confort a été ensuite déclinée en plusieurs variantes par Émery et Schachermayer [5] dont le I-confort que nous allons présenter.

Définition 17 *Soit $\mathcal{F} = (\mathcal{F}_n)_{n \leq 0}$ une filtration d'un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. On dit que \mathcal{F} est I-confortable si pour toute variable aléatoire ξ mesurable pour \mathcal{F}_0 et à valeurs dans un ensemble fini et pour tout réel $\delta > 0$, il existe un espace de probabilité $(\bar{\Omega}, \bar{\mathcal{A}}, \bar{\mathbb{P}})$ muni de deux filtrations \mathcal{F}' et \mathcal{F}'' telles que :*

- les filtrations \mathcal{F}' et \mathcal{F}'' sont isomorphes à la filtration \mathcal{F} ;
- les filtrations \mathcal{F}' et \mathcal{F}'' sont immergées dans $\mathcal{F}' \vee \mathcal{F}''$;
- il existe un entier $n_0 < 0$ tel que les tribus \mathcal{F}'_{n_0} et \mathcal{F}''_{n_0} sont indépendantes;
- les copies ξ' et ξ'' de ξ par les isomorphismes de la première condition vérifient $\bar{\mathbb{P}}[\xi' \neq \xi''] < \delta$.

Intuitivement, ces quatre conditions traduisent le fait qu'il est possible de coupler de manière non anticipative deux copies de \mathcal{F} de sorte que des conditions initiales indépendantes dans le passé lointain influent très peu sur la valeur de ξ .

Émery et Schachermayer ont ensuite montré que cette notion de confort est équivalente au second critère de Vershik :

Théorème 3 (Émery et Schachermayer [5]) *Soit $\mathcal{F} = (\mathcal{F}_n)_{n \leq 0}$ une filtration sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. Alors \mathcal{F} est I-confortable si et seulement si elle satisfait le critère de Vershik de second niveau.*

On en déduit immédiatement que le I-confort est une condition nécessaire et suffisante de standardité. Laurent a montré [15] que l'équivalence entre le second critère de Vershik et le I-confort a lieu variable aléatoire par variable aléatoire : les variables aléatoires ξ qui vérifient les conditions de Vershik sont exactement les variables aléatoires ξ qui vérifient les conditions demandées par le I-confort. Plus précisément, Laurent montre que pour toute variable aléatoire \mathcal{F}_0 -mesurable ξ à valeurs dans un espace métrique compact (K, ρ) ,

$$\text{disp}(\pi_n^{\mathcal{F}} \xi) = \inf \overline{\mathbb{E}}[\rho(\xi', \xi'')],$$

où l'infimum est pris sur l'ensemble des variables associées aux copies de \mathcal{F} vérifiant les trois premiers points de la définition précédente.

Les critères de Vershik et les critères de confort permettent d'étudier le caractère standard ou non standard de certaines filtrations. Un exemple intéressant est celui fourni par les processus des mots découpés.

1.6 Processus des mots découpés

Dans la première partie de cette thèse nous étudions les filtrations des processus des mots découpés. Ces processus sont inspirés d'un exemple donné par Vershik [24]. Ils ont été étudiés en langage probabiliste par Smorodinsky [22] dans le cas dyadique, puis par Émery et Schachermayer [5] toujours dans le cas dyadique et par Laurent [12], dans le cas général.

La distribution d'un processus des mots découpés dépend d'un alphabet A de taille $N \geq 2$ et d'une suite d'entiers positifs $(\ell_n)_{n \leq 0}$ tels que $\ell_0 = 1$ et, pour tout $n \leq 0$, le rapport $r_n = \ell_{n-1}/\ell_n$ est un entier ≥ 2 . Le processus $(X_n)_{n \leq 0}$ des mots découpés est indexé par les entiers négatifs. Pour tout $n \leq 0$, la loi de X_n est uniforme sur l'ensemble des mots de longueur ℓ_n sur A . De plus, si l'on coupe le mot X_{n-1} , de longueur ℓ_{n-1} , en r_n sous-mots de longueur ℓ_n , alors le mot X_n est choisi de façon uniforme parmi ces sous-mots, indépendamment de tout le processus jusqu'au temps $n - 1$. Plus précisément, la position du sous-mot X_n dans X_{n-1} est une variable aléatoire V_n indépendante de $\mathcal{F}_{n-1}^{X,V}$ et de loi uniforme dans $\{1, 2, \dots, r_n\}$. Le processus des mots découpés est $(X_n, V_n)_{n \leq 0}$.

La suite $(V_n)_{n \leq 0}$ est clairement un système d'innovations pour la filtration $\mathcal{F}^{X,V}$ qui est donc $(r_n)_{n \leq 0}$ -adique. La question que l'on se pose est de savoir dans quels cas la filtration $\mathcal{F}^{X,V}$ est standard, (ou de type produit, ce qui est équivalent ici puisque la filtration est (r_n) -adique).

De façon surprenante, la standardité de la filtration $\mathcal{F}^{X,V}$ dépend de la suite de longueurs $(\ell_n)_{n \leq 0}$. En effet, dans [12], Laurent pose les deux conditions suivantes :

$$\text{Il existe } \alpha < 1 \text{ tel que } r_n^\alpha \gg N^{\ell_n} \text{ quand } n \rightarrow -\infty. \quad (\nabla_N)$$

$$\text{La série } \sum_n \frac{\ln(r_n)}{\ell_n} \text{ converge.} \quad (\Delta)$$

Les conditions ∇_N et Δ sont incompatibles. Laurent montre que sous la condition ∇_N , la filtration du processus des mots découpés est de type produit, en construisant explicitement une suite d'innovations $(U_n)_{n \leq 0}$ qui engendre $\mathcal{F}^{X,V}$. Il montre ensuite que sous la condition Δ , cette filtration n'est pas standard. Laurent indique que les conditions ∇_N sont plus faibles que la condition

$$\frac{\ln(r_n)}{\ell_n} \rightarrow \infty \text{ quand } n \rightarrow -\infty, \quad (\nabla)$$

qui ne dépend pas de N .

Il remarque que les conditions ∇ et Δ avaient déjà été introduites par Vershik [26, paragraphe 1.3] pour étudier l'entropie de certaines suites décroissantes (r_n) -adiques de partitions mesurables. Vershik [26, théorème 4.4] a montré que l'entropie est toujours nulle sous la condition ∇ . Sous la condition Δ , il a construit (toujours dans [26, théorème 4.2]) un exemple où l'entropie est strictement positive, ce qui entraîne le caractère non standard.

Puisque les conditions ∇ et Δ ne couvrent pas tous les cas, Vershik et Laurent ont tous deux posé la question de ce qui se passe entre ∇ et Δ . Nous allons répondre à la question de Laurent : la condition Δ est en fait nécessaire et suffisante pour que la filtration du processus des mots découpés soit non standard. Ce qui se résume par le théorème suivant.

Théorème 4 *La filtration naturelle de $(X_n, V_n)_{n \leq 0}$ est standard si et seulement si la condition $\neg \Delta$ est vérifiée, c'est-à-dire si et seulement si la série $\sum_n \frac{\ln(r_n)}{\ell_n}$ diverge.*

Après l'écriture de cette thèse, Anatoly Vershik a attiré notre attention sur l'article [9], où Heicklen obtient un résultat équivalent au Théorème 4. La filtration (r_n) -adique étudiée par Heicklen a été introduite par Vershik [24] comme suit. (La seule différence entre ce que nous allons présenter et la description de Vershik et Heicklen est qu'ils indexent les suites par les entiers strictement positifs alors que nous allons les indexer par les entiers négatifs.)

Soit A un alphabet fini et $(G_n)_{n \leq 0}$ la suite décroissante de groupes définie par

$$G_n = \sum_{k=n+1}^0 \mathbb{Z}/r_k \mathbb{Z}.$$

Notons G l'union sur $n \leq 0$ des groupes G_n . Pour chaque $n \leq 0$, le groupe G_n agit sur A^G par translation à gauche. Autrement dit, pour tout $g \in G_n$ et $f \in A^G$, on définit $g \cdot f \in A^G$ par $(g \cdot f)(x) = f(xg)$ pour tout $x \in G$. Pour $f \in A^G$, notons $\text{Orb}_{G_n}(f)$ l'orbite de f sous l'action de G_n , c'est-à-dire,

$$\text{Orb}_{G_n}(f) = \{g \cdot f; g \in G_n\}.$$

Soit $F = (F(x))_{x \in G} \in A^G$ une fonction aléatoire dont les valeurs $F(x)$ sont indépendantes et uniformément distribuées sur A . La filtration $(\mathcal{F}_n)_{n \leq 0}$ étudiée par Vershik and Hecklen et la filtration naturelle du processus $(O_n)_{n \leq 0}$, défini par

$$O_n = \text{Orb}_{G_n}(F).$$

Pour tout $n \leq 0$, O_{n-1} est la réunion de r_n orbites sous l'action de G_n , une orbite pour chaque élément de G_{n-1}/G_n . Presque sûrement, ces orbites sont toutes différentes puisque les fonctions décalées $g \cdot F$ sont différentes. De plus, conditionnellement à $(O_{n-1}, O_{n-2}, \dots)$, la variable aléatoire O_n est uniforme sur ces r_n orbites. Ceci montre que $(\mathcal{F}_n)_{n \leq 0}$ est une filtration (r_n) -adique.

On peut montrer que la tribu asymptotique $\mathcal{F}_{-\infty}$ est triviale et que la filtration du processus des mots découpés est immergée dans $(\mathcal{F}_n)_{n \leq 0}$. De façon informelle, le mot X_n au temps n est donné par les valeurs en e (l'élément nul du groupe G) des éléments de O_n . On déduit X_n de X_{n-1} en séparant l'orbite sous G_{n-1} en r_n orbites sous G_n et en choisissant une de ces orbites au hasard uniformément.

Donc, la standardité de $(\mathcal{F}_n)_{n \leq 0}$ sous la condition $\neg \Delta$ implique que la filtration du processus des mots découpés est standard et, donc qu'elle est de type produit. La preuve de Hecklen repose sur le second critère de Vershik et est écrite dans le langage de la théorie ergodique. Bien que le résultat de Hecklen et notre Théorème 4 soient logiquement équivalents, nous pensons que notre preuve est intéressante puisqu'elle repose sur une méthode constructive, directe et probabiliste.

Grâce à la théorie de Vershik, Laurent [14] a déduit de ce théorème que la filtration du processus des mots découpés sur n'importe quel alphabet séparable est standard sous la condition $\neg \Delta$.

Ce théorème possède une application intéressante en théorie ergodique, comme nous allons le voir. On rappelle que l'entropie est un invariant bien connu associé à un automorphisme d'un espace probabilisé (c'est-à-dire une application bimesurable qui préserve la mesure). Dans [25], Vershik a défini un invariant

plus élaboré, appelé l'échelle. Mais calculer cet invariant est difficile, même dans les cas les plus simples. Cependant Laurent a montré que notre résultat énoncé au théorème 4 donne l'échelle d'un décalage de Bernoulli [15]: plus précisément, l'échelle de cette transformation est l'ensemble des suites $(r_n)_{n \leq 0}$ qui vérifient la condition Δ .

Dans le cas où la suite $(\ell_n)_{n \leq 0}$ vérifie la condition Δ , la filtration $(\mathcal{F}_n^{X,V})_{n \leq 0}$ est non standard, on peut donc se demander quelles filtrations extraites sont standard. Si $B \subset \mathbb{Z}^-$ est infini, la filtration $(\mathcal{F}_n^{X,V})_{n \in B}$ est celle du processus des mots découpés associé à la suite de longueurs $(\ell_n)_{n \in B}$. Elle est donc standard si et seulement si la suite $(\ell_n)_{n \in B}$ vérifie $\neg\Delta$. Ce résultat fournit une filtration non standard très proche d'être standard : l'exemple 2 de la première partie de cette thèse donne une filtration \mathcal{F} non standard, telle que $(\mathcal{F}_n)_{n \in B}$ est standard dès que B est infini et que son complémentaire dans \mathbb{Z}^- est infini.

Cet exemple confirme que la limite entre standardité et non standardité est subtile.

1.7 Processus stationnaires à valeurs dans un ensemble fini

Les filtrations naturelles d'un processus stationnaire à valeurs dans un ensemble fini forment une classe très étudiée de filtrations. Étant donné un processus X stationnaire, indexé par les entiers relatifs \mathbb{Z} et à valeurs dans un ensemble fini A , à quelle condition la filtration naturelle de ce processus est-elle standard ?

Le I-confort fournit une condition nécessaire et suffisante de standardité, cependant cette condition est peu maniable et l'on aimerait bien avoir des conditions plus simples à vérifier. Nous allons établir de telles conditions suffisantes (mais non nécessaires) de standardité.

Comme le I-confort de la filtration \mathcal{F}^X impose au passé lointain d'avoir peu d'influence sur la variable X_0 , nous allons chercher à quantifier l'influence du passé à long terme sur son présent. On note $X_n^{\triangleleft} := (X_k)_{k \leq n}$ pour tout n . Pour quantifier l'influence du passé à long terme on cherche à mesurer à quel point la loi de X_0 sachant X_{-1}^{\triangleleft} dépend du passé $(X_k)_{k \leq n-1}$ lorsque (X_n, \dots, X_{-1}) est connu, en comparant les lois conditionnelles $p(\cdot|x) = \mathcal{L}(X_0|X_{-1}^{\triangleleft} = x)$ pour différentes suites $x = (x_k)_{k \leq -1}$ qui coïncident sur leurs n derniers termes.

Différentes quantités pour mesurer cette influence ont été introduites. Dans [2], X. Bressaud, A. Maass, S. Martinez et J. San Martin montrent qu'un contrôle de cette influence, mesurée par des réels γ_n , est une condition suffisante de standardité pour un processus stationnaire à valeurs dans $\{0, 1\}$. Cependant leur résultat possède des restrictions importantes. Tout d'abord ils se limitent au cas

où A a deux éléments. De plus les écarts utilisés sont une mesure de l'influence maximale du passé sur le présent plutôt qu'une mesure de l'influence moyenne.

Une autre façon de mesurer l'influence est introduite dans [3] par Comets, Fernández et Ferrari. Un contrôle de cette quantité leur permet d'obtenir un résultat plus fort : l'existence d'un algorithme de simulation parfaite. Cependant ils utilisent eux aussi des écarts maximaux.

1.7.1 Une condition sur les écarts moyens

L'article «*Sufficient conditions of standardness for filtrations of stationary processes taking values in a finite space*» qui constitue la deuxième partie de cette thèse donne une condition suffisante de standardité souvent plus faible (et donc meilleure) que celle donnée dans l'article [2]. En effet, la condition du théorème suivant est un contrôle des influences moyennes du passé à long terme sur le présent alors que toutes les autres conditions connues utilisent des influences maximales.

Définition 18 (Influence moyenne) *Pour chaque $n \geq 0$, on note η_n l'influence moyenne à distance n , définie par*

$$\eta_n = \mathbb{E}[\|p(\cdot|Y_{-n:-1}) - p(\cdot|X_{-n-1}^{\triangleleft} Y_{-n:-1})\|_{VT}],$$

où $\|\mu - \nu\|_{VT}$ désigne la distance en variation totale entre deux probabilités μ et ν sur A , Y est une copie indépendante de X , $Y_{-n:-1} = (Y_{-n}, \dots, Y_{-1})$ et $X_{-n-1}^{\triangleleft} Y_{-n:-1}$ est la suite obtenue en concaténant $(X_k)_{k \leq -n-1}$ et (Y_{-n}, \dots, Y_{-1}) .

Théorème 5 *Si*

$$\sum_{n=0}^{+\infty} \eta_n < +\infty. \quad (\mathcal{H}'(\eta))$$

et si la loi $p(\cdot|X_{-1}^{\triangleleft})$ charge presque sûrement tout A (condition d'amorce) alors \mathcal{F}^X est standard.

Ce résultat possède un corollaire intéressant pour les chaînes à mémoire variable.

1.7.2 Chaînes à mémoire variable

Ces chaînes, étudiées par exemple dans [7] et largement utilisées dans les applications (par exemple dans [1] [16]), sont des processus stationnaires X à valeurs dans un alphabet fini A , tels que la loi de X_0 sachant le passé X_{-1}^{\triangleleft} ne dépend

que du passé $X_{-L:-1}$ à distance L , où L est une variable aléatoire à valeurs dans \mathbb{N} et mesurable pour \mathcal{F}_{-1}^X .

D'un façon plus précise, pour $x \in A^{\triangleleft}$, soit

$$\begin{aligned}\ell(x) &= \inf\{n \geq 0; y \mapsto p(\cdot | yx_{-n:-1}) \text{ est constant sur } A^{\triangleleft}\} \\ &= \inf\{n \geq 0; \forall y \in A^{\triangleleft}, p(\cdot | yx_{-n:-1}) = p(\cdot | x)\}.\end{aligned}$$

Par définition le processus stationnaire X est une chaîne de Markov à mémoire variable si $\ell(X_{-1}^{\triangleleft})$ est presque sûrement fini. Le théorème 5 fournit le résultat suivant.

Corollaire 1.7.1 *Si X vérifie la condition d'amorce, c'est-à-dire si presque sûrement la loi de X_0 sachant le passé X_{-1}^{\triangleleft} charge tout l'alphabet A et si $\ell(X_{-1}^{\triangleleft})$ est intégrable, alors la filtration naturelle \mathcal{F}^X est standard.*

Remarquons que la recherche de conditions suffisantes de standardité est intéressante puisque la filtration naturelle d'un processus stationnaire à valeurs dans un ensemble fini n'est pas toujours standard : la marche aléatoire en scène aléatoire fournit un exemple de chaîne à mémoire variable dont la filtration naturelle est non standard.

1.7.3 Marche aléatoire en scène aléatoire

On se donne $X = (X_n)_{n \in \mathbb{Z}}$ et $C = (C_s)_{s \in \mathbb{Z}}$ deux suites indépendantes de variables aléatoires i.i.d. de loi uniforme sur $\{-1, 1\}$ qui sont respectivement la suite des pas de la marche, et la scène aléatoire. On pose

$$\begin{aligned}S_n &= X_1 + \cdots + X_n \text{ si } n \geq 0 \\ S_n &= -X_{n+1} - \cdots - X_0 \text{ si } n < 0.\end{aligned}$$

On a donc $S_{n+1} = S_n + X_{n+1}$ pour tout $n \in \mathbb{Z}$ et $S_0 = 0$.

Notons $Y_n = C_{S_n}$ pour tout $n \in \mathbb{Z}$ la couleur du site visité par la marche aléatoire au temps n . Le processus stationnaire Z défini par $Z_n = (X_n, Y_n)$ pour tout $n \in \mathbb{Z}$ et à valeurs dans $A = \{-1, 1\}^2$ est appelé marche aléatoire en scène aléatoire.

Le processus $((X_{k+n})_{k \leq 0}, (C_{S_n+s})_{s \in \mathbb{Z}})$, dont dérive la marche aléatoire en scène aléatoire, est un exemple simple de processus $[T, T^{-1}]$. Rappelons que si T est le décalage de $\{-1, 1\}^{\mathbb{Z}}$ défini par $T((c_s)_{s \in \mathbb{Z}}) = (c_{s-1})_{s \in \mathbb{Z}}$, la transformation $[T, T^{-1}]$ est l'application de $\{-1, 1\}^{\mathbb{Z}^-} \times \{-1, 1\}^{\mathbb{Z}}$ dans lui-même définie par

$$[T, T^{-1}]((x_k)_{k \leq 0}, (c_s)_{s \in \mathbb{Z}}) = ((x_{k-1})_{k \leq 0}, (c_{s-x_0})_{s \in \mathbb{Z}}).$$

Pour tout $n \in \mathbb{Z}^-$, on a donc

$$((X_{k+n})_{k \leq 0}, (C_{S_n+s})_{s \in \mathbb{Z}}) = [T, T^{-1}]^{|n|}((X_k)_{k \leq 0}, (C_s)_{s \in \mathbb{Z}}).$$

Heicklen et Hoffman [10] ont montré que les filtrations naturelles de ces processus (et donc celle de la marche aléatoire en scène aléatoire) ne sont pas standard, bien que leurs tribus asymptotiques en $-\infty$ soient triviales. Cependant on ne sait pas si la filtration \mathcal{F}^Y du processus des couleurs des sites visités par la marche est standard ou non. L'étude de cette filtration est très délicate. Matzinger [19] montre que la tribu \mathcal{F}_0^Y permet de reconstruire la scène C à translation et symétrie près. Cependant cette information n'est pas exploitable d'un point de vue probabiliste puisque connaître la scène à translation et symétrie près ne fournit que des événements de probabilité 0 ou 1.

1.7.4 Une variante de la marche aléatoire en scène aléatoire

Nous allons voir qu'en modifiant très légèrement le processus nous pouvons obtenir une filtration standard. Soit $(X_n, Y_n)_{n \in \mathbb{Z}}$ une marche aléatoire en scène aléatoire classique. Nous pouvons la modifier de sorte que, à chaque temps n , on change la couleur de la scène au point visité S_n avec probabilité $q \in]0, 1/2[$ indépendamment du passé. Ce nouveau processus Z' (noté Z'' dans le paragraphe 4.4 de la deuxième partie) est appelé marche aléatoire en scène aléatoire renouvelée.

Pour des valeurs très petites de q , les probabilités de transition du processus Z' sont arbitrairement proches de celles de Z , l'influence du passé à long terme sur le présent des deux processus semble donc être à peu près la même. Cependant, nous montrons à la fin de cette thèse que la filtration $\mathcal{F}^{Z'}$ est standard (contrairement à \mathcal{F}^Z), en construisant explicitement une paramétrisation U_n qui génère Z' .

Ce dernier exemple montre bien la limite de nos conditions : la standardité est une propriété très fine, sensible à de très petites variations des probabilités. A contrario, les mesures de l'influence du passé que nous avons introduites font peu de différence entre les processus Z et Z' si q est proche de 0. D'ailleurs, les conditions de contrôle de l'influence du passé que nous donnons ne sont que des conditions suffisantes de standardité.

1.8 Pistes de recherche

Comme nous l'avons annoncé au début de cette introduction, l'étude des filtrations est loin d'être terminée. Voici une liste - non exhaustive - de prolongements possibles des travaux de cette thèse.

- **Existence d'une paramétrisation génératrice pour une filtration standard**

Une question toujours ouverte qui a fait l'objet de plusieurs affirmations hâtives [6] et de plusieurs tentatives de démonstrations [20] est la suivante : toute filtration standard admet-elle une paramétrisation génératrice ? Si la réponse à cette question était positive, ce résultat relierait la notion de standardité à la notion plus simple d'existence de paramétrisation génératrice et fournirait la réponse à certains problèmes de représentation des processus stationnaires [13].

- **Mots découpés avec lettres markoviennes**

Si l'on modifie le processus des mots découpés en choisissant les lettres de chaque mot à l'aide d'un processus de Markov stationnaire, peut-on donner une condition nécessaire et suffisante de standardité ? La réponse à cette question permettrait de déterminer l'échelle d'un décalage de Markov.

- **Processus stationnaire à valeurs dans un alphabet infini**

Nous avons vu que plusieurs conditions suffisantes de standardité ont été établies pour les filtrations naturelles des processus stationnaires à valeurs dans un ensemble fini. L'extension de ces résultats à des alphabets infinis serait une grande avancée dans le domaine. Certains résultats, comme celui de Comets, Fernández et Ferrari [3], semblent bien s'étendre au cas d'un alphabet dénombrable. Obtenir des conditions suffisantes de standardité pour des alphabets quelconques semble nettement plus difficile.

- **Existence d'un processus stationnaire à valeurs dans un ensemble à deux éléments et dont la filtration naturelle n'est pas standard**

La marche aléatoire en scène aléatoire est un processus stationnaire à valeurs dans un ensemble à quatre éléments dont la filtration naturelle n'est pas standard. Une transformation simple fournit un processus à valeurs dans un ensemble à trois éléments dont la filtration n'est pas standard. Cependant on ne connaît pas de processus à valeurs dans un ensemble à deux éléments possédant cette propriété. Le second facteur de la marche aléatoire en scène aléatoire, c'est-à-dire le processus des couleurs observées, semble être un bon candidat. Mais la filtration de ce processus est complexe, et malgré des résultats remarquables, comme celui de Matzinger [19], on ne sait toujours pas si elle est standard ou non.

- **Nature des filtrations extraites de la filtration $[T, T^{-1}]$, lorsque T est un décalage de Bernoulli**

La filtration $\mathcal{F} = (\mathcal{F}_n)_{n \in \mathbb{Z}}$ du processus $[T, T^{-1}]$, est aussi la filtration d'une marche aléatoire en scène aléatoire. D'après le théorème d'isomorphisme lacunaire, pour des parties $Q \subset \mathbb{Z}$ convenables, la filtration extraite $(\mathcal{F}_n)_{n \in Q}$ est standard. Caractériser ces parties Q permettrait de quantifier à quel point la filtration \mathcal{F} n'est pas standard et de mieux connaître la frontière entre filtration standard et filtrations non standard.

Chapter 2

Introduction for experts

2.1 Sigma-fields and filtrations

A filtration of a given probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is a non-decreasing family, indexed by a totally ordered set, of sub- σ -fields of \mathcal{A} . The main objects of study in this thesis are discrete-time filtrations $\mathcal{F} = (\mathcal{F}_n)_n$ indexed by the set \mathbb{Z} of relative integers or by the set \mathbb{Z}^- of nonpositive integers, or, equivalently, filtered probability spaces $(\Omega, \mathcal{A}, \mathbb{P}, \mathcal{F})$.

The most usual filtrations are the natural filtrations of processes: for every process $X = (X_n)_n$ indexed by \mathbb{Z} or by \mathbb{Z}^- , the natural filtration of X is the sequence \mathcal{F}^X defined by $\mathcal{F}_n^X = \sigma((X_k)_{k \leq n})$. In words, the σ -field \mathcal{F}_n^X contains the information of the trajectory of X until time n .

Here is another example coming from ergodic theory: let T be a measure-preserving transformation of a probability space (E, \mathcal{E}, π) , and let ξ be a random variable defined on $(\Omega, \mathcal{A}, \mathbb{P})$, with values in E and with law π . The σ -fields $(\sigma(T^n \circ \xi))_{n \geq 0}$ form a nonincreasing sequence. One gets a filtration indexed by nonpositive integers by reversing the time and setting $\mathcal{F}_n = \sigma(T^{|n|} \circ \xi)$ for $n \leq 0$. Some properties of the transformation T can be read on the filtration $(\mathcal{F}_n)_{n \leq 0}$. For example, T has the property K if the tail σ -field $\mathcal{F}_{-\infty}$ is trivial, that is, if $\mathcal{F}_{-\infty}$ contains only events of probability 0 or 1.

The notion of filtration is easy to define, but the classification up to isomorphism of filtrations is, for now, out of reach.

A naive definition of an isomorphism from a given filtered probability space $(\Omega, \mathcal{A}, \mathbb{P}, \mathcal{F})$ into another filtered probability space $(\Omega', \mathcal{A}', \mathbb{P}', \mathcal{F}')$ is that such an isomorphism should be a bimeasurable application $\Phi : \Omega \rightarrow \Omega'$ such that $\mathcal{F}_n = \Phi^{-1}(\mathcal{F}'_n)$ for every n , and $\mathbb{P}'(A') = \mathbb{P}(\Phi^{-1}(A'))$ for every $A' \in \mathcal{A}'$.

But this definition is too restrictive for at least two reasons: first, one needs to

work up to the negligible events, and second, the σ -fields \mathcal{A} and \mathcal{A}' may not contain singletons.

Émery and Schachermayer gave in [5] a looser definition: an isomorphism of a filtered probability space $(\Omega, \mathcal{A}, \mathbb{P}, \mathcal{F})$ into a filtered probability space $(\Omega', \mathcal{A}', \mathbb{P}', \mathcal{F}')$ is a linear application from the space $\mathbf{L}^0(\Omega, \mathcal{F}_\infty, \mathbb{P})$ of the real random variables on $(\Omega, \mathcal{F}_\infty, \mathbb{P})$ into $\mathbf{L}^0(\Omega', \mathcal{F}'_\infty, \mathbb{P}')$ which preserves the laws of the random variables, commutes with borelian applications, and sends \mathcal{F} on \mathcal{F}' .

We now define three types of filtrations which are invariant by isomorphism: product type filtrations, filtrations with a generating parametrization, and standard filtrations.

From now on, all the σ -fields considered are complete.

2.2 Product type filtrations

The simplest filtrations are product type filtrations.

Definition 1 (Product type filtrations and (r_n) -adic filtrations) *A filtration $\mathcal{F} = (\mathcal{F}_n)_n$ is of product type if there exists a family $I = (I_n)_n$ of independent random variables such that $\mathcal{F} = \mathcal{F}^I$.*

If moreover each I_n is uniform on $\{1, \dots, r_n\}$ for some positive integer r_n , then \mathcal{F} is said to be (r_n) -adic.

One can easily show that if \mathcal{F} is of product type, then any filtration isomorphic to \mathcal{F} is of product type as well.

A weaker property for a filtration is to possess innovations.

Definition 2 (Innovations) *A sequence of random variables $I = (I_n)_n$ is called a system of innovations for \mathcal{F} if for every n , $\mathcal{F}_n = \mathcal{F}_{n-1} \vee \sigma(I_n)$ and I_n is independent of \mathcal{F}_{n-1} .*

This definition forces the random variables I_n to be independent.

Definition 3 (Filtrations with independent increments) *One says that a filtration \mathcal{F} has independent increments if it possesses a system of innovations.*

This notion was introduced in this formulation by Émery and Schachermayer [5] and is also studied by Laurent, who uses the terminology “locally of product type”.

The following proposition links product type filtrations to systems of innovations.

Proposition 2.2.1 *If \mathcal{F} is a product type filtration generated by a sequence of independent random variables (I_n) , then the sequence (I_n) constitutes a system of innovations for \mathcal{F} and the tail σ -field $\mathcal{F}_{-\infty} = \bigcap_n \mathcal{F}_n$ is trivial.*

The converse of this result is false, that is, it may happen that \mathcal{F} contains strictly more information than the filtration \mathcal{F}^I generated by the innovations. Moreover, as we shall see, a more surprising situation can occur, of which Vershik gave several examples: there exists some filtrations with independent increments and trivial tail σ -field which cannot be generated by *any* independent sequence.

Since many filtrations do not possess a sequence of innovations, we now introduce a looser notion, called parametrization.

2.3 Filtrations with a generating parametrization

Definition 4 (Parametrizations) *Consider a filtration $\mathcal{F} = (\mathcal{F}_n)_n$ and a sequence $U = (U_n)_n$ of random variables. Then U is a parametrization of \mathcal{F} if for every n ,*

- U_n is independent of $\mathcal{F}_{n-1} \vee \mathcal{F}_{n-1}^U$,
- $\mathcal{F}_n \subset \mathcal{F}_{n-1} \vee \sigma(U_n)$.

If furthermore $\mathcal{F} \subset \mathcal{F}^U$, then U is a generating parametrization of \mathcal{F} .

In particular, any sequence of innovations of \mathcal{F} is a parametrization of \mathcal{F} . To state our next result, recall that a σ -algebra is called essentially separable if it is countably generated.

Proposition 2.3.1 *Any filtration $(\mathcal{F}_n)_{n \leq 0}$ such that \mathcal{F}_0 is essentially separable possesses a parametrization.*

Definition 5 (Inclusions and immersions) *Let $\mathcal{F} = (\mathcal{F}_n)_n$ and $\mathcal{G} = (\mathcal{G}_n)_n$ be two filtrations. One says that \mathcal{F} is included into \mathcal{G} , and one writes $\mathcal{F} \subset \mathcal{G}$, if $\mathcal{F}_n \subset \mathcal{G}_n$ for every n . One says that \mathcal{F} is immersed into \mathcal{G} , and one writes $\mathcal{F} \hookrightarrow \mathcal{G}$, if $\mathcal{F} \subset \mathcal{G}$ and if furthermore, for every n , \mathcal{F}_n is independent of \mathcal{G}_{n-1} conditionally on \mathcal{F}_{n-1} .*

Intuitively, the independence of \mathcal{F}_n and \mathcal{G}_{n-1} conditionally on \mathcal{F}_{n-1} means that \mathcal{G}_{n-1} gives no more information on \mathcal{F}_n than \mathcal{F}_{n-1} does. In particular a filtration

$(\mathcal{F}_n)_{n \in \mathbb{Z}}$ is not immersed into the filtration $(\mathcal{F}_{n+1})_{n \in \mathbb{Z}}$ unless it is a constant filtration.

The simplest non-trivial example of an immersion is obtained by independent enlargement.

Notation 1 Let $\mathcal{F} = (\mathcal{F}_n)_n$ and $\mathcal{G} = (\mathcal{G}_n)_n$ be two filtrations. Denote by $\mathcal{F} \vee \mathcal{G}$ the filtration defined by $(\mathcal{F} \vee \mathcal{G})_n = \mathcal{F}_n \vee \mathcal{G}_n = \sigma(\mathcal{F}_n \cup \mathcal{G}_n)$ for every n .

Proposition 2.3.2 If \mathcal{F} and \mathcal{H} are two independent filtrations, then \mathcal{F} is immersed into $\mathcal{F} \vee \mathcal{H}$.

The following proposition gives other examples of immersions.

Proposition 2.3.3 If \mathcal{F} possesses a generating parametrization U , then \mathcal{F} is immersed into \mathcal{F}^U .

2.4 Standard filtrations

The notion of immersion must be slightly weakened to provide a notion invariant by isomorphism.

Definition 6 A filtration \mathcal{F} is immersible into a filtration \mathcal{G} if there exists a filtration \mathcal{F}' isomorphic to \mathcal{F} such that \mathcal{F}' is immersed into \mathcal{G} .

We can now define the standardness of filtrations.

Definition 7 A filtration is standard if it is immersible into a product type filtration.

The notion of standardness was first introduced by Vershik [26] for decreasing sequences of measurable partitions. It was then adapted to continuous-time filtrations by Tsirelson [23] and studied by Dubins et al. [4] for continuous-time filtrations and by Émery and Schachermayer [5] for both continuous-time and discrete-time filtrations.

Proposition 2.3.1 has an interesting corollary for standard filtrations.

Corollary 2.4.1 (of proposition 2.3.1) . Let $\mathcal{F} = (\mathcal{F}_n)_{n \leq 0}$ be a filtration such that \mathcal{F}_0 is essentially separable. For every $k \leq 0$, the filtration $(\mathcal{F}_n)_{n \leq k}$ is standard if and only if the complete filtration \mathcal{F} is standard.

This corollary shows that the standardness of a filtration is an asymptotical property, which only depends of the behaviour of the filtration towards $-\infty$.

To summarize we defined three families of filtrations, related as follows:

$$\begin{aligned} & \{\text{Product type filtrations}\} \\ & \subset \{\text{Filtrations possessing a generating parametrization}\} \\ & \subset \{\text{Standard filtrations}\} \end{aligned}$$

In some specific contexts some of these notions coincide. Vershik's criterion in [26] (theorem 3.2, equivalency between points 2 and 3) shows that any (r_n) -adic filtration is standard if and only if it is of product type.

The boundary between standard and non standard filtrations is subtle. An important result of the theory of standard filtrations is the following one.

Theorem 1 (Lacunary isomorphism [5]) *Let $(\mathcal{F}_n)_{n \leq 0}$ be a filtration. Assume that \mathcal{F}_0 is essentially separable and that the σ -field $\mathcal{F}_{-\infty}$ is trivial. Then there exists an increasing application $\varphi : \mathbb{Z}^- \rightarrow \mathbb{Z}^-$ such that the extracted filtration $(\mathcal{F}_{\varphi(n)})_{n \leq 0}$ is standard.*

In [26], Vershik gives an example of a non standard filtration $(\mathcal{F}_n)_{n \leq 0}$ such that $(\mathcal{F}_{2n})_{n \leq 0}$ is standard. In [8] Gorbulsky gives another example with this property. We will see that split-word processes yield non standard filtrations which become standard when time is accelerated by omitting infinitely many instants of time.

2.5 Criterion of standardness

Although the notion of standardness is relatively easy to define, to know if a given filtration is standard or not is complex. This is why various necessary and sufficient conditions of standardness were established. We now recall some of them.

2.5.1 Vershik's first level criterion

In its first formulation ([26] theorem 3.2, equivalence between points 1 and 2), Vershik's first level criterion was established for (r_n) -adic decreasing sequences of measurable partitions. Later on it was extended to the separable case by Feldman and Smorodinsky [6]. The probabilistic formulation below is due to Laurent [12, définition 3.1.3]:

Definition 8 Let $\mathcal{F} = (\mathcal{F}_n)_{n \leq 0}$ be a filtration with independent increments such that \mathcal{F}_0 is essentially separable. One says that \mathcal{F} fulfills Vershik's first level criterion if for any random variable ξ measurable for \mathcal{F}_0 with values in a finite set F and any real number $\delta > 0$, there exists an integer $n < 0$, some innovations $(V_k)_{n \leq k \leq 0}$ of \mathcal{F} at times $n, \dots, 0$, and a random variable ξ' measurable for $\sigma(\{V_k; n \leq k \leq 0\})$ taking values in F such that $\mathbb{P}[\xi \neq \xi'] \leq \delta$.

Proposition 2.5.1 Let $\mathcal{F} = (\mathcal{F}_n)_{n \leq 0}$ be a filtration with independent increments. Then \mathcal{F} is of product type if and only if \mathcal{F} fulfills Vershik's first level criterion.

2.5.2 Vershik's second level criterion

Vershik's second level criterion gives a necessary and sufficient condition for a filtration to be standard. Once again, the original formulation ([26] theorem 3.2, equivalence between points 1 and 3) of this criterion was established for (r_n) -adic decreasing sequences of measurable partitions. The formulation below of this criterion is from Émery and Schachermayer [5] and requires some notations that we now introduce.

Definition 9 (Dispersion of a random variable) Let ξ be a random variable on a separable metric space (E, ρ) . The dispersion of ξ is the average distance between two independent copies ξ' and ξ'' of ξ , that is,

$$\text{disp}(\xi) = \mathbb{E}[\rho(\xi', \xi'')].$$

Definition 10 (Iterated Kantorovich-Rubinstein spaces) To any compact metric space (K, ρ) , one associates the set K' of probabilities on K and ρ' the Kantorovich-Rubinstein distance on K' , defined by

$$\rho'(\mu, \nu) = \inf \int \rho(x, y) d\Lambda(x, y),$$

where the infimum above is taken over the set of probabilities Λ on $K \times K$ of marginals μ and ν . Then (K', ρ') is a compact metric space.

One then defines recursively a sequence $(K_n, \rho_n)_{n \leq 0}$ of compact metric spaces by $(K_0, \rho_0) = (K, \rho)$ and $(K_{n-1}, \rho_{n-1}) = (K'_n, \rho'_n)$, for every $n \leq 0$.

Definition 11 (Tower of measures associated to a filtration)

Let $\mathcal{F} = (\mathcal{F}_n)_{n \leq 0}$ be a filtration and ξ a \mathcal{F}_0 -measurable random variable with values in the compact metric space (K, ρ) . Define recursively the variables $\pi_n^{\mathcal{F}} \xi$ with values in K_n by $\pi_0^{\mathcal{F}} \xi = \xi$ and for $n \leq 0$,

$$\pi_{n-1}^{\mathcal{F}} \xi = \mathcal{L}(\pi_n^{\mathcal{F}} \xi \mid \mathcal{F}_{n-1}).$$

Vershik's second level criterion can now be stated, using the terminology of Émery and Schachermayer [5].

Definition 12 (Vershik's second level criterion) *One says that the filtration $\mathcal{F} = (\mathcal{F}_n)_{n \leq 0}$ fulfills Vershik's second level criterion if for any \mathcal{F}_0 -measurable random variable ξ with values in $[0, 1]$,*

$$\text{disp}(\pi_n^{\mathcal{F}} \xi) \rightarrow 0 \quad \text{when } n \rightarrow -\infty.$$

Theorem 2 ([26]) *Standardness is equivalent to Vershik's second level criterion.*

2.5.3 I-cosiness

Vershik's second level criterion was adapted by Tsirelson [23] into a "cosiness" criterion for continuous-time filtrations. Later on, various notions of cosiness were introduced by Émery and Schachermayer [5], including the I-cosiness, that we now present.

Definition 13 (I-cosiness) *Let $\mathcal{F} = (\mathcal{F}_n)_{n \leq 0}$ be a filtration on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. One says that \mathcal{F} is I-cosy if for any \mathcal{F}_0 -measurable random variable ξ with values in a finite set, and any real $\delta > 0$, there exists a probability space $(\bar{\Omega}, \bar{\mathcal{A}}, \bar{\mathbb{P}})$ supplied with two filtrations \mathcal{F}' and \mathcal{F}'' such that:*

- *the filtrations \mathcal{F}' and \mathcal{F}'' are isomorphic to the filtration \mathcal{F} ;*
- *the filtrations \mathcal{F}' and \mathcal{F}'' are immersed into $\mathcal{F}' \vee \mathcal{F}''$;*
- *there exists an integer $n_0 < 0$ such that the σ -fields \mathcal{F}'_{n_0} and \mathcal{F}''_{n_0} are independent;*
- *the copies ξ' and ξ'' of ξ associated to the isomorphisms of the first condition are such that $\bar{\mathbb{P}}[\xi' \neq \xi''] \leq \delta$.*

Émery and Schachermayer showed that this notion of cosiness is equivalent to Vershik's second level criterion.

Theorem 3 (Émery and Schachermayer [5]) *I-cosiness and Vershik's second level criterion are equivalent.*

As a consequence, I-cosiness and standardness are equivalent.

2.6 Split-word processes

The definition of split-word processes was inspired by examples given by Vershik [1970]. These processes were introduced and studied probabilistically by Smorodinsky [22] in the dyadic case and by Laurent [12] in the general case.

The distribution of a split-word process depends on an alphabet A of size $N \geq 2$ and on a sequence of positive integers $(\ell_n)_{n \leq 0}$ such that $\ell_0 = 1$ and, for every $n \leq 0$, the ratio $r_n = \ell_{n-1}/\ell_n$ is an integer $r_n \geq 2$. The sequence $(X_n)_{n \leq 0}$ of split words is indexed by the nonpositive integers. For every $n \leq 0$, the law of X_n is uniform on the set of words of length ℓ_n on A . Moreover, if one splits the word X_{n-1} , whose length is ℓ_{n-1} , into r_n subwords of length ℓ_n , then the word X_n is chosen uniformly among these subwords, independently of everything up to time $n - 1$. More precisely, denote by V_n the location of the subword X_n in X_{n-1} . Then V_n is uniform in $\{1, 2, \dots, r_n\}$. The split-word process is $(X_n, V_n)_{n \leq 0}$.

The sequence $(V_n)_{n \leq 0}$ is clearly a system of innovations for the filtration $\mathcal{F}^{X,V}$ which is therefore $(r_n)_{n \leq 0}$ -adic. One could be tempted to deduce that $\mathcal{F}_n^{X,V} = \mathcal{F}_n^V$ for every $n \leq 0$, but this equality is clearly false since X_0 is independent of $(V_n)_{n \leq 0}$. However, one still can wonder whether the filtration $\mathcal{F}^{X,V}$ is a product type filtration (generated by some other independent sequence) or not.

Surprisingly, the answer depends on the sequence $(\ell_n)_{n \leq 0}$. Indeed, in [12], Laurent considers the following two conditions ∇_N and Δ .

- **Condition ∇_N :** There exists $\alpha < 1$ such that $r_n^\alpha \gg N^{\ell_n}$ when $n \rightarrow -\infty$.
- **Condition Δ :** The series $\sum_n \frac{\ln(r_n)}{\ell_n}$ converges.

Conditions ∇_N and Δ are mutually exclusive. Laurent showed that ∇_N implies that the filtration of the split-word process is of product type, by constructing an explicit independent sequence generating $\mathcal{F}^{X,V}$. He also showed that Δ implies that $\mathcal{F}^{X,V}$ is not of product type and he noted that conditions ∇_N are weaker than the following condition ∇ , which does not depend on N .

- **Condition ∇ :** $\frac{\ln(r_n)}{\ell_n} \rightarrow +\infty$ when $n \rightarrow -\infty$.

Laurent noted that conditions ∇ and Δ were previously introduced by Vershik [26, example 1] in the context of decreasing sequences of measurable partitions.

Since conditions ∇ and Δ do not exhaust all cases, both Vershik and Laurent asked what happens “between ∇ and Δ ”. We answer their question: condition Δ

is in fact necessary and sufficient for the filtration of a split-word process to be non standard.

Theorem 4 *The natural filtration of $(X_n, V_n)_{n \leq 0}$ is of product type if and only if condition Δ fails.*

After the completion of this thesis, Anatoly Vershik drew our attention to the paper [9], where Hecklen obtains a result equivalent to Theorem 4. The (r_n) -adic filtrations studied by Hecklen were introduced by Vershik [24] as follows. (We mention that both Vershik and Hecklen index sequences by the set of positive integers whereas we index them by the set of nonpositive integers, and that this is the only difference between their presentation and ours, given below.)

Let A be a finite alphabet and $(G_n)_{n \leq 0}$ the decreasing sequence of groups defined by

$$G_n = \sum_{k=n+1}^0 \mathbb{Z}/r_k \mathbb{Z}.$$

Let G denote the union over $n \leq 0$ of the groups G_n . For every $n \leq 0$, the group G_n acts on A^G (on the left) by canonical shifts. Namely, for every $g \in G_n$ and $f \in A^G$, one defines $g \cdot f \in A^G$ by $(g \cdot f)(x) = f(xg)$ for every $x \in G$. Let $\text{Orb}_{G_n}(f)$ denote the orbit of a given $f \in A^G$ under the action of G_n , that is, $\text{Orb}_{G_n}(f) = \{g \cdot f; g \in G_n\}$.

Let $F = (F(x))_{x \in G} \in A^G$ be a random function whose coordinates $F(x)$ are independent and uniformly distributed in A . The filtration $(\mathcal{F}_n)_{n \leq 0}$ studied by Vershik and Hecklen is the natural filtration of $(O_n)_{n \leq 0}$, where

$$O_n = \text{Orb}_{G_n}(F).$$

For each $n \leq 0$, O_{n-1} is the union of r_n orbits under the action of G_n , one orbit for each element of G_{n-1}/G_n . Almost surely, these orbits are all different since the shifted functions $g \cdot F$ are different. Furthermore, conditionally on $(O_{n-1}, O_{n-2}, \dots)$, the random variable O_n is uniformly distributed on these r_n orbits. This shows that $(\mathcal{F}_n)_{n \leq 0}$ is an (r_n) -adic filtration.

One can show that the tail σ -field $\mathcal{F}_{-\infty}$ is trivial and that the filtration of the split-words process is immersed in $(\mathcal{F}_n)_{n \leq 0}$. Informally, the word X_n at time n is given by the values at e (the identity of the group G) of the elements of O_n . One gets X_n from X_{n-1} by splitting the orbits under G_{n-1} into r_n orbits under G_n and by choosing one of these orbits uniformly randomly.

Thus, the standardness of $(\mathcal{F}_n)_{n \leq 0}$ when condition Δ fails implies that the natural filtration of the split-words process is standard and, therefore, that it is of product type. Hecklen's proof relies on Vershik's standardness criterion and uses the language of ergodic theory. Although Hecklen's result and our Theorem 4 are

logically equivalent, we believe that our proof is interesting because it relies on a constructive, direct and probabilistic method.

Using Vershik's theory, one can deduce from theorem 4 that the filtration of the split-word process on any separable alphabet (endowed with an arbitrary measure) is standard when condition Δ fails, see [14] for a proof.

This result has interesting applications in ergodic theory, as we now explain. Recall that entropy is a well known invariant associated to an automorphism of a probability space (that is, a bimeasurable application preserving the measure). Vershik [25] defined a much more elaborate invariant, which he named the scale. But computing this invariant is a very difficult task, even in simple cases. However, Laurent showed that our result provides the exact scale of a dyadic transformation [15], more precisely, the scale of this transformation is the set of sequences $(r_n)_{n \leq 0}$ is such that condition Δ holds.

If $\varphi : \mathbb{Z}^- \rightarrow \mathbb{Z}^-$ is increasing, the filtration $(\mathcal{F}_{\varphi(n)}^{X,V})_{n \leq 0}$ is the filtration of the split-word process associated to the sequence of lengths $(\ell_{\varphi(n)})_{n \leq 0}$. Therefore it is standard if and only if the sequence $(\ell_{\varphi(n)})_{n \leq 0}$ is such that condition Δ fails. This provides an example of a non standard filtration in which the transition from the non standard case to the standard case is sharp since $\mathcal{F}_{\phi(n)}$ is standard for every subsequence $\phi : \mathbb{Z}^- \rightarrow \mathbb{Z}^-$ which omits infinitely many (negative) integers.

2.7 Stationary processes with values in a finite set

Natural filtrations of stationary processes with values in a finite set are a well-studied class of filtrations. Given a stationary processes X , indexed by \mathbb{Z} and valued in a finite set, what condition ensures the standardness of the natural filtration of this process?

A necessary and sufficient condition of standardness is I-cosiness, nevertheless this criterion is difficult to handle and simpler conditions would be appreciable. We now provide such sufficient (but not necessary) conditions of standardness.

Since the I-cosiness of the filtration \mathcal{F}^X forces the random variable X_0 to be almost independent of the long-term past given the recent past, we will measure the influence of the long-term past of the process on its present.

For every n in \mathbb{Z} , let $X_n^{\triangleleft} := (X_k)_{k \leq n}$. For $n \geq 0$, we quantify the influence of X_{-n-1}^{\triangleleft} on the conditional law $\mathcal{L}(X_0 | X_{-1}^{\triangleleft})$ by comparing the conditional laws $p(\cdot | x) = \mathcal{L}(X_0 | X_{-1}^{\triangleleft} = x)$ for different sequences $x = (x_k)_{k \leq -1}$ which coincide on their last n terms.

Several quantities were introduced to measure this influence. In [2], Bressaud et al. showed that a control of this influence is a sufficient condition of standardness for a stationary processes with values in $\{0, 1\}$. But their result has some im-

portant restrictions since, first, the process must be valued in a set A with only two elements, and, second, the quantities used are a measure of the maximum influence of the past on the present whereas an average measure would be more flexible.

We give a sufficient condition which is often weaker (and therefore better) than the condition considered in [2], since it is based on a control of the average influence of the long term past on the present.

Definition 14 (Average influences) *For every $n \geq 0$, the average influence at distance n is*

$$\eta_n = \mathbb{E}[\|p(\cdot|Y_{-n:-1}) - p(\cdot|X_{-n-1}^{\triangleleft} Y_{-n:-1})\|_{VT}].$$

Here $\|\mu - \nu\|_{VT}$ is the total variation distance between the probabilities μ and ν on A , Y is an independent copy of X , $Y_{-n:-1} = (Y_{-n}, \dots, Y_{-1})$ and $X_{-n-1}^{\triangleleft} Y_{-n:-1}$ is the sequence obtained by the concatenation of $(X_k)_{k \leq -n-1}$ and (Y_{-n}, \dots, Y_{-1}) .

Theorem 5 *If the series $\sum_n \eta_n$ converges and if the law of $p(\cdot|X_{-1}^{\triangleleft})$ charges almost surely every element of A , then \mathcal{F}^X is standard.*

Chapter 3

Filtrations of split-word processes

M. Smorodinsky and S. Laurent have initiated the study of the filtrations of split-word processes, in the framework of discrete negative time. For these filtrations, we show that Laurent's sufficient condition for non standardness is also necessary, thus yielding a practical standardness criterion. In turn, this criterion enables us to exhibit a non standard filtration which becomes standard when time is accelerated by omitting infinitely many instants of time.

3.1 Introduction

We shall be interested in filtrations, in the setting of discrete, negative time: given a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, a filtration is an increasing family $\mathcal{F} = (\mathcal{F}_n)_{n \leq 0}$ of sub- σ -fields of \mathcal{A} ; observe that the time n ranges over all negative integers. (Equivalently, one could consider decreasing families of σ -fields indexed by positive integers, known as reverse filtrations; but we find it more convenient to let time run forward, with $n + 1$ posterior to n , at the mild cost of dealing with negative instants.) As discovered by Vershik [26], in this framework very subtle phenomena occur in the vicinity of time -1 .

For a simple example, suppose that the σ -field $\cap_n \mathcal{F}_n$ is degenerate and that, for each n , \mathcal{F}_n is generated by \mathcal{F}_{n-1} and by some Bernoulli random variable U_n which is independent of \mathcal{F}_{n-1} and uniformly distributed on the 2-set $\{0, 1\}$. Under these hypotheses, it may happen that \mathcal{F} contains more information than the natural filtration of the Bernoulli process $U = (U_n)_{n \leq 0}$ (this is similar to weak solutions in SDEs); but something more surprising is also possible: that \mathcal{F} is not generated by any Bernoulli process whatsoever. Such filtrations have been called non standard by Vershik, who has given in [26] a necessary and sufficient criterion for standardness, and several examples of non standard filtrations. The rigorous definition of a standard filtration will be recalled later, in section 3.4.

All filtrations considered in this study have an additional property: for each n , \mathcal{F}_n is generated by \mathcal{F}_{n-1} and by some random variable U_n which is independent from \mathcal{F}_{n-1} and uniformly distributed on some finite set with r_n elements. Such a filtration is called (r_n) -adic. For these filtrations, as shown by Vershik [26], standardness turns out to be tantamount to a simpler, much more intuitive property: an (r_n) -adic filtration \mathcal{F} is standard if and only if \mathcal{F} is of product type, that is, \mathcal{F} is the natural filtration of some process $V = (V_n)_{n \leq 0}$ where the V_n are independent random variables (in this case, it is easy to see that the process V can be chosen with the same law as U .) So, at first reading, ‘standard’ can be replaced with ‘of product type’ in this introduction.

When time is accelerated by extracting a subsequence, that is, when $(\mathcal{F}_n)_{n \in \mathbb{Z}^-}$ is replaced with $(\mathcal{F}_n)_{n \in Q}$ where Q is some infinite subset of the time-axis \mathbb{Z}^- , a standard filtration always remains standard, but a non standard one may become standard (or not). Examples of this phenomenon were first studied by Vershik in the framework of ergodic theory, and then, in a probabilistic setting, by Laurent [12]. Lacunary isomorphism theorem [5] states that, from any filtration $(\mathcal{F}_n)_{n \in \mathbb{Z}^-}$ (such that \mathcal{F}_0 is essentially separable), there exists $Q \subset \mathbb{Z}^-$ such that $(\mathcal{F}_n)_{n \in Q}$ is standard.

By varying the parameters in an example initially due to Vershik [26] and later modified (in the dyadic case) by M. Smorodinsky [22], Laurent [12] has described a family of filtrations, the split-word filtrations; he has shown some of them (the fastest ones) to be standard, and some other ones (the slowest ones) to be non standard; but an intermediate class was left undecided. We continue his study, and show that all these intermediate filtrations are in fact standard. This yields an easily verifiable necessary and sufficient condition for a split-word filtration to be standard.

As the family of split-word filtrations is stable by extracting subsequences, this criterion makes it simple to observe on these examples the transition from non-standardness to standardness when time is accelerated. We find that this transition is, in some sense, sharp: in Example 2, we exhibit a non standard filtration \mathcal{F} such that, for every infinite subset Q of \mathbb{Z}^- with infinite complementary, the corresponding extracted filtration $(\mathcal{F}_n)_{n \in Q}$ is standard. This \mathcal{F} is as close to being standard as possible, for, if $Q \subset \mathbb{Z}^-$ is cofinite and if \mathcal{G} is any filtration, the extracted filtration $(\mathcal{G}_n)_{n \in Q}$ clearly has the same asymptotic properties (standardness, product type, etc.) as \mathcal{G} . To our knowledge, in the earlier literature, the best result in this direction was the existence of a non standard filtration \mathcal{F} such that $(\mathcal{F}_{2n})_{n \leq 0}$ is standard (examples are given by Vershik [26], Gorbulsky [8] and Tsirelson in an unpublished paper).

In this paper we study the filtrations of split-word processes. These processes are inspired by examples given by Vershik [26, example 2,3,4] and have been introduced and studied in terms of probability theory by Smorodinsky [22] in the

dyadic case and by Laurent [12] in the general case.

The distribution of a split-word process depends on an alphabet A of size $N \geq 2$ and on a sequence of positive integers $(\ell_n)_{n \leq 0}$ such that $\ell_0 = 1$ and, for every $n \leq 0$, the ratio $r_n = \ell_{n-1}/\ell_n$ is an integer $r_n \geq 2$. The sequence $(X_n)_{n \leq 0}$ of split words is indexed by the nonpositive integers. For every $n \leq 0$, the law of X_n is uniform on the set of words of length ℓ_n on A . Moreover, if one splits the word X_{n-1} , whose length is ℓ_{n-1} , into r_n subwords of length ℓ_n , then the word X_n is chosen uniformly among these subwords, independently of everything up to time $n - 1$. More precisely, denote by V_n the location of the subword X_n in X_{n-1} . Then V_n is uniform in $\{1, 2, \dots, r_n\}$. The split-word process is $(X_n, V_n)_{n \leq 0}$.

Call $\mathcal{F}^{X,V} = (\mathcal{F}_n^{X,V})_{n \leq 0}$ the natural filtration of $(X_n, V_n)_{n \leq 0}$. Clearly, every subsequence $(\mathcal{F}_n^{X,V})_{n \in Q}$ with $Q \subset \mathbb{Z}^-$ is the natural filtration of a split-word process with lengths $(\ell_n/\ell_m)_{n \in Q}$ on the alphabet A^{ℓ_m} , where $m = \max Q$.

The filtration $\mathcal{F}^{X,V}$ is (r_n) -adic since for every $n \leq 0$,

$$\mathcal{F}_n^{X,V} = \mathcal{F}_{n-1}^{X,V} \vee \sigma(V_n) \text{ with } V_n \text{ independent of } \mathcal{F}_{n-1}^{X,V}.$$

Moreover, the tail σ -field $\mathcal{F}_{-\infty}^{X,V}$ is trivial, thanks to proposition 6.2.1 in [12]. Yet, the inclusion $\mathcal{F}_n^V \subset \mathcal{F}_n^{X,V}$ is clearly strict since X_0 is independent of $(V_n)_{n \leq 0}$. However, the filtration $\mathcal{F}^{X,V}$ may still be a product type filtration (generated by some other independent sequence).

3.1.1 Results

Surprisingly, the nature of the filtration $\mathcal{F}^{X,V}$ depends on the sequence $(\ell_n)_{n \leq 0}$:

Theorem 6 *The filtration $\mathcal{F}^{X,V}$ is standard (or equivalently, is of product type) if and only if, the series $\sum_n \frac{\ln(r_n)}{\ell_n}$ diverges.*

The ‘if’ part of the theorem, which is new, will be proved in section 3.3, and the ‘only if’ part in section 3.4.

Note that the convergence of the series $\sum_n \frac{\ln(r_n)}{\ell_n}$ is equivalent to the condition Δ of Laurent [12], who established the ‘only if’ part of theorem 6.

The condition that we call $\neg\Delta$ (the divergence of the series $\sum_n \frac{\ln(r_n)}{\ell_n}$), which implies standardness, improves on Laurent’s sufficient condition of standardness:

(∇_N) There exists $\alpha < 1$ such that $r_n^\alpha \gg N^{\ell_n}$ as $n \rightarrow -\infty$.

Laurent notices that conditions ∇_N are weaker than the condition

$$\frac{\ln(r_n)}{\ell_n} \rightarrow \infty \text{ as } n \rightarrow -\infty, \quad (\nabla)$$

which does not depend on N . Laurent indicates that conditions ∇ and Δ were previously introduced by Vershik [26, example 1], in the context of decreasing sequences of measurable partitions.

Since conditions ∇ and Δ do not exhaust all possible situations, both Vershik and Laurent asked what happens “between ∇ and Δ ”. In theorem 6, we solve Laurent’s question: condition Δ is in fact necessary and sufficient for the filtration of the split-word process to be non standard.

After the completion of this paper, Anatoly Vershik drew our attention to the paper [9], where Heicklen obtains a result equivalent to Theorem 6. The (r_n) -adic filtrations studied by Heicklen were introduced by Vershik [24] as follows. (We mention that both Vershik and Heicklen index sequences by the set of nonnegative integers whereas we index them by the set of nonpositive integers, and that this is the only difference between their presentation and ours, given below.)

Let A be a finite alphabet and $(G_n)_{n \leq 0}$ the decreasing sequence of groups defined by

$$G_n = \sum_{k=n+1}^0 \mathbb{Z}/r_k \mathbb{Z}.$$

Let G denote the union over $n \leq 0$ of the groups G_n . For every $n \leq 0$, the group G_n acts on A^G (on the left) by canonical shifts. Namely, for every $g \in G_n$ and $f \in A^G$, one defines $g \cdot f \in A^G$ by $(g \cdot f)(x) = f(xg)$ for every $x \in G$. Let $\text{Orb}_{G_n}(f)$ denote the orbit of a given $f \in A^G$ under the action of G_n , that is, $\text{Orb}_{G_n}(f) = \{g \cdot f; g \in G_n\}$.

Let $F = (F(x))_{x \in G} \in A^G$ be a random function whose coordinates $F(x)$ are independent and uniformly distributed in A . The filtration $(\mathcal{F}_n)_{n \leq 0}$ studied by Vershik and Heicklen is the natural filtration of $(O_n)_{n \leq 0}$, where

$$O_n = \text{Orb}_{G_n}(F).$$

For each $n \leq 0$, O_{n-1} is the union of r_n orbits under the action of G_n , one orbit for each element of G_{n-1}/G_n . Almost surely, these orbits are all different since the shifted functions $g \cdot F$ are different. Furthermore, conditionally on $(O_{n-1}, O_{n-2}, \dots)$, the random variable O_n is uniformly distributed on these r_n orbits. This shows that $(\mathcal{F}_n)_{n \leq 0}$ is an (r_n) -adic filtration.

One can show that the tail σ -field $\mathcal{F}_{-\infty}$ is trivial and that the filtration of the split-words process is immersed in $(\mathcal{F}_n)_{n \leq 0}$. Informally, the word X_n at time n is given by the values at e (the identity of the group G) of the elements of O_n .

One gets X_n from X_{n-1} by splitting the orbits under G_{n-1} into r_n orbits under G_n and by choosing one of these orbits uniformly randomly.

Thus, the standardness of $(\mathcal{F}_n)_{n \leq 0}$ when condition Δ fails implies that the natural filtration of the split-words process is standard and, therefore, that it is of product type. Heicklen's proof relies on Vershik's standardness criterion and uses the language of ergodic theory. Although Heicklen's result and our Theorem 6 are logically equivalent, we believe that our proof is interesting because it relies on a constructive, direct and probabilistic method.

This result has interesting applications in ergodic theory as we now explain. Recall that entropy is a well known invariant associated to an automorphism of a probability space (that is, a bimeasurable application preserving the measure). Vershik [25] defined a much more elaborate invariant, named the scale. But computing this invariant is a very difficult task, even in simple cases. However, Laurent showed that our result provides the exact scale of a dyadic transformation [15]: more precisely, the scale of this transformation is the set of sequences $(r_n)_{n \leq 0}$ fulfilling condition Δ .

Using Vershik's theory, one can deduce from theorem 6 that the filtration of the split-word process on any separable alphabet (endowed with an arbitrary measure) is standard under condition $\neg\Delta$, see [14] for a proof.

3.1.2 Examples

Condition $\neg\Delta$ forces the length ℓ_n to grow very quickly as n goes to $-\infty$. If a given sequence $(\ell_n)_{n \leq 0}$ is $\neg\Delta$, then every sequence $(\ell'_n)_{n \leq 0}$ such that $(\ln r'_n)/\ell'_n \geq (\ln r_n)/\ell_n$ is $\neg\Delta$ as well. No similar property holds for the sequence $(\ell_n)_{n \leq 0}$ only, nor for the sequence $(r_n)_{n \leq 0}$ only. Indeed, the first example of this section provides a sequence $(r_n)_n$ which is $\neg\Delta$ and such that $(r_n^2)_{n \leq 0}$ is Δ .

Example 1 (A standard split-word filtration) *Let $\ell_0 = 1$ and $\ell_{n-1} = 2^{\ell_n}$ for every $n \leq 0$. That is to say*

$$\ell_n = 2^{2^{\dots^2}} \quad \text{where the figure 2 appears } |n| \text{ times.}$$

Then for every $n < 0$,

$$r_n = \ell_{n-1}/\ell_n = 2^{\ell_n - \ell_{n+1}}.$$

Therefore

$$\log_2(r_n)/\ell_n = \frac{\ell_n - \ell_{n+1}}{\ell_n} \rightarrow 1,$$

which proves that (r_n) is $\neg\Delta$.

Theorem 6 has another interesting consequence which we now explain. Recall that, by the lacunary isomorphism theorem [5], from any filtration $(\mathcal{F}_n)_{n \leq 0}$ (such that \mathcal{F}_0 is essentially separable), one can extract a filtration $(\mathcal{F}_n)_{n \in Q}$ which is standard. In [26], Vershik provides an example where $(\mathcal{F}_n)_{n \leq 0}$ is non standard whereas $(\mathcal{F}_{2n})_{n \leq 0}$ is standard. In [8], Gorbulsky also gives such an example. Theorem 6 provides an example of a non standard filtration (example 2 below) in which the transition from the non standard case to the standard case is very sharp: $(\mathcal{F}_n)_{n \in Q}$ is standard for every infinite subset Q of \mathbb{Z}^- with infinite complementary.

Example 2 (A non standard filtration close to standardness) Set $\ell_0 = 1$ and $\ell_{n-1} = 4^{\sqrt{\ell_n}}$ for every $n \leq 0$. That is to say

$$\ell_n = 4^{2^{\cdot^{\cdot^{\cdot^2}}}} \quad \text{where the figure 2 appears } |n| - 1 \text{ times.}$$

Then the filtration $(\mathcal{F}_n^{(X,V)})_{n \leq 0}$ is not of product type. Yet, if ϕ is a strictly increasing application from $-\mathbb{N}$ into $-\mathbb{N}$ such that $\phi(n) - n \rightarrow -\infty$ as $n \rightarrow -\infty$, then the filtration $(\mathcal{F}_{\phi(n)}^{(X,V)})_{n \leq 0}$ is of product type.

Proof. On the one hand

$$\frac{\log_2 r_n}{\ell_n} \leq \frac{\log_2 \ell_{n-1}}{\ell_n} = \frac{2}{\sqrt{\ell_n}},$$

which is the general term of a convergent series, thus $(\ell_n)_{n \leq 0}$ is Δ . On the other hand the filtration $(\mathcal{F}_{\phi(n)}^{(X,V)})_{n \leq 0}$ is the filtration of a split-word process of length process $(\ell'_n)_{n \leq 0} = (\ell_{\phi(n)})_{n \leq 0}$. The ratios between successive lengths are, for $n \leq 0$,

$$r'_n = \ell'_{n-1} / \ell'_n = \ell_{\phi(n-1)} / \ell_{\phi(n)}.$$

If $\phi(n) - n \rightarrow -\infty$ when $n \rightarrow -\infty$, then $\phi(n-1) \leq \phi(n) - 2$ infinitely often. For these n ,

$$r'_n \geq \frac{\ell_{\phi(n)-2}}{\ell_{\phi(n)}} = \frac{4^{\sqrt{\ell_{\phi(n)-1}}}}{\ell_{\phi(n)}} = \frac{4^{2^{\sqrt{\ell_{\phi(n)}}}}}{\ell_{\phi(n)}}$$

hence

$$\frac{\log_2 r'_n}{\ell'_n} \geq 2 \frac{2^{\sqrt{\ell_{\phi(n)}}}}{\ell_{\phi(n)}} - \frac{\log_2 \ell_{\phi(n)}}{\ell_{\phi(n)}}.$$

This shows that a subsequence of $(\log_2(r'_n) / \ell'_n)_n$ converges to infinity, hence that $(\ell'_n)_n$ is $\neg\Delta$. \square

3.2 Laurent's method and tools

In this section, we introduce the tools used by Laurent to prove that under condition ∇ , the filtration of the split-word process is of product type. Laurent used a canonical coupling to build explicitly a sequence of innovations $(V'_n)_{n \leq 0}$ which generates the process $(X_n, V_n)_{n \leq 0}$.

Definition 15 *If $(\mathcal{F}_n)_{n \leq 0}$ is a filtration, and $(U_n)_{n \leq 0}$ is a sequence of random variables such that for every $n \leq 0$,*

$$\mathcal{F}_n = \mathcal{F}_{n-1} \vee \sigma(U_n) \quad \text{with } U_n \text{ independent of } \mathcal{F}_{n-1},$$

one says that $(U_n)_{n \leq 0}$ is a sequence of innovations for $(\mathcal{F}_n)_{n \leq 0}$.

This method is strengthened in section 3.3, where we consider a partial canonical coupling to improve on condition ∇ .

We remind the reader that sequences are indexed by the nonpositive integers.

3.2.1 Change of innovations

We start with a complete definition of split-word processes.

Definition 16 (Split-word process) *Let $(r_n)_{n \leq 0}$ denote a sequence of integers such that $r_n \geq 2$ for every $n \leq 0$. Set $l_0 = 1$ and, for every $n \leq 0$, $l_{n-1} = r_n l_n$. Let A denote a finite set, called the alphabet, with cardinal $N \geq 2$.*

A split-word process is any process $(X_n, V_n)_{n \leq 0}$ such that, for every $n \leq 0$,

- X_n is uniformly distributed on A^{l_n} ,
- V_n is uniformly distributed on $\{1, \dots, r_n\}$ and independent of the σ -algebra $\mathcal{F}_{n-1}^{(X, V)} = \sigma(X_m, V_m; m \leq n-1)$,
- if the word X_{n-1} (with length $l_{n-1} = l_n r_n$) is partitioned into r_n subwords of length l_n , X_n is the V_n -th among those r_n subwords.

The sequence $(X_n)_{n \leq 0}$ is a inhomogeneous Markov process indexed by the negative integers and generated by the innovations $(V_n)_{n \leq 0}$. The existence of such a process $(X_n, V_n)_{n \leq 0}$ is guaranteed by Kolmogorov's theorem.

To prove that, under some conditions, the filtration of the split-word process is of product type, one has to switch from one set of innovations to another. Lemma 3.2.1 provides a general method to build new innovations.

Lemma 3.2.1 (Change of innovations) For every $n \leq -1$, let $\{\varphi_w^n\}_w$ denote a family of permutations of $\{1, \dots, r_{n+1}\}$, indexed by the elements w of A^{ℓ_n} , and let

$$V'_{n+1} = \varphi_{X_n}^n(V_{n+1}).$$

Then $(V'_n)_{n \leq 0}$ is a sequence of generating innovations for $(X_n)_{n \leq 0}$. This means that, for every negative integer $n \leq -1$, the following properties hold:

- The random variable V'_{n+1} is uniformly distributed on $\{1, \dots, r_{n+1}\}$.
- The random variable V'_{n+1} is independent of $\mathcal{F}_n^{X,V}$ and therefore also of $\mathcal{F}_n^{X,V'}$.
- The random variable X_{n+1} is a measurable function of X_n and V'_{n+1} .

Proof of lemma 3.2.1. For every negative integer n and every v such that $1 \leq v \leq r_{n+1}$, a simple computation proves that

$$\mathbb{P}[V'_{n+1} = v | \mathcal{F}_n^{X,V}] = \mathbb{P}[V_{n+1} = (\varphi_{X_n}^n)^{-1}(v) | \mathcal{F}_n^{X,V}] = 1/r_{n+1}.$$

This shows the first two properties. The third property follows from the fact that X_{n+1} is the k -th subword of X_n , where $k = (\varphi_{X_n}^n)^{-1}(V'_n)$. \square

3.2.2 Canonical word and coupling

To build the innovations which generate the processus $(X_n)_{n \leq 0}$, one can use, and improve on, Laurent's construction under the stronger condition ∇ . This uses the notions of canonical word and canonical coupling, which we recall below.

Definition 17 (Canonical alphabets and canonical words) For every integer $M \geq 2$, the canonical alphabet on M letters is $A_M = \{1, \dots, M\}$. Canonical words on A_M are the words whose i -th letter is congruent to i modulo M . Hence, the letters of A_M appear in order and are repeated periodically.

Canonical words will usually be denoted by the letter c . For example the canonical word of length 11 on A_3 is 12312312312.

Notation 2 (General notations) To simplify the definition of the canonical coupling, one identifies any ordered alphabet B of size $M \geq 2$ with A_M according to the rank of each letter in the alphabet B .

The i -th letter of a word w is denoted by $w(i)$. Let $w = (w(i))_{1 \leq i \leq r}$ denote a word of length r . For every $1 \leq i \leq r$, $H(w, i)$ denotes the number of instances of the letter $w(i)$ among the $(i-1)$ first letters of w :

$$H(w, i) = \sum_{1 \leq j < i} \mathbf{1}_{\{w(j)=w(i)\}}.$$

Definition 18 (Canonical coupling) *Let w denote a word of length r on an ordered alphabet B of size $M \geq 2$. The canonical coupling associated to w is the permutation φ_w of $\{1, \dots, r\}$ defined as follows: for every $i \leq r$,*

$$\varphi_w(i) = w(i) + H(w, i)M \quad \text{if} \quad w(i) + H(w, i)M \leq r.$$

After this process has been applied to every i , one chooses $\varphi_w(j)$ for the integers j such that $w(j) + H(w, j)M > r$, in an increasing way and in order to make φ_w a bijection. (So $\varphi_w(j)$ is the smallest k which does not belong yet to the range of φ_w .)

Later on, we apply the notions of canonical word and canonical coupling to some alphabets A^ℓ with $\ell \geq 1$.

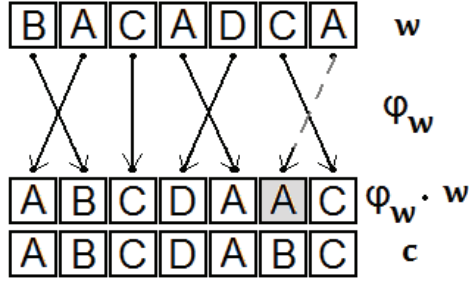


Figure 3.1: Example of canonical coupling

By construction φ_w is one of the permutations φ such that $\varphi \cdot w := w \circ \varphi^{-1}$ is as close as possible to a canonical word. Lemma 3.2.2 makes this statement more precise.

Lemma 3.2.2 (Comparison of w and $c \circ \varphi_w$) *Let c be the canonical word of length $r \geq 1$ on an ordered alphabet B of size $M \geq 2$. Then for every $1 \leq i \leq r$ and $w \in B^r$,*

$$c(\varphi_w(i)) = w(i) \quad \text{if and only if} \quad w(i) + H(w, i)M \leq r.$$

Proof of lemma 3.2.2. By definition of c , the number of instances of j in c is

$$K_j = \max\{k \geq 0 : j + (k - 1)M \leq r\} = \lfloor \frac{r - j}{M} \rfloor + 1 = \lceil \frac{r + 1 - j}{M} \rceil.$$

If $w(i) + H(w, i)M \leq r$, then $\varphi_w(i) = w(i) + H(w, i)M$, thus $c(\varphi_w(i)) = w(i)$ by definition of c .

Otherwise, $w(i) + H(w, i)M > r$, hence the number $H(w, i)$ of instances of $w(i)$ among the $i - 1$ first letters of w is at least the number $K_{w(i)}$ of instances of the letter $w(i)$ in c . According to the first case, φ_w sends the ranks of the $K_{w(i)}$ first instances of $w(i)$ in w on the ranks of the instances of $w(i)$ in c . Since φ_w is bijective, the rank of every instance of the letter $w(i)$ in c has an antecedent by φ_w that is less than i . Therefore the letter in c with rank $\varphi_w(i)$ cannot be $w(i)$, that is, $c(\varphi_w(i)) \neq w(i)$. \square

Lemma 3.2.2 implies lemma 3.2.3 below, which is a slight improvement on Laurent's lemma 6.3.2 of [12].

Lemma 3.2.3 *Let X denote a uniform random word of length r on an ordered alphabet B of size $M \geq 2$, V a uniform random variable on $\{1, \dots, r\}$, independent of X , and c the canonical word of length r on B . Then,*

$$\mathbb{P}[X(V) \neq c(\varphi_X(V))] \leq M/r + 2(M/r)^{1/3}.$$

Proof of lemma 3.2.3. Since $X(V) \leq M$ and $\{X(V) = c(\varphi_X(V))\} = \{X(V) + H(X, V)M \leq r\}$,

$$\{X(V) = c(\varphi_X(V))\} \supset \{H(X, V)M \leq r - M\},$$

hence, for every positive s ,

$$\{X(V) = c(\varphi_X(V))\} \supset \{H(X, V)M \leq V - 1 + Ms\} \cap \{V - 1 + Ms \leq r - M\}.$$

Taking complements, one gets

$$\{X(V) \neq c(\varphi_X(V))\} \subset \{H(X, V)M > V - 1 + Ms\} \cup \{V - 1 + Ms > r - M\},$$

which yields

$$\mathbb{P}[X(V) \neq c(\varphi_X(V))] \leq \mathbb{P}[H(X, V)M > V - 1 + Ms] + \mathbb{P}[V - 1 + Ms > r - M].$$

Since $r + 1 - V$ and V are both uniform on $\{1, \dots, r\}$,

$$\mathbb{P}[V - 1 + Ms > r - M] = \mathbb{P}[V < M(s + 1)] \leq M(s + 1)/r.$$

On the other hand,

$$\mathbb{P}[H(X, V)M > V - 1 + Ms] \leq \mathbb{P}[|H(X, V) - (V - 1)/M| > s].$$

Since V and X are independent, conditionally on V the distribution of $H(X, V)$ is binomial $\text{Bin}(V - 1, 1/M)$. The conditional expectation of $H(X, V)$ is $(V - 1)/M$ and the conditional variance of $H(X, V)$ is

$$(V - 1)(M - 1)/M^2 \leq V/M,$$

hence the Bienaymé-Chebychev inequality yields

$$\mathbb{P}[H(X, V)M > V - 1 + Ms] \leq \mathbb{E}[V]/(Ms^2) = (r + 1)/(2Ms^2) \leq r/(Ms^2).$$

Finally,

$$\mathbb{P}[X(V) \neq c(\varphi_X(V))] \leq M(s + 1)/r + r/(Ms^2).$$

This upper bound for $s = (r/M)^{2/3}$ implies the statement of the lemma. \square

3.2.3 Adapting Laurent's proof

Let us state a slight improvement on Laurent's result [12, proposition 6.3.3]:

Theorem 7 *If $r_n \gg N^{\ell_n}$ as $n \rightarrow -\infty$, then the natural filtration of $(X_n, V_n)_{n \leq 0}$ is of product type.*

Let us introduce notations for the proof. From now on, a word of length ℓ_n on the alphabet A will often be seen as a word of length r_{n+1} on the alphabet $A^{\ell_{n+1}}$.

Proof of theorem 7.

- Choice among sub-words: for every x in $A^{\ell_{n-1}}$ and v in $\{1, \dots, r_n\}$, denote by $f_n(x, v) = x(v)$ the v -th letter of x seen as a word of length r_n on the alphabet A^{ℓ_n} . So $X_n = f_n(X_{n-1}, V_n)$.
- New innovations: let c_{n-1} be the canonical word of length r_n on the alphabet A^{ℓ_n} (for some fixed order on A^{ℓ_n}). Let $\varphi_{X_{n-1}}$ be the canonical coupling associated to X_{n-1} seen as a word of length r_n on A^{ℓ_n} . Set $V'_n = \varphi_{X_{n-1}}(V_n)$.
- Construction of a sequence $(X'_n)_{n \leq 0}$ approximating $(X_n)_{n \leq 0}$: set

$$X'_n = f_n(c_{n-1}, V'_n) = f_n(c_{n-1}, \varphi_{X_{n-1}}(V_n)).$$

The key point is to show that

$$\mathbb{P}[X_n = X'_n] \rightarrow 1 \text{ as } n \rightarrow -\infty$$

by bounding above

$$\mathbb{P}[X_{n-1}(V_n) \neq c_{n-1}(\varphi_{X_{n-1}}(V_n))].$$

Applying lemma 3.2.3 to X_{n-1} seen as a word of length r_n on the alphabet A^{ℓ_n} , one gets that,

$$\mathbb{P}[X_n \neq X'_n] \leq N^{\ell_n}/r_n + 2(N^{\ell_n}/r_n)^{1/3}.$$

Hence $\mathbb{P}[X_n \neq X'_n] \rightarrow 0$ since $N^{\ell_n} \ll r_n$.

Define an application $f'_n : A^{\ell_{n-1}} \times \{1, \dots, r_n\} \rightarrow A^{\ell_n}$ by $f'_n(x, v') = f_n(x, \varphi_x^{-1}(v'))$. Then $f'_n(X_{n-1}, V'_n) = X_n$ and $f'_n(\cdot, V'_n) \circ \dots \circ f'_{m+1}(\cdot, V'_{m+1})(X_m) = X_n$ for $m \leq n \leq 0$. Therefore, under the assumption ∇ , one has,

$$\begin{aligned} \mathbb{P}[X_n \neq f'_n(\cdot, V'_n) \circ \dots \circ f'_{m+1}(\cdot, V'_{m+1})(X'_m)] &\leq \mathbb{P}[X_m \neq X'_m] \\ &\rightarrow 0 \text{ as } m \rightarrow -\infty. \end{aligned}$$

This implies the convergence in probability:

$$\begin{aligned} X_n &= \lim_{m \rightarrow -\infty} f'_n(\cdot, V'_n) \circ f'_{n-1}(\cdot, V'_{n-1}) \circ \dots \circ f'_{m+1}(\cdot, V'_{m+1})(X'_m) \\ &= \lim_{m \rightarrow -\infty} f'_n(\cdot, V'_n) \circ \dots \circ f'_{m+1}(\cdot, V'_{m+1}) \circ f_m(\cdot, V'_m)(c_{m-1}) \end{aligned}$$

and proves that the innovations $(V'_k)_{k \leq n}$ determine the words $(X_k)_{k \leq n}$ and the innovations $(V_k = \varphi_{X_{k-1}}^{-1}(V'_k))_{k \leq n}$. Thus, the filtration $(\mathcal{F}_n^{(X,V)})_{n \leq 0}$ is generated by the innovations $(V'_n)_{n \leq 0}$. \square

3.3 Improving on condition ∇

3.3.1 Statement of the main result

As said before, there is a gap between the conditions ∇ and Δ under which the problem has been solved by Laurent. Our next theorem bridges the gap between the two conditions.

Theorem 8 *If the series $\sum_n \ln(r_n)/\ell_n$ diverges (condition $\neg\Delta$), then the filtration $(\mathcal{F}_n^{(X,V)})_{n \leq 0}$ is of product type.*

Remark 1 *Laurent [12] states condition Δ as the convergence of the series*

$$\sum_n \frac{\ln(r_n!)}{\ell_{n-1}}.$$

The inequalities $\frac{1}{2}r \ln(r) \leq \ln(r!) \leq r \ln(r)$, valid for every $r \geq 2$, ensure that the series

$$\sum_n \frac{\ln(r_n)}{\ell_n} \quad \text{and} \quad \sum_n \frac{\ln(r_n!)}{\ell_{n-1}}$$

both converge or both diverge. Hence Laurent's condition Δ and the condition Δ which we stated in our introduction and used since, are indeed equivalent.

Condition $\neg\Delta$ is easy to express, but less handy to prove things. The equivalent wording below, which is closer to theorem 7, is more convenient.

Proposition 3.3.1 (Rewording of condition $\neg\Delta$) *Condition Δ fails if and only if there exists a sequence $(\alpha_n)_{n \leq 0}$ of nonnegative real numbers and an increasing application $\phi : -\mathbb{N} \rightarrow -\mathbb{N}$, such that the following properties hold:*

1. *For every $n \leq 0$, $r_{\phi(n)} \geq N^{2\alpha_n \ell_{\phi(n)}}$.*
2. *The series $\sum_n \alpha_n$ diverges.*

Furthermore, when they exist, the sequence $(\alpha_n)_{n \leq 0}$ and the application ϕ can be chosen in such a way that the additional properties below hold:

3. When $n \rightarrow -\infty$, $r_{\phi(n)} \gg N^{2\alpha_n \ell_{\phi(n)}}$ (in particular, $r_{\phi(n)} \rightarrow +\infty$).
4. The series $\sum_n \alpha_{2n}$ diverges.
5. For every $n \leq 0$, $0 < \alpha_n \leq 1$.
6. For every $n \leq -2$, the ratio $\alpha_n \ell_{\phi(n)} / \ell_{\phi(n+1)}$ is an integer.

The proof of this result can be found in section 3.3.4.

3.3.2 Construction of the new innovations

The construction of the new innovations uses a *partial canonical coupling*. This tool sharpens the canonical coupling that was introduced in section 3.2.2 and which has to be kept in mind.

Under the hypotheses of theorem 8, the ratios $(r_n)_{n \leq 0}$ are no longer big enough for the innovations associated to the canonical coupling to approach the entire word X_n in only one step. Therefore several steps are necessary to get a good information on the word.

Definition 19 (Partial canonical coupling) *Let w be a word of length ℓr on the alphabet A and $\lambda \in \{1, \dots, \ell\}$ an integer. Denote by \tilde{w} the word extracted from w by splitting w into r sub-words of length ℓ and keeping only the first λ letters of each sub-word. In other words, if $w = (w_1, \dots, w_{\ell r})$, then*

$$\tilde{w} = (w_1, \dots, w_\lambda, w_{\ell+1}, \dots, w_{\ell+\lambda}, \dots, w_{(r-1)\ell+1}, \dots, w_{(r-1)\ell+\lambda})$$

is the word constituted of the letters w_i such that $i = j \bmod \ell$ with $1 \leq j \leq \lambda$.

*Let $\varphi_{\tilde{w}}$ be the canonical coupling of \tilde{w} towards the canonical word \tilde{c} of length r on the alphabet A^λ . Since $\varphi_{\tilde{w}}$ belongs to \mathfrak{S}_r (the symmetric group on r letters), one can apply it to the r sub-words of w of length ℓ . The permutation $\varphi_w^{\lambda/\ell} = \varphi_{\tilde{w}}$ is called the *partial canonical coupling of rate λ/ℓ associated to w .**

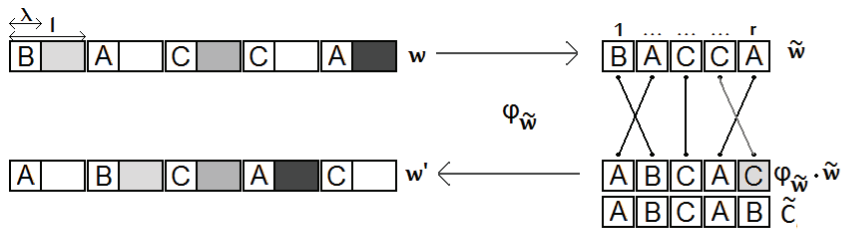


Figure 3.2: Example of a partial canonical coupling. In the proof, this coupling will be considered with $\ell = \ell_{\phi(2n)}$, $\lambda = \alpha_{2n}$, $r = r_{\phi(2n)}$, $w = X_{\phi(2n)-1}$ and $\tilde{w} = \tilde{X}_{\phi(2n)-1}$.

Proof of theorem 8. Assume that $\neg\Delta$ holds. Fix ϕ and $(\alpha_n)_{n \leq 0}$ which fulfill conditions 1,2,3,4,5 and 6 of proposition 3.3.1. We now construct new innovations $(V'_n)_{n \leq 0}$ which generate the same filtration than the process X .

Let us define new innovations $(V'_k)_{k \leq 0}$ as follows. For every $k \leq 0$, define $V'_k = \varphi_{X_{k-1}}^{\alpha_{2n}}(V_k) = \varphi_{\tilde{X}_{k-1}}(V_k)$ if there exists an integer n (necessarily unique) such that $k = \phi(2n)$, and $V'_k = V_k$ otherwise.

Lemma 2.3 will be used to show that with probability close to 1, the first $\alpha_{2n} \ell_{\phi(2n)}$ letters of the words $X_{\phi(2n)}$ and those of $f_{\phi(2n)}(C_{\phi(2n)-1}, V'_{\phi(2n)})$ coincide.

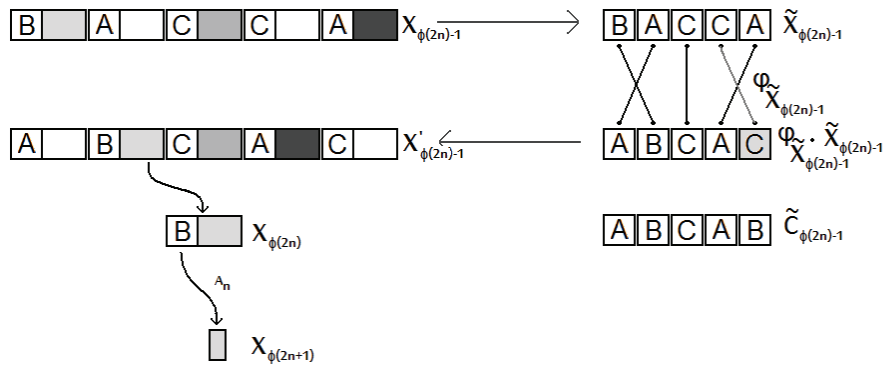


Figure 3.3: Description of the method.

3.3.3 Proof of the main result

This proof is split into three steps.

First step: $X_{\phi(2n+1)}$ comes from the beginning of $X_{\phi(2n)}$ infinitely often

One focuses on the events

$$A_n = \{X_{\phi(2n+1)} \text{ comes from the first } \alpha_{2n} \ell_{\phi(2n)} \text{ letters of } X_{\phi(2n)}\}.$$

One computes $\mathbb{P}[A_n]$ by counting: the number of possible choices for the innova-

tions V_k for $\phi(2n) + 1 \leq k \leq \phi(2n + 1)$ is

$$\prod_{k=\phi(2n)+1}^{\phi(2n+1)} r_k = \ell_{\phi(2n)} / \ell_{\phi(2n+1)}.$$

The number of cases such that A_n occurs is the number of sub-words of length $\ell_{\phi(2n+1)}$ entirely included in the first $\alpha_{2n}\ell_{\phi(2n)}$ letters of $X_{\phi(2n)}$: this number is $\alpha_{2n}\ell_{\phi(2n)} / \ell_{\phi(2n+1)}$ thanks to the additional hypothesis that $\alpha_{2n}\ell_{\phi(2n)} / \ell_{\phi(2n+1)}$ is an integer. Therefore $\mathbb{P}[A_n] = \alpha_{2n}$ and the series $\sum_n \mathbb{P}[A_n]$ diverges.

Moreover A_n is a (deterministic) function of V_k for $\phi(2n) + 1 \leq k \leq \phi(2n + 1)$, hence the events A_n are independent and the Borel-Cantelli lemma ensures that almost surely, A_n occurs for infinitely many n .

Note that $A_n \in \mathcal{F}_{\phi(2n+1)}^{V'}$ for every n since $V'_k = V_k$ for every time k which is not one of the integers $\phi(2n)$.

Second step: Use of lemma 3.2.3

Our purpose is to prove lemma 3.3.2.

Lemma 3.3.2 *For every $n \leq 0$, set $I_{\phi(2n)} = \{1, \dots, \alpha_{2n}\ell_{\phi(2n)}\}$ and fix a word $C_{\phi(2n)-1}$ of length $\ell_{\phi(2n)-1} = r_{\phi(2n)}\ell_{\phi(2n)}$ on A such that $\tilde{C}_{\phi(2n)-1}$ is the canonical word of length $r_{\phi(2n)}$ on the alphabet $A^{\alpha_{2n}\ell_{\phi(2n)}}$. The probability for $X_{\phi(2n)}(I_{\phi(2n)})$ (the first $\alpha_{2n}\ell_{\phi(2n)}$ letters of $X_{\phi(2n)}$) to be the $V'_{\phi(2n)}$ -th sub-word of $\tilde{C}_{\phi(2n)-1}$ converges to 1 as n tends towards $-\infty$. That is to say,*

$$\mathbb{P} [X_{\phi(2n)}(I_{\phi(2n)}) = f_{\phi(2n)}(C_{\phi(2n)-1}, V'_{\phi(2n)})(I_{\phi(2n)})] \rightarrow 1, \text{ as } n \rightarrow -\infty.$$

Proof of lemma 3.3.2. Note that $X_{\phi(2n)}(J_{\phi(2n)})$ is the $V_{\phi(2n)}$ -th letter of $\tilde{X}_{\phi(2n)-1}$ seen as a word of length $r_{\phi(2n)}$ on the alphabet $A^{\alpha_{2n}\ell_{\phi(2n)}}$ (where $\tilde{X}_{\phi(2n)-1}$ is built from $X_{\phi(2n)-1}$ according to definition 19 and the caption of figure 3.3.2). Hence

$$\begin{aligned} \mathbb{P} [X_{\phi(2n)}(I_{\phi(2n)}) = f_{\phi(2n)}(C_{\phi(2n)-1}, V'_{\phi(2n)})(I_{\phi(2n)})] \\ = \mathbb{P} [\tilde{X}_{\phi(2n)-1}(V_{\phi(2n)}) = \tilde{C}_{\phi(2n)-1}(\varphi_{\tilde{X}_{\phi(2n)-1}}(V_{\phi(2n)}))] \end{aligned}$$

Lemma 3.2.3 applied to $\tilde{X}_{\phi(2n)-1}$ seen as a word of length $r_{\phi(2n)}$ on the alphabet $A^{\alpha_{2n}\ell_{\phi(2n)}}$ provides

$$\begin{aligned} \mathbb{P} [\tilde{X}_{\phi(2n)-1}(V_{\phi(2n)}) \neq f_{\phi(2n)}(C_{\phi(2n)-1}, V'_{\phi(2n)})(I_{\phi(2n)})] \\ \leq N^{\alpha_{2n}\ell_{\phi(2n)}} / r_{\phi(2n)} + 2(N^{\alpha_{2n}\ell_{\phi(2n)}} / r_{\phi(2n)})^{1/3}. \end{aligned}$$

From proposition 3.3.1, each term converges to 0, hence lemma 3.3.2 holds.

Third step: Use of the innovations to recover $(X_n)_{n \leq 0}$

Our aim is to show that, for every $m \leq 0$, X_m is a function of the innovations $(V_k)_{k \leq m}$. Consider once again the events

$$A_n = \{X_{\phi(2n+1)} \text{ comes from the first } \alpha_{2n} \ell_{\phi(2n)} \text{ letters of } X_{\phi(2n)}\}.$$

If the event A_n occurs and if

$$X_{\phi(2n)}(I_{\phi(2n)}) = f_{\phi(2n)}(C_{\phi(2n)-1}, V'_{\phi(2n)})(I_{\phi(2n)}),$$

then

$$X_{\phi(2n+1)} = f_{\phi(2n+1)}(\cdot, V'_{\phi(2n+1)}) \circ \dots \circ f_{\phi(2n)}(\cdot, V'_{\phi(2n)})(C_{\phi(2n)-1}).$$

Moreover, since A_n depends only on $(V'_{\phi(2n)+1}, \dots, V'_{\phi(2n+1)})$, it is independent of $\{X_{\phi(2n)}(I_{\phi(2n)}) = f_{\phi(2n)}(C_{\phi(2n)-1}, V'_{\phi(2n)})(I_{\phi(2n)})\}$. Thus

$$\begin{aligned} \mathbb{P}[X_{\phi(2n+1)} = f_{\phi(2n+1)}(\cdot, V'_{\phi(2n+1)}) \circ \dots \circ f_{\phi(2n)}(\cdot, V'_{\phi(2n)})(C_{\phi(2n)-1}) \mid A_n] \\ \geq \mathbb{P}[X_{\phi(2n)}(I_{\phi(2n)}) = f_{\phi(2n)}(C_{\phi(2n)-1}, V'_{\phi(2n)})(I_{\phi(2n)})], \end{aligned}$$

which tends to 1 as n goes to $-\infty$ by lemma 3.3.2.

Some formulas below will be easier to read thanks to the introduction of the function g_n which associates X_{n+1} to (X_n, V'_{n+1}) . Namely, for every integer $n \leq 0$, every word x in A^{ℓ_n} and every integer $1 \leq v \leq r_n$, define

$$\begin{aligned} g_n(x, v) &= f_n(x, \varphi_{\bar{x}}(v)) \text{ if } n \text{ is one of the integers } \phi(2k), \\ &= f_n(x, v) \text{ otherwise.} \end{aligned}$$

Let $X'_{\phi(2n)}$ be the word of length $\ell_{\phi(2n)}$ whose first $\alpha_{2n} \ell_{\phi(2n)}$ letters are those of the word $f_{\phi(2n)}(C_{\phi(2n)-1}, V'_{\phi(2n)})$ and the others are set to 1. Then, by lemma 3.3.2,

$$\mathbb{P}[X_{\phi(2n)}(I_{\phi(2n)}) = X'_{\phi(2n)}(I_{\phi(2n)})] \rightarrow 1 \text{ as } n \rightarrow -\infty.$$

For $n < m \leq 0$, call $X'_{m,n}$ the offspring of $X'_{\phi(2n)}$ at time m , that is,

$$X'_{m,n} = g_{m-1}(\cdot, V'_m) \circ g_{m-2}(\cdot, V'_{m-1}) \circ \dots \circ g_{\phi(2n)+1}(\cdot, V'_{\phi(2n)})(X'_{\phi(2n)-1}).$$

Then,

$$\mathbb{P}[X_m \neq X'_{m,n} \mid A_n] \leq \mathbb{P}[X_{\phi(2n)}(I_{\phi(2n)}) \neq X'_{\phi(2n)}(I_{\phi(2n)}) \mid A_n],$$

hence $\mathbb{P}[X_m \neq X'_{m,n} \mid A_n] \rightarrow 0$, as n goes to $-\infty$.

Lemma 3.3.3 below, which will be proved at the end of section 3.3.4, enables us to use Borel-Cantelli's lemma twice.

Lemma 3.3.3 *Let $(a_n)_{n \geq 0}$ and $(b_n)_{n \geq 0}$ denote two bounded sequences of non-negative real numbers such that the series $\sum_n b_n$ diverges and such that $a_n \ll b_n$. Then there exists an increasing application $\theta : \mathbb{N} \rightarrow \mathbb{N}$ such that the series $\sum_n a_{\theta(n)}$ converges and the series $\sum_n b_{\theta(n)}$ diverges.*

Continuation of the proof of theorem 8

Since the series $\sum_n \mathbb{P}(A_n)$ diverges and, for every fixed $m \leq 0$,

$$\mathbb{P}[X_m \neq X'_{m,n} | A_n] \rightarrow 0, \text{ when } n \rightarrow -\infty,$$

our last lemma applied to the sequences $(\mathbb{P}[A_n])_{n \leq m}$ and $(\mathbb{P}[X_m \neq X'_{m,n} | A_n])_{n \leq m}$ provides a deterministic increasing application $\theta : -\mathbb{N} \rightarrow -\mathbb{N}$ such that

$$\sum_{n \leq m} \mathbb{P}[X_m \neq X'_{m,\theta(n)} | A_{\theta(n)}] < \infty \text{ and } \sum_{n \leq m} \mathbb{P}[A_{\theta(n)}] = \infty.$$

By Borel-Cantelli's lemma, the events $\{X_m \neq X'_{m,\theta(n)}\} \cap A_{\theta(n)}$ occur only for a finite number of times n , whereas the independent events $A_{\theta(n)}$ occur infinitely often. Thus for every word x in A^m , almost surely,

$$\{X_m = x\} = \limsup_{n \rightarrow -\infty} A_{\theta(n)} \cap \{X'_{m,\theta(n)} = x\}.$$

This proves that $\{X_m = x\}$ belongs to $\mathcal{F}_m^{V'}$, hence X_m is a function of the innovations $(V'_m)_{m \leq 0}$. \square

3.3.4 Proof of some auxiliary facts

Proof of proposition 3.3.1. Assume that $(\alpha_n)_{n \leq 0}$ and ϕ exist such that 1 and 2 hold. Then, for every $n \leq 0$, $r_{\phi(n)} \geq N^{2\alpha_n \ell_{\phi(n)}}$ hence $\log_N r_{\phi(n)} \geq 2\alpha_n \ell_{\phi(n)}$. Therefore

$$\sum_n \frac{\log_N r_n}{\ell_n} \geq \sum_n \frac{\log_N r_{\phi(n)}}{\ell_{\phi(n)}} \geq 2 \sum_n \alpha_n,$$

and the last series diverges hence condition $\neg\Delta$ holds.

Conversely, assume that condition $\neg\Delta$ holds. Let $\beta_n = \frac{1}{4} \log_N(r_n) / \ell_n$. Then, $r_n = N^{4\beta_n \ell_n}$ hence $r_n / N^{2\beta_n \ell_n} = N^{2\beta_n \ell_n}$.

Since $\sum_{n \leq 0} \beta_n$ diverges and $\sum_{n \leq 0} |n|2^n$ converges, there exists an increasing application ϕ such that the series $\sum_n \beta_{\phi(n)}$ diverges and such that $\beta_{\phi(n)} \geq 2^{\phi(n)} |\phi(n)|$ for every $n \leq 0$.

Replacing, if necessary, ϕ by φ given by $\varphi(n) = \phi(n-1)$, one can ensure that the series $\sum_n \beta_{\phi(2n)}$ diverges as well.

Since $r_n \geq 2$ for every n , $\ell_{\phi(n)} \geq 2^{-\phi(n)}$, hence $\beta_{\phi(n)} \geq 2^{\phi(n)}|\phi(n)|$ implies that $\beta_{\phi(n)}\ell_{\phi(n)} \geq |\phi(n)| \geq |n|$. Hence,

$$r_{\phi(n)}/N^{2\beta_{\phi(n)}\ell_{\phi(n)}} = N^{2\beta_{\phi(n)}\ell_{\phi(n)}} \geq N^{2|n|} \geq 1,$$

and this sequence converges to $+\infty$. Furthermore, defining $\alpha_n = \min(\beta_{\phi(n)}, 1)$, one sees that $(\alpha_n)_n$ fulfills conditions 1-2-4-5 (condition 3 being a consequence of condition 4).

We now show how to build from $(\alpha_n)_n$ a sequence $(\alpha'_n)_n$ such that condition 6 holds as well.

Recall that from the construction above, $r_n = N^{4\beta_n\ell_n}$ and $\beta_{\phi(n)}\ell_{\phi(n)} \geq |\phi(n)|$, hence $r_{\phi(n)} \geq N^{4|\phi(n)|} \geq 2^{-4\phi(n)}$. Also $\alpha_n \geq 2^{\phi(n)}|\phi(n)|$.

These inequalities implies that the ratio $\varrho_n = \alpha_n\ell_{\phi(n)}/\ell_{\phi(n+1)}$ is such that

$$\begin{aligned} \varrho_n &= \alpha_n \left(\prod_{k=\phi(n)+1}^{\phi(n+1)-1} r_k \right) r_{\phi(n+1)} \\ &\geq 2^{\phi(n)}|\phi(n)| 2^{\phi(n+1)-\phi(n)-1} 2^{-4\phi(n+1)} \\ &= 2^{-3\phi(n+1)-1}|\phi(n)|. \end{aligned}$$

Since $\phi(n+1) \leq -1$ and $|\phi(n)| \geq 2$ for every $n \leq -2$, this shows that $\varrho_n \geq 8$ for every $n \leq -2$.

Thus, $8\varrho_n/9 \leq \lfloor \varrho_n \rfloor \leq \varrho_n$ and the sequence $(\alpha'_n)_n$ defined by

$$\alpha'_n = \frac{\lfloor \alpha_n\ell_{\phi(n)}/\ell_{\phi(n+1)} \rfloor}{\ell_{\phi(n)}/\ell_{\phi(n+1)}},$$

is such that $\frac{8}{9}\alpha_n \leq \alpha'_n \leq \alpha_n$ for every $n \leq -2$. Therefore it fulfills the conditions already satisfied by the sequence $(\alpha_n)_{n \leq 0}$ and the additional condition that $\alpha'_n\ell_{\phi(n)}/\ell_{\phi(n+1)}$ is an integer for every $n \leq 0$. \square

Proof of lemma 3.3.3. Call B any finite upper bound of the sequence $(b_n)_{n \in \mathbb{N}}$. Since $a_n \ll b_n$, for any positive integer k , there exists an integer N_k such that for every $n \geq N_k$, $a_n \leq b_n 2^{-k}$. We now define a sequence of disjoint intervals of integers $J_k = \{i_k, \dots, j_k\}$, as follows. Set $j_{-1} = -1$ and let $k \geq 0$.

Once j_{k-1} is defined, let $i_k = \max\{j_{k-1} + 1, N_k\}$. Since the series $\sum_n b_n$ diverges and $0 \leq b_n \leq B$ for every $n \in \mathbb{N}$, one can choose an integer $j_k \geq i_k$ such

that $B \leq \sum_{n=i_k}^{j_k} b_n < 2B$ and let $I_k = \{i_k, \dots, j_k\}$. Note that $\sum_{n \in J_k} b_n \geq B$ and $\sum_{n \in J_k} a_n \leq 2B/2^k$.

Calling Q the set $\bigcup_{k \in \mathbb{N}} I_k$, one gets

$$\sum_{n \in Q} a_n = \sum_{k \in \mathbb{N}} \sum_{n \in J_k} a_n \leq 2B \sum_{k \in \mathbb{N}} 1/2^k = 4B,$$

and

$$\sum_{n \in Q} b_n = \sum_{k \in \mathbb{N}} \sum_{n \in J_k} b_n \geq \sum_{k \in \mathbb{N}} B = \infty,$$

which completes the proof of the lemma. \square

3.4 Proof of non standardness under Δ

The non standardness of the split-word process was established by Laurent in his thesis [12] and a similar result was obtained by Vershik in [26] in the context of decreasing sequences of measurable partitions. In this section we give a simplified presentation of Laurent's proof.

The proof of this result involves a subtle notion on filtrations, which is standardness. The notion of standardness was first introduced by Vershik for decreasing measurable partitions. This notion has been adapted to continuous time filtrations by Tsirelson [23], it has been formulated by Dubins, Feldman, Smorodinsky and Tsirelson [4] for continuous time filtrations and by Emery and Schachermayer [5] for discrete time filtrations.

Many necessary and sufficient conditions for standardness have been established, for instance Vershik's self-joining criterion and various notions of cosiness. All these criteria are based on coupling methods. Checking them in specific cases is often a technical task. Yet, these criteria are the key tool to solve some difficult problems. For example, Tsirelson defines and uses a notion of cosiness to prove that the filtration of Walsh's Brownian motion is not Brownian since it is non standard.

By definition, a filtration $\mathcal{F} = (\mathcal{F}_n)_{n \leq 0}$ indexed by $n \leq 0$ is *standard* if, modulo an enlargement of the probability space, one can immerse \mathcal{F} in a filtration generated by an i.i.d. process. Recall that a filtration $\mathcal{F} = (\mathcal{F}_n)_{n \leq 0}$ is *immersed* in a filtration $\mathcal{G} = (\mathcal{G}_n)_{n \leq 0}$ if, for every $n \leq 0$, $\mathcal{F}_n \subset \mathcal{G}_n$ and \mathcal{F}_n and \mathcal{G}_{n-1} are independent conditionally on \mathcal{F}_{n-1} . Roughly speaking, this means that \mathcal{G}_{n-1} gives

no further information on \mathcal{F}_n than \mathcal{F}_{n-1} does. Equivalently, \mathcal{F} is immersed in \mathcal{G} if every \mathcal{F} -martingale is a \mathcal{G} -martingale.

Laurent negates the so-called I-cosiness property to prove that under Δ , the filtration is non standard and therefore non of product type.

We follow the method that Smorodinsky [22] used to prove the non-existence of a “generating parametrization” in the case where $r_n = 2$ for every $n \leq 0$. This method still works in the general case and provides the non standard behaviour of the filtration.

The purpose of this section is to show that if the sequence $(r_n)_{n \leq 0}$ is Δ , then the filtration $\mathcal{F}^{(X,V)}$ is non standard. Note that Δ holds for every bounded sequence $(r_n)_n$.

3.4.1 Preliminary notions

We first recall the notion of I-cosiness, due to Émery and Schachermayer [5].

Definition 20 (Immersion and co-immersion) *Let $\mathcal{F} = (\mathcal{F}_n)_{n \leq 0}$ and $\mathcal{G} = (\mathcal{G}_n)_{n \leq 0}$ denote two filtrations defined on the same probability space.*

- \mathcal{F} is immersed in \mathcal{G} if every martingale in \mathcal{F} is a martingale in \mathcal{G} .
- \mathcal{F} and \mathcal{G} are co-immersed if \mathcal{F} and \mathcal{G} are both immersed in $\mathcal{F} \vee \mathcal{G}$.

Definition 21 (I-cosiness) *Let \mathcal{F} be a filtration on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. One says that \mathcal{F} satisfies the I-cosiness criterion if for every random variable Y measurable for \mathcal{F}_0 with values in a finite set and for every real $\delta > 0$, there exists a probability space $(\bar{\Omega}, \bar{\mathcal{A}}, \bar{\mathbb{P}})$ and two filtrations \mathcal{F}' and \mathcal{F}'' on $(\bar{\Omega}, \bar{\mathcal{A}}, \bar{\mathbb{P}})$, such that the following properties hold.*

- The filtrations \mathcal{F}' and \mathcal{F}'' are both isomorphic to \mathcal{F} .
- The filtrations \mathcal{F}' and \mathcal{F}'' are co-immersed.
- There exists an integer n_0 such that \mathcal{F}'_{n_0} and \mathcal{F}''_{n_0} are independent.
- The copies Y' and Y'' of Y by the isomorphisms of the first condition, verify $\bar{\mathbb{P}}[Y' \neq Y''] < \delta$.

The proof of the non standardness of the filtration $\mathcal{F}^{(X,V)}$ uses the easy part of the equivalence between I-cosiness and standardness.

Theorem 9 (Corollary 5 [5]) *A filtration is standard if and only if it satisfies the I-cosiness criterion and is essentially separable.*

To prove the non standardness of $\mathcal{F}^{(X,V)}$, it is therefore sufficient to show that $\mathcal{F}^{(X,V)}$ does not satisfies the I-cosiness criterion. The tools of the proof are introduced just below.

Let n_0 be a negative integer which will be fixed later.

Definition 22 (Definition of Aut_n) *The intervals of integers $\{1, \dots, \ell_k\}$, $\{\ell_k + 1, \dots, 2\ell_k\}, \dots, \{\ell_n - \ell_k + 1, \dots, \ell_n\}$ are called blocks of $\{1, \dots, \ell_n\}$ of length ℓ_k . Every permutation of the ℓ_n/ℓ_{n_0} blocks of length ℓ_{n_0} which induces for every $k \in \{n, \dots, n_0\}$ a bijection between the blocks of length ℓ_k is called an automorphism of $\{1, \dots, \ell_n\}$ adapted to $\{r_{n+1}, \dots, r_{n_0}\}$. One denotes by Aut_n the set of those permutations.*

One can enumerate the automorphisms adapted to $(r_n)_{n \leq 0}$, by induction, as follows. By definition, an automorphism a in Aut_{n-1} is built from a permutation σ of $\{1, \dots, r_n\}$ and from r_n automorphisms $(a_k)_{1 \leq k \leq r_n}$ in Aut_n . One gets a from σ and $(a_k)_{1 \leq k \leq r_n}$ by setting, for every $1 \leq j \leq r_n$ and every $1 \leq k \leq \ell_n$,

$$a((j-1)\ell_n + k) = (\sigma(j) - 1)\ell_n + a_j(k).$$

Therefore $\#(\text{Aut}_{n-1}) = \#(\mathfrak{S}_{r_n})(\#(\text{Aut}_n))^{r_n} = r_n!(\#(\text{Aut}_n))^{r_n}$. By induction

$$\#(\text{Aut}_n) = \prod_{k=n+1}^{n_0} (r_k!)^{r_{n+1} \dots r_{k-1}} = \prod_{k=n+1}^{n_0} (r_k!)^{\ell_n/\ell_{k-1}}.$$

Note that

$$\#(\text{Aut}_n) = \exp(\ell_n S_n) \quad \text{where} \quad S_n = \sum_{k=n+1}^{n_0} \frac{\ln r_k!}{\ell_{k-1}}.$$

We now define a semi-metrics based on the Hamming distance.

Definition 23 (Semi-metrics e_n on A^{ℓ_n}) *Recall that the Hamming distance on A^{ℓ_n} is defined by*

$$d_n^H(x, x') = \#\{k \in \{1, \dots, \ell_n\} : x(k) \neq x'(k)\}.$$

One defines an action of the group Aut_n on A^{ℓ_n} , seen as $(A^{\ell_{n_0}})^{\ell_n/\ell_{n_0}}$, by

$$a \cdot x = x \circ a^{-1},$$

and a semi-metrics e_n on A^{ℓ_n} , by

$$e_n(x, x') = \frac{1}{\ell_n} \min\{d_n^H(a \cdot x, x'); a \in \text{Aut}_n\}.$$

The quantity $e_n(x, x')$ is the smallest proportion of letters which are different between the words x' and $x \circ a^{-1}$ as a goes through Aut_n . Our next result is a recursion relation, useful to compute e_n .

Lemma 3.4.1 *For every $n \leq 0$, $x = (w_1, \dots, w_{r_n})$ and $x' = (w'_1, \dots, w'_{r_n})$ where w_i and w'_j belong to A^{ℓ_n} ,*

$$e_{n-1}(x, x') = \min_{\sigma \in \mathfrak{S}_{r_{n-1}}} \left(\frac{1}{r_n} \sum_{j=1}^{r_n} e_n(w_j, w'_{\sigma(j)}) \right).$$

Proof of lemma 3.4.1.

Using the decomposition of every automorphism a of Aut_{n-1} into a permutation σ of $\{1, \dots, r_n\}$ and r_n automorphisms a_k , $1 \leq k \leq r_n$ in Aut_n , and the additivity of the restricted Hamming distance, one sees that $\ell_{n-1}e_{n-1}(x, x')$ is the minimum over these σ and a_k of the sums

$$\sum_{j=1}^{r_n} d_n^H(a_j \cdot w_j, w'_{\sigma(j)}),$$

hence

$$\ell_{n-1}e_{n-1}(x, x') = \min_{\sigma \in \mathfrak{S}_{r_n}} \left(\ell_n \sum_{j=1}^{r_n} e_n(w_j, w'_{\sigma(j)}) \right).$$

□

To prove that the filtration $\mathcal{F}^{(X,V)}$ is non standard under the assumption Δ , by denying the I-cosiness criterion, one considers X' and X'' two copies of $(X_n)_{n \leq 0}$ such that:

- The associated filtrations $\mathcal{F}' = \mathcal{F}^{(X')}$ and $\mathcal{F}'' = \mathcal{F}^{(X'')}$ are co-immersed;
- There exists an integer $m < n_0$ such that $(X'_k)_{k \leq m}$ and $(X''_k)_{k \leq m}$ are independent.

Our proof of the non standardness of $\mathcal{F}^{(X,V)}$ includes three steps:

- We prove the inequality $\mathbb{P}[X'_{n_0} \neq X''_{n_0}] \geq \mathbb{E}[e_n(X'_n, X''_n)]$.
- We bound below $\mathbb{E}[e_n(X'_n, X''_n)]$ when X'_n and X''_n are independent.
- We negate the I-cosiness criterion.

3.4.2 Proof of the inequality $\mathbb{P}[X'_{n_0} \neq X''_{n_0}] \geq \mathbb{E}[e_n(X'_n, X''_n)]$ for $n \leq n_0$

For $n \leq n_0$ one denotes by

$$M_n = \mathbb{P}[X'_{n_0} \neq X''_{n_0} | \mathcal{F}'_n \vee \mathcal{F}''_n], \quad L_n = e_n(X'_n, X''_n).$$

By construction, $(M_n)_{n \leq n_0}$ is a martingale. Loosely speaking, this martingale measures the influence of the past before time n on the word at time n_0 . The key step is to prove that $(M_n)_{n \leq n_0}$ is bounded below by $(L_n)_{n \leq n_0}$.

Let us prove that $(L_n)_{n \leq n_0}$ is a sub-martingale in the filtration $\mathcal{F}' \vee \mathcal{F}''$. Let $n \leq n_0$. Using the co-immersion of the filtrations \mathcal{F}' and \mathcal{F}'' , the conditional law of X'_n given $\mathcal{F}'_{n-1} \vee \mathcal{F}''_{n-1}$ is uniform on the r_n sub-words of X'_{n-1} of length ℓ_n :

$$\mathcal{L}(X'_n | \mathcal{F}'_{n-1} \vee \mathcal{F}''_{n-1}) = \frac{1}{r_n} \sum_{i=1}^{r_n} \delta_{f_n(X'_{n-1}, i)},$$

where $f_n(x, v)$ denotes the v -th sub-word (of length ℓ_n) of x , for $x \in A^{\ell_{n-1}}$ and $1 \leq v \leq r_n$ (as in the proof of theorem 7). Furthermore,

$$\mathcal{L}(X''_n | \mathcal{F}'_{n-1} \vee \mathcal{F}''_{n-1}) = \frac{1}{r_n} \sum_{j=1}^{r_n} \delta_{f_n(X''_{n-1}, j)}.$$

Therefore, the conditional law of (X'_n, X''_n) given $\mathcal{F}'_{n-1} \vee \mathcal{F}''_{n-1}$ can be written as follows

$$\mathcal{L}((X'_n, X''_n) | \mathcal{F}'_{n-1} \vee \mathcal{F}''_{n-1}) = \frac{1}{r_n} \sum_{1 \leq i, j \leq r_n} C_{i,j} \delta_{(f_n(X'_{n-1}, i), f_n(X''_{n-1}, j))},$$

where $(C_{i,j})_{1 \leq i, j \leq r_n}$ is a bistochastic matrix measurable for $\mathcal{F}'_{n-1} \vee \mathcal{F}''_{n-1}$. In particular,

$$\mathbb{E}[L_n | \mathcal{F}'_{n-1} \vee \mathcal{F}''_{n-1}] = \frac{1}{r_n} \sum_{1 \leq i, j \leq r_n} C_{i,j} e_n(f_n(X'_{n-1}, i), f_n(X''_{n-1}, j)).$$

This quantity is the image of the matrix $(C_{i,j})_{1 \leq i, j \leq r_n}$ by a linear form. Since bistochastic matrices belong to the convex hull of the permutation matrices, one gets

$$\mathbb{E}[L_n | \mathcal{F}'_{n-1} \vee \mathcal{F}''_{n-1}] \geq \frac{1}{r_n} \inf_{\sigma \in \mathfrak{S}_{r_n}} \sum_{i=1}^{r_n} e_n(f_n(X'_{n-1}, i), f_n(X''_{n-1}, \sigma(i))),$$

hence

$$\mathbb{E}[L_n | \mathcal{F}'_{n-1} \vee \mathcal{F}''_{n-1}] \geq e_{n-1}(X'_{n-1}, X''_{n-1}) = L_{n-1},$$

thanks to the recursion relation verified by the semi-metrics e_n . This shows that $(L_n)_{n \leq n_0}$ is a sub-martingale in the filtration $\mathcal{F}' \vee \mathcal{F}''$.

Since $M_{n_0} = \mathbf{1}_{\{X'_{n_0} \neq X''_{n_0}\}} \geq L_{n_0}$, one gets, for every $n \leq n_0$

$$M_n = \mathbb{E}[M_{n_0} | \mathcal{F}'_n \vee \mathcal{F}''_n] \geq \mathbb{E}[L_{n_0} | \mathcal{F}'_n \vee \mathcal{F}''_n] \geq L_n,$$

yielding the inequality $\mathbb{P}[X'_{n_0} \neq X''_{n_0}] \geq \mathbb{E}[e_n(X'_n, X''_n)]$ by taking the expectations. This inequality is going to be used to prove that the probability $\mathbb{P}[X'_{n_0} \neq X''_{n_0}]$ can not be made as small as one wishes if the processes X' and X'' are independent until a given time n .

3.4.3 Bounding below $\mathbb{E}[e_n(X'_n, X''_n)]$ when X'_n and X''_n are independent

The purpose of this subsection is to prove the next inequality:

Lemma 3.4.2 *Let $(X'_k)_{k \leq 0}$ and $(X''_k)_{k \leq 0}$ be two copies of the split-word process that are independent until time n . Then, for every $\alpha > 0$ and $n \leq n_0$,*

$$\mathbb{P}[e_n(X'_n, X''_n) \leq 1 - N^{-1} - \alpha] \leq \exp(\ell_n(S_n - 2\alpha^2)), \text{ where } S_n = \sum_{k=n+1}^{n_0} \frac{\ln(r_k!)}{\ell_{k-1}}$$

The proof is based on Hoeffding's large deviations inequality, see [21] for a proof of this inequality.

Lemma 3.4.3 (Hoeffding) *Let q in $]0, 1[$ and $(\epsilon_k)_{1 \leq k \leq n}$ denote n independent Bernoulli random variables of parameter q , and Z_n their average. Then, for every real $\alpha > 0$,*

$$\mathbb{P}[Z_n \leq q - \alpha] \leq \exp(-2n\alpha^2).$$

Proof of lemma 3.4.2. One applies lemma 3.4.3 to the variables

$$\epsilon_i = \mathbf{1}_{\{X'_n(i) \neq X''_n(i)\}}, \quad 1 \leq i \leq \ell_n.$$

Since X'_n and X''_n are independent and uniform on A^{ℓ_n} , the ϵ_i are independent and Bernoulli with parameter $q_N = 1 - 1/N$. Thus, denoting by d_n^H the Hamming distance on A^{ℓ_n} ,

$$\mathbb{P}\left[\frac{1}{\ell_n} d_n^H(X'_n, X''_n) \leq q_N - \alpha\right] \leq \exp(-2\ell_n \alpha^2).$$

For every a in Aut_n , $a \cdot X''$ has the same law as X'' and is independent of X' . Since, by definition $e_n(X'_n, X''_n) = \min\{d_n^H(X'_n, a \cdot X''_n); a \in \text{Aut}_n\}$, one gets

$$[e_n(X'_n, X''_n) \leq q_N - \alpha] = \left[\exists a \in \text{Aut}_n, \frac{1}{\ell_n} d_n^H(X'_n, a \cdot X''_n) \leq q_N - \alpha \right],$$

hence

$$\begin{aligned} \mathbb{P}[e_n(X'_n, X''_n) \leq q_N - \alpha] &\leq \sum_{a \in \text{Aut}_n} \mathbb{P}\left[\frac{1}{\ell_n} d_n^H(X'_n, a \cdot X''_n) \leq q_N - \alpha\right] \\ &= \sharp(\text{Aut}_n) \mathbb{P}\left[\frac{1}{\ell_n} d_n^H(X'_n, X''_n) \leq q_N - \alpha\right] \\ &\leq \exp(\ell_n S_n) \exp(-2\ell_n \alpha^2), \end{aligned}$$

since $\sharp(\text{Aut}_n) = \exp(\ell_n S_n)$. □

Choice of the integer n_0 Under assumption Δ , by remark 1 after theorem 8, one can choose $n_0 \leq 0$ such that

$$S_{-\infty} < 2(1 - 1/N)^2, \text{ where } S_{-\infty} = \sum_{k=-\infty}^{n_0} \frac{\ln(r_k!)}{\ell_{k-1}}.$$

Once n_0 is fixed, choose a real α such that $\sqrt{S_{-\infty}/2} < \alpha < 1 - 1/N$.

Since $S_n - 2\alpha^2 \leq S_{-\infty} - 2\alpha^2 < 0$ and $\ell_n \geq \ell_{n_0}$ for every $n \leq n_0$, lemma 3.4.2 yields

$$\begin{aligned} \mathbb{P}[e_n(X'_n, X''_n) \leq 1 - 1/N - \alpha] &\leq \exp(\ell_n(S_n - 2\alpha^2)) \\ &\leq \exp(\ell_{n_0}(S_{-\infty} - 2\alpha^2)) = \beta, \end{aligned}$$

with $\beta < 1$. Hence, $\mathbb{E}[e_n(X'_n, X''_n)] \geq (1 - \beta)(1 - 1/N - \alpha)$, which is positive.

3.4.4 Negation of the I-cosiness criterion

If (X', V') and (X'', V'') are two copies of the split-word process whose filtrations are co-immersed and independent until time $n \leq n_0$,

$$\mathbb{P}[X'_{n_0} \neq X''_{n_0}] \geq \mathbb{E}[e_n(X'_n, X''_n)] \geq (1 - \beta)(1 - 1/N - \alpha),$$

and this lower bound is positive, contradicting the I-cosiness criterion. Thus the filtration of the split-word process is non standard. □

Chapter 4

Sufficient conditions of standardness for filtrations of stationary processes taking values in a finite space

Let X be a stationary process with finite state-space A . Bressaud et al. recently provided a sufficient condition for the natural filtration of X to be standard when A has size 2. Their condition involves the conditional laws $p(\cdot|x)$ of X_0 conditionally on the whole past $(X_k)_{k \leq -1} = x$ and controls the strength of the influence of the “old” past of the process on its present X_0 . It involves the maximal gaps between $p(\cdot|x)$ and $p(\cdot|y)$ for infinite sequences x and y which coincide on their n last terms. In this paper, we first show that a slightly stronger result holds for any finite state-space. Then, we provide sufficient conditions for standardness based on average gaps instead of maximal gaps.

4.1 Introduction

4.1.1 Setting

In this paper we study stationary processes $X = (X_n)_{n \in \mathbb{Z}}$ indexed by the integer line \mathbb{Z} and with values in a finite set A . We assume that X is defined recursively as follows: for every $n \in \mathbb{Z}$, X_n is a function of the “past” $X_{n-1}^{\triangleleft} = (X_k)_{k \leq n-1}$ of X and of a “fresh” random variable U_n , which brings in some “new” randomness. In particular the process $U = (U_n)_{n \in \mathbb{Z}}$ is independent. To be more specific, we introduce some notations and definitions about σ -algebras.

All σ -fields are assumed to be complete. For every process $\xi = (\xi_n)_{n \in \mathbb{Z}}$ and

every $n \in \mathbb{Z}$, let $\xi_n^{\triangleleft} = (\xi_k)_{k \leq n}$ and $\mathcal{F}_n^\xi = \sigma(\xi_n^{\triangleleft})$. The natural filtration of ξ is the nondecreasing sequence $\mathcal{F}^\xi = (\mathcal{F}_n^\xi)_{n \in \mathbb{Z}}$. Furthermore, $\mathcal{F}_\infty^\xi = \sigma(\xi_k; k \in \mathbb{Z})$ and $\mathcal{F}_{-\infty}^\xi$ is the tail σ -algebra $\mathcal{F}_{-\infty}^\xi = \bigcap_{k \in \mathbb{Z}} \mathcal{F}_k^\xi$.

We say that a process U is a *governing process* for X , or that U *governs* X if, for every $n \in \mathbb{Z}$, (i) U_{n+1} is independent of $\mathcal{F}_n^{X,U}$, and (ii) X_{n+1} is measurable with respect to $\sigma(U_{n+1}) \vee \mathcal{F}_n^X$. In particular any governing process is independent. If moreover the U_n are uniform on $[0, 1]$, the process (U, X) is – according to Schachermayer’s definition [20] and up to a time reversal – a *parametrization* of the process X .

Likewise, we say that a process U is a *generating process* for X , or that U *generates* X if, for every $n \in \mathbb{Z}$, X_n is measurable with respect to \mathcal{F}_n^U . This is equivalent to the condition that $\mathcal{F}_n^X \subset \mathcal{F}_n^U$ for every $n \in \mathbb{Z}$, a property which, from now on, we write as $\mathcal{F}^X \subset \mathcal{F}^U$.

One could be led to believe that when $\mathcal{F}_{-\infty}^X$ is trivial, any process governing X generates X as well. But, although notoriously used by Wiener and Kallianpur in [11] (not published but see a discussion in [18]), this argument is false. As a simple counterexample, assume that X is i.i.d., that every X_n is uniform on $\{-1, 1\}$, and set $U_n = X_n X_{n-1}$ for every $n \in \mathbb{Z}$. Then $\mathcal{F}_{-\infty}^X$ is trivial and U governs X but X_0 is independent of \mathcal{F}_∞^U hence U does not generate X .

Governing and generating processes are related to *immersions* of filtrations. Recall that the filtration \mathcal{F}^X is *immersed* in the filtration \mathcal{F}^U if $\mathcal{F}^X \subset \mathcal{F}^U$ and if, for every $n \in \mathbb{Z}$, \mathcal{F}_{n+1}^X and \mathcal{F}_n^U are independent conditionally on \mathcal{F}_n^X . Roughly speaking, this means that \mathcal{F}_n^U gives no further information on X_{n+1} than \mathcal{F}_n^X does. Equivalently, \mathcal{F}^X is immersed in \mathcal{F}^U if every \mathcal{F}^X -martingale is an \mathcal{F}^U -martingale. The following easy fact holds (see a proof in section 4.5.2).

Lemma 4.1.1 *If U is a governing and generating process for X , then \mathcal{F}^X is immersed in \mathcal{F}^U .*

Another notable property of filtrations is *standardness*. Recall that \mathcal{F}^X is *standard* if, modulo an enlargement of the probability space, one can immerse \mathcal{F}^X in a filtration generated by an i.i.d. process. Vershik introduced standardness in the context of ergodic theory. Examples of non standard filtrations include the filtrations of $[T, T^{-1}]$ transformations, introduced in [10]. Split-word processes, inspired by Vershik’s (r_n) -adic sequences of decreasing partitions [26] and studied in [22] and [12], for instance, also provide non standard filtrations.

Obviously, lemma 4.1.1 above implies that if X has a generating and governing process, then \mathcal{F}^X is standard. Whether the converse holds is not known.

Necessary and sufficient conditions for standardness include Vershik’s self-joining criterion and Tsirelson’s notion of I -cosiness. Both notions are discussed in [5]

and are based on conditions which are subtle and not easy to use nor to check in specific cases.

Our goal in this paper is to provide sufficient conditions of standardness that are easier to use than the ones mentioned above. Each of our conditions involves a measure of the influence of the “old” past of the process on its present. We introduce them in the next section.

4.1.2 Statement of the results

We now introduce some measures of the influence of the past of a process on its present. To conveniently state these definitions and, later on, our results, we first introduce some notations.

Recall that X is a stationary process indexed by the integer line \mathbb{Z} with values in some finite set A and with natural filtration \mathcal{F}^X .

Notation 3 (1) *Slabs:* For any sequence $(\xi_n)_{n \in \mathbb{Z}}$ in $A^{\mathbb{Z}}$, deterministic or random, and any integers $i \leq j$, $\xi_{i:j}$ is the $(j - i + 1)$ -uple $(\xi_n)_{i \leq n \leq j}$ in A^{j-i+1} .

(2) *Shifts:* If $k - i = \ell - j$, $\xi_{i:k} = \zeta_{j:\ell}$ means that $\xi_{i+n} = \zeta_{j+n}$ for every integer n such that $0 \leq n \leq k - i$.

Infinite case: Let A^{\triangleleft} denote the space of sequences $(\xi_n)_{n \leq -1}$. For every i in \mathbb{Z} , a sequence $(\xi_n)_{n \leq i}$ is also considered as an element of A^{\triangleleft} since, similarly to the finite case, one identifies $\xi_i^{\triangleleft} = (\xi_n)_{n \leq i}$ and $\zeta_j^{\triangleleft} = (\zeta_n)_{n \leq j}$ if $\xi_{i+n} = \zeta_{j+n}$ for every integer $n \leq 0$.

(3) *Concatenation:* For all $i \geq 0$, $j \geq 0$, $x = (x_n)_{1 \leq n \leq i}$ in A^i and $y = (y_n)_{1 \leq n \leq j}$ in A^j , xy denotes the concatenation of x and y , defined as

$$xy = (x_1, \dots, x_i, y_1, \dots, y_j), \quad xy \in A^{i+j}.$$

Infinite case: $i \geq 0$, $y = (y_n)_{1 \leq n \leq i}$ in A^i and $x = (x_n)_{n \leq -1}$ in A^{\triangleleft} , xy denotes the concatenation of x and y , defined as

$$xy = (\dots, x_{-2}, x_{-1}, y_1, \dots, y_i), \quad xy \in A^{\triangleleft}.$$

Notation 4 For each $n \geq 0$, $x \in A^n$ and $a \in A$, set

$$p(a|x) = \mathbb{P}(X_0 = a \mid X_{-n:-1} = x),$$

with the convention

$$p(a|x) = \mathbb{P}(X_0 = a) \quad \text{if } \mathbb{P}[X_{-n:-1} = x] = 0.$$

In the following,

$$p(\cdot|x) = \mathbb{P}(X_0 = \cdot \mid X_{-1}^{\triangleleft} = x), \quad x \in A^{\triangleleft},$$

denotes a regular version of the conditional law of X_0 given X_{-1}^{\triangleleft} .

We now introduce three quantities γ_n , α_n and δ_n measuring the pointwise influence at distance n .

Definition 24 For every $n \geq 0$, let

$$\gamma_n = 1 - \inf \left\{ \frac{p(a|xz)}{p(a|yz)} ; a \in A, x \in A^{\triangleleft}, y \in A^{\triangleleft}, z \in A^n, p(a|yz) > 0 \right\},$$

$$\alpha_n = 1 - \inf_{z \in A^n} \sum_{a \in A} \inf \{p(a|yz) ; y \in A^{\triangleleft}\},$$

$$\delta_n = \sup \{ \|p(\cdot|xz) - p(\cdot|yz)\| ; x \in A^{\triangleleft}, y \in A^{\triangleleft}, z \in A^n \},$$

where, for all probabilities μ and ν on A , $\|\mu - \nu\|$ is the distance in total variation between μ and ν , defined as

$$\|\mu - \nu\| = \frac{1}{2} \sum_{a \in A} |\mu(a) - \nu(a)| = \sum_{a \in A} [\mu(a) - \nu(a)]_+.$$

Note that the definitions of γ_n , α_n and δ_n depend on the choice of the regular version $(p(\cdot|x))_{x \in A^{\triangleleft}}$ of the conditional law of X_0 given X_{-1}^{\triangleleft} . One needs a "good" version to get small influences for applying the theorems below.

The sequences $(\gamma_n)_{n \geq 0}$, $(\alpha_n)_{n \geq 0}$ and $(\delta_n)_{n \geq 0}$ are non-increasing, $[0, 1]$ -valued, and $\delta_n \leq \gamma_n$, $\delta_n \leq \alpha_n$ for every $n \geq 0$ (see the proof in section 4.5.1).

For every $[0, 1]$ -valued sequence $(\varepsilon_n)_{n \geq 0}$, we consider the condition

$$\sum_{k=0}^{+\infty} \prod_{n=0}^k (1 - \varepsilon_n) = +\infty. \quad (\mathcal{H}(\varepsilon))$$

For instance, $\mathcal{H}(\gamma)$ and $\mathcal{H}(2\delta)$ are respectively

$$\sum_{k=0}^{+\infty} \prod_{n=0}^k (1 - \gamma_n) = +\infty, \quad \text{and} \quad \sum_{k=0}^{+\infty} \prod_{n=0}^k (1 - 2\delta_n) = +\infty.$$

Observe that if two $[0, 1]$ -valued sequences $(\varepsilon_n)_{n \geq 0}$ and $(\zeta_n)_{n \geq 0}$ are such that $\varepsilon_n \leq \zeta_n$ for every $n \geq 0$, then $\mathcal{H}(\zeta)$ implies $\mathcal{H}(\varepsilon)$. Hence condition $\mathcal{H}(\varepsilon)$ asserts that $(\varepsilon_n)_{n \geq 0}$ is "small enough" in a way.

The definition of $(\gamma_n)_{n \geq 0}$ and the assumption $\mathcal{H}(\gamma)$ are both stated in [2]. The main result of [2] is the following.

Theorem 10 (X.Bressaud, A.Maass, S.Martinez and J.San Martin [2])

Assume that the size of A is 2, then $\mathcal{H}(\gamma)$ implies that \mathcal{F}^X is standard.

The scope of theorem 10 is restricted by the following three conditions. First, the size of A must be 2. Second, one must control the *ratios* of probabilities which define γ_n . Third, $\mathcal{H}(\gamma)$ implies that $\gamma_0 < 1$, therefore one can show that $\mathcal{H}(\gamma)$ implies the existence of $c > 0$ such that $p(a|x) \geq c$ for every x in A^\triangleleft and a in A such that $\mathbb{P}[X_0 = a] > 0$ (see the proof in section 4.5.4).

Our first result allows to get rid of the two first restrictions.

Theorem 11 (1) *Assume that A is finite, that $2\delta_0 < 1$ and that $\mathcal{H}(2\delta)$ holds. Then \mathcal{F}^X is standard.*

(2) *If the size of A is 2, $\mathcal{H}(\delta)$ alone implies that \mathcal{F}^X is standard.*

Theorem 11 generalizes and improves on theorem 10 of [2], since $\delta_n \leq \gamma_n$ for every n . Note that the straight adaptation of the proof of [2] to sizes of A at least 3 leads to the more stringent condition $\mathcal{H}(2\gamma)$.

Another measure of influence, based on the quantities α_n defined before, is introduced and used in [3] (actually the notation there is $a_n = 1 - \alpha_n$). The authors show that if $\mathcal{H}(\alpha)$ holds, there exists a perfect sampling algorithm for the process X , a result which implies that \mathcal{F}^X is standard. But since $\delta_n \leq \alpha_n$ for every $n \geq 0$, the result of [3] does not imply theorem 10.

Theorems 10 and 11 and the exact sampling algorithm of [3] all require an upper bound of some pointwise influence sequence. Our next result uses a less restrictive hypothesis based on some average influences η_n , defined below.

Definition 25 *For every $n \geq 0$, let η_n denote the average influence at distance n , defined as*

$$\eta_n = \sum_{z \in A^n} \mathbb{E}[\|p(\cdot|z) - p(\cdot|X_{-n-1}^\triangleleft z)\|] \cdot \mathbb{P}[X_{-n:-1} = z],$$

and call $\mathcal{H}'(\eta)$ the condition

$$\sum_{k=0}^{+\infty} \eta_k < +\infty. \quad (\mathcal{H}'(\eta))$$

Note that η_n is also

$$\eta_n = \mathbb{E}[\|p(\cdot|Y_{-n:-1}) - p(\cdot|X_{-n-1}^\triangleleft Y_{-n:-1})\|],$$

where Y is an independent copy of X .

Definition 26 (Priming condition)

We say that the process X fulfills the priming condition if for every a in A , $p(a|X_{-1}^\triangleleft) > 0$ almost surely.

Theorem 12 *Assume that A is finite and that X fulfills the priming condition. Then, $\mathcal{H}'(\eta)$ implies that \mathcal{F}^X is standard.*

The sequence $(\eta_n)_{n \geq 0}$ is $[0, 1]$ -valued. If $\eta_n < 1$ for every $n \leq 0$, then $\mathcal{H}'(\eta)$ clearly implies $\mathcal{H}(\eta)$. Yet, since $\eta_n \leq \delta_n$ for every $n \geq 0$ (see the proof in section 4.5.1), the condition $\mathcal{H}'(\eta)$ cannot be compared to the conditions $\mathcal{H}(\delta)$ and $\mathcal{H}(2\delta)$.

Theorem 12 gives a remarkable result for chains with memory of variable length. These chains, studied notably in [7] and widely used for mathematical models, are stationary processes X taking values in a finite alphabet A , such that the distribution of X_0 given the past X_{-1}^{\triangleleft} depends only on a past $X_{-\ell:-1}$ of length ℓ , where ℓ is random and measurable with respect to \mathcal{F}_{-1}^X .

More precisely, for $x \in A^{\triangleleft}$, let

$$\begin{aligned} \ell(x) &= \inf\{n \geq 0; y \mapsto p(\cdot | yx_{-n:-1}) \text{ is constant on } A^{\triangleleft}\} \\ &= \inf\{n \geq 0; \forall y \in A^{\triangleleft}, p(\cdot | yx_{-n:-1}) = p(\cdot | x)\}. \end{aligned}$$

Then X is a variable length Markov chain if $\ell(X_{-1}^{\triangleleft})$ is almost surely finite. The following result holds.

Corollary 4.1.2 *If X fulfills the priming condition and if $\ell(X_{-1}^{\triangleleft})$ is integrable, then the natural filtration \mathcal{F}^X is standard.*

Once again we refer the reader to section 4.5.3 for the proof.

Here is a plan of the rest of the paper. In section 4.2, we prove theorem 11. In section 4.3, we prove theorem 12. In section 4.4, we compare theorems 11 and 12 through examples. Finally in section 4.5, we prove some facts stated without proof in the introduction, namely lemma 4.1.1, corollary 4.1.2, a consequence of the assumption $\mathcal{H}(\gamma)$ and some inequalities involving the quantities α_n , γ_n , δ_n and η_n .

4.2 Pointwise influence

4.2.1 Construction of a governing sequence

We construct a governing sequence with values in the standard simplex on $\#A$ vertices.

Notation 5 *Let H be the hyperplane in \mathbb{R}^A defined by*

$$H = \{x = (x_a)_{a \in A} \in \mathbb{R}^A : \sum_{a \in A} x_a = 1\}.$$

Let S be the simplex in H defined by

$$S := (\mathbb{R}_+)^A \cap H = \left\{ x \in (\mathbb{R}_+)^A : \sum_{a \in A} x_a = 1 \right\}.$$

In other words, $S = \text{Conv}(E_A)$ is the convex envelope of the canonical basis $E_A = (E_a)_{a \in A}$ of \mathbb{R}^A .

Let λ denote the Lebesgue measure on H and $\mu = (\mathbf{1}_S / \lambda(S))\lambda$ the uniform distribution on S .

Notation 6 For any probability p on A , let

$$G(p) = (p(a))_{a \in A} = \sum_{a \in A} p(a) E_a, \quad G(p) \in S.$$

For a in A , denote by $f_a(\cdot, p)$ the affine map from H to H which sends E_a on $G(p)$ and lets invariant E_b for every b in A , $b \neq a$. Let

$$S_a(p) = f_a(S, p) = \text{Conv}\left(\{G(p)\} \cup E_A \setminus \{E_a\}\right).$$

A short computation yields the interpretation of $p(a)$ below.

Lemma 4.2.1 For any a in A , $\det(f_a(\cdot, p)) = p(a)$. Therefore, for any measurable $B \subset S$,

$$\lambda[f_a(B, p)] = \lambda[B] p(a).$$

In particular $\lambda(S_a(p)) = \lambda(S) p(a)$, hence $\mu(S_a(p)) = p(a)$.

We now characterize $S_a(p)$.

By convention, for every $r > 0$, we set $r/0 = \infty$, and $0/0 = 0$.

Lemma 4.2.2 Let p be a probability on A , a in A , then

$$\begin{aligned} S_a(p) &= \left\{ x = (x_a)_{a \in A} \in S : \frac{x_a}{p(a)} = \min_{b \in A} \frac{x_b}{p(b)} \right\} \\ &= \left\{ x \in S : \forall b \in A, \frac{x_a}{p(a)} \leq \frac{x_b}{p(b)} \right\}. \end{aligned}$$

Corollary 4.2.3 S is the union of the simplices $S_a(p)$, with a in A and that, if $a \neq b$, the simplices $S_a(p)$ and $S_b(p)$ meet only at their boundary.

Proof: it's a straight corollary of the lemmas 4.2.1 and 4.2.2.

Proof. Call $\Sigma_a(p)$ the right-hand side. Since $\Sigma_a(p)$ is a convex polyedron and contains the points $G(p)$ and E_b for every $b \neq a$, $S_a(p) \subset \Sigma_a(p)$.

As regards the other inclusion, let $x = (x_a)_{a \in A}$ in $\Sigma_a(p)$. Then $x_a/p(a)$ is finite, and

$$x = \frac{x_a}{p(a)}G(p) + \sum_{b \neq a} \left(x_b - p(b) \frac{x_a}{p(a)} \right) E_b.$$

From the definition of $\Sigma_a(p)$, $x_b/p(b) \geq x_a/p(a)$ for every $b \neq a$, hence one has $x_b - p(b)x_a/p(a) \geq 0$ for every $b \neq a$. Furthermore,

$$\frac{x_a}{p(a)} + \sum_{b \neq a} \left(x_b - p(b) \frac{x_a}{p(a)} \right) = \frac{x_a}{p(a)} + \sum_{b \in A} \left(x_b - p(b) \frac{x_a}{p(a)} \right) = \sum_{b \in A} x_b = 1,$$

hence x is indeed a barycenter of the points $G(p)$ and E_b for $b \neq a$. This concludes the proof. \square

One knows that the simplices $(S_a(p))_{a \in A}$ cover S and intersect only on a set of measure zero. Hence, for almost every s in S , there exists a unique a in A such that $s \in S_a(p)$. Our next definition deals with the tie cases.

Definition 27 Fix once and for all a total ordering of A . For every s in S and every probability p on A with full support, define

$$g(s, p) = \min\{a \in A : s \in S_a(p)\}.$$

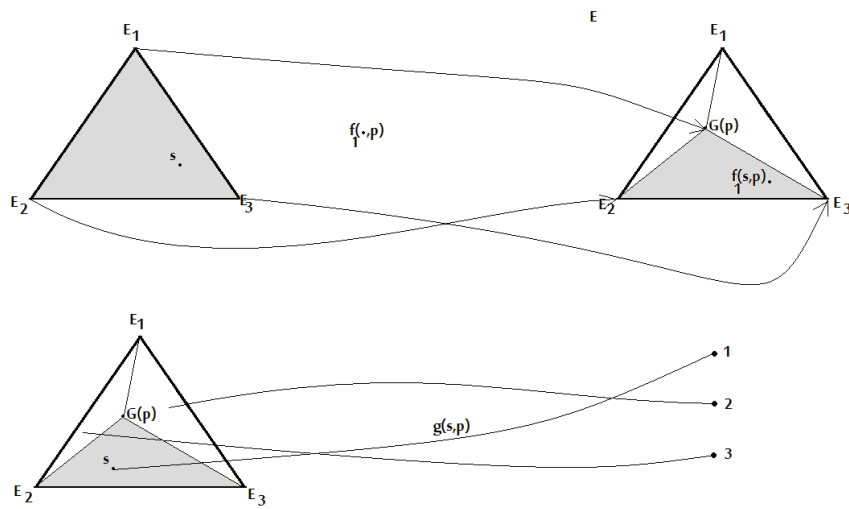


Figure 4.1: $f(\cdot, p)$ and $g(\cdot, p)$

Lemma 4.2.4 *Let U denote a random variable uniformly distributed on S . Then the distribution of $g(U, p)$ is p .*

Indeed, up to negligible events, $\{g(U, p) = a\} = \{U \in S_a(p)\}$, hence

$$\mathbb{P}[g(U, p) = a] = \mu(S_a(p)) = p(a).$$

The following lemma is our main tool to construct governing sequences.

Lemma 4.2.5 *Let X be a random variable with distribution p on A . Let W be a random variable with uniform distribution on S and independent of X . Introduce*

$$U = f_X(W, p) = \sum_{a \in A} f_a(W, p) \mathbf{1}_{\{X=a\}}.$$

Then U is uniformly distributed on S and $X = g(U, p)$ almost surely.

Proof. Since $U \in f_X(S, p) = S_X(p)$, $X = g(U, p)$ almost surely. We now prove that U is uniformly distributed on S .

The sets $S_a(p)$ for a in A cover S and their pairwise intersections are negligible for λ . Hence, for every Borel subset B of S ,

$$\begin{aligned} \mathbb{P}[U \in B] &= \sum_{a \in A} \mathbb{P}[X = a ; f_a(W, p) \in B] = \sum_{a \in A} \mathbb{P}[X = a] \cdot \mathbb{P}[W \in f_a(\cdot, p)^{-1}(B)] \\ &= \sum_{a \in A} p(a) \frac{\lambda(f_a(\cdot, p)^{-1}(B) \cap S)}{\lambda(S)} = \sum_{a \in A} \frac{\lambda(B \cap S_a)}{\lambda(S)} = \frac{\lambda(B)}{\lambda(S)} = \mu(B), \end{aligned}$$

where the second equality stems from the independence of X and W and the fourth equality stems from lemma 4.2.1. This concludes the proof. \square

4.2.2 Upper bound of the error

In this section we study the dependence of the random variable $g(U, p)$ with respect to p . The following result will be used twice.

Proposition 4.2.6 (Upper bound of the error) *Let U be a random variable uniformly distributed on S . Let p and q be two probabilities on A . Then,*

$$\mathbb{P}[g(U, p) \neq g(U, q)] \leq 2 \|p - q\|.$$

In the special case $\#A = 2$,

$$\mathbb{P}[g(U, p) \neq g(U, q)] = \|p - q\|.$$

Remark 2 *The better result when $\#A = 2$ is the reason why theorem 11 involves weaker hypotheses on $(\delta_n)_n$ in this case.*

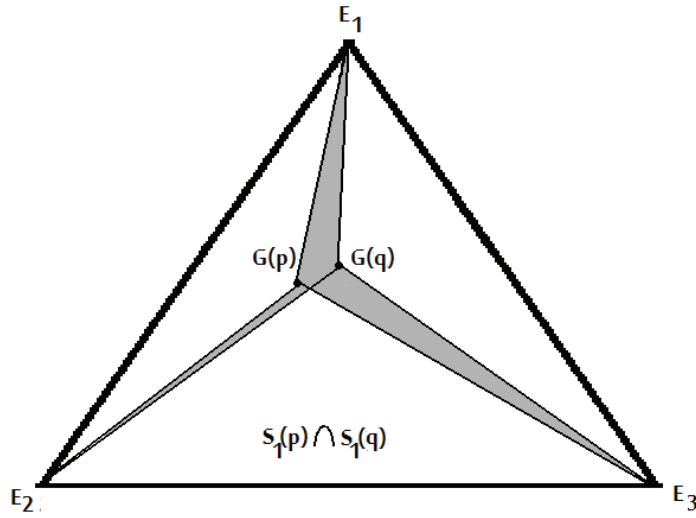


Figure 4.2: Computation of $\mathbb{P}[g(U, p) \neq g(U, q)]$. The grey area shows the $s \in S$ such that $g(s, p) \neq g(s, q)$.

Proof of proposition 4.2.6. Assume without loss of generality that U is constructed from i.i.d. random variables $(\varepsilon_a)_{a \in A}$ exponentially distributed with parameter 1, as follows. For every a in A ,

$$U_a = \frac{\varepsilon_a}{\sum_{b \in A} \varepsilon_b}.$$

The event $\{g(U, p) \neq g(U, q)\}$ depends on $(\varepsilon_a)_a$, as follows. By definition of g , up to negligible events,

$$\{g(U, p) = g(U, q)\} = \bigcup_{a \in A} C_a, \quad \text{with } C_a = \{U \in S_a(p) \cap S_a(q)\}.$$

Furthermore, since for every $a \in A$, $\mathbb{P}[C_a] = 0$ if $p(a) > 0$ or $q(a) > 0$, and since $\mu(C_a \cap C_b) = 0$ for $a \neq b$, one gets

$$\mathbb{P}[g(U, p) = g(U, q)] = \sum_{a \in A} \mathbb{P}[C_a] \mathbf{1}_{\{p(a) > 0, q(a) > 0\}}.$$

For every $a \in A$ such that $p(a) > 0$ and $q(a) > 0$, lemma 4.2.2 gives

$$C_a = \left\{ \frac{\varepsilon_a}{p(a)} = \min_b \frac{\varepsilon_b}{p(b)} ; \frac{\varepsilon_a}{q(a)} = \min_b \frac{\varepsilon_b}{q(b)} \right\} = \left\{ \varepsilon_a \leq \min_b \left(p(a) \frac{\varepsilon_b}{p(b)}, q(a) \frac{\varepsilon_b}{q(b)} \right) \right\},$$

hence

$$C_a = \bigcap_{b \neq a} \{ \varepsilon_b \geq \lambda_{b/a} \varepsilon_a \}, \quad \lambda_{b/a} = \max \left(\frac{p(b)}{p(a)}, \frac{q(b)}{q(a)} \right).$$

Conditioning on ε_a and using that the random variables $(\varepsilon_b)_{b \neq a}$ are i.i.d., exponentially distributed and independent of ε_a , one gets

$$\mathbb{P}[C_a | \varepsilon_a] = \mathbb{P} \left[\bigcap_{b \neq a} \{ \varepsilon_b \geq \lambda_{b/a} \varepsilon_a \} \mid \varepsilon_a \right] = \prod_{b \neq a} \exp(-\lambda_{b/a} \varepsilon_a),$$

hence

$$\mathbb{P}[C_a] = \mathbb{E} \left(\exp \left(- \left(\sum_{b \neq a} \lambda_{b/a} \right) \varepsilon_a \right) \right) = \frac{1}{1 + \sum_{b \neq a} \lambda_{b/a}}.$$

Therefore

$$\mathbb{P}[g(U, p) = g(U, q)] = \sum_{a \in A} \mathbb{P}[C_a] \mathbf{1}_{\{p(a) > 0, q(a) > 0\}} = \sum_{a \in A} \frac{\mathbf{1}_{\{p(a) > 0, q(a) > 0\}}}{1 + \sum_{b \neq a} \lambda_{b/a}}.$$

This last expression is not so easy to compute because each $\lambda_{b/a}$ is defined as a maximum. However,

$$\sum_{a \in A} \frac{\mathbf{1}_{\{p(a) > 0\}}}{1 + \sum_{b \neq a} p(b)/p(a)} = \sum_{a \in A} \frac{p(a) \mathbf{1}_{\{p(a) > 0\}}}{p(a) + \sum_{b \neq a} p(b)} = \sum_a p(a) = 1.$$

Subtracting the expression for $\mathbb{P}[g(U, p) = g(U, q)]$ to this, one gets

$$\mathbb{P}[g(U, p) \neq g(U, q)] = \sum_{a \in A} \frac{\mathbf{1}_{\{p(a) > 0\}}}{1 + \sum_{b \neq a} p(b)/p(a)} - \sum_{a \in A} \frac{\mathbf{1}_{\{p(a) > 0, q(a) > 0\}}}{1 + \sum_{b \neq a} \lambda_{b/a}}.$$

Coming back to the definition of $\lambda_{b/a}$ and using simple algebraic manipulations, one gets for any $a \in A$ such that $p(a) > 0$ and $q(a) > 0$,

$$\begin{aligned} \frac{1}{1 + \sum_{b \neq a} p(b)/p(a)} - \frac{1}{1 + \sum_{b \neq a} \lambda_{b/a}} &= \frac{\sum_{b \neq a} (\lambda_{b/a} - p(b)/p(a))}{\left(1 + \sum_{b \neq a} p(b)/p(a)\right) \left(1 + \sum_{b \neq a} \lambda_{b/a}\right)} \\ &= p(a) \frac{r(a)}{q(a) + r(a)}, \end{aligned}$$

where

$$r(a) = \sum_b [q(b)p(a) - p(b)q(a)]_+.$$

Furthermore, for any $a \in A$ such that $p(a) > 0$ and $q(a) = 0$ one gets

$$\frac{1}{1 + \sum_{b \neq a} p(b)/p(a)} = p(a) = p(a) \frac{r(a)}{q(a) + r(a)}.$$

Summing on every a , one gets finally

$$\mathbb{P}[g(U, p) \neq g(U, q)] = \sum_{a \in A} p(a) \frac{r(a)}{q(a) + r(a)},$$

If $A = \{a, a'\}$ and, for example, $q(a) < p(a)$, then $r(a) = p(a) - q(a)$ and $r(a') = 0$, hence

$$\mathbb{P}[g(U, p) \neq g(U, q)] = r(a) = p(a) - q(a) = \|p - q\|.$$

In the general case, note that

$$q(a) + r(a) \geq q(a) + \sum_b (q(b)p(a) - p(b)q(a)) = p(a),$$

hence

$$\begin{aligned} \mathbb{P}[g(U, p) \neq g(U, q)] &\leq \sum_{a \in A} r(a) = \sum_{a \in A} \sum_{b \neq a} [p(a)q(b) - q(a)p(b)]_+ \\ &\leq \sum_{a \in A} \sum_{b \neq a} p(a) \left[[q(b) - p(b)]_+ + p(b) [p(a) - q(a)]_+ \right], \end{aligned}$$

where the last inequality stems from the fact that $(u + v)_+ \leq (u)_+ + (v)_+$ for every u and v . Finally, the last double sum is at most $2\|p - q\|$, which ends the proof in the general case. \square

Recall that if p and q are two *fixed* probabilities on A , then for every random variables Z_p and Z_q with laws p and q defined on the same probability space,

$$\mathbb{P}[Z_p \neq Z_q] \geq \|p - q\|.$$

Conversely, a standard construction in coupling theory provides random variables Z_p and Z_q with laws p and q such that $\mathbb{P}[Z_p \neq Z_q] = \|p - q\|$.

The interest of proposition 4.2.6 is to provide a global coupling of *all* probabilities on A . One can wonder whether the constant 2 in this proposition can be improved. Our next result (not used in the sequel) shows that the constant 2 is optimal for the coupling $(g(U, p))_p$, and that it is not possible to do much better with any other global coupling.

Proposition 4.2.7 (Optimality of the upper bound of the error)

If $\#A \geq 3$, the constant 2 in the inequality $\mathbb{P}[g(U, p) \neq g(U, q)] \leq 2\|p - q\|$ of proposition 4.2.6 is optimal.

Furthermore, if $(Z_p)_p$ is a family of random variables indexed by probabilities on A , where each Z_p follows the law p , then there exist two probabilities $p \neq q$ such that

$$\mathbb{P}[Z_p \neq Z_q] \geq 2(1 - 1/\#A)\|p - q\|.$$

Proof. The first part of the proposition follows from the explicit example where $\{a, b, c\} \subset A$, $p(a) = q(a) = 1 - \varepsilon$ and $p(b) = q(c) = \varepsilon$ in the limit $\varepsilon \rightarrow 0$.

As regards the second part, let $N = \#A$ and for every a in A , let Z^a denote the random variable of $(Z_p)_p$ with uniform distribution on $A \setminus \{a\}$. Choose $a_0 \in A$, and consider the random set D of the elements a of A such that $Z^a = Z^{a_0}$. For every $a, b \in A$,

$$\mathbf{1}_{[Z^a \neq Z^b]} \geq \mathbf{1}_{[a \in D, b \notin D]} + \mathbf{1}_{[a \notin D, b \in D]}.$$

By summing over $a, b \in A$ and by taking expectations, one gets

$$\sum_{a,b} \mathbb{P}[Z^a \neq Z^b] \geq \mathbb{E}[2\#D(N - \#D)].$$

Of course $a_0 \in D$, and $Z^{a_0} \notin D$, since $Z^a \neq a$ almost surely for every a . Thus $1 \leq \#D \leq N - 1$. Hence,

$$\sum_{a,b} \mathbb{P}[Z^a \neq Z^b] \geq 2(N - 1).$$

There are at most $N(N - 1)$ nonzero terms in the sum above, hence there exist $a \neq b$ such that

$$\mathbb{P}[Z^a \neq Z^b] \geq 2/N.$$

Since $\|p_a - p_b\| = 1/(N - 1)$, this yields $\mathbb{P}[Z^a \neq Z^b] \geq 2(1 - 1/N)\|p_a - p_b\|$, which ends the proof. \square

4.2.3 Proof of the result with maximum influences

We now prove 11. Let $W = (W_n)_{n \in \mathbb{Z}}$ be an i.i.d. sequence of random variables, uniformly distributed on S , independent of the process X . Applying the construction of governing sequences in section 4.2.1, we introduce, for every n in \mathbb{Z} ,

$$U_n = f_{X_n}(W_n, P_{n-1}) \quad \text{with} \quad P_{n-1} = p(\cdot | X_{n-1}^{\triangleleft}).$$

Let $n \in \mathbb{Z}$. Thanks to the stationarity of the process X and to the independence of X and W , P_n is the conditional law of X_{n+1} given $\mathcal{F}_n^{X,W}$. Since W_{n+1} is independent of $\mathcal{F}_n^{X,W}$ and X_{n+1} , lemma 4.2.5 yields that

1. U_{n+1} is independent of $\mathcal{F}_n^{X,W}$, and therefore of $\mathcal{F}_n^{X,U}$,
2. U_{n+1} is uniformly distributed on S ,
3. $X_{n+1} = g(U_{n+1}, P_n)$ almost surely.

For every T in \mathbb{Z} , we now define a process X^T which is a function of $(U_n)_{n \geq T+1}$ in such a way that X^T approximates X when $T \rightarrow -\infty$.

Let $X_n^T = a_0$ for $n \leq T$ with $a_0 \in A$ fixed, and assume that X_n^T is defined up to time $n \geq T$. Define

$$X_{n+1}^T = g(U_{n+1}, P_n^T) \quad \text{where} \quad P_n^T = p(\cdot | (X^T)_n^{\triangleleft}).$$

Proposition 4.2.6 implies that for $n \geq T$,

$$\mathbb{P}[X_{n+1} \neq X_{n+1}^T \mid \mathcal{F}_n^{X,U}] = \mathbb{P}[g(U_{n+1}, P_n) \neq g(U_{n+1}, P_n^T) \mid \mathcal{F}_n^{X,U}] \leq 2\|P_n - P_n^T\|,$$

because P_n and P_n^T are measurable for $\mathcal{F}_n^{X,U}$ and U_{n+1} is independent of $\mathcal{F}_n^{X,U}$.

For n in \mathbb{Z} , let L_n^T count the number of consecutive times before n such that X^T and X coincide, that is,

$$L_n^T = \max\{k \geq 0 : X_{n-k+1:n}^T = X_{n-k+1:n}\}.$$

On the event $\{L_n^T = \ell\}$, the sequences X_n^{\triangleleft} and $(X^T)_n^{\triangleleft}$ coincide on their last ℓ terms. Hence, on the event $\{L_n^T = \ell\}$,

$$\|P_n - P_n^T\| \leq \sup \left\{ \|p(\cdot | xz) - p(\cdot | yz)\| ; x \in A^{\triangleleft}, y \in A^{\triangleleft}, z \in A^{\ell} \right\} = \delta_{\ell}.$$

One gets

$$\mathbb{P}[X_{n+1} \neq X_{n+1}^T \mid \mathcal{F}_n^{X,U}] \leq 2\delta_{L_n^T}.$$

The end of our proof follows the method in [2]: consider a \mathbb{Z}^+ -valued Markov chain, $Z = (Z_n)_{n \geq 0}$ starting from $Z_0 = 0$, with transition probabilities

$$p_{i,i+1} = 1 - 2\delta_i, \quad p_{i,0} = 2\delta_i, \quad \text{for every } i \geq 0.$$

For any $n \geq T$, it happens that L_n^T dominates stochastically Z_{n-T} , in the sense of the following lemma.

Lemma 4.2.8 *For every $k \geq 0$ and $n \geq T$, $\mathbb{P}[L_n^T \geq k] \geq \mathbb{P}[Z_{n-T} \geq k]$.*

Proof of lemma 4.2.8. The result is obvious for $n = T$ since $Z_0 = 0$. Assume that the result holds for $n \geq T$. Then,

$$\begin{aligned} \mathbb{P}[L_{n+1}^T \geq k+1] &= \mathbb{P}[L_n^T \geq k, X_{n+1} = X_{n+1}^T] \\ &= \mathbb{E}[\mathbf{1}_{\{L_n^T \geq k\}} \mathbb{P}[X_{n+1} = X_{n+1}^T \mid \mathcal{F}_n^{X,U}]] \\ &\geq \mathbb{E}[\mathbf{1}_{\{L_n^T \geq k\}} (1 - 2\delta_{L_n^T})]. \end{aligned}$$

Since $(\delta_i)_i$ is non-increasing and since $2\delta_i < 1$ for every i , the sequence indexed by i of general term $\mathbf{1}_{\{i \geq k\}}(1 - 2\delta_i)$ is non-decreasing. By induction, one obtains

$$\begin{aligned} \mathbb{P}[L_{n+1}^T \geq k + 1] &\geq \mathbb{E}[\mathbf{1}_{\{Z_{n-T} \geq k\}}(1 - 2\delta_{Z_{n-T}})] \\ &= \mathbb{P}[Z_{n-T} \geq k, Z_{n-T+1} = Z_{n-T} + 1] \\ &= \mathbb{P}[Z_{n-T+1} \geq k + 1], \end{aligned}$$

which ends the recurrence over $n \geq T$ and the proof of the lemma. \square

Using this to estimate $\mathbb{P}[L_n^T = 0]$, one gets

$$\mathbb{P}[X_n^T \neq X_n] = \mathbb{P}[L_n^T = 0] \leq \mathbb{P}[Z_{n-T} = 0].$$

Let μ be the measure defined on \mathbb{Z}^+ by

$$\nu(k) = \prod_{n=0}^k (1 - 2\delta_n),$$

for every $k \geq 0$. the hypothesis of theorem 11 ensure that μ has infinite mass.

If $\prod_{n=0}^{+\infty} (1 - 2\delta_n) = 0$, then μ is invariant and the state 0 is recurrent. But the chain Z is irreducible since the hypothesis of theorem 11 forces the positivity of the probabilities $(1 - 2\delta_n)$. Hence Z is null recurrent.

If $\prod_{n=0}^{+\infty} (1 - 2\delta_n) > 0$, then Z is transient since for every i , the probability of never returning to i from i is $\prod_{n=i}^{+\infty} (1 - 2\delta_n) > 0$. In both cases

$$\mathbb{P}[X_n^T \neq X_n] \rightarrow 0 \text{ when } T \rightarrow -\infty.$$

In other words, X_n^T converges in probability to X_n when $T \rightarrow -\infty$, in particular X_n is measurable for \mathcal{F}_n^U , which proves that U generate X . Using lemma 4.1.1 one gets that the filtration \mathcal{F}^X is immersed in \mathcal{F}^U , therefore \mathcal{F}^X is standard. This ends the proof of theorem 11.

4.3 Average influences

This section is devoted to the proof of theorem 12.

4.3.1 Priming lemma

Recall that the governing sequence U with values in S and based on lemma 4.2.5 is defined by

$$U_n = f_{X_n}(W_n, P_{n-1}) \quad \text{where} \quad P_{n-1} = p(\cdot | X_{n-1}^{\leq}),$$

where $W = (W_n)_{n \in \mathbb{Z}}$ is a sequence of i.i.d. random variables uniform on S , independent of X . Recall also that, from lemma 4.2.5, $X_n = g(U_n, P_{n-1})$ almost surely for every $n \geq 0$.

Let $\ell > 0$. Let us show that with probability close to 1, for each x in A^ℓ , $X_{1:\ell} = x$ as soon as $U_{1:\ell} \in B_x$ where B_x is a measurable subset of S^ℓ with μ -measure independent of x .

Recall that X satisfies the priming condition if for every a in A , $p(a|X_{-1}^\triangleleft) > 0$ almost surely.

Lemma 4.3.1 (Priming lemma) *Set $\ell > 0$. If X verifies the priming condition, then for every $\varepsilon \in]0, 1[$, there exist a real number $\beta_\ell > 0$ and a collection $(B_x)_{x \in A^\ell}$ of Borel sets of S^ℓ such that for every $x \in A^\ell$,*

$$\mu^{\otimes \ell}[B_x] = \beta_\ell \quad \text{and} \quad \mathbb{P}[X_{1:\ell} = x \mid U_{1:\ell} \in B_x] \geq 1 - \varepsilon.$$

Therefore if Y is a random variable valued in A^ℓ independent of $(X_n, U_n)_{n \in \mathbb{Z}}$,

$$\mathbb{P}[U_{1:\ell} \in B_Y] = \beta_\ell \quad \text{and} \quad \mathbb{P}[X_{1:\ell} = Y \mid U_{1:\ell} \in B_Y] \geq 1 - \varepsilon.$$

Proof. For every $n \in \mathbb{Z}$, let $P_n = p(\cdot | X_n^\triangleleft)$. Thanks to the stationarity of the process X , P_n is the conditional law of X_{n+1} given \mathcal{F}_n^X and the priming condition ensures that the support of P_n is A almost surely.

Let $\varepsilon \in]0, 1[$ and $\ell > 0$. For any fixed $x \in A^\ell$ let us construct by induction Borel sets B_1^x, \dots, B_ℓ^x of S with positive measure such that for every $m \in \{1, \dots, \ell\}$,

$$\mathbb{P}(C_m) \geq \left(1 - \frac{\varepsilon}{\ell}\right) \mu(B_m^x) \mathbb{P}(C_{m-1}) > 0,$$

where $C_0 = \Omega$ and for every $m \in \{1, \dots, \ell\}$,

$$C_m = \{X_{1:m} = x_{1:m} ; U_{1:m} \in B_1^x \times \dots \times B_m^x\}.$$

Let $m \in \{1, \dots, \ell\}$. Assume that B_1^x, \dots, B_{m-1}^x are constructed verifying the induction hypothesis. Since $\mathbb{P}[C_{m-1}] > 0$, one gets, thanks to the priming condition,

$$\mathbb{P}[P_{m-1}(x_m) = 0 \mid C_{m-1}] = 0.$$

Therefore one can choose a real number $q \in]0, 1]$ such that

$$\mathbb{P}[P_{m-1}(x_m) \leq q \mid C_{m-1}] < \frac{\varepsilon}{\ell}.$$

Set $U_m = (U_{m,1}, \dots, U_{m,N})$. Since $X_m = g(U_m, P_{m-1})$ almost surely, one gets up to negligible events,

$$\begin{aligned} \{X_m = x_m\} &\supset \{X_m = x_m ; P_{m-1}(x_m) > q\} \\ &= \{g(U_m, P_{m-1}) = x_m ; P_{m-1}(x_m) > q\} \\ &= \left\{ \frac{U_{m,x_m}}{P_{m-1}(x_m)} = \min_{k \in A} \frac{U_{m,k}}{P_{m-1}(k)} ; P_{m-1}(x_m) > q \right\} \\ &\supset \left\{ \frac{U_{m,x_m}}{q} \leq \min_{k \neq x_m} U_{m,k} ; P_{m-1}(x_m) > q \right\}. \end{aligned}$$

Set

$$B_m^x = \left\{ (y_1, \dots, y_N) \in S ; \frac{y_{x_m}}{q} \leq \min_{k \neq x_m} y_k \right\}.$$

Then $\mu(B_m^x) > 0$ and

$$\{X_m = x_m ; U_m \in B_m^x\} \supset \{U_m \in B_m^x ; P_{m-1}(x_m) > q\}.$$

Since

$$C_m = \{X_m = x_m ; U_m \in B_m^x\} \cap C_{m-1},$$

the independence of U_m and $\mathcal{F}_{m-1}^{X,U}$ and the choice of q yield

$$\begin{aligned} \mathbb{P}[C_m] &\geq \mathbb{P}[U_m \in B_m^x ; P_{m-1}(x_m) > q ; C_{m-1}] \\ &= \mu(B_m^x) \mathbb{P}[P_{m-1}(x_m) > q ; C_{m-1}] \\ &\geq \mu(B_m^x) \left(1 - \frac{\varepsilon}{\ell}\right) \mathbb{P}[C_{m-1}]. \end{aligned}$$

Therefore $\mathbb{P}[C_m] > 0$.

By reducing the Borel set B_ℓ^x at the last step of the induction, one can make the measure $\mu^{\otimes \ell}[B_1^x \times \dots \times B_\ell^x]$ independent of $x \in A^\ell$. Denote by β_ℓ this measure, then set $B_x = B_1^x \times \dots \times B_\ell^x$. One gets

$$\mathbb{P}[X_{1:\ell} = x \mid U_{1:\ell} \in B_x] = \frac{\mathbb{P}[X_{1:\ell} = x ; U_{1:\ell} \in B_x]}{\mathbb{P}[U_{1:\ell} \in B_x]}.$$

By independence,

$$\mathbb{P}[X_{1:\ell} = x, U_{1:\ell} \in B_x] \geq \prod_{k=1}^{\ell} \mu(B_k^x) \left(1 - \frac{\varepsilon}{\ell}\right), \quad \text{and} \quad \mathbb{P}[U_{1:\ell} \in B_x] = \prod_{k=1}^{\ell} \mu(B_k^x),$$

hence

$$\mathbb{P}[X_{1:\ell} = x \mid U_{1:\ell} \in B_x] \geq \left(1 - \frac{\varepsilon}{\ell}\right)^\ell \geq 1 - \varepsilon,$$

which ends the proof. \square

4.3.2 Approximation until a given time

Choose $\varepsilon > 0$ and $\ell \geq 1$ such that $\sum_{n \geq \ell} \eta_n \leq \varepsilon$, let $J = [s, t]$ be an interval of integers such that $t - s + 1 = \ell$.

Then, let Y be a random variable taking values in A^ℓ , independent of $(X_n, U_n)_{n \in \mathbb{Z}}$ and distributed like X_J .

Lemma 4.3.1 provides a real number β_ℓ and Borel sets $(B_x)_{x \in A^\ell}$, such that

$$\mathbb{P}[X_J = Y \mid U_J \in B_Y] \geq 1 - \varepsilon \quad \text{and} \quad \mathbb{P}[U_J \in B_Y] = \beta_\ell.$$

Using Y and the governing sequence $(U_n)_{n \geq t+1}$, let us construct random variables $(X'_n)_{n \geq s}$ by taking $X'_J = Y$ and for every $n > t$

$$X'_n = g(U_n, P'_{n-1}) \quad \text{where} \quad P'_{n-1} = p(\cdot \mid X'_{s:n-1}).$$

The random variable Y is useful in the proof of our following result.

Lemma 4.3.2 *For every $n \geq s$, the law of $X'_{s:n}$ is the law of $X_{s:n}$.*

Proof. For every $n \geq t + 1$, $y \in A^{n-s}$ and all $x \in A$,

$$\mathbb{P}[X'_n = x \mid X'_{s:n-1} = y] = p(x|y) = \mathbb{P}[X_n = x \mid X_{s:n-1} = y].$$

Since the law of $X'_J = Y$ is the same as the law of X_J , the result follows by induction. \square

Lemma 4.3.3 *One has $\mathbb{P}[X' \neq X \text{ on } [s, +\infty[\mid U_J \in B_Y] \leq 3\varepsilon$.*

Proof. Since $X_n = g(U_n, P_{n-1})$ and $X'_n = g(U_n, P'_{n-1})$, proposition 4.2.6 yields for $n > t$,

$$\mathbb{P}[X'_n \neq X_n \mid \mathcal{F}_{n-1}^{X,U} \vee \sigma(Y)] \leq 2\|P'_{n-1} - P_{n-1}\|.$$

Let

$$(\star) = \mathbb{P}[X'_n \neq X_n ; X'_{s:n-1} = X_{s:n-1} ; U_J \in B_Y].$$

Since

$$\{X'_{s:n-1} = X_{s:n-1} ; U_J \in B_Y\} \in \mathcal{F}_{n-1}^{X,U} \vee \sigma(Y),$$

one gets

$$\begin{aligned}
(\star) &\leq \mathbb{E}\left[2\|P_{n-1} - P'_{n-1}\| \mathbf{1}_{\{X'_{s:n-1}=X_{s:n-1}\}} \mathbf{1}_{\{U_J \in B_Y\}}\right] \\
&= 2 \sum_{\substack{y \in A^\ell \\ z \in A^{n-t-1}}} \mathbb{E}\left[\|p(\cdot|X_{s-1}^\triangleleft yz) - p(\cdot|yz)\| \mathbf{1}_{\{yz=X'_{s:n-1}=X_{s:n-1}\}} \mathbf{1}_{\{U_J \in B_Y\}}\right] \\
&\leq 2 \sum_{y,z} \mathbb{E}\left[\|p(\cdot|X_{s-1}^\triangleleft yz) - p(\cdot|yz)\| \mathbf{1}_{\{U_J \in B_Y\}} \mathbf{1}_{\{yz=X'_{s:n-1}\}}\right], \\
&= 2 \sum_{y,z} \mathbb{E}\left[\|p(\cdot|X_{s-1}^\triangleleft yz) - p(\cdot|yz)\| \mu^{\otimes \ell_k}(B_Y) \mathbb{P}[X'_{s:n-1} = yz]\right] \\
&= 2\beta_\ell \sum_{x \in A^{n-s}} \mathbb{E}\left[\|p(\cdot|X_{s-1}^\triangleleft x) - p(\cdot|x)\| \mathbb{P}[X_{s:n-1} = x]\right] \\
&= 2\beta_\ell \eta_{n-s},
\end{aligned}$$

where the last three equations stem from the independence of X_{s-1}^\triangleleft , U_J , $U_{t+1:n-1}$ and Y , from lemma 4.3.2 and from the definition of η_n . Hence,

$$\mathbb{P}[X'_n \neq X_n ; X'_{s:n-1} = X_{s:n-1} \mid U_J \in B_Y] \leq 2\eta_{n-s},$$

therefore,

$$\mathbb{P}[X'_{s:n} = X_{s:n} \mid U_J \in B_Y] \geq \mathbb{P}[X'_{s:n-1} = X_{s:n-1} \mid U_J \in B_Y] - 2\eta_{n-s}.$$

By induction, one gets

$$\mathbb{P}[X'_{s:n} = X_{s:n} \mid U_J \in B_Y] \geq \mathbb{P}[X'_J = X_J \mid U_J \in B_Y] - 2 \sum_{m=\ell}^{n-s} \eta_m.$$

Since $X'_J = Y$ and $\mathbb{P}[X_J = Y \mid U_J \in B_Y] \geq 1 - \varepsilon$, this yields

$$\mathbb{P}[X' = X \text{ on } [s, +\infty[\mid U_J \in B_Y] \geq 1 - \varepsilon - 2 \sum_{m=\ell}^{\infty} \eta_m \geq 1 - 3\varepsilon,$$

which ends the proof. \square

4.3.3 Successive approximations and end of the proof of the result with average influences

Our next step in the proof of theorem 12 is to approach the random variable X_0 by measurable functions of the governing sequence. To this aim, we group the innovations by intervals of times. For every $m > 0$ one chooses L_m such that

$$\sum_{n \geq L_m} \eta_n \leq 1/m.$$

For each m , lemma 4.3.1 (the priming lemma) applied to $\ell = L_m$ and $\varepsilon = 1/m$ provides a real number $\beta_{L_m} > 0$ and Borel sets $(B_x)_{x \in A^{L_m}}$ of S^{L_m} with measure β_{L_m} such that

$$\mathbb{P}[X_{1:L_m} = x \mid U_{1:L_m} \in B_x] \geq 1 - 1/m.$$

Choose an integer $M_m \geq 1/\beta_{L_m}$. Split \mathbb{Z}_+ into M_1 intervals of length L_1 , M_2 intervals of length L_2 , ... More precisely set, for every $n \geq 1$,

$$\ell_n = L_m \text{ if } M_1 + \cdots + M_{m-1} < n \leq M_1 + \cdots + M_m$$

and

$$\varepsilon_n = 1/m \text{ if } M_1 + \cdots + M_{m-1} < n \leq M_1 + \cdots + M_m$$

Therefore, for every $k \geq 0$ one gets

$$\sum_{n \geq \ell_k} \eta_n \leq \varepsilon_k.$$

At last, for every $k \geq 0$, set

$$t_k = - \sum_{1 \leq n \leq k} \ell_n$$

that is to say $t_0 = 0$ and $t_k = t_{k-1} - \ell_k$ for $k \geq 1$. Define, for $k \geq 0$, the interval of integers

$$J_k = [t_k, t_k + \ell_k - 1] = [t_k, t_{k-1} - 1] \text{ and } X_{J_k} = X_{t_k:t_{k-1}-1}.$$

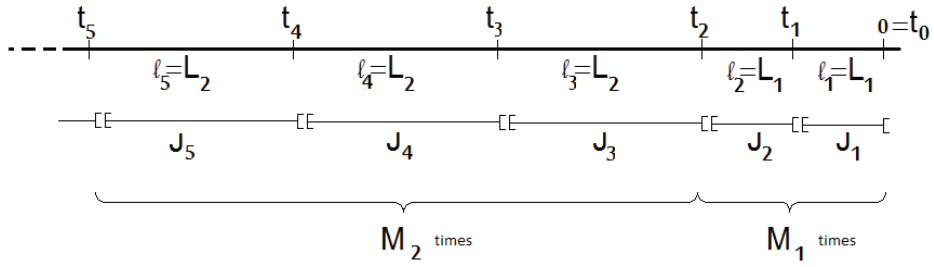


Figure 4.3: Splitting \mathbb{Z}_- in intervals of times.

Let $Y = (Y_k)_{k \geq 1}$ be a sequence of random variables, independent of $(X_n, U_n)_{n \in \mathbb{Z}}$ and such that for every $k \geq 1$, the law of Y_k is the law of $X_{1:\ell_k}$.

For every $k \geq 0$, let us use the construction of section 4.3.2: set $X_{J_k}^k = Y_k$, then for every $n \geq t_k + \ell_k = t_{k-1}$

$$X_n^k = g(U_n, P_{n-1}^k) \text{ where } P_{n-1}^k = p(\cdot | X_{t_k:n-1}^k).$$

Therefore lemma 4.3.3 yields the inequality

$$\mathbb{P}[X_{t_k:0}^k \neq X_{t_k:0}^k \mid U_{J_k} \in B_{Y_k}] \leq 3\varepsilon_k,$$

and

$$\mathbb{P}[X_0^k \neq X_0 \mid U_{J_k} \in B_{Y_k}] \rightarrow 0 \text{ when } k \rightarrow +\infty.$$

Moreover each event $\{U_{J_k} \in B_{Y_k}\}$ is independent of the others (indeed they are functions of random variables U_k for disjoint sets of indices k) and

$$\sum_{k \geq 1} \mathbb{P}[U_{J_k} \in B_{Y_k}] = \sum_{k \geq 1} \beta_{\ell_k} = \sum_{m=1}^{+\infty} M_m \beta_{L_m} = +\infty$$

since $M_m \beta_{L_m} \geq 1$ by choice of M_m .

Lemma 4.3.4, stated below, provides a deterministic increasing function θ such that

$$\sum_{k \geq 1} \mathbb{P}[X_0^{\theta(k)} \neq X_0 ; U_{J_{\theta(k)}} \in B_{Y_{\theta(k)}}] < +\infty$$

and $\sum_{k \geq 1} \mathbb{P}[U_{J_{\theta(k)}} \in B_{Y_{\theta(k)}}] = +\infty.$

Using Borel-Cantelli's lemma, one deduces that almost surely,

- $\{X_0^{\theta(k)} \neq X_0\} \cap \{U_{J_{\theta(k)}} \in B_{Y_{\theta(k)}}\}$ is realized for a finite number of k only,
- $\{U_{J_{\theta(k)}} \in B_{Y_{\theta(k)}}\}$ is realized for an infinite number of k .

Thus, for every $a \in A$,

$$\{X_0 = a\} = \limsup_{k \rightarrow \infty} \{U_{J_{\theta(k)}} \in B_{Y_{\theta(k)}}\} \cap \{X_0^{\theta(k)} = a\}.$$

Therefore, $\{X_0 = a\}$ belongs to $\mathcal{F}_0^U \vee \sigma(Y)$. Since the sequence $Y = (Y_k)_{k \geq 0}$ is independent of $\mathcal{F}_0^{U, X}$, one gets

$$\mathbb{P}[X_0 = a | \mathcal{F}_0^U] = \mathbb{P}[X_0 = a | \mathcal{F}_0^U \vee \sigma(Y)] = \mathbf{1}_{\{X_0 = a\}} \text{ a.s.,}$$

therefore $\{X_0 = a\} \in \mathcal{F}_0^U$. By stationarity of the process (X, U) , one gets the inclusion of the filtration \mathcal{F}^X into the filtration \mathcal{F}^U . Therefore lemma 4.1.1 yields that \mathcal{F}^X is immersed in \mathcal{F}^U , which ends the proof. \square

Lemma 4.3.4 *Let $(a_n)_{n \geq 0}$ and $(b_n)_{n \geq 0}$ denote two bounded sequences of nonnegative real numbers such that the series $\sum_n b_n$ diverges and such that $a_n \ll b_n$.*

Then there exists an increasing function $\theta : \mathbb{N} \rightarrow \mathbb{N}$ such that the series $\sum_n a_{\theta(n)}$ converges and the series $\sum_n b_{\theta(n)}$ diverges.

4.4 Examples

In this section we study some examples showing the advantages and the limitations of our results.

- Our first example (section 4.4.1) is a chain with memory of variable length which fulfills the hypotheses of theorem 12 but not those of theorem 10 nor theorem 11. Its natural filtration is standard.
- Our second example (section 4.4.2) is derived from the well known $[T, T^{-1}]$ transformation. It provides a stationary process with values in a finite space, whose natural filtration is non standard. This example does not fulfill any of the two conditions of theorem 12 (namely the priming condition and the summability of the gaps).
- Our third example (section 4.4.3) is a slight adaptation of the second one, where the filtration of the process is still non standard although the priming condition is fulfilled.
- Our fourth and last example (section 4.4.4) is another adaptation of the second example in which the filtration is standard although the condition of summability of the gaps is not fulfilled and the related conditional probabilities are close to those of the second example.

4.4.1 Parity of the number of ones in a row

This first example provides a setting where one proves standardness using theorem 12.

Let $(X_n)_{n \in \mathbb{Z}}$ be a stationary process taking values in $\{0, 1\}$ such that

$$\mathbb{P}[X_0 = 0 \mid \mathcal{F}_{-1}^X] = \frac{1}{3} + \frac{1}{3} \mathbf{1}_{\{T \text{ is even or } T = -\infty\}} \text{ where } T := \sup\{k < 0 : X_k = 0\}.$$

The existence of such a process is ensured by proposition 2.10 in [7]. A simple computation gives, for every $n \geq 0$,

$$\gamma_n = \frac{1}{2}, \quad \delta_n = \frac{1}{3}, \quad \alpha_n = \frac{1}{3}, \quad \eta_n \leq \frac{1}{3} \mathbb{P}[T < -n] \leq \left(\frac{2}{3}\right)^{n+1}.$$

Therefore this process fulfills the hypotheses of theorem 12 (and its corollary 4.1.2) but neither those of theorem 10 nor those of theorem 11. The filtration \mathcal{F}^X is standard.

4.4.2 Random walk in random scenery

The second example is a process whose filtration is non standard.

Let $X = (X_n)_{n \in \mathbb{Z}}$ and $C = (C_s)_{s \in \mathbb{Z}}$ be two independent sequences of i.i.d. random variables with uniform law on $\{-1, 1\}$. Set

$$\begin{aligned} S_n &= X_1 + \cdots + X_n \text{ if } n \geq 0 \\ S_n &= -X_{n+1} - \cdots - X_0 \text{ if } n < 0 \end{aligned} .$$

Therefore $S_{n+1} = S_n + X_{n+1}$ for every $n \in \mathbb{Z}$. Set $C_{S_n} = Y_n$. The stationary process Z defined by $Z_n = (X_n, Y_n)$ for every $n \in \mathbb{Z}$ and taking values in $A = \{-1, 1\}^2$ is called random walk in random scenery.

This process is derived from $((X_{n+}, C_{S_{n+}}))_{n \in \mathbb{Z}}$ where $X_{n+} = (X_{n+m})_{m \leq 0}$ is the trajectory of X until time n and $C_{S_{n+}} = (C_{S_{n+s}})_{s \in \mathbb{Z}}$ is the scenery seen from S_n . It is easy to prove that the processes Z and $(X_{n+}, C_{S_{n+}})_{n \in \mathbb{Z}}$ generate the same filtration. Indeed, given (X_k, C_{S_k}) for every $k \leq n$, one knows the trajectory X_{n+} and one can deduce the increments $(S_n - S_k)_{k \leq n}$. Since those increments visit almost surely every integer, one can recover the scenery seen from S_n .

The process $((X_{n+}, C_{S_{n+}}))_{n \in \mathbb{Z}}$ is the most famous $[T, T^{-1}]$ process. Indeed the $[T, T^{-1}]$ transformation is the application from $\{-1, 1\}^{\mathbb{Z}^-} \times \{-1, 1\}^{\mathbb{Z}}$ into itself defined by

$$[T, T^{-1}]((x_n)_{n \leq 0}, (c_s)_{s \in \mathbb{Z}}) = ((x_{n-1})_{n \leq 0}, (c_{s-x_0})_{s \in \mathbb{Z}}).$$

One checks that for every $n \leq 0$

$$(X_{n+}, C_{S_{n+}}) = [T, T^{-1}]^{-n}(X, C).$$

According to [10], the natural filtration of the process $((X_{n+}, C_{S_{n+}}))_{n \in \mathbb{Z}}$ is non standard though its asymptotic σ -field at $-\infty$ is trivial. Therefore the same holds for the natural filtration of Z .

Let $n \geq 0$. Let us study the probabilities $p(a|z)$ for $a \in A$ and $z \in A^n$. Note $z = (z_{-n}, \dots, z_{-1})$, $z_k = (x_k, y_k)$ and $a = (x_0, y_0)$.

First, note that some of the events

$$\left\{ Z_{-n:-1} = z \right\} = \left\{ X_{-n:-1} = x_{-n:-1} ; Y_{-n:-1} = y_{-n:-1} \right\}$$

are impossible. Indeed, by definition of the process Y for $i < j$, $Y_i = Y_j$ on the event $X_{i+1} + \cdots + X_j = 0$. When the event $\{Z_{-n:-1} = z\}$ is impossible, one says that z is not admissible. Note that to compute η_n , one only needs to consider probabilities $p(\cdot|z)$ and $p(\cdot|wz)$ for admissible $z \in A^n$ and $w \in A^\triangleleft$. Yet, wz may be non-admissible even if z and w are admissible.

Assume that z is admissible. Then $\mathbb{P}[X_0 = x_0 | Z_{-n:-1} = z] = 1/2$. If for some $i \in \{-n, \dots, -1\}$, $x_{i+1} + \cdots + x_0 = 0$, then the conditions $Z_{-n:-1} = z$ and $X_0 = x_0$

imply that $Y_0 = y_i$. Otherwise, the color Y_0 is independent of $Z_{-n:-1}$ and X_0 . Thus for any admissible word $z \in A^n$,

$$p(a|z) = \begin{cases} 1/2 & \text{if there exists } i \text{ such that } x_{i+1} + \cdots + x_0 = 0 \text{ and } y_0 = y_i, \\ 0 & \text{if there exists } i \text{ such that } x_{i+1} + \cdots + x_0 = 0 \text{ and } y_0 \neq y_i, \\ 1/4 & \text{otherwise.} \end{cases}$$

Therefore for every $n \geq 0$, $\gamma_n = 1$ and $\delta_n = \alpha_n = 1/2$.

Furthermore, for almost every admissible word $w = (x_n, y_n)_{n < 0}$ in A^\triangleleft , there exists $t < 0$ such that $x_{t+1} + \cdots + x_0 = 0$ and the same argument gives that

$$p(a|w) = \frac{1}{2} \mathbf{1}_{\{y_t=y_0\}}.$$

For non-admissible $w \in A^\triangleleft$, the value of $p(a|w)$ can be chosen arbitrarily. Set

$$p(a|w) = \frac{1}{2} \mathbf{1}_{\{y_d=y_0\}} \text{ where } d = \sup\{t \leq -1 : x_{t+1} + \cdots + x_0 = 0\}$$

if d is well defined, and $p(a|w) = 1/4$ otherwise.

With this convention, one gets that, for almost any admissible $w \in A^\triangleleft$ and $z = ((x_{-n}, y_{-n}), \cdots, (x_{-1}, y_{-1})) \in A^n$,

$$|p(a|z) - p(a|wz)| = \begin{cases} 0 & \text{if there exists } t \in [-n, -1] \text{ such that } x_t + \cdots + x_0 = 0 \\ 1/4 & \text{otherwise.} \end{cases}$$

Therefore, for every $n \geq 0$,

$$\begin{aligned} \eta_n &= \frac{1}{4} \left(\mathbb{P}[\forall i \in \{-n, \dots, -1\}, X_{i+1} + \cdots + X_{-1} \geq 0] \right. \\ &\quad \left. + \mathbb{P}[\forall i \in \{-n, \dots, -1\}, X_{i+1} + \cdots + X_{-1} \leq 0] \right) \\ &= \frac{1}{2} \mathbb{P}[\forall i \in \{-n, \dots, -1\}, X_{i+1} + \cdots + X_{-1} \geq 0]. \end{aligned}$$

One sees that $\eta_n \sim C/\sqrt{n}$ with $C \in \mathbb{R}_*^+$. Hence the process Z does not fulfill any of the hypotheses of theorem 12.

4.4.3 Random walk in random scenery with misreadings

We construct a variant of the random walk in random scenery which fulfills the priming condition but whose natural filtration is non standard.

Construct $Z_n = (X_n, Y_n)$ as in section 4.4.2. Fix $q \in]0, 1/2[$. Let $(\xi_n)_{n \in \mathbb{Z}}$ be a sequence of i.i.d. random variables taking values in $\{-1, 1\}$, independent of \mathcal{F}^Z and such that $\mathbb{P}[\xi_0 = 1] = 1 - q$. Define a process $(Z'_n)_{n \in \mathbb{Z}}$ by

$$Z'_n = (X_n, Y_n \xi_n).$$

The process $(Z'_n)_{n \in \mathbb{Z}}$ is a random walk in random scenery in which at each time, one misreads the color of the site Y_n with probability q .

The processes Z' and $(Z_n, \xi_n)_{n \in \mathbb{Z}}$ generate the same filtration. Indeed, the random variables ξ_m associated to the times $m < n$ where $S_m = S_n$ are independent and take the value 1 with probability $1 - q > 1/2$, therefore the color Y_n is the most common color among the colors $Y_m \xi_m$ seen at those times. Therefore, for almost every $z \in A^\triangleleft$ and $a \in A$, the corresponding conditional probability $p'(a|z)$ is equal to $q/2$ or to $(1 - q)/2$ depending on these colors.

Moreover, by independent enlargement, \mathcal{F}^Z is immersed into $\mathcal{F}^{Z, \xi} = \mathcal{F}^{Z'}$. Since \mathcal{F}^Z is non standard, one deduces that $\mathcal{F}^{Z'}$ is non standard either.

By a short calculation, one gets for every $n > 0$,

$$\gamma_n = \frac{q}{1 - q}, \quad \delta_n = \frac{1 - 2q}{4}, \quad \alpha_n = 1 - 2q.$$

Since the probabilities $p'(a|z)$ related to this process satisfy $p'(a|z) \geq q/2$, for every $a \in A$ and $z \in A^\triangleleft$, the priming condition is fulfilled. The exact value of $p'(a|z)$ for $z \in A^n$ is difficult to compute, but the corresponding gaps η'_n verify

$$\begin{aligned} \eta'_n &\geq \left(\frac{1}{4} - \frac{q}{2} \right) \left(\mathbb{P}[\forall i \in \{-l, \dots, -1\}, X_{i+1} + \dots + X_{-1} \geq 0] \right. \\ &\quad \left. + \mathbb{P}[\forall i \in \{-l, \dots, -1\}, X_{i+1} + \dots + X_{-1} \leq 0] \right) \\ &= (1 - 2q)\eta_n. \end{aligned}$$

Therefore for $q < 1/2$, the sequence $(\eta'_n)_n$ is not summable, thus the process Z' does not verify the condition of summability of the gaps.

4.4.4 Random walk in renewed random scenery

We construct another variant of the random walk in random scenery in which the natural filtration is standard although the condition of summability of the gaps of theorem 12 is not fulfilled.

We consider a variant of the process $(X_n, C_{S_n+})_{n \in \mathbb{Z}}$ in which at each time n the color at 0 of the scenery seen from S_n is changed with probability $q \in]0, 1/2[$. For every $g \in \{-1, 1\}^{\mathbb{Z}}$, denote $\bar{g} \in \{-1, 1\}^{\mathbb{Z}}$ the application defined by

$$\bar{g}(s) = g(s) \text{ for } s \neq 0 \text{ and } \bar{g}(0) = -g(0).$$

Let (X_n, G_n) be a stationary Markov chain with values in $\{-1, 1\} \times \{-1, 1\}^{\mathbb{Z}}$, with transition probabilities

$$p((x, g), (x', g')) = \begin{cases} (1-q)/2 & \text{if } g' = g(x' + \cdot) \\ q/2 & \text{if } g' = \bar{g}(x' + \cdot) \end{cases}.$$

The random walk in renewed random scenery is the process $Z'' = (Z''_n)_{n \in \mathbb{Z}}$ defined by $Z''_n = (X_n, G_n(0))$.

The corresponding probabilities $p''(a|z)$ are close to the probabilities $p(a|z)$. Indeed,

$$p''(a|z) = \begin{cases} (1-q)/2 & \text{if } p(a|z) = 1/2 \\ q/2 & \text{if } p(a|z) = 0 \\ 1/4 & \text{if } p(a|z) = 1/4 \end{cases}.$$

Therefore the corresponding gaps verify

$$\eta''_n = (1 - 2q)\eta_n.$$

To show that the filtration $\mathcal{F}^{Z''}$ is standard, one can use the following trick: instead of changing the color at 0 of the scenery G_n with probability q , one draws at random this color with probability $2q$. One needs a random variable ε_n taking the value 1 if this drawing occurs and 0 otherwise, and a random variable κ_n giving the color obtained if the drawing occurs.

To construct these random variables, consider two independent sequences of random variables $(\beta_n)_{n \in \mathbb{Z}}$ and $(V_n)_{n \in \mathbb{Z}}$, independent of $\mathcal{F}^{X, G}$ such that

- the β_n are i.i.d. Bernoulli variables of parameter $(1 - 2q)/(1 - q)$,
- the V_n are i.i.d. and uniform on $\{-1, 1\}$.

Let, for every $n \in \mathbb{Z}$,

$$\varepsilon_n = 1 - \beta_n \mathbf{1}_{\{G_n(0) = G_{n-1}(X_n)\}} \quad \text{and} \quad \kappa_n = G_n(0) \mathbf{1}_{\{\varepsilon_n = 1\}} + V_n \mathbf{1}_{\{\varepsilon_n = 0\}}.$$

Let us show that the random variables $U_n = (X_n, \varepsilon_n, \kappa_n)$ constitute a governing sequence for the process Z'' . Given Z''_{n-1} and U_n , one deduces Z''_n thanks to the equalities

$$\begin{aligned} G_n(s) &= G_{n-1}(X_n + s) & \text{if } \varepsilon_n = 0 \text{ or } s \neq 0 \\ G_n(0) &= \kappa_n & \text{if } \varepsilon_n = 1. \end{aligned}$$

It remains to check that U_n is independent of the σ -field $\mathcal{G}_{n-1} = \mathcal{F}_{n-1}^{X,G,\beta,V}$ and *a fortiori* of $\mathcal{F}_{n-1}^{Z'',U}$. Thanks to the independence of the processes β , V and (X, G) one gets for every $x \in \{-1, 1\}$,

$$\mathbb{P}[X_n = x ; G_n = G_{n-1}(X_n + \cdot) | \mathcal{G}_{n-1}] = (1 - q)/2,$$

$$\mathbb{P}[X_n = x ; G_n \neq G_{n-1}(X_n + \cdot) | \mathcal{G}_{n-1}] = q/2.$$

Therefore, for every c and x in $\{-1, 1\}$,

$$\begin{aligned} \mathbb{P}[\varepsilon_n = 1 ; \kappa_n = c ; X_n = x | \mathcal{G}_{n-1}] &= \mathbb{P}[\varepsilon_n = 1 ; G_n(0) = c ; X_n = x | \mathcal{G}_{n-1}] \\ &= (1) + (2) + (3), \end{aligned}$$

with

$$(1) = \mathbb{P}[\beta_n = 0 ; G_n \neq G_{n-1}(X_n + \cdot) ; G_{n-1}(X_n) = -c ; X_n = x | \mathcal{G}_{n-1}],$$

$$(2) = \mathbb{P}[\beta_n = 0 ; G_n = G_{n-1}(X_n + \cdot) ; G_{n-1}(X_n) = c ; X_n = x | \mathcal{G}_{n-1}],$$

$$(3) = \mathbb{P}[\beta_n = 1 ; G_n \neq G_{n-1}(X_n + \cdot) ; G_{n-1}(X_n) = -c ; X_n = x | \mathcal{G}_{n-1}].$$

One gets

$$\begin{aligned} (1) &= \frac{q}{1-q} \times \mathbb{P}[G_n \neq G_{n-1}(X_n + \cdot) ; G_{n-1}(x) = -c ; X_n = x | \mathcal{G}_{n-1}] \\ &= \frac{q}{2} \times \frac{q}{1-q} \times \mathbb{P}[G_{n-1}(x) = -c | \mathcal{G}_{n-1}] \\ &= \frac{q}{2} \times \frac{q}{1-q} \times \mathbf{1}_{\{G_{n-1}(x)=-c\}}, \end{aligned}$$

$$\begin{aligned} (2) &= \frac{q}{1-q} \times \mathbb{P}[G_n = G_{n-1}(X_n + \cdot) ; G_{n-1}(x) = c ; X_n = x | \mathcal{G}_{n-1}] \\ &= \frac{1-q}{2} \times \frac{q}{1-q} \times \mathbb{P}[G_{n-1}(x) = c | \mathcal{G}_{n-1}] \\ &= \frac{q}{2} \times \mathbf{1}_{\{G_{n-1}(x)=c\}}, \end{aligned}$$

and

$$\begin{aligned} (3) &= \frac{1-2q}{1-q} \times \mathbb{P}[G_n \neq G_{n-1}(X_n + \cdot) ; G_{n-1}(x) = -c ; X_n = x | \mathcal{G}_{n-1}] \\ &= \frac{q}{2} \times \frac{1-2q}{1-q} \times \mathbb{P}[G_{n-1}(x) = -c | \mathcal{G}_{n-1}] \\ &= \frac{q}{2} \times \frac{1-2q}{1-q} \times \mathbf{1}_{\{G_{n-1}(x)=-c\}}. \end{aligned}$$

Thus, for every c and x in $\{-1, 1\}$,

$$\mathbb{P}[\varepsilon_n = 1 ; \kappa_n = c ; X_n = x \mid \mathcal{G}_{n-1}] = \frac{q}{2}.$$

Moreover, by independence of β_n , V_n and \mathcal{G}_{n-1} ,

$$\begin{aligned} & \mathbb{P}[\varepsilon_n = 0 ; \kappa_n = c ; X_n = x \mid \mathcal{G}_{n-1}] \\ &= \mathbb{P}[\beta_n = 1 ; G_n = G_{n-1}(X_n + \cdot) ; V_n = c ; X_n = x \mid \mathcal{G}_{n-1}] \\ &= \frac{1-2q}{1-q} \times \frac{1}{2} \times \frac{1-q}{2} \\ &= \frac{1-2q}{4}. \end{aligned}$$

This shows that the random variables $U_n = (X_n, \varepsilon_n, \kappa_n)$ constitute a governing sequence for the process Z'' .

Let us show the inclusion $\mathcal{F}_n^{Z''} \subset \mathcal{F}_n^U$ for any $n \in \mathbb{Z}$, that is to say, that the sequence $(U_k)_{k \leq n}$ is sufficient to recover the scenery G_n seen from S_n . The variables $(X_k)_{k \leq n}$ determine the increments $(S_n - S_k)_{k \leq n}$ and for every $s \in \mathbb{Z}$, $S_n - S_k = s$ for an infinite number of times $k \leq n$. Among those times, there is an infinite number of times such that $\varepsilon_k = 1$. The value of κ_k at the last time $k \leq n$ such that $S_n - S_k = s$ and $\varepsilon_k = 1$ is equal to $G_n(s)$. Therefore $\mathcal{F}^G \subset \mathcal{F}^U$, and since $\mathcal{F}^X \subset \mathcal{F}^U$, one gets $\mathcal{F}^{Z''} \subset \mathcal{F}^U$. Finally, lemma 4.1.1 yields that $\mathcal{F}^{Z''}$ is immersed in \mathcal{F}^U , therefore the natural filtration of the process $(Z''_n)_{n \in \mathbb{Z}}$ is standard.

4.5 Proofs of auxiliary facts

4.5.1 Inequalities involving α_n , δ_n , γ_n and η_n

To prove that $\delta_n \leq \gamma_n$ for every $n \geq 0$, consider x and y in A^\triangleleft and $z \in A^n$. Then,

$$\begin{aligned} \|p(\cdot|xz) - p(\cdot|yz)\| &= \sum_{a \in A} [p(a|xz) - p(a|yz)]_+ \\ &= \sum_{a \in A} p(a|xz) \left(1 - \frac{p(a|yz)}{p(a|xz)}\right)_+ \leq \sum_{a \in A} p(a|xz) \gamma_n = \gamma_n. \end{aligned}$$

Taking the supremum over x , y and z , one gets $\delta_n \leq \gamma_n$.

To prove that $\eta_n \leq \delta_n$ for every $n \geq 0$, consider for every $z \in A^n$, the law Q_z of X_{-n-1}^\triangleleft conditionally on $X_{-n:-1} = z$. Then,

$$p(\cdot|z) = \int_{A^\triangleleft} p(\cdot|yz) Q_z(dy).$$

Thus, for every x in A^\triangleleft and z in A^n ,

$$\begin{aligned} \|p(\cdot|z) - p(\cdot|xz)\| &= \left\| \int_{y \in A^\triangleleft} (p(\cdot|yz) - p(\cdot|xz)) Q_z(dy) \right\| \\ &\leq \int_{y \in A^\triangleleft} \|p(\cdot|yz) - p(\cdot|xz)\| Q_z(dy) \\ &\leq \sup_{y \in A^\triangleleft} \|p(\cdot|yz) - p(\cdot|xz)\| \leq \delta_n. \end{aligned}$$

For every z in A^n , $\|p(\cdot|z) - p(\cdot|X_{-n-1}^\triangleleft z)\| \leq \delta_n$ almost surely. Taking the expectation and the average over z , one gets $\eta_n \leq \delta_n$.

To prove that $\delta_n \leq \alpha_n$ for every $n \geq 0$, consider $z \in A^n$ and $y, y' \in A^\triangleleft$. Then,

$$\begin{aligned} \|p(\cdot|yz) - p(\cdot|y'z)\| &= \sum_{a \in A} |p(a|yz) - p(a|y'z)|_+ \\ &= \sum_{a \in A} \left(p(a|yz) - \min(p(a|yz), p(a|y'z)) \right) \\ &\leq 1 - \inf_{z \in A^n} \sum_{a \in A} \inf \{ p(a|yz) : y \in A^\triangleleft \} \\ &= \alpha_n. \end{aligned}$$

This ends the proof.

4.5.2 Proof of lemma 4.1.1

Assume that X is a process valued in a measurable space (E, \mathfrak{E}) and that U is a governing and generating process of X . Let $n \in \mathbb{Z}$. Since U governs X , there exists a measurable function ψ_n such that $X_{n+1} = \psi_n(U_{n+1}, X_n^\triangleleft)$ (axiom (ii)). Let $B \in \mathfrak{E}$. We try to estimate

$$\rho = \mathbb{P}[X_{n+1} \in B \mid \mathcal{F}_n^{X,U}] = \mathbb{P}[\psi_n(U_{n+1}, X_n^\triangleleft) \in B \mid \mathcal{F}_n^{X,U}].$$

Since U governs X , U_{n+1} and $\mathcal{F}_n^{X,U}$ are independent (axiom (i)) hence ρ is a function of X_n^\triangleleft only, that is,

$$\rho = \mathbb{P}[\psi(U_{n+1}, X_n^\triangleleft) \in B \mid \mathcal{F}_n^X] = \mathbb{P}[X_{n+1} \in B \mid \mathcal{F}_n^X].$$

Hence \mathcal{F}_{n+1}^X is independent of \mathcal{F}_n^U conditionally on \mathcal{F}_n^X . This shows that \mathcal{F}^X is immersed in \mathcal{F}^U .

4.5.3 Proof of corollary 4.1.2

Assume that X is a chain with memory of variable length and Y be an independent copy of X . As $p(\cdot|X_{-n-1}^\triangleleft Y_{-n:-1}) = p(\cdot|Y_{-n:-1})$. on the event $\{\ell(Y_{-1}^\triangleleft) \leq n\}$,

$$\|p(\cdot|X_{-n-1}^\triangleleft Y_{-n:-1}) - p(\cdot|Y_{-n:-1})\| \leq \mathbf{1}_{\{\ell(Y_{-1}^\triangleleft) \geq n+1\}}.$$

Taking expectations, one gets $\eta_n \leq \mathbb{P}[\ell(Y_{-1}^\triangleleft) \geq n+1]$, hence

$$\sum_{n \geq 0} \eta_n \leq \mathbb{E}[\ell(Y_{-1}^\triangleleft)] < +\infty.$$

This ends the proof.

4.5.4 Proof that $\mathcal{H}(\gamma)$ provides a positive lower bound for $p(a|x)$

We show that $\mathcal{H}(\gamma)$ implies the existence of $c > 0$ such that $p(a|x) \geq c$ for every x in A^\triangleleft and a such that $\mathbb{P}[X_0 = a] > 0$.

Assume that $\mathcal{H}(\gamma)$, that is

$$\sum_{k=0}^{+\infty} \prod_{n=0}^k (1 - \gamma_n) = +\infty.$$

Therefore, $1 - \gamma_0 > 0$. By definition of γ_0 , for every $a \in A$, $x, y \in A^\triangleleft$,

$$p(a|x) \geq (1 - \gamma_0)p(a|y).$$

Integrating this inequality with respect to the law of X^\triangleleft , one gets

$$p(a|x) \geq (1 - \gamma_0)\mathbb{P}[X_0 = a].$$

Since A is finite, this ends the proof.

Chapter 5

Annexe : preuve de quelques faits simples

5.0.5 Preuve de la proposition 1.3.1

Nous allons montrer que toute filtration $(\mathcal{F}_n)_{n \leq 0}$ telle que \mathcal{F}_0 est essentiellement séparable possède une paramétrisation. La construction d'une paramétrisation génératrice se base sur le lemme élémentaire suivant.

Lemme 5.0.1 *Soit Z une variable aléatoire réelle de fonction de répartition F . Pour $z \in \mathbb{R}$, notons $\Delta F(z) = F(z) - F(z-) = \mathbb{P}[Z = z]$. Soit ϕ l'inverse de F continu à gauche : pour tout $u \in]0, 1[$,*

$$\phi(u) = \inf\{z \in \mathbb{R} ; F(z) \geq u\}.$$

Soit V une variable aléatoire à valeurs dans $]0, 1[$, de loi uniforme sur $[0, 1]$, indépendante de Z . On pose

$$U = (1 - V)F(Z-) + VF(Z) = F(Z-) + V\Delta F(Z).$$

Alors (i) U suit la loi uniforme sur $[0, 1]$ et (ii) $\phi(U) = Z$ presque sûrement.

Preuve du lemme 5.0.1. Remarquons tout d'abord que pour tout $z \in \mathbb{R}$ et $u \in]0, 1[$,

$$F(z) < u \Leftrightarrow z < \phi(u)$$

$$u \leq F(z) \Leftrightarrow \phi(u) \leq z$$

On en déduit que pour tout $u \in]0, 1[$, $F(\phi(u)-) \leq u \leq F(\phi(u))$ et que pour tout $z \in \mathbb{R}$ et $u \in]0, 1[$,

$$\begin{aligned} F(z-) < u \leq F(z) &\Rightarrow \forall \varepsilon > 0, F(z - \varepsilon) < u \leq F(z) \\ &\Rightarrow \forall \varepsilon > 0, z - \varepsilon < \phi(u) \leq z \\ &\Rightarrow \phi(u) = z \\ &\Rightarrow u \leq F(z). \end{aligned}$$

Par conséquent

$$F(z-) < u \leq F(z) \Leftrightarrow (z = \phi(u) \text{ et } F(z-) < u).$$

Par définition de U , $F(Z-) \leq U \leq F(Z)$. De plus, si $F(Z-) < F(Z)$ alors $F(Z-) < U$ puisque $V > 0$. On a donc pour tout $u \in]0, 1[$,

$$\begin{aligned} \{U < u\} &= \{F(Z) < u\} \cup \{F(Z-) < u \leq F(Z) ; V\Delta F(Z) < u - F(Z-)\} \\ &= \{F(Z) < u\} \cup \{Z = \phi(u) ; F(Z-) < u ; V\Delta F(Z) < u - F(Z-)\} \\ &= \{Z < \phi(u)\} \cup \left\{ Z = \phi(u) ; F(\phi(u)-) < u ; V < \frac{u - F(\phi(u)-)}{\Delta F(\phi(u))} \right\} \end{aligned}$$

Par indépendance de Z et V , on a donc

$$\begin{aligned} \mathbb{P}[U < u] &= F(\phi(u)-) + \Delta F(\phi(u)) \mathbf{1}_{\{F(\phi(u)-) < u\}} \frac{u - F(\phi(u)-)}{\Delta F(\phi(u))} \\ &= F(\phi(u)-) + \mathbf{1}_{\{F(\phi(u)-) < u\}} (u - F(\phi(u)-)) \\ &= u, \end{aligned}$$

puisque $F(\phi(u)-) \leq u$. On a montré (i): U suit la loi uniforme sur $[0, 1]$.

Pour montrer (ii) introduisons l'ensemble

$$B = \{z \in \mathbb{R} : \exists \varepsilon > 0, F(z - \varepsilon) = F(z)\}.$$

Soient $z \in B$ et I la composante connexe de B contenant z . Par définition de B , il existe $\varepsilon > 0$ tel que F est constante sur $[z - \varepsilon, z]$, et donc $]z - \varepsilon, z] \subset I$. On en déduit que les composantes de B sont des intervalles de la forme $]a, b[$ ou $]a, b]$ avec $-\infty \leq a < b \leq +\infty$. Elles sont donc en quantité dénombrable.

Montrons que F est constante sur chaque composante connexe. Soit I une composante connexe de B , il suffit de montrer que pour tout $c \in I$, F est constante sur $I \cap]-\infty, c]$. Soit $c \in I$, et $m = \inf\{z \in \mathbb{R} : F(z) = F(c)\}$. Alors par continuité à droite de F , on a $F(m) = F(c)$, donc F est constante sur $[m, c]$. On ne peut pas avoir $m \in I$, sinon cela contredirait la définition de m . Comme I est un intervalle contenant c , on a $I \subset]m, +\infty[$ ce qui entraîne que F est constante sur $I \cap]-\infty, c]$. Donc F est constante sur I .

Pour toute composante connexe I de B , on a donc $\mathbb{P}[Z \in I] = 0$. Comme B n'a qu'un nombre dénombrable de composantes connexes, on a donc $\mathbb{P}[Z \in B] = 0$. Et comme sur l'évènement $\{Z \notin B\}$, on a pour tout $\varepsilon > 0$, $F(Z - \varepsilon) < U \leq F(Z)$, d'où $Z - \varepsilon < \phi(U) \leq Z$, on obtient $\phi(U) = Z$ presque sûrement. Le lemme 5.0.1 est démontré.

Nous allons maintenant finir la preuve de la proposition 1.3.1. Pour tout $n \leq 0$, on se donne Z_n une variable aléatoire réelle qui engendre \mathcal{F}_n . Quitte à grossir l'espace $(\Omega, \mathcal{A}, \mathbb{P})$ on peut se donner une suite $(V_n)_{n \leq 0}$ de variables aléatoires i.i.d. de loi uniforme sur $[0, 1]$, indépendante de \mathcal{F}_0 . On note F_n la fonction de répartition conditionnelle de Z_n sachant \mathcal{F}_{n-1} et on pose

$$U_n = (1 - V_n)F_n(Z_n -) + V_n F_n(Z_n).$$

D'après le lemme 5.0.1 on sait que (i) la loi de U_n sachant \mathcal{F}_{n-1} est uniforme sur $[0, 1]$ et (ii) $Z_n = \phi_n(U_n)$ presque sûrement, où ϕ_n est l'inverse continu à gauche de F_n .

De (i) on déduit que U_n est indépendante de \mathcal{F}_{n-1} et de loi uniforme sur $[0, 1]$. De plus pour tout $z \in \mathbb{R}$, on déduit de (ii) que

$$\{Z_n \leq z\} = \{\phi_n(U_n) \leq z\} = \{U_n \leq F_n(z)\} \in \sigma(U_n) \vee \mathcal{F}_{n-1}.$$

Donc $\mathcal{F}_n = \sigma(Z_n) \subset \sigma(U_n) \vee \mathcal{F}_{n-1}$, ce qui achève la preuve de la proposition.

5.0.6 Preuve de la proposition 1.3.3

Soit $\mathcal{F} = (\mathcal{F}_n)_n$ une filtration et $U = (U_n)_n$ une paramétrisation génératrice de \mathcal{F} . On veut montrer que si $A \in \mathcal{F}_{n+1}$, alors $\mathbb{P}[A \mid \mathcal{F}_n] = \mathbb{P}[A \mid \mathcal{F}_n^U]$. Comme $\mathcal{F}_{n+1} \subset \mathcal{F}_n \vee \sigma(U_{n+1})$, il suffit de le montrer pour des évènements de la forme $A = B \cap C$ où $B \in \mathcal{F}_n$ et $C \in \sigma(U_{n+1})$. Pour un tel évènement, on a

$$\mathbb{P}[A \mid \mathcal{F}_n^U] = \mathbf{1}_B \mathbb{P}[C \mid \mathcal{F}_n^U] = \mathbf{1}_B \mathbb{P}[C \mid \mathcal{F}_n] = \mathbb{P}[A \mid \mathcal{F}_n].$$

On a montré que pour tout n , \mathcal{F}_{n+1} est indépendant de \mathcal{F}_n^U conditionnellement à \mathcal{F}_n . Ce qui montre que \mathcal{F} est immergé dans \mathcal{F}^U .

5.0.7 Preuve du corollaire 1.4.1

Il suffit de montrer que pour toute filtration $\mathcal{F} = (\mathcal{F}_n)_{n \leq 0}$ telle que \mathcal{F}_0 est essentiellement séparable, si $(\mathcal{F}_n)_{n \leq -1}$ est standard, alors \mathcal{F} est standard puisque la réciproque est immédiate.

Soit $\mathcal{F} = (\mathcal{F}_n)_{n \leq 0}$ une filtration telle que \mathcal{F}_0 est essentiellement séparable et telle que $(\mathcal{F}_n)_{n \leq -1}$ est standard. Soit $\mathcal{F}' = (\mathcal{F}'_n)_{n \leq -1}$ une filtration isomorphe à $(\mathcal{F}_n)_{n \leq -1}$ immergée dans une filtration de type produit $(\mathcal{G}_n)_{n \leq -1}$. La proposition 1.3.1 fournit une variable aléatoire U indépendante de \mathcal{F}_{-1} telle que $\mathcal{F}_0 \subset \mathcal{F}_{-1} \vee \sigma(U)$.

Soient X_0 et X_{-1} des variables aléatoires réelles engendrant \mathcal{F}_0 et \mathcal{F}_{-1} . Il existe une application borélienne $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ telle que $X_0 = f \circ (X_{-1}, U)$ presque sûrement.

Soit X'_{-1} l'image de X_{-1} par l'isomorphisme de $(\Omega, \mathcal{A}, \mathbb{P}, (\mathcal{F}_n)_{n \leq -1})$ dans $(\Omega', \mathcal{A}', \mathbb{P}', (\mathcal{F}'_n)_{n \leq -1})$.

Soit U' une variable aléatoire indépendante de \mathcal{G}_{-1} et de même loi que U . Notons $\mathcal{G}_0 = \mathcal{G}_{-1} \vee \sigma(U)$, $X'_0 = f \circ (X'_{-1}, U')$ et $\mathcal{F}'_0 = \sigma(X'_0)$.

La standardité de \mathcal{F} découle des trois points suivants :

- (1) : $(\mathcal{F}'_n)_{n \leq 0}$ est une filtration et est isomorphe à $(\mathcal{F}_n)_{n \leq 0}$,
- (2) : $(\mathcal{F}'_n)_{n \leq 0}$ est immergée dans $(\mathcal{G}_n)_{n \leq 0}$,
- (3) : $(\mathcal{G}_n)_{n \leq 0}$ est de type produit.

Le point (3) est évident par construction de \mathcal{G}_0 , le point (2) se démontre comme la proposition 1.3.3. Montrons le point (1).

Toute variable aléatoire de $\mathcal{L}^0(\Omega, \mathcal{F}_0, \mathbb{P})$ s'écrit sous la forme $g \circ X_0$, avec g de \mathbb{R} dans \mathbb{R} borélienne. Comme X_0 a même loi que X'_0 , l'application Φ de $L^0(\Omega, \mathcal{F}_0, \mathbb{P})$ dans $L^0(\Omega', \mathcal{F}'_0, \mathbb{P}')$ qui à la classe de $g \circ X_0$ associe la classe de $g \circ X'_0$ est bien définie et est un isomorphisme pour les opérations boréliennes. En effet, comme $\mathcal{F}_0 = \sigma(X_0)$ et $\mathcal{F}'_0 = \sigma(X'_0)$, Φ envoie \mathcal{F}_0 sur \mathcal{F}'_0 .

Comme $\mathcal{F}_{-1} \subset \mathcal{F}_0$, il existe une fonction borélienne $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ telle que

$$X_{-1} = \varphi \circ X_0 = \varphi \circ f \circ (X_{-1}, U) \text{ presque sûrement.}$$

Comme (X_{-1}, U) a même loi que (X'_{-1}, U') , on a aussi presque sûrement

$$X'_{-1} = \varphi \circ X'_0 = \varphi \circ f \circ (X'_{-1}, U').$$

Par définition de Φ , on a donc $\Phi(X_{-1}) = X'_{-1}$. La restriction de Φ à $L^0(\Omega, \mathcal{F}_{-1}, \mathbb{P})$ est donc l'isomorphisme de l'hypothèse, ce qui montre que Φ envoie \mathcal{F}_n sur \mathcal{F}'_n pour tout $n \leq -1$. De plus $\mathcal{F}'_{-1} \subset \mathcal{F}'_0$ car $X'_{-1} = \varphi \circ X'_0$. Donc $(\mathcal{F}'_n)_{n \leq 0}$ est une filtration et est isomorphe à $(\mathcal{F}_n)_{n \leq 0}$.

Bibliography

- [1] G. Bejerano and G. Yona. Variations on probabilistic suffix trees: statistical modeling and prediction of protein families. *Bioinformatics*, 17(1):23–43, 2001.
- [2] X. Bressaud, A. Maass, S. Martinez, and J. San Martin. Stationary processes whose filtrations are standard. *The Annals of Probability*, 34(4):1589–1600, 2006.
- [3] F. Comets, R. Fernández, and P.A. Ferrari. Processes with long memory: regenerative construction and perfect simulation. *The Annals of Applied Probability*, 12(3):921–943, 2002.
- [4] L. Dubins, J. Feldman, M. Smorodinsky, and B. Tsirelson. Decreasing sequences of σ -fields and a measure change for Brownian motion. *The Annals of Probability*, 24(2):882–904, 1996.
- [5] M. Émery and W. Schachermayer. On Vershik’s standardness criterion and Tsirelson’s notion of cosiness. *Séminaire de Probabilités, XXXV*, 1755:265–305, 2001.
- [6] J. Feldman and M. Smorodinsky. Decreasing sequences of measurable partitions: product type, standard and prestandard. *Ergodic Theory and Dynamical Systems*, 20(4):1079–1090, 2000.
- [7] A. Galves and E. Löcherbach. Stochastic chains with memory of variable length. *TICSP series*, 38:117–133, 2008.
- [8] A. Gorbulsky. About one property of entropy of decreasing sequence of measurable partitions. *Zapiski Nauchnykh Seminarov POMI*, 256:19–24, 1999.
- [9] D. Hecklen. Bernoullis are standard when entropy is not an obstruction. *Israel Journal of Mathematics*, 107:141–155, 1998.
- [10] D. Hecklen and C. Hoffman. T, T^{-1} is not standard. *Ergodic Theory and Dynamical Systems*, 18(4):875–878, 1998.

- [11] G. Kallianpur and N. Wiener. Non-linear prediction. 1956.
- [12] S. Laurent. Filtrations à temps discret négatif. 2004.
- [13] S. Laurent. Further comments on the representation problem for stationary processes. *Statistics and Probability Letters*, 80(7-8):592–596, 2009.
- [14] S. Laurent. On standardness and I-cosiness. Technical Report TR09017, IAP Network, 2009.
- [15] S. Laurent. On Vershikian and I-cosy random variables and filtrations. Technical Report TR09018, IAP Network, 2009.
- [16] F. Leonardi. Rate of convergence of penalized likelihood context tree estimators. Technical Report math/0701810v2, ArXiv, 2007.
- [17] C. Leuridan. Filtration d’une marche aléatoire stationnaire sur le cercle. *Séminaire de Probabilités*, XXXVI:335–347, 2003.
- [18] P. Masani. Wiener’s contributions to generalized harmonic analysis, prediction theory and filter theory. *Bulletin of the American Mathematical Society*, 72(1):73–125, 1966.
- [19] H. Matzinger. Reconstructing a 2-color scenery by observing it along a simple random walk path. *The Annals of Applied Probability*, 15(1B):778–819, 2005.
- [20] W. Schachermayer. On certain probabilities equivalent to Wiener measure, d’après Dubins, Feldman, Smorodinsky and Tsirelson. *Séminaire de Probabilités XXXIII*, pages 221–239, 2001.
- [21] A.N. Shiryaev. *Probability*, volume 95 of *Graduate Texts in Mathematics*. Springer-Verlag, 1996.
- [22] M. Smorodinsky. Processes with no standard extension. *Israel Journal of Mathematics*, 107:327–331, 1998.
- [23] B. Tsirelson. Triple points: from non-Brownian filtrations to harmonic measures. *Geometric and Functional Analysis*, 7(6):1096–1142, 1997.
- [24] A.M. Vershik. Decreasing sequences of measurable partitions, and their applications. *Soviet mathematics Doklady*, 11:1007–1011, 1970.
- [25] A.M. Vershik. Four definitions of the scale of an automorphism. *Functional Analysis and Its Applications*, 7(3):169–181, 1973.
- [26] A.M. Vershik. Theory of decreasing sequences of measurable partitions. *Saint Petersburg Mathematical Journal*, 6:705–761, 1995.

- [27] V.G. Vinokourov. Two nonisomorphic exact endomorphisms of a Lebesgue space with isomorphic decomposition sequence. *Stochastic Processes and Related Problems*, 1:43–45, 1970.

Abstract Standardness is an important invariant in the theory of filtrations indexed by negative integer times. The main purpose of this thesis is to determine whether some filtrations are standard or not.

We first focus on the filtrations of split-word processes, introduced and studied by Smorodinsky and by Laurent. We prove that Laurent's sufficient condition for non standardness is also necessary. This yields a practical criterion of standardness. In turn, this criterion enables us to exhibit non standard filtrations which become standard when time is accelerated by omitting infinitely many instants of time.

Secondly, we study the natural filtrations of stationary processes on finite state-spaces. Recently Bressaud et al. provided a sufficient condition for the natural filtration of such a process $(X_k)_k$ to be standard when the state-space has size 2. Their condition involves the conditional laws $p(\cdot|x)$ of X_0 conditionally on $(X_k)_{k \leq -1} = x$ and controls the influence of the remote past of the process on its present X_0 . Bressaud and al. measure the maximal strength of this influence. We provide sufficient conditions for standardness based on some average gaps between these conditional laws, instead of the maximal gaps.

Résumé Le caractère standard ou non standard est un invariant important dans la théorie des filtrations à temps discret négatif. Le but principal de cette thèse est de déterminer si certaines filtrations sont standard ou non.

Nous nous intéressons tout d'abord aux filtrations des processus des mots découpés, introduits et étudiés par Smorodinsky et par Laurent. Nous montrons que la condition suffisante de non standardité donnée par Laurent est aussi une condition nécessaire. Il en découle un critère simple de standardité, qui nous permet de donner un exemple de filtration non standard qui devient standard dès que le temps est accéléré par l'omission d'un nombre infini d'instant.

Nous étudions ensuite les filtrations naturelles de processus stationnaires à valeurs dans un ensemble fini. Récemment Bressaud et al. ont donné une condition nécessaire pour que la filtration naturelle d'un tel processus $(X_k)_k$ soit standard quand le cardinal de l'espace d'états vaut 2. Leur condition utilise les lois conditionnelles $p(\cdot|x)$ de X_0 sachant tout le passé $(X_k)_{k \leq -1} = x$ et contrôle l'influence du passé lointain sur l'évolution présente du processus. Elle utilise les écarts maximaux entre $p(\cdot|x)$ et $p(\cdot|y)$ pour des suites infinies x et y qui coïncident sur leurs n derniers termes. Nous donnons une condition suffisante de standardité faisant intervenir les écarts moyens entre ces lois conditionnelles plutôt que les écarts maximaux.