



**HAL**  
open science

## Filtrations associated to some two-to-one transformations

Christophe Leuridan

► **To cite this version:**

Christophe Leuridan. Filtrations associated to some two-to-one transformations. Séminaire de Probabilités LI, 2301, Springer, pp.29–89, 2022, Lecture Notes in Mathematics. hal-01987993v2

**HAL Id: hal-01987993**

**<https://hal.science/hal-01987993v2>**

Submitted on 6 Jan 2020

**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.

# Filtrations associated to some two-to-one transformations

Christophe Leuridan

January 6, 2020

## Abstract

The aim of the present paper is the study of filtrations indexed by the non-positive integers associated to (non-invertible) measure-preserving maps. We establish a necessary and sufficient condition for the filtration associated to some skew-products to be Kolmogorovian, i.e. to have a trivial tail  $\sigma$ -field at time  $-\infty$ . This condition improves on Meilijson's result.

More specifically, we focus on dyadic filtrations associated to two-to-one maps provided by skew-products, like  $[\text{Id}, T]$  or  $[T^{-1}, T]$ . Determining whether these filtrations can or cannot be generated by some sequence of independent random variables is often difficult, although Vershik's criteria provide tools to investigate this question.

In this paper, we revisit many classical examples of filtrations associated to two-to-one maps provided by skew-products. The first examples are rather simple and are given as an illustration of Vershik's intermediate criterion. The last two ones are much more involved and yield non-product-type filtrations. Our purpose is to give a more complete and readable presentation of the proofs already existing.

MSC Classification : 37A05, 60J05.

Keywords : measure-preserving maps, skew products, dyadic filtrations, product-type filtration, standardness, Vershik's criteria, split-word process, nibbled-word process.

## 1 Introduction

The classification up to isomorphism of the measure-preserving maps in probability spaces is one of the main topics in ergodic theory and the notion entropy is a key tool in this theory. Classical ergodic properties (ergodicity, mixing property, exactness, 'Bernoulliness') are also invariant by isomorphisms, and can be used to prove that two dynamical systems are not isomorphic.

When one works with non-invertible measure-preserving maps, it is also interesting to study the standardness or non-standardness of 'the' associated filtration (indexed by the non-positive integers). Let us explain how to construct this filtration.

### 1.1 Filtration associated to a measure-preserving map

Let  $(E, \mathcal{E}, \pi)$  be a probability space and  $R$  be a measure-preserving map on  $(E, \mathcal{E}, \pi)$ . The sequence of  $\sigma$ -fields  $(R^{-n}\mathcal{E})_{n \geq 0}$  is non-increasing. Reverting time yields a filtration (namely a non-decreasing family of  $\sigma$ -fields) indexed by the non-positive integers.

To give a probabilistic interpretation, fix a random variable  $Z$  taking values in  $E$ , with law  $\pi$ , defined on some probability space  $(\Omega, \mathcal{A}, \mathbf{P})$ . Assume the existence of a transition kernel  $K$  such that

$$\forall y \in E, \quad \forall B \in \mathcal{E}, \quad K(y, B) = \pi[Z \in B | R(Z) = y].$$

One gets a stationnary Markov process with transition kernel  $K$  by setting

$$\forall n \leq 0, \quad Z_n := R^{|n|} \circ Z.$$

Its natural filtration (indexed by the non-positive integers) is given by

$$\forall n \leq 0, \quad \mathcal{F}_n^Z := \sigma((Z_k)_{k \leq n}) = \sigma(Z_n) = Z^{-1}(R^n \mathcal{E}).$$

Of course, looking at this filtration is interesting only when the measure-preserving map  $R$  is not invertible. In this case, many ergodic properties of  $R$  can be viewed as properties of the filtration  $(\mathcal{F}_n^Z)_{n \leq 0}$  thus defined.

For example, one can check <sup>1</sup> that the tail  $\sigma$ -field at time  $-\infty$  is

$$\mathcal{F}_{-\infty}^Z := \bigcap_{n \leq 0} \mathcal{F}_n^Z = Z^{-1} \left( \bigcap_{n \leq 0} R^n \mathcal{E} \right).$$

Therefore, the filtration  $(\mathcal{F}_n^Z)_{n \leq 0}$  is *Kolmogorovian* (namely  $\mathcal{F}_{-\infty}^Z$  is trivial under  $\mathbf{P}$ ) if and only if the transformation  $R$  is *exact* (namely  $\bigcap_{n \geq 0} R^{-n} \mathcal{E}$  is trivial under  $\pi$ ).

Another interesting property is *dyadicity*. By definition, one says that a filtration  $(\mathcal{F}_n)_{n \leq 0}$  on  $(\Omega, \mathcal{A}, \mathbf{P})$  has *independent increments* if for every  $n \leq 0$ , one can find a random variable  $\xi_n$ , independent of  $\mathcal{F}_{n-1}$  such that  $\mathcal{F}_n = \mathcal{F}_{n-1} \vee \sigma(\xi_n)$ . Such an independent complement is called an *innovation at time  $n$*  since it carries ‘the new information’ at time  $n$ . Furthermore, if each innovation  $\xi_n$  is uniform on some finite set possibly depending on  $n$  (respectively some set finite of constant size  $r$ ), the filtration  $(\mathcal{F}_n)_{n \leq 0}$  is *poly-adic*, (respectively *r-adic*).

One checks that the filtration  $(\mathcal{F}_n^Z)_{n \leq 0}$  associated to the transformation  $R$  is dyadic if and only if  $R$  is *two-to-one*, namely if for  $\pi$ -almost every  $y \in E$ , the probability measure  $K(y, \cdot)$  has two atoms, each one having measure  $1/2$ .

For example, the map  $R_2 : x \mapsto [2x]$  from  $\mathbf{I} := [0, 1[$  to itself preserves the uniform measure on  $\mathbf{I}$ , and is two-to-one since the associated transition kernel  $K$  is given by

$$\forall y \in \mathbf{I}, \quad K(y, \cdot) = \frac{1}{2}(\delta_{y/2} + \delta_{(1+y)/2}).$$

Actually,  $R_2$  acts as a Bernoulli shift on the dyadic expansions. Indeed, if the dyadic expansion of  $x \in \mathbf{I}$  is

$$x = \sum_{n=1}^{\infty} \frac{a_n(x)}{2^n},$$

then the dyadic expansion of  $R_2(x)$  is

$$[2x] = \sum_{n=2}^{\infty} \frac{a_n(x)}{2^{n-1}} = \sum_{n=1}^{\infty} \frac{a_{n+1}(x)}{2^n}.$$

---

<sup>1</sup>The  $\sigma$ -field  $Z^{-1} \left( \bigcap_{n \leq 0} R^n \mathcal{E} \right)$  is contained in  $\mathcal{F}_n^Z = Z^{-1}(R^n \mathcal{E})$  for every  $n \leq 0$ , hence in  $\mathcal{F}_{-\infty}^Z$ . Conversely, if  $A \in \mathcal{F}_{-\infty}^Z$ , then for each  $n \leq 0$ , one can find  $B_n \in R^n \mathcal{E}$  such that  $A = Z^{-1}(B_n)$ . The set  $B := \liminf_{n \rightarrow -\infty} B_n$  belongs to  $\bigcap_{n \leq 0} R^n \mathcal{E}$  and one has  $A = Z^{-1}(B)$ . The reverse inclusion follows.

Furthermore, if  $Z$  is a uniform random variable with values on  $\mathbf{I}$ , the natural filtration of the process  $(Z_n)_{n \leq 0}$  defined by  $Z_n = R_2^{|n|}(Z)$  is also the natural filtration of the i.i.d. sequence  $(\xi_n)_{n \leq 0}$  defined by  $\xi_n = a_{1-n}(Z)$ . By Kolmogorov's zero-one law, the filtration  $(\mathcal{F}_n^Z)_{n \leq 0}$  is Kolmogorovian since it is *product-type* (i.e. generated by some sequence independent random variables).

Yet, Vershik [23, 24] discovered that Kolmogorovian dyadic filtrations are not necessarily product-type. He provided counterexamples with the help of criteria characterizing the product-type filtrations among the poly-adic filtrations. Vershik's theory has been transcribed into the language of stochastic processes by Émery & Schachermayer [4], and later by Laurent [11, 12, 13].

In the present paper, we will apply these criteria to well-known examples of dyadic filtrations associated to two-to-one measure-preserving maps. The existing proofs of their productness or non-productness are scattered in the literature, are sometimes incompletely given and difficult to read, at least for probabilists. All these examples are provided by skew products.

## 1.2 Skew products

Let  $F$  be a countable set with size  $\geq 2$  and  $\mu$  a probability measure on  $F$  giving positive mass to every point of  $F$ . Call  $S$  the Bernoulli shift on  $F^{\mathbf{Z}^-}$  defined by

$$S(f)(n) := f(n-1) \text{ for every } f \in F^{\mathbf{Z}^-} \text{ and } n \in \mathbf{Z}_-.$$

Here, we define Bernoulli shifts on  $F^{\mathbf{Z}^-}$  instead of  $F^{\mathbf{Z}^+}$  because we will work with filtrations indexed by the non-positive integers.

Given a family  $(T_k)_{k \in F}$  of invertible measure-preserving maps on some probability space  $(G, \mathcal{G}, Q)$ , we define the skew product  $R = S \times (T_k)_{k \in F}$  by

$$R(f, g) := (S(f), T_{f(0)}(g)).$$

The transformation  $R$  thus defined preserves the measure  $\mu^{\otimes \mathbf{Z}^-} \otimes Q$  and is not invertible. More precisely, the inverse images by  $R$  of any element  $(\tilde{f}, \tilde{g}) \in F^{\mathbf{Z}^-} \times G$  are the elements  $(\tilde{f}x, T_x^{-1}\tilde{g})$ , where  $\tilde{f}x \in F^{\mathbf{Z}^-}$  denotes the sequence obtained by concatenation of  $\tilde{f}$  and  $x$ , defined by  $(\tilde{f}x)(n) = \tilde{f}(n+1)$  for every  $n \leq -1$  and  $(\tilde{f}x)(0) = x$ .

Fix an  $F^{\mathbf{Z}^-} \times E$ -valued random variable  $Z = (\xi, \gamma)$  with law  $\mu^{\otimes \mathbf{Z}^-} \otimes Q$ . Then the law of  $Z$  given  $R(Z)$  is given by the following transition kernel

$$\forall (\tilde{f}, \tilde{g}) \in F^{\mathbf{Z}^-} \times G, \quad K((\tilde{f}, \tilde{g}), \cdot) = \sum_{x \in F} \mu\{x\} \delta_{(\tilde{f}x, T_x^{-1}\tilde{g})}.$$

Moreover, the random variable  $\xi_0$  has law  $\mu$  and is independent of  $R(Z) = (S(\xi), T_{\xi_0}(\gamma))$ .

We can go farther and define two processes  $(\gamma_n)_{n \leq 0}$  and  $(Z_n)_{n \leq 0}$  by

$$\forall n \leq 0, \quad \gamma_n := T_{\xi_{n+1}} \circ \cdots \circ T_{\xi_0}(\gamma) \text{ and } Z_n := R^{|n|}(Z) = (S^{|n|}(\xi), \gamma_n),$$

From the remarks above and from the recursion relation  $\gamma_n = T_{\xi_n}^{-1}(\gamma_{n-1})$ , we derive:

- the filtration  $(\mathcal{F}_n^Z)_{n \leq 0}$  has independent increments and the process  $(\xi_n)_{n \leq 0}$  is a sequence of innovations, i.e.

$$\forall n \leq 0, \mathcal{F}_n^Z = \mathcal{F}_{n-1}^Z \vee \sigma(\xi_n) \text{ with } \xi_n \text{ independent of } \mathcal{F}_{n-1}^Z;$$

In particular, if  $F$  is finite and  $\mu$  is uniform, the filtration  $(\mathcal{F}_n^Z)_{n \leq 0}$  is poly-adic.

- the processes  $(\gamma_n)_{n \leq 0}$  and  $(Z_n)_{n \leq 0}$  are stationary Markov chains whose evolutions are governed by the sequence  $(\xi_n)_{n \leq 0}$ , since for every  $n \leq 0$ ,  $\gamma_n = T_{\xi_n}^{-1} \circ \gamma_{n-1}$ ;
- if for  $Q$ -almost every  $z \in G$ , the images  $(T_k(z))_{k \in F}$  are pairwise distinct, then the processes  $(\gamma_n)_{n \leq 0}$  and  $(Z_n)_{n \leq 0}$  generate the same filtration since for every  $n \leq 0$ , the knowledge of  $(\gamma_k)_{k \leq n}$  is sufficient to recover  $(\xi_k)_{k \leq n}$ .

This last remark will be used to give simpler descriptions of the filtration  $(\mathcal{F}_n^Z)_{n \geq 0}$ .

Two questions naturally arise. First, is the tail  $\sigma$ -field  $\mathcal{F}_{-\infty}^Z$  trivial, or equivalently, is the transformation  $R$  exact? Answering this question is not always simple. The next question is more involved: is the filtration  $(\mathcal{F}_n^Z)_{n \leq 0}$  *product-type*, i.e. can it be generated by some sequence of independent random variables? By Kolmogorov zero-one law, the triviality of the tail  $\sigma$ -field  $\mathcal{F}_{-\infty}^Z$  is necessary. But this condition is far from being sufficient although  $(\mathcal{F}_n^Z)_{n \leq 0}$  admits a sequence of innovations, namely  $(\xi_n)_{n \leq 0}$ . Actually, the process  $(\xi_n)_{n \leq 0}$  may generate a filtration which is strictly included in  $(\mathcal{F}_n^Z)_{n \leq 0}$  even if  $\mathcal{F}_{-\infty}^Z$  is trivial. However, in some cases, the filtration  $(\mathcal{F}_n^Z)_{n \leq 0}$  can be generated by some other sequence of independent random variables, whereas in other cases it cannot. Many examples are given in subsection 1.4.

### 1.3 Exactness of Meilijson's skew products

In section 2, we focus on the particular case considered by Meilijson where the transformations  $T_k$  are the powers of a same invertible transformation  $T$ , namely when  $F \subset \mathbf{Z}$  and  $T_k = T^k$  for every  $k \in F$ . In this case, the skew product  $R$  will be denoted by  $S \times T$ . We get the following characterization.

**Theorem 1** *Call  $d$  the greatest common divisor of all differences  $k - \ell$  where  $k$  and  $\ell$  range over  $F$ . Then  $S \times T$  is exact (namely  $\mathcal{F}_{-\infty}$  is trivial) if and only if  $T^d$  is ergodic.*

The difficult part in theorem 1 is the implication  $T^d$  ergodic  $\Rightarrow S \times T$  exact. The proof involves the spectral measure of  $T$ .

A weaker statement was stated by Meilijson and proved by a different method in [17]. Actually, Meilijson worked with the bilateral shift  $\bar{S}$  on  $F^{\mathbf{Z}}$ ; we use the notation  $\bar{S}$  because it is natural extension  $\bar{S}$  of 'our' unilateral shift  $S$ . The transformation  $\bar{S} \times T$  considered by Meilijson is invertible and is the the natural extension of 'our' transformation  $S \times T$ . With these notations, Meilijson stated that (see [17]):

$$(d = 1 \text{ and } T \text{ ergodic}) \Rightarrow \bar{S} \times T \text{ is a K-automorphism,}$$

$$(T^n \text{ ergodic for every } n \geq 1) \Rightarrow \bar{S} \times T \text{ is a K-automorphism.}$$

Our theorem 1 brings two improvements: our assumption ( $T^d$  is ergodic) is weaker than Meilijson's ones, and our conclusion ( $S \times T$  exact) is stronger than Meilijson's one. Indeed, the natural extension of an exact transformation is always a K-automorphism, but the converse is not true.

J.P. Thouvenot and the referee drew my attention to the fact that an application of theorem 1 greatly simplifies the proof of a theorem of Lindenstrauss, Peres and Schlag (Theorem 1.4 in [16]), which states that any Bernoulli system of entropy  $h \in ]0, \infty]$  has  $K$ -partitions of conditional entropy  $\eta$  for any  $\eta \in ]0, h]$ . In section 2, we give a more complete presentation of this theorem before proving it with the help of theorem 1.

## 1.4 Examples of two-to-one skew products

Consider again skew products  $S \times (T_k)_{k \in \mathbf{Z}}$ . These skew products are two-to-one when  $F$  is a pair  $\{k_1, k_2\}$  and  $\mu$  is the uniform measure on  $F$ . Then, when one only looks at the second component in the product  $F^{\mathbf{Z}^-} \times E$ , the apparent effect of  $S \times (T_k)_{k \in F}$  is to apply at random  $T_{k_1}$  or  $T_{k_2}$ , that is why the skew product is simply denoted by  $[T_{k_1}, T_{k_2}]$ . In this case, the innovations  $(\xi_n)_{n \geq 0}$  are uniform on the pair  $F$ , so the filtration  $(\mathcal{F}_n)_{n \geq 0}$  is dyadic. We now present classic examples of such transformations, from the simplest to the hardest, where the associated filtration  $(\mathcal{F}_n^{\mathbf{Z}})_{n \in \mathbf{Z}}$  is known to be product-type or not product-type.

The first two examples are given in Vershik's original paper [23], the former being only mentioned at page 744, without any proof (a proof is given in [15] in a slightly more general context).

**Theorem 2** (*Vershik*) *If  $T$  is an irrational rotation on the circle, then the filtration associated to  $[T, T^{-1}]$  are product-type. Hence, the natural filtration of an irrational random walk on the circle indexed by the non-positive integers is product-type.*

We shall see that the same result holds with  $[T, \text{Id}]$ , or more generally with  $[T_1, T_2]$  when the  $T_i$  are the translations  $x \mapsto x + \alpha_i$  on  $\mathbf{R}/\mathbf{Z}$  provided that  $\alpha_1 - \alpha_2$  is irrational.

Let us mention two stronger results concerning  $[T, \text{Id}]$ .

In [5], Feldman and Rudolph prove that the filtration associated to  $[T, \text{Id}]$  is product-type when  $T$  is rank-1. This condition includes all pure-point spectrum transformations and in particular irrational rotations of the circle.

Furthermore, Rudolph et Hoffman prove in [10] (part 6) that if  $T$  is an irrational rotation on the circle, then  $[T, \text{Id}]$  is isomorphic to a unilateral Bernoulli shift.

**Theorem 3** (*Vershik*) *Let  $\Gamma = \langle a, b \rangle$  be the free group with two generators  $a, b$ . Set  $G = \{0, 1\}^\Gamma$  endowed with the uniform measure, and call  $T_a, T_b$  the translations on  $G$  defined by  $T_a g(x) = g(a^{-1}x)$ ,  $T_b g(x) = g(b^{-1}x)$ . Then the filtration associated to  $[T_a, T_b]$  is not product-type. Hence, the natural filtration of a random walk on  $G$  with steps given by  $T_a^{-1}, T_b^{-1}$  and indexed by the non-positive integers is not product-type.*

The next result was conjectured by Vershik and proved by Heicklen and Hoffman in [7].

**Theorem 4** (*Heicklen - Hoffman*) *If  $T$  is a non-trivial bilateral Bernoulli shift, then the filtrations associated to  $[T, T^{-1}]$  are not product-type. Hence, the natural filtration of a symmetric random walk in a random scenery on  $\mathbf{Z}$  indexed by the non-positive integers is not product-type.*

When  $T$  is a non-trivial bilateral Bernoulli shift, the filtration associated to  $[T, T^{-1}]$  is closely related to the random walk in random scenery. Our presentation of the proof of theorem 4 relies on a non-stationary sub-process, called the nibbled-word process by Laurent [11], which generates a dyadic but not product-type filtration. Since the study of  $[T, \text{Id}]$  also involves (a copy of) the nibbled-word process, a very similar proof shows that the filtration associated to  $[T, \text{Id}]$  is not product-type. Using Sinai's theorem stating that every automorphism of a Lebesgue space with positive entropy admits a non-trivial bilateral Bernoulli shift as a factor, one can deduce the following generalisation.

**Corollary 5** *Let  $T$  be any automorphism of a Lebesgue space with positive entropy. Then the filtrations associated to  $[T, \text{Id}]$  and  $[T, T^{-1}]$  are not product-type.*

What happens for automorphisms with null entropy? We saw that when  $T$  is an irrational rotation,  $[T, \text{Id}]$  and  $[T, T^{-1}]$  provide product-type filtrations. Yet, Hoffman constructed an automorphism  $T$  with null entropy such that  $[T, \text{Id}]$  provides a not-product-type filtration [8].

**Theorem 6** (Hoffman) *One can construct an automorphism of a Lebesgue space with null entropy such that the filtration associated to  $[T, \text{Id}]$  is not product-type.*

The proofs of these results rely on Vershik's criteria. The original proofs are written in the language of ergodic theory and in an allusive way (some notations are not defined and some proofs are omitted), so they are difficult to read, at least for probabilists. The purpose of the present paper is to explain Vershik's criteria with a probabilist point of view and to provide proofs that are more accessible to non-specialists.

### Plan of the paper

In section 2, we prove theorem 1 and apply it to simplify the proof of theorem 1.4. of Lindenstrauss, Peres and Schlag [16].

In section 3, we introduce Vershik's criteria to characterize product-type filtrations among dyadic filtrations indexed by the non-positive integers.

In section 4, we prove theorems 2 and 3 to warm-up.

Sections 5 and 6 are devoted to the proof of theorem 4, corollary 5 and theorem 6.

Last, section 7 provides the proof of some auxiliary results.

## 2 Proof and applications of theorem 1

### 2.1 Proof of theorem 1

Let  $F \subset \mathbf{Z}$  be a countable set with size  $\geq 2$ ,  $\mu$  be a probability measure on  $F$  giving positive mass to every point of  $F$  and  $d$  be the greatest common divisor of all differences  $k - \ell$  where  $k$  and  $\ell$  range over  $F$ . Call  $S$  the Bernoulli shift on  $F^{\mathbf{Z}^-}$  (endowed with the probability measure  $\mu^{\otimes \mathbf{Z}^-}$ ) defined by

$$S(f)(n) := f(n - 1) \text{ for every } f \in F^{\mathbf{Z}^-} \text{ and } n \in \mathbf{Z}_-$$

and let  $T$  be an automorphism of the probability space  $(G, \mathcal{G}, Q)$ .

To prove that  $S \times T$  is exact if and only if  $T^d$  is ergodic, we use a characterization of the ergodicity of  $T^d$  involving the spectral measures associated to  $T$ . Let us recall the definition of these measures. More details can be found in [3], chapter 1, section 4.

Since  $T$  is an automorphism of the probability space  $(G, \mathcal{G}, Q)$ , the Koopman operator  $U_T : f \mapsto f \circ T$  from  $L^2(Q)$  to  $L^2(Q)$  is unitary. For every  $f \in L^2(Q)$ , there exists a unique finite measure  $\sigma_f$  on the unit circle  $\mathbf{U}$  of  $\mathbf{C}$  such that

$$\forall k \in \mathbf{Z}, \int_{\mathbf{U}} z^k d\sigma_f(z) = \langle f, U_T^k f \rangle_{L^2(Q)} = \int_G \bar{f} \times (f \circ T^k) dQ.$$

The measure  $\sigma_f$  thus defined is called the spectral measure of  $U_T$  associated to  $f$ .

**Lemma 7** Let  $d \geq 1$  be an integer,  $\mathbf{U}_d = \{z \in \mathbf{U} : z^d = 1\}$  and

$$L^2(Q)_0 = \{g \in L^2(Q) : \int_{\mathbf{U}} g dQ = 0\}.$$

Then  $T^d$  is ergodic if and only if  $\sigma_g(\mathbf{U}_d) = 0$  for every  $g \in L^2(Q)_0$ .

Proof. Call  $P_d$  the orthogonal projection on  $\text{Ker}(U_T^d - \text{Id})$ . Let  $f \in L^2(Q)$ . Then

$$\|P_d f\|_{L^2(Q)}^2 = \langle f, P_d f \rangle_{L^2(Q)} \text{ and } P_d f = \lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^{kd} \text{ in } L^2(Q),$$

by Von Neumann ergodic theorem (see for example [18], chapitre 2) applied to  $T^d$ . Thus

$$\|P_d f\|_{L^2(Q)}^2 = \lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{k=0}^{n-1} \langle f, f \circ T^{kd} \rangle_{L^2(Q)} = \lim_{n \rightarrow +\infty} \int_{\mathbf{U}} \left( \frac{1}{n} \sum_{k=0}^{n-1} z^{kd} \right) d\sigma_f(z) = \sigma_f(\mathbf{U}_d)$$

by Lebesgue dominated convergence theorem, since

$$\forall z \in \mathbf{U}, \quad \frac{1}{n} \sum_{k=0}^{n-1} z^{kd} \xrightarrow[n \rightarrow +\infty]{} \mathbf{1}_{\mathbf{U}_d}(z).$$

But  $T^d$  is ergodic if and only if  $\text{Ker}(U_T^d - \text{Id})$  contains only the  $Q$ -almost surely constant functions, namely if  $P_d f = 0$  for every  $f \in L^2(Q)_0$ . Lemma 7 follows.

We will also use the following result on characteristic functions.

**Lemma 8** Let  $\mu$  be a probability measure on  $\mathbf{Z}$ . Call  $F$  its support,  $d$  the greatest common divisor of all differences  $k - \ell$  where  $k$  and  $\ell$  range over  $F$ . For every  $z \in \mathbf{U}$ , set

$$\forall z \in \mathbf{U}, \quad \varphi(z) = \sum_{k \in \mathbf{Z}} \mu\{k\} z^k = \sum_{k \in F} \mu\{k\} z^k.$$

The function  $\varphi$  thus defined is the characteristic function of  $\mu$ . Then  $|\varphi(z)| = 1$  if and only if  $z$  belongs to  $\mathbf{U}_d$ .

Proof. Fix  $k_0 \in F$  and let  $z \in \mathbf{U}$ .

For every  $z \in \mathbf{U}_d$  and  $k \in F$ ,  $z^{k-k_0} = 1$  since  $d$  divides  $k - k_0$ , hence we get  $\varphi(z) = z^{k_0}$ , so  $|\varphi(z)| = 1$ .

Conversely, if  $|\varphi(z)| = 1$ , then equality holds in the triangle inequality

$$|\varphi(z)| \leq \sum_{k \in F} |\mu(k) z^k| = 1,$$

so the complex numbers  $\mu(k) z^k$  lie on a same half-line with origin 0, namely  $\mathbf{R}_+ z^{k_0}$ . Since  $|z| = 1$ , we get that  $z^k = z^{k_0}$  for every  $k \in F$ , so  $z^{k-\ell} = 1$  for every  $(k, \ell) \in F^2$ . But Bézout's lemma tell us that  $d$  can be written as some (finite) linear combination with integer coefficients of the differences  $k - \ell$  where  $(k, \ell)$  ranges over  $F^2$ . Hence  $z^d = 1$ , which achieves the proof.

We now prove theorem 1.



Proof. As we have done in the introduction, we fix a random variable  $Z = (\xi, \gamma)$  with law  $\pi = \mu^{\otimes \mathbf{Z}^-} \otimes Q$ , and we set for every  $n \leq 0$ ,

$$Z_n = (S \times T)^{|n|}(Z) = (S^{|n|}(\xi), T^{\xi_{n+1} + \dots + \xi_0}(\gamma)).$$

Then  $S \times T$  is exact if and only if  $\mathcal{F}_{-\infty}^Z$  is trivial, namely if  $\mathbf{E}[R|\mathcal{F}_{-\infty}^Z] = \mathbf{E}[R]$  a.s. for every  $R \in L^2(\mathcal{F}_0^Z)$ . But the set of all  $R \in L^2(\mathcal{F}_0^Z)$  satisfying this equality is a closed linear subspace of  $L^2(\mathcal{F}_0^Z)$ , so one needs only to consider the case where

$$R = \mathbf{I}_{\{(\xi_{m+1}, \dots, \xi_0) = (x_{m+1}, \dots, x_0)\}} f(\gamma),$$

with  $m \leq 0$ ,  $(x_{m+1}, \dots, x_0) \in F^{|m|}$  and  $f \in L^2(Q)$ .

Set  $S_n = -\xi_{n+1} - \dots - \xi_0$  for every  $n \leq 0$ <sup>2</sup> and  $s_m = -x_{m+1} - \dots - x_0$ . Then the random variable  $R$  above can be written

$$R = \mathbf{I}_{\{(\xi_{m+1}, \dots, \xi_0) = (x_{m+1}, \dots, x_0)\}} f\left(T^{s_m}(T^{-S_m}(\gamma))\right).$$

Since  $T^{-S_m}(\gamma)$  is  $\mathcal{F}_m^Z$ -measurable whereas  $(\xi_{m+1}, \dots, \xi_0)$  is independent of  $\mathcal{F}_m^Z$ , one has

$$\mathbf{E}[R|\mathcal{F}_m^Z] = cf\left(T^{s_m}(T^{-S_m}(\gamma))\right) \text{ with } c := \mu\{x_{m+1}\} \cdots \mu\{x_0\}.$$

In the same way, for every  $n \geq 0$ , the random variable  $T^{-S_{m-n}}(\gamma)$  is  $\mathcal{F}_{m-n}^Z$ -measurable whereas the random variable  $S_m - S_{m-n} = \xi_{m-n+1} + \dots + \xi_m$  is independent of  $\mathcal{F}_{m-n}^Z$  and has distribution  $\mu^{*n}$ , so

$$\mathbf{E}[R|\mathcal{F}_{m-n}^Z] = c \sum_{k \in \mathbf{Z}} \mu^{*n}(k) (f \circ T^{s_m - k})(T^{-S_{m-n}}(\gamma)).$$

Call  $g$  the orthogonal projection of  $f$  on  $L^2(Q)_0$  and  $\varphi$  the characteristic function of  $\mu$ , like in lemma 8. Since the law of  $T^{-S_{m-n}}(\gamma)$  is  $Q$ , and since  $T$  preserves  $Q$ , we get

$$\begin{aligned} \text{Var}(\mathbf{E}[R|\mathcal{F}_{m-n}^Z]) &= c^2 \sum_{k, \ell \in \mathbf{Z}} \mu^{*n}(k) \mu^{*n}(\ell) \text{Cov}(f \circ T^{s_m - k}(\gamma), f \circ T^{s_m - \ell}(\gamma)) \\ &= c^2 \sum_{k, \ell \in \mathbf{Z}} \mu^{*n}(k) \mu^{*n}(\ell) \langle g \circ T^{s_m - k}, g \circ T^{s_m - \ell} \rangle_{L^2(Q)} \\ &= c^2 \int_{\mathbf{U}} \left( \sum_{k, \ell \in \mathbf{Z}} \mu^{*n}(k) \mu^{*n}(\ell) z^{k - \ell} \right) d\sigma_g(z) \\ &= c^2 \int_{\mathbf{U}} \left| \sum_{k \in \mathbf{Z}} \mu^{*n}(k) z^k \right|^2 d\sigma_g(z) \\ &= c^2 \int_{\mathbf{U}} |\varphi(z)|^{2n} d\sigma_g(z). \end{aligned}$$

On the one hand,  $\mathbf{E}[R|\mathcal{F}_{m-n}^Z] \xrightarrow{n \rightarrow +\infty} \mathbf{E}[R|\mathcal{F}_{-\infty}^Z]$  in  $L^2(P)$  by the backwards martingale convergence theorem. On the other hand and  $|\varphi(z)| \leq 1$  for every  $z \in \mathbf{U}$  with equality if and only if  $z \in \mathbf{U}_d$ . Hence, letting  $n$  go to infinity, we get

$$\text{Var}(\mathbf{E}[R|\mathcal{F}_{-\infty}^Z]) = c^2 \sigma_g(\mathbf{U}_d).$$

<sup>2</sup>We choose this definition with the minus signs to have the same recursion relation as in the usual random walks, namely  $S_n = S_{n-1} + \xi_n$ .

Hence theorem 1 follows, by lemma 7.

Remark : the implication ‘ $\mathcal{F}_{-\infty}^Z$  trivial  $\implies T^d$  ergodic’ can be proved directly as follows. Assume that  $\mathcal{F}_{-\infty}^Z$  is trivial. By definition of  $d$ , the support  $F$  of  $\mu$  is contained in some coset  $r + d\mathbf{Z}$ , so the random variables  $\xi_{-n+1} + \dots + \xi_0$  takes values in  $nr + d\mathbf{Z}$ .

Let  $B \in \mathcal{G}$  be a subset such that  $T^{-d}(B) = B$ . Since the subsets  $T^k(B)$ ,  $k \in \mathbf{Z}$ , are also invariant by  $T^d$ , we get that for every  $n \geq 0$ ,

$$\{\gamma \in B\} = \{T^{\xi_{-n+1} + \dots + \xi_0}(\gamma) \in T^{rn}(B)\} = \{(S \times T)^n(\xi, \gamma) \in F^{\mathbf{Z}^-} \times T^{rn}(B)\}.$$

Hence  $\{\gamma \in B\} \in \mathcal{F}_{-\infty}^Z$ , so  $Q(B) = \mathbf{P}[\gamma \in B] \in \{0, 1\}$ . The ergodicity of  $T^d$  follows.

## 2.2 A question of Sinai on $K$ -partitions

In the whole section,  $(E, \mathcal{E}, \pi)$  denotes a non-trivial Lebesgue probability space and  $\tau$  an automorphism of  $(E, \mathcal{E}, \pi)$ .

To any random variable  $X$  on  $(E, \mathcal{E}, \pi)$  with values in  $\mathbf{R}$  (or more generally in any Polish space), we associate the measurable partition  $(X^{-1}(r))_{r \in \mathbf{R}}$  of  $E$ .

Rokhlin’s theory shows that there is a bijective correspondance between the measurable partitions up to the null sets and the sub- $\sigma$ -fields of  $\mathcal{E}$  up to the null sets.

By definition,  $\tau$  is a  $K$ -automorphism if there exists some sub- $\sigma$ -field  $\mathcal{S}_0$  of  $\mathcal{E}$  such that

- $\tau^{-1}\mathcal{S}_0 \subset \mathcal{S}_0$  (equivalently, the family  $(\tau^k\mathcal{S}_0)_{k \in \mathbf{Z}}$  is non-decreasing);
- $\bigvee_{n \in \mathbf{Z}} \tau^{-n}\mathcal{S}_0 = \mathcal{E}$  modulo  $\pi$  (equivalently  $\tau^k\mathcal{S}_0 \rightarrow \mathcal{E}$  as  $k \rightarrow +\infty$ );
- $\bigcap_{n \in \mathbf{Z}} \tau^{-n}\mathcal{S}_0 = \{\emptyset, E\}$  modulo  $\pi$  (equivalently  $\tau^k\mathcal{S}_0 \rightarrow \{\emptyset, E\}$  as  $k \rightarrow -\infty$ ).

Such a  $\sigma$ -subalgebra  $\mathcal{S}_0$  is called a  $K$ -sub- $\sigma$ -field, and the measurable partition  $\xi_0$  corresponding to  $\mathcal{S}_0$  is called a  $K$ -partition.

Bernoulli automorphisms are  $K$ -automorphisms. Indeed, if  $\gamma$  is an independent generator of a Bernoulli automorphism, then the partition  $\bigvee_{n \leq 0} \tau^{-n}\gamma$  is a  $K$ -partition.

The following question was raised by Sinai: *fix a (non-trivial)  $K$ -automorphism  $\tau$  on a Lebesgue space  $(E, \mathcal{E}, \pi)$ . Given a  $K$ -partition  $\xi_0$ , what are the possible values for the conditional entropy*

$$H_c(\xi_0) := H(\xi_0 | \bigvee_{n \geq 1} \tau^{-n}\xi_0)?$$

Before going on, it is worth giving a few explanations. Call  $\mathcal{S}_0$  the  $\sigma$ -field corresponding to  $\xi_0$ .

- Since the family  $(\tau^k\mathcal{S}_0)_{k \in \mathbf{Z}}$  is non-decreasing, one has  $H_c(\xi_0) = H(\xi_0 | \tau^{-1}\xi_0)$ .
- If  $\gamma$  is a countable generator of  $\tau$  with finite entropy, then  $\bigvee_{n \leq 0} \tau^{-n}\gamma$  is a  $K$ -partition whose conditional entropy equals  $h(\tau)$ .
- Even if  $\xi_0$  is uncountable,  $H_c(\xi_0)$  can still be finite; in this case the conditional law of  $\xi_0$  given  $\bigvee_{n \geq 1} \tau^{-n}\xi_0$  is almost surely carried by some random countable set.

- One has  $H_c(\xi_0) > 0$ . To see this, we argue by contradiction, assuming that  $H_c(\xi_0)$  equals 0. Then  $\xi_0$  would be measurable with regard to  $\bigvee_{n \geq 1} \tau^{-n} \mathcal{S}_0$ ; by translation,  $\xi_{-1}$  would be measurable with regard to  $\bigvee_{n \geq 2} \tau^{-n} \mathcal{S}_0$ , and so on, hence  $\xi_0$  would be measurable with regard to the tail  $\sigma$ -field  $\bigcap_{n \in \mathbf{Z}} \tau^{-n} \mathcal{S}_0$ , which is assumed to be trivial. By translation, each  $\sigma$ -field  $\tau^{-n} \mathcal{S}_0$  would be trivial, so  $\mathcal{E}$  itself would be trivial.
- One has  $H_c(\xi_0) \leq h(\tau)$ . Indeed, one may assume that  $\xi$  is the partition  $(X^{-1}(r))_{r \in \mathbf{R}}$  associated to some real random variable  $X$ . Then

$$H_c(\xi_0) = H(X | (X \circ \tau^n)_{n \geq 1}) = \sup_f H(f \circ X | (X \circ \tau^n)_{n \geq 1}),$$

where the supremum above is taken over all Borel functions  $f : \mathbf{R} \rightarrow \mathbf{R}$  taking countably many values. But given such a function  $f$ , the discrete random variable  $f \circ X$  yields a countable partition  $\xi^f$  of  $E$ , so

$$\begin{aligned} h(\tau) \geq h(\tau, \xi_f) &= H(\xi^f | \bigvee_{n \geq 1} \tau^{-n} \xi^f) \\ &= H(f \circ X | (f \circ X \circ \tau^n)_{n \geq 1}) \\ &\geq H(f \circ X | (X \circ \tau^n)_{n \geq 1}). \end{aligned}$$

The statement follows.

- The inequality  $H_c(\xi_0) \leq h(\tau)$  may be strict.

Using results obtained by den Hollander and Steif [9] on generalizations of the  $[T, T^{-1}]$  process, Vershik shows in [26] that if  $\tau$  is Bernoulli with infinite entropy, then  $H_c(\xi_0)$  can attain any value in  $]0, +\infty[$ .

In [16] (Theorem 1.4), Lindenstrauss, Peres and Schlag show that when  $\tau$  is Bernoulli (with finite or infinite entropy),  $H_c(\xi_0)$  can attain any value in  $]0, h(\tau)[$ .

Using Austin's theorem [1] (every ergodic automorphism is isomorphic to the direct product of some Bernoulli shift and some automorphism with arbitrary small entropy), one sees that this conclusion still holds when  $\tau$  is only assumed to be a  $K$ -automorphism. Let us explain why. Given  $\varepsilon > 0$ , Austin's theorem ensures that  $\tau$  is isomorphic to  $S \times T$ , where  $S$  is Bernoulli and  $h(T) \leq \varepsilon$ , and both factors  $S$  and  $T$  are necessarily  $K$ -automorphisms. Fix a  $K$ -partition  $\zeta_0$  of  $T$ . Then  $H_c(\zeta_0) \leq h(T) \leq \varepsilon$ . For every  $K$ -partition  $\eta_0$  of  $S$ , the partition  $\eta_0 \otimes \zeta_0$  is a  $K$ -partition of  $S \times T$  and  $H_c(\eta_0 \otimes \zeta_0) = H_c(\eta_0) + H_c(\zeta_0)$ . Since  $H_c(\eta_0)$  can attain any value in  $]0, h(S)[$ ,  $H_c(\eta_0 \otimes \zeta_0)$  can attain any value in  $]H_c(\zeta_0), h(S) + H_c(\zeta_0)[$ , therefore any value in  $]\varepsilon, H_c(\tau) - \varepsilon[$ .

Let us come back to the particular case of Bernoulli shifts. The proof given by Lindenstrauss, Peres and Schlag relies on the following theorem (Theorem 1.1 in [16]).

**Theorem 9 (Lindenstrauss, Peres, Schlag)** *Endow the space  $\{-1, 1\}^{\mathbf{Z}}$  with the probability measure  $\tilde{\mu}_p^{\otimes \mathbf{Z}}$ , where  $\tilde{\mu}_p = (1-p)\delta_{-1} + p\delta_1$ . For every  $\lambda \in [1/2, 1[$ , define a stationary process  $(Y_n^{(\lambda)})_{n \in \mathbf{Z}}$  on the probability space  $(\{-1, 1\}^{\mathbf{Z}}, \tilde{\mu}_p^{\otimes \mathbf{Z}})$  by*

$$\forall \omega \in \{-1, 1\}^{\mathbf{Z}}, \quad Y_n^{(\lambda)}(\omega) := \sum_{k \geq 0} \omega_{n+k} \lambda^k.$$

*Then there exists a Lebesgue-negligible subset  $N$  of  $[1/2, 1[$ , which does not depend on  $p$  such that for all  $p \in [1/3, 2/3]$  and  $\lambda \in [1/2, 1[ \setminus N$  such that  $H(p) + \log_2 \lambda > 0$ ,*

- the left tail  $\sigma$ -field  $\bigcap_{n \in \mathbf{Z}} \sigma((Y_k)_{k \leq n})$  is trivial,
- $H(Y_0^{(\lambda)} | \sigma((Y_n^{(\lambda)})_{n \leq -1})) = H(p) + \log_2 \lambda$ .

Hence, if  $T$  denotes the shift operator defined on  $\{-1, 1\}^{\mathbf{Z}}$  by  $T(\omega)_n = \omega_{n-1}$ , the partition  $\xi_0^{(\lambda)}$  of  $\{-1, 1\}^{\mathbf{Z}}$  associated to the random variable  $Y_0^{(\lambda)}$  is a  $K$ -partition of  $T$  and  $H_c(\xi_0^{(\lambda)}) = H(p) + \log_2 \lambda$ .

The proof of this theorem involves very technical analytic estimates. The major difficulty is to show that the left tail  $\sigma$ -field is trivial.

As suggested by Thouvenot and the referee, we show below that an application of our theorem 1 yields a simpler proof. Our construction involves Meilijson's skew products and uses Ornstein's theorem stating that Bernoulli automorphisms having the same entropy are isomorphic.

### 2.3 Alternative proof using Meilijson's skew products

Let  $F = \{0, 1\}$  and  $\Lambda$  be a finite non-empty alphabet.

Let  $S$  be the unilateral shift on  $F^{\mathbf{Z}^-}$  defined by  $S(f)(n) = f(n-1)$ , and  $\bar{S}$  be the bilateral shift on  $F^{\mathbf{Z}}$  defined by the same formula.

Let  $T$  be the bilateral shift on  $\Lambda^{\mathbf{Z}}$  defined by  $T(g)(n) = g(n-1)$ .

Fix  $p \in ]0, 1[$  and set  $\mu_p = (1-p)\delta_0 + p\delta_1$ . Let  $\nu$  be any probability measure on  $\Lambda$ . We endow  $F^{\mathbf{Z}^-}$ ,  $F^{\mathbf{Z}}$  and  $\Lambda^{\mathbf{Z}}$  with their product  $\sigma$ -fields and with the probability measures

$$P = \mu^{\otimes \mathbf{Z}^-}, \quad \bar{P} = \mu^{\otimes \mathbf{Z}}, \quad \text{and} \quad Q = \nu^{\otimes \mathbf{Z}}.$$

Hence  $S$ ,  $\bar{S}$  and  $T$  are Bernoulli shifts.

The skew products  $S \rtimes T$  and  $\bar{S} \rtimes T$  are the transformations on  $F^{\mathbf{Z}^-} \times \Lambda^{\mathbf{Z}}$  and  $F^{\mathbf{Z}} \times \Lambda^{\mathbf{Z}}$  given by  $S \rtimes T(f, g) = (S(f), T^{f(0)}(g))$  and  $\bar{S} \rtimes T(f, g) = (\bar{S}(f), T^{f(0)}(g))$ . These transformations preserve the probability measures  $P \otimes Q$  and  $\bar{P} \otimes Q$ .

The next result is classical.

**Proposition 10** *The transformation  $\bar{S} \rtimes T$  is Bernoulli with entropy  $H(\mu_p) + pH(\nu)$ .*

Proof. One may assume that  $\Lambda = \llbracket 1, s \rrbracket$ . Set  $q_j = \nu\{j\}$  for each  $j \in \llbracket 1, s \rrbracket$ . Call  $\gamma = \{C_0, \dots, C_s\}$  the partition of  $F^{\mathbf{Z}} \times \Lambda^{\mathbf{Z}}$  defined by

$$\begin{aligned} C_0 &:= \{(f, g) \in F^{\mathbf{Z}} \times \Lambda^{\mathbf{Z}} : f(0) = 0\}, \\ C_j &:= \{(f, g) \in F^{\mathbf{Z}} \times \Lambda^{\mathbf{Z}} : (f(0), g(0)) = (1, j)\} \text{ if } j \neq 0. \end{aligned}$$

Then  $(P \otimes Q)(C_j) = r_j$ , where  $r_0 = 1 - p$  and  $r_j = pq_j$  for every  $j \in \llbracket 1, s \rrbracket$ . We claim that the partition  $\gamma$  is an independent generator of  $\bar{S} \rtimes T$ , so  $\bar{S} \rtimes T$  is Bernoulli and

$$h(\bar{S} \rtimes T) = H(\gamma) = (1-p) \log_2(1-p) + \sum_{j=1}^s pq_j \log_2(pq_j) = H(\mu_p) + pH(\nu).$$

Hence what we have to prove is that the translated partitions  $\gamma_n = (\bar{S} \rtimes T)^{-n} \gamma$  for  $n \in \mathbf{Z}$  are independent and generate the product  $\sigma$ -field on  $F^{\mathbf{Z}} \times \Lambda^{\mathbf{Z}}$ .

To check this, introducing random variables is helpful. Endow the space  $F^{\mathbf{Z}} \times \Lambda^{\mathbf{Z}}$  with the probability measure  $P \otimes Q$ . Call  $(X_n)_{n \in \mathbf{Z}}$  and  $(Y_n)_{n \in \mathbf{Z}}$  the coordinate processes associated to the factors  $F^{\mathbf{Z}}$  and  $\Lambda^{\mathbf{Z}}$ . Under the probability measure  $P \otimes Q$ , these two processes are independent, the random variables  $(X_n)_{n \in \mathbf{Z}}$  are i.i.d. with law  $\mu_p$  and the random variables  $(Y_n)_{n \in \mathbf{Z}}$  are i.i.d. with law  $\nu$ .

We define a ‘random walk’  $(\Sigma_n)_{n \in \mathbf{Z}}$  on  $\mathbf{Z}$  and indexed by  $\mathbf{Z}$  by

$$\Sigma_n = \sum_{1 \leq i \leq n} X_i \text{ if } n \geq 0, \quad \Sigma_n = - \sum_{n+1 \leq i \leq 0} X_i \text{ if } n \leq 0.$$

Since  $\Sigma_n - \Sigma_{n-1} = X_n \in \{0, 1\}$  for every  $n \in \mathbf{Z}$ , the process  $(\Sigma_n)_{n \in \mathbf{Z}}$  is non-decreasing and almost surely visits every integer.

To the partition  $\gamma = \{C_0, \dots, C_s\}$ , we associate a discrete random variable, still denoted by  $\gamma$  and taking values in  $\llbracket 0, s \rrbracket$ , as follows: for each  $(f, g) \in F^{\mathbf{Z}} \times \Lambda^{\mathbf{Z}}$ , we call  $\gamma(f, g)$  the index of the only block of the partition  $\gamma$  which contains  $(f, g)$ , so  $\gamma(f, g) = j$  whenever  $(f, g) \in C_j$ . Observe that  $\gamma = X_0 Y_0$ , i.e.  $\gamma = 0$  on the event  $[X_0 = 0]$ , whereas  $\gamma = Y_0$  on the event  $[X_0 \neq 0]$ .

Given  $n \in \mathbf{Z}$ , the random variable associated partition  $\gamma_n$  is  $\gamma \circ (\bar{S} \times T)^n = X_{-n} Y_{\Sigma_{-n}}$ . Indeed, a recursion shows that for every  $(f, g) \in F^{\mathbf{Z}} \times \Lambda^{\mathbf{Z}}$ ,

$$\begin{aligned} (\bar{S} \times T)^n(f, g) &= (\bar{S}^n(f), T^{f(0)+\dots+f(-n+1)}(g)) \text{ if } n \geq 0, \\ (\bar{S} \times T)^n(f, g) &= (\bar{S}^n(f), T^{-f(1)-\dots-f(|n|)}(g)) \text{ if } n \leq 0, \end{aligned}$$

so

$$\begin{aligned} (X_0, Y_0)(\bar{S} \times T)^n(f, g) &= (f(-n), g(-f(0) - \dots - f(-n+1))) \text{ if } n \geq 0, \\ (X_0, Y_0)(\bar{S} \times T)^n(f, g) &= (f(-n), g(f(1) + \dots + f(-n))) \text{ if } n \leq 0, \end{aligned}$$

Now, fix  $N \geq 0$  and  $j_{-N}, \dots, j_N$  in  $\llbracket 0, q \rrbracket$ . Call  $I$  the subset of all  $i \in \llbracket -N, N \rrbracket$  such that  $j_i \neq 0$  and  $I^c$  its complement in  $\llbracket -N, N \rrbracket$ . For every  $n \in \llbracket -N, N \rrbracket$ , set

$$\begin{aligned} s_n &= \sum_{1 \leq i \leq n} \mathbf{1}_I(i) = |I \cap \llbracket 1, n \rrbracket| \text{ if } n \geq 0, \\ s_n &= - \sum_{n+1 \leq i \leq 0} \mathbf{1}_I(i) = -|I \cap \llbracket n+1, 0 \rrbracket| \text{ if } n \leq 0. \end{aligned}$$

On the event  $[(X_{-n} Y_{\Sigma_{-n}}, \dots, X_n Y_{\Sigma_n} = (j_{-n}, \dots, j_n)]$ , we have  $\Sigma_i = s_i$  for every  $i \in I$ . Thus

$$[(X_{-n} Y_{\Sigma_{-n}}, \dots, X_n Y_{\Sigma_n} = (j_{-n}, \dots, j_n)] = \bigcap_{i \in I^c} [X_i = 0] \cap \bigcap_{i \in I} [X_i = 1; Y_{s_i} = j_i].$$

Since the integers  $(s_i)_{i \in I}$  are all different, we get

$$\begin{aligned} (P \otimes Q)[(X_{-n} Y_{\Sigma_{-n}}, \dots, X_n Y_{\Sigma_n} = (j_{-n}, \dots, j_n)] &= \prod_{i \in I^c} (1-p) \times \prod_{i \in I} (pq_j) \\ &= \prod_{i \in \llbracket -N, N \rrbracket} r_j. \end{aligned}$$

The independence of the random variables  $(X_n Y_{\Sigma_n})_{n \in \mathbf{Z}}$  follows.

Moreover, the sequences  $(X_n)_{n \in \mathbf{Z}}$  and  $(Y_m)_{m \in \mathbf{Z}}$  can be recovered from  $(X_n Y_{\Sigma_n})_{n \in \mathbf{Z}}$ . Indeed, for every  $n \in \mathbf{Z}$ ,  $X_n = \mathbf{1}_{[X_n Y_{\Sigma_n} \neq 0]}$ . For every  $m \in \mathbf{Z}$ , the random variable  $I_m = \inf\{n \in \mathbf{Z} : \Sigma_n \geq m\}$  is completely determined by sequence  $(X_n)_{n \in \mathbf{Z}}$  and almost surely finite; on the almost sure event  $[I_m \in \mathbf{Z}]$ , one has  $\Sigma_{I_m} = m$  and  $X_{I_m} = 1$ , so  $Y_m = X_{I_m} Y_{\Sigma_{I_m}}$ .

The proof is complete.  $\square$

We now use Theorem 1 to get a  $K$ -partition of  $\bar{S} \times T$  and we compute its conditional entropy.

**Proposition 11** *Keep the notations of proposition 10. Let  $Z$  be the canonical projection  $(f, g) \mapsto (f|_{\mathbf{Z}_-}, g)$  from  $F^{\mathbf{Z}} \times \Lambda^{\mathbf{Z}}$  to  $F^{\mathbf{Z}_-} \times \Lambda^{\mathbf{Z}}$ . Call  $\mathcal{C}$  the product  $\sigma$ -field on  $F^{\mathbf{Z}_-} \times \Lambda^{\mathbf{Z}}$ ,  $\mathcal{Z}_0 = Z^{-1}\mathcal{C}$  and  $\zeta_0$  the measurable partition of  $F^{\mathbf{Z}} \times \Lambda^{\mathbf{Z}}$  associated to  $Z$ . Then  $\zeta_0$  is a  $K$ -partition and*

$$H_c(\zeta_0) = H(\zeta_0 | (\bar{S} \times T)^{-1}\zeta_0) = H(\mu_p).$$

*Proof.* Define the coordinate processes  $(X_n)_{n \in \mathbf{Z}}$  and  $(Y_n)_{n \in \mathbf{Z}}$ , and the ‘random walk’  $(\Sigma_n)_{n \in \mathbf{Z}}$  as above. Then  $Z = ((X_k)_{k \leq 0}, (Y_n)_{n \in \mathbf{Z}})$  can be viewed as a random variable on  $F^{\mathbf{Z}_-} \times \Lambda^{\mathbf{Z}}$  whose law is  $P \times Q$ .

By construction,  $(S \times T) \circ Z = Z \circ (\bar{S} \times T)$ . Since  $Z \circ (\bar{S} \times T)$  is a measurable function of  $Z$ , we derive  $(\bar{S} \times T)^{-1}(\mathcal{Z}_0) \subset \mathcal{Z}_0$ .

The sequence  $(Y_n)_{n \in \mathbf{Z}}$  is the second component of  $Z$  and for each  $n \in \mathbf{Z}$ ,  $X_{-n}$  is the 0-coordinate of the first component of  $Z \circ (\bar{S} \times T)^n$ . Therefore, the  $\sigma$ -fields  $((\bar{S} \times T)^{-n}(\mathcal{Z}_0))_{n \in \mathbf{Z}}$  generate the product  $\sigma$ -field  $\mathcal{C}$  on  $F^{\mathbf{Z}} \times \Lambda^{\mathbf{Z}}$ .

Now, define a stationary Markov process  $(Z_n)_{n \in \mathbf{Z}}$  by  $Z_n = Z \circ (\bar{S} \times T)^{-n}$  and call  $(\mathcal{F}_n^Z)_{n \leq 0}$  its natural filtration. Then for all  $n \in \mathbf{Z}$ ,  $(\bar{S} \times T)^n \mathcal{Z} = Z_n^{-1} \mathcal{C} \subset \mathcal{F}_n^Z$  and for  $n \leq 0$ , we have  $Z_n = (S \times T)^{-n} \circ Z$ . By Theorem 1, the measure-preserving map  $S \times T$  is exact since  $T$  is ergodic, hence the tail  $\sigma$ -field  $\mathcal{F}_{-\infty}^Z$  is trivial. But

$$\bigcap_{n \in \mathbf{Z}} (\bar{S} \times T)^n \mathcal{Z} \subset \mathcal{F}_{-\infty}^Z.$$

Hence,  $\zeta_0$  is a  $K$ -partition of  $\bar{S} \times T$ .

Last,

$$H(\zeta_0 | (\bar{S} \times T)^{-1}\zeta_0) = H(Z | Z \circ (\bar{S} \times T)) = H(Z | (S \times T) \circ Z) = H(X_0) = H(\mu_p),$$

since the knowledge of  $Z$  is equivalent to the knowledge of  $X_0$  and  $(S \times T)(Z)$  and since  $X_0$  is independent of  $(S \times T)(Z)$ .  $\square$

Let us now derive a simpler proof of Theorem 1.4 of Lindenstrauss, Peres, Schlag [16].

First, consider the case where  $\tau$  has a finite entropy. By Ornstein’s theorem, (bilateral) Bernoulli shifts with same entropy are isomorphic. Hence,  $\tau$  is isomorphic to the Cartesian product of  $n$  skew products  $S \times T$  like above, provided that  $n \times (H(\mu_p) + pH(\nu)) = h(\tau)$ . Let  $\xi_0$  be the partition corresponding to  $\zeta_0^{\otimes n}$  via this isomorphism. Then  $\xi_0$  is a  $K$ -partition and  $H_c(\xi_0) = H_c(\xi_0 | \tau^{-1}\xi_0) = n \times H(\mu_p)$ . Since the only constraint on the integer  $n \geq 1$ , the probability  $p \in ]0, 1[$  and the probability measure  $\nu$  is the relation  $n \times (H(\mu_p) + pH(\nu)) = h(\tau)$ , the value of  $H_c(\xi_0)$  can attain any real number in  $]0, h(\tau)[$ .

The case where  $\tau$  has an infinite entropy can be treated in a similar way by choosing a sequence  $(p_n)_{n \geq 1}$  of parameters in  $]0, 1[$  and a sequence  $(\nu_n)_{n \geq 1}$  of probability measures on finite alphabets  $(\Lambda_n)_{n \geq 1}$  in such a way that  $\sum_{n \geq 1} H(\mu_{p_n})$  is the target value and  $\sum_{n \geq 1} p_n H(\nu_n) = +\infty$ , and by considering an infinite Cartesian product of Meiljson's skew products  $(S_n \times T_n)_{n \geq 1}$  like above.  $\square$

### 3 Vershik's tools

In this section, we introduce Vershik's standardness criteria. Most of the material of this section is abridged from [14].

#### 3.1 Immersion, productness and standardness

Unless otherwise specified, the filtrations  $(\mathcal{F}_n)_{n \leq 0}$  considered here are defined on a given probability space  $(\Omega, \mathcal{A}, \mathbf{P})$ , are indexed by the non-positive integers, and have essentially separable  $\sigma$ -field  $\mathcal{F}_0$ . This means that  $\mathcal{F}_0$  can be generated by countably many events (modulo the null sets), or equivalently that the Hilbert space  $L^2(\mathcal{F}_0)$  is separable.

An important notion in the theory of filtrations is the notion of immersion.

**Definition 12** *Let  $(\mathcal{F}_n)_{n \leq 0}$  and  $(\mathcal{G}_n)_{n \leq 0}$  be two filtrations. One says that  $(\mathcal{F}_n)_{n \leq 0}$  is immersed in  $(\mathcal{G}_n)_{n \leq 0}$  if for every  $n \leq 0$ ,  $\mathcal{F}_n \subset \mathcal{G}_n$  and  $\mathcal{F}_0$  and  $\mathcal{G}_n$  are conditionally independent with regard to  $\mathcal{F}_n$ . An equivalent definition is that every martingale in  $(\mathcal{F}_n)_{n \leq 0}$  is still a martingale in  $(\mathcal{G}_n)_{n \leq 0}$ .*

We refer the reader to [4] or [12] to find more details on this notion. In the present paper, the immersions will follow from the next lemma.

**Lemma 13** *Let  $(\mathcal{F}_n)_{n \leq 0}$  and  $(\mathcal{G}_n)_{n \leq 0}$  be two filtrations such that  $\mathcal{F}_n \subset \mathcal{G}_n$  for every  $n \leq 0$ . If  $(\mathcal{F}_n)_{n \leq 0}$  and  $(\mathcal{G}_n)_{n \leq 0}$  admit a common sequence of innovations, then  $(\mathcal{F}_n)_{n \leq 0}$  is immersed in  $(\mathcal{G}_n)_{n \leq 0}$ .*

We now introduce the notion of standard filtration.

**Definition 14** *A filtration  $(\mathcal{F}_n)_{n \leq 0}$  is standard if it is isomorphic to another filtration (possibly defined on another probability space) which is immersed in some product-type filtration (having also essentially a separable final  $\sigma$ -field).*

Actually, when Vershik defined standardness, he considered only poly-adic filtrations, and he defined standardness as productness: according to Vershik's definition, standard poly-adic filtrations are product-type poly-adic filtrations. Fortunately, these two definitions of standardness coincide on poly-adic filtrations.

**Theorem 15** *Every poly-adic filtration immersed in some product-type filtration is also product-type.*

Actually, this non-trivial statement is a key result in Vershik's theory and relies on Vershik's second level criterion, stated later in this section.

We now introduce the three Vershik's properties (first level, intermediate and second level, according to the terminology used by Émery, Schachermayer and Laurent) which leads to the three corresponding Vershik's criteria. Defining the three Vershik's properties is interesting since each of them helps to understand the other ones, although we essentially use Vershik's intermediate criteria in the present paper.

### 3.2 Vershik's first level property

Let  $(\mathcal{F}_n)_{n \leq 0}$  be a poly-adic filtration: for each  $n \leq 0$  one can find a uniform random variable  $\xi_n$  with values in some finite set  $F_n$  such that  $\xi_n$  is an *innovation at time  $n$*  of the filtration  $(\mathcal{F}_n)_{n \leq 0}$ , namely

$$\mathcal{F}_n = \mathcal{F}_{n-1} \vee \sigma(\xi_n) \quad \text{mod } \mathbf{P}, \quad \text{with } \xi_n \text{ independent of } \mathcal{F}_{n-1}.$$

One can check that the sequence of sizes  $(|F_n|)_{n \leq 0}$  is uniquely determined. This sequence is called the *adicity* of the filtration  $(\mathcal{F}_n)_{n \leq 0}$ .

When  $(\mathcal{F}_n)_{n \leq 0}$  is dyadic, the set  $F_n$  can be chosen to be independent of  $n$ . For example, we will take  $F = \{0, 1\}$  (respectively  $F = \{-1, 1\}$ ) when we work with the filtration associated to  $[\text{Id}, T]$  (respectively  $[T, T^{-1}]$ ).

A first important thing to understand is the way to get any sequence of innovations from the original one.

**Lemma 16** *From  $(\xi_n)_{n \leq 0}$ , one can get another sequence of innovations as follows: for each  $n \leq 0$ , fix any  $\mathcal{F}_{n-1}$ -measurable random permutation  $\Sigma_n$  on  $F_n$  and set  $\eta_n = \Sigma_n(\xi_n)$ . Conversely, every sequence of innovations of  $(\mathcal{F}_n)_{n \leq 0}$  with values in the sets  $(F_n)_{n \leq 0}$  can be obtained in this way.*

The next important thing to understand is that two different systems of innovations may carry a different information. The simplest example is the situation where  $(\xi_n)_{n \leq 0}$  are (independent and) uniform on  $\{-1, 1\}$  and  $\eta_n = \xi_{n-1}\xi_n$  for every  $n \leq 0$ . By lemma 16,  $(\eta_n)_{n \leq 0}$  is still a sequence of innovations, but it carries less information than the sequence  $(\xi_n)_{n \leq 0}$ , since  $\xi_0$  is independent of  $(\eta_n)_{n \leq 0}$ .

This remark opens the possibility for the filtration  $(\mathcal{F}_n)_{n \leq 0}$  to be product-type even if it is not generated by the original sequence of innovations  $(\xi_n)_{n \leq 0}$ . The example of the simple irrational random walk on the circle  $\mathbf{R}/\mathbf{Z}$ , indexed by  $\mathbf{Z}_-$ , whose filtration is product-type although it is not generated by the sequence of steps, shows that this situation can actually occur, as we shall see in section 4.

The possibility or the impossibility of choosing a good sequence of innovations to approach a given random variable leads to Vershik's first level property.

Fix a separable complete metric space  $(A, d)$ , endowed with the Borel  $\sigma$ -field. Denote by  $L^1(\mathcal{F}_0; A)$  the set of all classes modulo almost sure equality of  $\mathcal{F}_0$ -measurable random variables  $X$  taking values in  $A$ , such that for some (equivalently, for all)  $a \in A$ , the real-random variable  $d(a, X)$  is integrable. Endow  $L^1(\mathcal{F}_0; A)$  with the distance defined by  $D(X, Y) = \mathbf{E}[d(X, Y)]$ .

**Definition 17 (First level Vershik property)** *Let  $X \in L^1(\mathcal{F}_0; A)$ . One says that  $X$  satisfies Vershik's first level property if for every  $\varepsilon > 0$ , one can approach  $X$  by some measurable function of finitely many innovations of the filtration  $(\mathcal{F}_n)_{n \leq 0}$  so that the distance in  $L^1(\mathcal{F}_0; A)$  is at most  $\varepsilon$ .*



### 3.3 Vershik's intermediate property

For the sake of simplicity, we focus on  $r$ -adic filtrations only, although the definitions and theorems below can be extended to all poly-adic filtrations. Actually, the simplifications occur essentially in the notations, since we work with the powers  $F^h$  of a fixed set  $F$  instead of products like  $\prod_{-h+1 \leq k \leq 0} F_k$ .

In the whole subsection, we fix an  $r$ -adic filtration  $(\mathcal{F}_n)_{n \leq 0}$  and a sequence  $(\xi_n)_{n \leq 0}$  of innovations taking values in a set  $F$  of size  $r \geq 2$ . As before,  $(A, d)$  denotes a separable complete metric space, endowed with the Borel  $\sigma$ -field.

The definition of Vershik's intermediate criterion relies on split-words processes, on the quotients of  $\ell^1$ -metrics on the sets  $A^{F^h}$ ,  $h \geq 0$  by the action of  $r$ -ary tree automorphisms, and on the notion of dispersion.

**Definition 18 (Split-word processes with given final value and innovations)**

Let  $X \in L^1(\mathcal{F}_0; A)$ . For every  $n \leq -1$ , there exists an  $\mathcal{F}_n$ -measurable random map  $W_n$  from  $F^{|n|}$  to  $A$  such that for each  $(x_{n+1}, \dots, x_0) \in A^{|n|}$ ,  $X = W_n(x_{n+1}, \dots, x_0)$  almost surely on the event  $\{(\xi_{n+1}, \dots, \xi_0) = (x_{n+1}, \dots, x_0)\}$ . Such a random map is almost surely unique. The process  $(W_n, \xi_n)_{n \geq 0}$  thus defined is the split-word process associated to  $X$ , to the filtration  $(\mathcal{F}_n)_{n \leq 0}$  and to the innovations  $(\xi_n)_{n \leq 0}$ .

The existence and the essential uniqueness that legitimise the definition above will be established in section 7, lemma 49.

Note that  $W_0$  is the map which sends the empty sequence  $() \in F^0$  on  $X$ . Informally, if we want at time  $n \leq 0$  to predict the future value of  $X$ , there are  $2^{|n|}$  possible (non necessarily distinct) values, one for each possible value of  $(\xi_{n+1}, \dots, \xi_0)$ . By definition,  $W_n(x_{n+1}, \dots, x_0)$  is the value of  $X$  that we will get if  $(\xi_{n+1}, \dots, \xi_0) = (x_{n+1}, \dots, x_0)$ .

The recursion formula  $W_n(x_{n+1}, \dots, x_0) = W_{n-1}(\xi_n, x_{n+1}, \dots, x_0)$  shows that the process  $(W_n, \xi_n)_{n \leq 0}$  is an inhomogeneous Markov chain and that  $(\xi_n)_{n \leq 0}$  is a sequence of innovations of the filtration  $(\mathcal{F}_n^{W, \xi})_{n \leq 0}$ . Hence the filtration  $(\mathcal{F}_n^{W, \xi})_{n \leq 0}$  is immersed in the filtration  $(\mathcal{F}_n)_{n \leq 0}$ .

If one fixes a total order on the set  $F$  and endows each  $F^{|n|}$  with the lexicographic order, then each  $W_n$  can be viewed as a word of length  $2^{|n|}$  on the alphabet  $A$ , namely the word  $(W_n(x_{n+1}, \dots, x_0))_{(x_{n+1}, \dots, x_0) \in F^{|n|}}$ . Furthermore,  $W_n$  is the left half or the right half of  $W_{n-1}$  according to the value of  $\xi_n$ . This explains the terminology 'split-word process' used. Note that the alphabet  $A$  can be uncountable and that the successive letters are not assumed to be independent unlike in the standard split-word process considered by Smorodinsky [21], Laurent [11] and Ceillier [2].

We now give a formal model of the automorphism group  $G_h$  of the  $r$ -ary tree with given height  $h \geq 0$ . Call

$$T_h = \bigcup_{i=0}^h F^i$$

the set of all sequences of elements of  $F$  with length  $\leq h$ . The set  $T_h$  can be viewed as the set of all vertices of a  $r$ -ary tree with height  $h \geq 0$ : the root is the empty sequence  $()$ , and the  $r$  children of a given vertex  $(x_0, \dots, x_{i-1}) \in F^i$  with  $i \leq h-1$  are the vertices  $(x_0, \dots, x_{i-1}, x_i)$  where  $x_i$  ranges over  $F$ .

Assume now that  $h \geq 1$ . To each family of permutations  $\sigma \in \mathfrak{S}(F)^{T_{h-1}}$ , we associate

a permutation  $g_\sigma \in \mathfrak{S}(T_h)$  preserving this tree structure by setting

$$g_\sigma(x_1, \dots, x_i) = (\sigma() (x_1), \sigma(x_1) (x_2), \dots, \sigma(x_1, \dots, x_{i-1}) (x_i))$$

for every  $(x_1, \dots, x_i) \in T_h$ . In this formula, the permutation  $\sigma(x_1, \dots, x_{i-1})$  acts on the subtrees under the vertex  $(x_1, \dots, x_{i-1})$  and the permutations associated to the shortest sequences are performed first.

Note that for every  $\sigma, \tau \in \mathfrak{S}(F)^{T_{h-1}}$ ,  $g_\tau \circ g_\sigma = g_{\tau\sigma}$  if one defines  $\tau\sigma$  by

$$\tau\sigma(x_1, x_2, \dots, x_i) = \tau(g_\sigma(x_1, \dots, x_i)) \circ \sigma(x_1, x_2, \dots, x_i).$$

This justifies the following definition.

**Definition 19 (Automorphism group of the  $r$ -ary tree  $T_h$ )** *The set  $G_h := \mathfrak{S}(F)^{T_{h-1}}$  endowed with the multiplication thus defined is a group and is isomorphic to the group  $(\{g_\sigma : \sigma \in G_h\}, \circ)$ , so we view  $G_h$  as the automorphism group of the  $r$ -ary tree  $T_h$ . We get an action of the group  $G_h$  on the set  $A^{F^h}$  by setting*

$$\forall(\sigma, w) \in G_h \times A^{F^h}, \quad \sigma \cdot w := w \circ g_\sigma^{-1}.$$

When  $F = \{-1, 1\}$ , the set  $\mathfrak{S}(F)$  is the pair  $\{\text{Id}, -\text{Id}\}$ . Figure 1 below gives an example of the action of such an automorphism on  $T_3$ . The left part of figure 3.3 represents  $T_3$ , the values of  $\sigma$  on each vertex of  $T_2$  (the + and - stand for Id and  $-\text{Id}$ ) and  $w$ : the images of the elements of  $\{-1, 1\}^3$  ordered in lexicographic order are denoted by a,b,c,d,e,f,g,h. The right part indicates the images of each vertex of  $T_3$  and  $\sigma \cdot w$ .

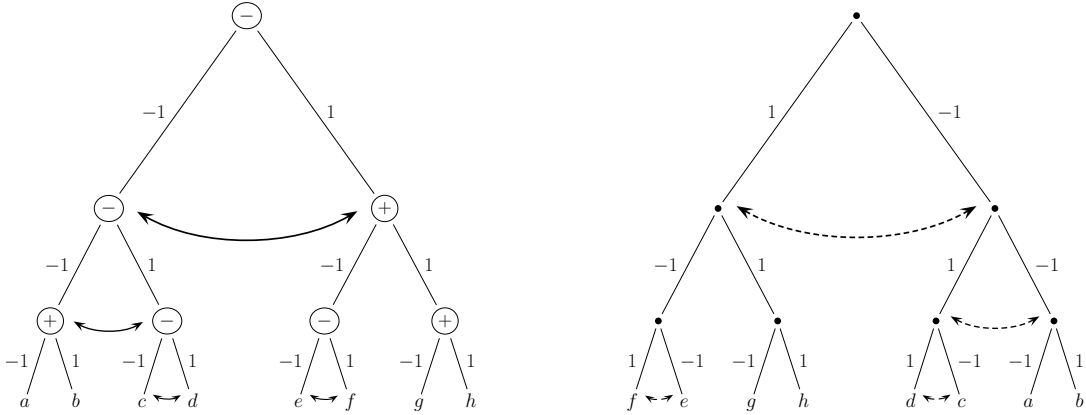


Figure 1: The map  $\sigma: T_2 \rightarrow \{-1, 1\}$  is represented by the symbols  $\ominus$  and  $\oplus$  on all vertices of the left tree but the leaves. The permutation  $g_\sigma: T_3 \rightarrow T_3$  sends each vertex of the left tree on the vertex on the right tree corresponding to the same label. The maps  $w: \{-1, 1\}^3 \rightarrow A$  and  $g_\sigma \cdot w: \{-1, 1\}^3 \rightarrow A$  can be identified with the words  $abcde fgh$  and  $feghdcab$  respectively.

We now define a metric on  $A^{F^h}$  and a metric modulo the  $r$ -ary tree automorphisms.

**Definition 20 (Metric and pseudo-metric on  $A^{F^h}$ )** *For every  $u$  and  $v$  in  $A^{F^h}$ , set*

$$\delta_{-h}(u, v) = \delta_h(u, v) = \frac{1}{2^h} \sum_{(x_1, \dots, x_h) \in F^h} d(u(x_1, \dots, x_h), v(x_1, \dots, x_h)),$$

$$d_{-h}(u, v) = d_h(u, v) = \min_{g \in G_h} \delta_h(u, g \cdot v).$$

In this definition, we use indifferently the index  $h$  or  $-h$  to handle lighter notations in the next subsection, in which the height  $h$  will be equal to  $-n$ , where  $n \leq 0$  is an instant prior to time 0.

The metric  $\delta_{-h} = \delta_h$  is just the Hamming distance, normalized to vary between 0 and 1. It is invariant under the action of the group  $G_h$  (each automorphism of the tree  $T_h$  induces a permutation on the leaves, i.e. on the set  $F^h$ ). The pseudo-metric  $d_{-h} = d_h$  measures the shortest distance between the orbits modulo the action of the group  $G_h$ .

Remind the definition  $G_h := \mathfrak{S}(F)^{T_{h-1}}$  given above. If  $h \geq 2$ , choosing an element  $\sigma$  in  $G_h$  is equivalent to choosing independently the permutation  $\sigma()$  in  $\mathfrak{S}(F)$  and the  $|F|$  elements  $(\sigma_{x_1})_{x_1 \in F}$  in  $G_{h-1} := \mathfrak{S}(F)^{T_{h-2}}$ , where  $\sigma_{x_1}(x_2, \dots, x_i) = \sigma(x_1, x_2, \dots, x_i)$  for every  $i \in \llbracket 0, h-1 \rrbracket$  and  $(x_1, \dots, x_i) \in F^i$ . From this observation, we derive the following recursion relation (which is still valid when  $h = 1$ ).

**Proposition 21** *Let  $h \geq 1$ . For every  $u$  and  $v$  in  $A^{F^h}$ , set*

$$d_h(u, v) = \min_{\varsigma \in \mathfrak{S}(F)} \frac{1}{|F|} \sum_{x_1 \in F} d_{h-1}(u(x_1, \cdot), v(\varsigma(x_1), \cdot)).$$

Last, we need to define the dispersion of random variables taking values in any pseudo-metric space.

**Definition 22 (Dispersion of an integrable random variable)** *Let  $(E, \mathcal{E})$  be any measurable space,  $e$  be a measurable pseudo-distance on  $(E, \mathcal{E})$ , and  $X$  be a random variable with values in  $(E, \mathcal{E})$ . By definition, the dispersion of a random variable  $X$  with regard to  $e$ , denoted by  $\text{disp}(X, e)$ , is the expectation of  $e(X', X'')$  where  $X'$  and  $X''$  are two independent copies of  $X$  defined on a same probability space.*

We can now define Vershik's intermediate property.

**Definition 23 (Vershik's intermediate property)** *Let  $X \in L^1(\mathcal{F}_0; A)$ . Denote by  $(W_n, \xi_n)_{n \geq 0}$  the split-word process associated to  $X$ , to the filtration  $(\mathcal{F}_n)_{n \leq 0}$  and to the sequence of innovations  $(\xi_n)_{n \leq 0}$ . One says that  $X$  satisfies Vershik's intermediate property if  $\text{disp}(W_n, d_n) \rightarrow 0$  as  $n \rightarrow -\infty$ .*

The next proposition shows that Vershik's intermediate property can be rephrased  $\inf_{n \leq 0} \text{disp}(W_n, d_n) = 0$  and that it depends only on the random variable  $X$  and the filtration  $(\mathcal{F}_n)_{n \leq 0}$ .

**Proposition 24** *The quantities  $\text{disp}(W_n, d_n)$  are a non-decreasing function of  $n$  and they do not depend on the sequence of innovations  $(\xi_n)_{n \leq 0}$ .*

Let us explain informally why the second part is true. By lemma 16, replacing  $(\xi_n)_{n \leq 0}$  with another sequence of innovations interchanges the letters in each word  $W_n$  according to a random automorphism of the  $r$ -ary tree with height  $|n|$ . This operation preserves the dispersion of the words  $W_n$  since the pseudo-metrics  $d_n$  identify words which are in the same orbit modulo the tree isomorphisms.

Proposition 24 will be proved at subsection 7.2 and will help us to show that Vershik's first level and intermediate properties are equivalent.

### 3.4 Vershik's second level property

Keep the notations of the last subsection. Vershik's second level property relies on the construction of a tower of measures with the help of Kantorovich-Rubinstein metrics.

**Definition 25 (Kantorovich-Rubinstein metric)** *Let  $(E, \rho)$  be a non-empty separable metric space. Call  $E'$  the set of all probability measures on  $(E, \mathcal{B}(E))$  having a finite first moment, namely the set of all probability measures  $\mu$  on  $(E, \mathcal{B}(E))$  such that for some (equivalently, for all)  $a \in E$ ,*

$$\int_E \rho(a, x) \, d\mu(x) < +\infty.$$

The Kantorovich-Rubinstein metric on  $E'$  is defined by

$$\rho'(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \int_{E^2} \rho(x, y) \, d\pi(x, y),$$

where  $\Pi(\mu, \nu)$  denotes the set of all probability measures on  $(E^2, \mathcal{B}(E^2))$  with margins  $\mu$  and  $\nu$ .

One can check that the metric  $\rho'$  is measurable with regard to the topology of narrow convergence. The topology defined by  $\rho'$  is finer than the topology of narrow convergence, and these two topologies coincide when  $(E, \rho)$  is compact.

The space  $(E', \rho')$  is still a separable metric space, thus the construction above can be iterated. Moreover,  $(E', \rho')$  is complete (or compact) whenever  $(E, \rho)$  is.

#### Definition 26 (Progressive predictions and Vershik's second level property)

*Let  $X \in L^1(\mathcal{F}_0; A)$ . The Vershik's progressive predictions of  $X$  with regard to  $(\mathcal{F}_n)_{n \leq 0}$  are the random variables  $\pi_n X \in L^1(\mathcal{F}_n; A^{(n)})$  defined recursively by  $\pi_0 X = X$  taking values in  $(A^{(0)}, d^{(0)}) = (A, d)$ , and for every  $n \leq -1$ ,  $\pi_n X = \mathcal{L}(\pi_{n-1} X | \mathcal{F}_n)$ , taking values in  $(A^{(n)}, d^{(n)}) = ((A^{(n-1)})', (d^{(n-1)})')$ . One says that  $X$  satisfies Vershik's second level property if  $\text{disp}(\pi_n X, d^{(n)}) \rightarrow 0$  as  $n \rightarrow -\infty$ .*

Actually the quantities  $\text{disp}(\pi_n X, d^{(n)})$  considered in Vershik's second level property are the same as the quantities  $\text{disp}(W_n, d_n)$  considered in Vershik's intermediate property, so these two properties are equivalent. One can check that they are also equivalent to the first level property. The equality  $\text{disp}(\pi_n X, d^{(n)}) = \text{disp}(W_n, d_n)$  follows from the next proposition, which can be proved by recursion.

**Proposition 27** *Define recursively the map  $i_n : A^{F^{|n|}} \rightarrow A^{(n)}$  for every  $n \leq 0$  by  $i_0 = \text{Id}_A$  and for every  $n \leq 0$*

$$i_{n-1}(w) = \frac{1}{|F|} \sum_{x \in F} \delta_{i_n(w(x, \cdot))}.$$

*Then  $i_n$  is an isometry from the pseudo-metric space  $(A^{F^{|n|}}, d_n)$  to the metric space  $(A^{(n)}, d^{(n)})$ . Moreover, given  $X \in L^1(\mathcal{F}_0; A)$ , the Vershik's progressive predictions of  $X$  can be derived from the split-word process  $(W_n)_{n \leq 0}$  associated to  $X$  by the formula  $\pi_n X = i_n(W_n)$ .*

### 3.5 Vershik's standardness criteria

We keep the notations of the last subsection. Since the three Vershik's properties (first level, intermediate, and second level) are equivalent (see subsection 7.2 and Theorem 4.9 in [14]), we do not distinguish them below. We can now state Vershik's standardness criteria.

**Theorem 28 (Vershik standardness criteria)**, *Let  $(\mathcal{F}_n)_{n \leq 0}$  be an  $r$ -adic filtration such that  $\mathcal{F}_0$  is essentially separable. Then  $(\mathcal{F}_n)_{n \leq 0}$  is product-type if and only if for every separable complete metric space  $(A, d)$ , every random variable in  $L^1(\mathcal{F}_0; A)$  satisfies Vershik's property.*

Considering every integrable random variable with values in any separable complete metric space makes a lot of generality. Actually, the properties below simplify a bit the verification of Vershik's criteria.

#### Proposition 29 (Stability properties)

1. *The set of all random variables in  $L^1(\mathcal{F}_0; A)$  which satisfy Vershik's property is closed in  $L^1(\mathcal{F}_0; A)$ .*
2. *If  $X \in L^1(\mathcal{F}_0; A)$  satisfies Vershik's property then every measurable function of  $X$  with values in some separable complete metric space satisfies Vershik's property.*
3. *Let  $n \leq 0$  and  $X \in L^1(\mathcal{F}_n; A)$ . Endow  $F$  with the discrete metric and  $A \times F^{|n|}$  with the product metric. If  $X$  satisfies Vershik's property, then the random variable  $(X, \xi_{n+1}, \dots, \xi_0)$  with values in the product  $A \times F^{|n|}$  satisfies Vershik's property.*

These stability properties allow us to restrict the class of separable complete metric spaces considered. For example, one can consider only  $\mathbf{R}$  endowed with the usual metric, or the class of all finite subsets of  $\mathbf{N}$  endowed with the discrete metric. And in many cases, the checking work can be reduced much more.

**Proposition 30 (Natural filtrations of Markov processes)** *Let  $(X_n)_{n \leq 0}$  be a (possibly inhomogeneous) Markov process in which each random variables  $X_n$  takes values in some separable bounded complete metric space (possibly depending on  $n$ ). Assume that the filtration  $(\mathcal{F}_n^X)_{n \leq 0}$  is  $r$ -adic.*

1. *Then  $(\mathcal{F}_n^X)_{n \leq 0}$  is product-type if and only if each  $X_n$  satisfies Vershik's property.*
2. *When the Markov process  $(X_n)_{n \leq 0}$  is stationary,  $(\mathcal{F}_n^X)_{n \leq 0}$  is product-type if and only  $X_0$  satisfies Vershik's property.*

## 4 First examples of application of Vershik's criterion

In this section, we apply Vershik's criterion to two rather simple situations, namely  $[T, T^{-1}]$  where  $T$  is an irrational rotation on the circle, and  $[T_a, T_b]$  where  $T_a$  and  $T_b$  are shifts related to the free group with generators  $a$  and  $b$ .

#### 4.1 $[T, T^{-1}]$ when $T$ is an irrational rotation on the circle

Endow  $\mathbf{T} := \mathbf{R}/\mathbf{Z}$  with the quotient pseudo-metric  $d$  of the usual metric on  $\mathbf{R}$ , and with the uniform measure  $Q$ . Actually,  $d$  is a translation-invariant metric on  $\mathbf{T}$ , and is bounded above by  $1/2$ . Fix an irrational real number  $\alpha$ , denote by  $\bar{\alpha}$  its equivalence class in  $\mathbf{T}$  and call  $T$  the translation  $x \mapsto x + \bar{\alpha}$  on  $\mathbf{T}$ . Let  $F = \{-1, 1\}$ ,  $\mu$  be the uniform law on  $F$ , and set  $\pi := \mu^{\otimes \mathbf{Z}_-} \otimes Q$ .

The transformation  $[T, T^{-1}]$  preserves  $\pi$  since  $T$  is an automorphism of  $(\mathbf{T}, \mathcal{B}(\mathbf{T}), Q)$  and  $[T, T^{-1}]$  is exact by Theorem 1 ( $T^2$  is ergodic since  $2\alpha$  is irrational). Equivalently, ‘the’ filtration associated to  $[T, T^{-1}]$  is Kolmogorovian, and we have to show that it is standard. To study this filtration, we fix a random variable  $((\xi_k)_{k \in \mathbf{Z}}, \gamma)$  with law  $\mu^{\otimes \mathbf{Z}} \otimes Q$  and we define a ‘random walk’ on  $\mathbf{Z}$  and indexed by  $\mathbf{Z}$  by

$$\begin{aligned} S_n &:= -\xi_{n+1} - \cdots - \xi_0 & \text{if } n \leq 0, \\ S_n &:= \xi_1 + \cdots + \xi_n & \text{if } n \geq 0, \end{aligned}$$

with the convention  $S_0 = 0$ . Hence  $S_n - S_{n-1} = \xi_n$  for every  $n \in \mathbf{Z}$ .

For every  $n \in \mathbf{Z}$ , set  $\gamma_n = T^{-S_n}\gamma = \gamma - S_n\alpha$  and  $Z_n = ((\xi_{k+n})_{n \leq 0}, \gamma_n)$ . Then for every  $n \leq 0$ ,  $\gamma_n = \gamma_{n-1} - \xi_n\alpha$ . As observed in the introduction, the process  $(\gamma_n)_{n \in \mathbf{Z}}$  is a stationary Markov chain governed by the sequence  $(\xi_k)_{k \in \mathbf{Z}}$ , and ‘the’ filtration associated to  $[T, T^{-1}]$ , i.e.  $(\mathcal{F}_n^Z)_{n \leq 0}$ , is also the natural filtration of  $(\gamma_n)_{n \leq 0}$ . By proposition 30, one only needs to check that  $\gamma_0$  satisfies Vershik’s first level property.

Fix  $\varepsilon > 0$ . Since the subset  $\mathbf{Z} + 2\alpha\mathbf{Z}_+$  is dense in  $\mathbf{R}$ , the balls  $(B(2\ell\bar{\alpha}, \varepsilon/2))_{\ell \in \mathbf{Z}_+}$  cover  $\mathbf{T}$ . By compactness, one can extract a finite covering  $(B(2\ell\bar{\alpha}, \varepsilon/2))_{\ell \in \llbracket 0, L \rrbracket}$ . By translation, one can replace  $\llbracket 0, L \rrbracket$  by any interval  $\llbracket m, m + L \rrbracket$  with  $m \in \mathbf{Z}$ . Hence given an integer  $N \geq 0$ , the random balls  $(B(2S_k\bar{\alpha}, \varepsilon/2))_{k \in \llbracket 0, N \rrbracket}$  cover  $\mathbf{T}$  as soon as  $\max(S_0, \dots, S_N) - \min(S_0, \dots, S_N) \geq L$ . This happens with probability  $\geq 1 - \varepsilon$  provided the integer  $N$  is sufficiently large.

Fix such an integer  $N$  and consider the stopping time

$$\begin{aligned} \tau &:= \inf\{t \geq -N : d(\alpha(S_t - S_{-N}), \gamma_t) < \varepsilon/2\} \\ &= \inf\{t \geq -N : d(2\alpha(S_t - S_{-N}), \gamma_{-N}) < \varepsilon/2\}. \end{aligned}$$

Then the balls  $(B(2\alpha(S_t - S_{-N}), \varepsilon/2))_{t \in \llbracket -N, 0 \rrbracket}$  cover  $\mathbf{T}$  with probability  $\geq 1 - \varepsilon$  since  $(\xi_{-N+1}, \dots, \xi_0)$  has the same law as  $(\xi_1, \dots, \xi_N)$ . Hence  $P[\tau \leq 0] \geq 1 - \varepsilon$ .

For each  $t \in \llbracket -N + 1, 0 \rrbracket$ , the event  $\{\tau < t\} = \{\tau \leq t - 1\}$  belongs to  $\mathcal{F}_{t-1}^Z$ , thus the random variable  $\eta_t := (\mathbf{1}_{\{\tau \leq t\}} - \mathbf{1}_{\{\tau > t\}})\xi_t$  is an innovation at time  $t$  of the filtration  $(\mathcal{F}_n^Z)_{n \leq 0}$ . The random variable

$$\tilde{\gamma}_0 := \alpha \sum_{t=-N+1}^0 \eta_t$$

is a measurable function of  $\eta_{-N+1}, \dots, \eta_0$ . Observe that whereas  $\gamma_0 = \gamma_\tau + \alpha S_\tau$  whereas

$$\text{on the event } \{\tau \leq 0\}, \quad \tilde{\gamma}_0 = \alpha \sum_{t=-N+1}^{\tau} \xi_t - \alpha \sum_{t=\tau+1}^0 \xi_t = \alpha(S_\tau - S_{-N}) + \alpha S_\tau,$$

so  $d(\tilde{\gamma}_0, \gamma_0) = d(\alpha(S_\tau - S_{-N}), \gamma_\tau) < \varepsilon/2$  on this event. Hence

$$\mathbf{E}[d(\tilde{\gamma}_0, \gamma_0)] \leq \frac{\varepsilon}{2}P[\tau \leq 0] + \frac{1}{2}\mathbf{P}[\tau > 0] \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

The proof is complete.

**Alternative proof using Vershik's intermediate criterion.** The split-word process associated to the random variable  $\gamma_0$  and to the innovations  $(\xi_n)_{n \leq 0}$  is  $(W_n, \xi_n)_{n \leq 0}$ , where  $W_n$  is the map from  $\{-1, 1\}^{|n|}$  to  $\mathbf{T}$  defined by

$$\forall n \leq 0, W_n(x_{n+1}, \dots, x_0) = \gamma_n - \alpha(x_{n+1} + \dots + x_0).$$

To show that  $\text{disp}(W_n, d_n) \rightarrow 0$  as  $n \rightarrow -\infty$ , we consider two independent copies  $\gamma'_n$  and  $\gamma''_n$  of the random variable  $\gamma_n$ , defined on a same probability space  $(\bar{\Omega}, \bar{\mathcal{A}}, \bar{\mathbf{P}})$  and call  $W'_n$  and  $W''_n$  the corresponding copies of  $W_n$ .

Let  $\varepsilon > 0$ . Set

$$\tau_n(x_{n+1}, \dots, x_0) = \inf \{t \in \llbracket n, 0 \rrbracket : |\gamma'_n - \gamma''_n - 2\alpha(x_{n+1} + \dots + x_t)| \leq \varepsilon/2\},$$

with the convention  $\inf \emptyset = +\infty$ . Since the validity of the inequality  $\tau_n(x_{n+1}, \dots, x_0) \leq t$  depends only on  $(x_{n+1}, \dots, x_t)$ , we can define an automorphism of the binary tree with height  $|n|$  as follows: for every  $t \in \llbracket -n, -1 \rrbracket$  and  $(x_{n+1}, \dots, x_t) \in \{-1, 1\}^{t+n}$ ,

$$\begin{aligned} \sigma_n(x_{n+1}, \dots, x_t) &= -\text{Id} & \text{if } t < \tau_n(x_{n+1}, \dots, x_0), \\ \sigma_n(x_{n+1}, \dots, x_t) &= -\text{Id} & \text{if } t \geq \tau_n(x_{n+1}, \dots, x_0). \end{aligned}$$

If  $\tau_n(x_{n+1}, \dots, x_0) = t \in \llbracket -n, 0 \rrbracket$ , then

$$\begin{aligned} W'_n(x_{n+1}, \dots, x_0) - W''_n(g_\sigma(x_{n+1}, \dots, x_0)) &= \gamma'_n - \alpha(x_{n+1} + \dots + x_0) - \gamma''_n \\ &\quad + \alpha(-x_{n+1} - \dots - x_t + x_{t+1} + \dots + x_0) \\ &= \gamma'_n - \gamma''_n - 2\alpha(x_{n+1} + \dots + x_t). \end{aligned}$$

Hence  $d(W'_n(x_{n+1}, \dots, x_0), W''_n(g_\sigma(x_{n+1}, \dots, x_0))) \leq \varepsilon/2$  whenever  $\tau_n(x_{n+1}, \dots, x_0) \leq 0$ . Since  $d$  is bounded above by  $1/2$ , we get

$$\text{disp}(W_n, d_n) \leq d_n(W'_n, W''_n \circ g_\sigma) \leq \frac{\varepsilon}{2} + \frac{1}{2} \times \frac{1}{2^{|n|}} \sum_{(x_{n+1}, \dots, x_0) \in \{-1, 1\}^{|n|}} \mathbf{1}_{\{\tau_n(x_{n+1}, \dots, x_0) > 0\}}.$$

In this formula, the mean over all  $(x_{n+1}, \dots, x_0) \in \{-1, 1\}^{|n|}$  is the probability that a random walk on  $\mathbf{T}$  with uniform initial position and with uniformly steps in  $\{-\alpha, \alpha\}$  does not attain the ball  $B(0, \varepsilon/2)$  in at most  $|n|$  steps. This probability goes to 0, so  $\text{disp}(W_n, d_n) \rightarrow 0$  as  $n \rightarrow -\infty$ . Hence  $\gamma(0)$  satisfies Vershik's intermediate criterion.

**Remark 31** Let  $\alpha_1$  and  $\alpha_2$  be two real numbers  $\alpha_1$  and  $\alpha_2$  such that  $\alpha_1 - \alpha_2$  is irrational. Call  $T_i$  the translations  $x \mapsto x + \alpha_i$  on  $\mathbf{T}$ . Then the filtration associated to  $[T_1, T_2]$  is product-type.

*Proof.* Let  $\alpha = (\alpha_1 - \alpha_2)/2$  and  $\beta = (\alpha_1 + \alpha_2)/2$ . Consider again the process  $(\gamma_n)_{n \in \mathbf{Z}}$  defined above. The process  $(\gamma'_n)_{n \in \mathbf{Z}}$  defined by  $\gamma'_n = \gamma_n - n\beta$  generates the same filtration as  $(\gamma_n)_{n \in \mathbf{Z}}$ , which is product-type by theorem 2. But for every  $n \leq 0$ ,

$$\gamma'_n = \gamma_{n-1} - \xi_n \alpha - n\beta = \gamma'_{n-1} - \xi_n \alpha - \beta = \begin{cases} \gamma'_{n-1} - \alpha_1 & \text{if } \xi_n = 1 \\ \gamma'_{n-1} - \alpha_2 & \text{if } \xi_n = -1 \end{cases}.$$

Since  $T_1(x) \neq T_2(x)$  for every  $x \in \mathbf{R}/\mathbf{Z}$ , one deduces that the natural filtration of the process  $(\gamma'_n)_{n \leq 0}$  is 'the' filtration associated to  $[T_1, T_2]$ .

## 4.2 Filtration associated to $[T_a, T_b]$ , where $T_a$ and $T_b$ are shifts related to the free group with generators $a$ and $b$

We prove the following slight generalization of Vershik's theorem:

Let  $\Gamma = \langle a, b \rangle$  be the free group with two generators  $a, b$ . Let  $A$  be a countable alphabet endowed with the discrete metric  $d$  and with a probability measure  $\nu$  which gives a positive probability to each letter. The translations  $T_a$  and  $T_b$  on the set  $G = A^\Gamma$  defined by  $T_a g(x) = g(a^{-1}x)$  and  $T_b g(x) = g(b^{-1}x)$  preserve the measure  $Q := \nu^{\otimes \Gamma}$ . Moreover, the filtration associated to  $[T_a, T_b]$  is not product-type. Hence, the natural filtration of a random walk on  $G$  with steps given by  $T_a^{-1}, T_b^{-1}$  and indexed by the non-positive integers is not product-type.

Let  $\gamma$  be a random variable taking values in  $G$ , with law  $Q$  and  $\xi = (\xi_n)_{n \leq 0}$  be a sequence of independent uniform random variables taking values in  $F := \{a, b\}$ , independent of  $\gamma$ . Set  $Z = (\xi, \gamma)$ . Then 'the' filtration associated to  $[T_a, T_b]$  is the natural filtration of the process  $(Z_n)_{n \leq 0}$  defined by

$$Z_n = [T_a, T_b]^{|n|}(Z) = ((\xi_{k+n})_{n \leq 0}, \gamma_n) \text{ where } \gamma_n = T_{\xi_{n+1}} \circ \dots \circ T_{\xi_0}(\gamma).$$

A recursion shows that for every  $n \leq 0$  and  $y \in \Gamma$ ,

$$\gamma_n(y) = \gamma(\xi_0^{-1} \dots \xi_{n+1}^{-1} y).$$

We want to apply Vershik's intermediate criterion to the random variable  $\gamma(1)$ , where  $1$  is the identity element of the group  $\Gamma$ . Since  $\gamma(1) = \gamma_n(\xi_{n+1} \dots \xi_0)$  and since random map  $\gamma_n$  is  $\mathcal{F}_n^Z$ -measurable, the split-words process  $(W_n, \xi_n)_{n \leq 0}$  associated to the random variable  $\gamma(1)$  and to the innovations  $(\xi_n)_{n \leq 0}$  is given by

$$\forall n \leq 0, \quad \forall (x_{n+1}, \dots, x_0) \in F^{|n|}, \quad W_n(x_{n+1}, \dots, x_0) = \gamma_n(x_{n+1} \dots x_0).$$

Since  $\Gamma$  is the free group generated by  $a$  and  $b$ , the map  $(x_{n+1}, \dots, x_0) \mapsto x_{n+1} \dots x_0$  from  $F^{|n|}$  to  $\Gamma$  is injective, so the 'letters' of the 'word'  $W_n$ , namely the random variables  $W_n(x_{n+1}, \dots, x_0)$  where  $(x_{n+1}, \dots, x_0) \in F^{|n|}$  are independent and equidistributed. Therefore, up to the numbering of the positions, the process  $(W_n, \xi_n)_{n \leq 0}$  is the process studied by Smorodinsky in [21].

Let  $\delta_n$  be the Hamming distance on  $A^{F^{|n|}}$  (defined as the proportion of sites at which two maps in  $A^{F^{|n|}}$  disagree),  $G_n$  the automorphism group of the binary tree with height  $|n|$  and  $d_n$  be the quotient pseudo-distance of  $d_n$  by  $G_n$ . Let  $\gamma'$  and  $\gamma''$  be two independent copies of  $\gamma$ , defined on a same probability space  $(\bar{\Omega}, \bar{\mathcal{A}}, \bar{\mathbf{P}})$ . Then  $\gamma'$  and  $\gamma''$  are also independent copies of  $\gamma_n$ . Call  $W'_n$  and  $W''_n$  the corresponding copies of  $W_n$ , and  $S_n$  the number of sites in  $F^{|n|}$  at which  $W'_n$  and  $W''_n$  disagree.

Since  $\Gamma$  is the free group generated by  $a$  and  $b$ , the map  $(x_{n+1}, \dots, x_0) \mapsto x_{n+1} \dots x_0$  from  $F^{|n|}$  to  $\Gamma$  is injective. Since  $\gamma'$  and  $\gamma''$  are two independent i.i.d processes, we derive that  $S_n$  has a binomial distribution with parameters  $2^{|n|}$  and  $p$ , where

$$p = 1 - \sum_{z \in A} \nu\{z\}^2 > 0.$$

By the large deviation inequality (see lemma 50), one gets for every  $\varepsilon \in ]0, p[$ ,

$$P[\delta_n(W'_n, W''_n) \leq \varepsilon] = P[S_n \leq 2^{|n|}\varepsilon] \leq f_p(\varepsilon)^{2^{|n|}} \text{ where } f_p(\varepsilon) = \left(\frac{p}{\varepsilon}\right)^\varepsilon \left(\frac{1-p}{1-\varepsilon}\right)^{1-\varepsilon}.$$



For every  $\sigma \in G_n$ , the random map  $W_n'' \circ \sigma$  is independent of  $W_n'$  and has the same law as  $W_n$ . Since the size of  $G_n$  is  $2^{2^{|n|}-1}$ , we get

$$P[d_n(W_n', W_n'') \leq \varepsilon] = P[\exists \sigma \in G_n : \delta_n(W_n', W_n'' \circ g_\sigma) \leq \varepsilon] \leq \frac{1}{2} (2f_p(\varepsilon))^{2^{|n|}}.$$

The limit of  $2f_p(\varepsilon)$  as  $\varepsilon$  tends to 0 is  $2(1-p)$ .

If  $p > 1/2$ , then choosing  $\varepsilon$  sufficiently small yields  $P[d_n(W_n', W_n'') \leq \varepsilon] \rightarrow 0$  as  $n \rightarrow -\infty$ , so  $\text{disp}(W_n, d_n)$  remains bounded away from 0 since

$$\text{disp}(W_n, d_n) = \mathbf{E}[d_n(W_n', W_n'')] \geq \varepsilon P[d_n(W_n', W_n'') \leq \varepsilon].$$

Hence the random variable  $\gamma(1)$  does not satisfy Vershik's criterion.

If  $p \leq 1/2$ , one can fix a positive integer  $d$  and consider the split-word process associated to the random variable  $(\gamma(1), \dots, \gamma(a^{d-1}))$  with values in  $A^d$ , namely the process  $(\widetilde{W}_n)_{n \leq 0}$  where  $\widetilde{W}_n$  is the map from  $F^{|n|}$  to  $A$  given by

$$\widetilde{W}_n(x_{n+1}, \dots, x_0) = (\gamma_n(x_{n+1} \cdots x_0), \dots, \gamma_n(x_{n+1} \cdots x_0 a^{d-1})).$$

Replacing  $\gamma_n$  with  $\gamma'$  and  $\gamma''$  yields two independent copies  $\widetilde{W}_n'$  and  $\widetilde{W}_n''$ . Since the products  $x_{n+1} \cdots x_0 a^k$  with  $(x_{n+1}, \dots, x_0) \in F^{|n|}$  and  $k \in \llbracket 0, d-1 \rrbracket$  are all different, the number of sites in  $F^{|n|}$  at which  $\widetilde{W}_n'$  and  $\widetilde{W}_n''$  disagree has a binomial distribution with parameters  $2^{|n|}$  and  $p_d$ , where

$$p_d = 1 - \left( \sum_{z \in A} \nu\{z\}^2 \right)^d.$$

If  $d$  is chosen sufficiently large,  $p_d > 1/2$  so the same argument as above applies and the random variable  $(\gamma(1), \dots, \gamma(a^{d-1}))$  does not satisfy Vershik's criterion.

In all cases, the natural filtration of the process  $(Z_n)_{n \leq 0}$  is not product type. The proof is complete.

**Remark 32** *The same argument as the argument given at the end of subsection 5.3 shows that this conclusion still holds if the discrete alphabet  $(A, d)$  is replaced by any separable complete metric space.*

## 5 Proof of theorem 4 and corollary 5

We now focus with the transformation  $[T, T^{-1}]$ , when  $T$  is a Bernoulli shift. We will see at the end of the section why the situation is essentially the same if one looks at  $[T, \text{Id}]$ .

Thus, in the whole section,  $F$  denotes the pair  $\{-1, 1\}$ , so  $\mathfrak{S}(F) = \{-\text{Id}, \text{Id}\}$ , and  $\mu$  denotes the uniform law on the pair  $F$ . We fix a separable complete metric space  $(A, d)$ , called alphabet  $A$ , endowed with a non-trivial probability measure  $Q$ . Let  $T$  be the shift on  $G = A^{\mathbf{Z}}$  defined by  $T(g)(s) = g(s-1)$ . Then the transformation  $[T, T^{-1}]$  is the map from  $F^{\mathbf{Z}^-} \times A^{\mathbf{Z}}$  to itself defined by

$$[T, T^{-1}](f, g) = \left( (f(k-1))_{k \leq 0}, (g(s-f(0)))_{s \in \mathbf{Z}} \right).$$

This maps preserves the probability measure  $\pi = \mu^{\otimes \mathbf{Z}^-} \otimes \nu^{\otimes \mathbf{Z}}$ .

Let  $(\xi_n)_{n \in \mathbf{Z}}$  and  $\gamma = (\gamma(s))_{s \in \mathbf{Z}}$  be two independent random variables with respective laws  $\mu^{\otimes \mathbf{Z}}$  and  $\nu^{\otimes \mathbf{Z}}$ , defined on same probability space  $(\Omega, \mathcal{A}, \mathbf{P})$ . Then  $Z := ((\xi_n)_{n \leq 0}, \gamma)$  is a random variable with law  $\pi$ . By definition, ‘the’ filtration associated to  $[T, T^{-1}]$  is the filtration  $(\mathcal{F}_n^Z)_{n \leq 0}$  generated by the process  $(Z_n)_{n \leq 0}$  where

$$\forall n \leq 0, \quad Z_n = [T, T^{-1}]^{-n}(Z) = ((\xi_{k+n})_{k \leq 0}, (\gamma(s - \xi_0 - \dots - \xi_{n+1}))_{s \in \mathbf{Z}}).$$

## 5.1 Random walk in a random scenery and nibbled words process

Let  $(S_n)_{n \in \mathbf{Z}}$  be the ‘random walk’ on  $\mathbf{Z}$  and indexed by  $\mathbf{Z}$  given by

$$\begin{aligned} S_n &= -\xi_{n+1} - \dots - \xi_0 & \text{if } n \leq 0, \\ S_n &= \xi_1 + \dots + \xi_n & \text{if } n \geq 0, \end{aligned}$$

with the convention  $S_0 = 0$ . The random variables  $(\xi_n)_{n \in \mathbf{Z}}$  are the steps for this symmetric simple random walk since  $S_n - S_{n-1} = \xi_n$  for every  $n \in \mathbf{Z}$ .

The i.i.d. process  $\gamma$  is independent of this random walk and we view it as a random scenery: the random variable  $\gamma(s)$  is the color at the site  $s$ . At time  $n$ , the position of the symmetric random walk is  $S_n$ , and the color seen at this position is  $\gamma(S_n)$ . The process  $((\xi_n, \gamma(S_n)))_{n \in \mathbf{Z}}$  is called a *random walk in a random scenery* and the shifted map  $\gamma(S_n + \cdot)$  is the scenery viewed from the position  $S_n$ . A survey of the results involving  $[T, T^{-1}]$  and the random walk in a random scenery (in  $\mathbf{Z}^d$ ) can be found in Steif’s paper [22].<sup>3</sup>

From the process  $((\xi_n, \gamma(S_n + \cdot)))_{n \leq 0}$ , we derive a process  $(W_n, \xi_n)_{n \leq 0}$  by setting

$$W_n = (\gamma(S_n + i))_{i \in I_n} \text{ where } I_n = I_{|n|} = \{-|n|, 2 - |n|, \dots, |n| - 2, |n|\}.$$

Note that  $I_n$  is exactly the set of all possible values of the sum  $x_{n+1} + \dots + x_0$  when  $x_{n+1}, \dots, x_0$  range over  $F = \{-1, 1\}$ . The next properties of process  $(W_n, \xi_n)_{n \leq 0}$  follow immediatly from its definition. Such a process was studied by Laurent in [11] and called *nibbled-word process*.

**Proposition 33 (Properties of the process  $(W_n, \xi_n)_{n \leq 0}$ )** *For every  $n \leq 0$ ,*

- *the random word  $W_n$  is  $\mathcal{F}_n^Z$ -measurable since it is the image the random variable  $\gamma(S_n + \cdot)$  by the canonical projection on  $A^{I_n}$ .*
- *the random word  $W_n$  is made with  $|n| + 1$  letters chosen independently in the alphabet  $A$  according to the law  $\nu$  ;*
- *since  $W_n = (W_{n-1}(\xi_n + k))_{k \in I_n}$ , one gets  $W_n$  from  $W_{n-1}$  by suppressing the first letter if  $\xi_n = 1$  and the last letter if  $\xi_n = -1$  ;*
- *the random variable  $\xi_n$  is uniform on  $F = \{-1, 1\}$  and independent on  $\mathcal{F}_{n-1}^Z$ , therefore on  $\mathcal{F}_{n-1}^{W, \xi} = \sigma((W_k, \xi_k)_{k \leq n-1})$ .*

<sup>3</sup>For every  $n \leq 0$ ,  $Z_n = ((\xi_{k+n})_{k \leq 0}, \gamma(S_n + \cdot))$  so  $(\mathcal{F}_n^Z)_{n \leq 0}$  is the natural filtration of the process  $(\xi_n, \gamma(S_n + \cdot))_{n \leq 0}$ . Actually,  $(\mathcal{F}_n^Z)_{n \leq 0}$  is also the natural filtration of the process  $(\gamma(S_n + \cdot))_{n \leq 0}$  since  $T^{-1}(g) \neq T(g)$  for  $\nu^{\otimes \mathbf{Z}}$ -almost every  $g \in A^{\mathbf{Z}}$ . Using the recurrence of the symmetric simple random on  $\mathbf{Z}$ , one could check that  $(\mathcal{F}_n^Z)_{n \leq 0}$  is also the natural filtration of the process  $((\xi_n, \gamma(S_n)))_{n \in \mathbf{Z}}$ . We shall not use these refinements in the present paper.

As a result,  $(\xi_n)_{n \leq 0}$  is a sequence of innovations of the filtration  $(\mathcal{F}_n^{W, \xi})_{n \leq 0}$ , so  $(\mathcal{F}_n^{W, \xi})_{n \leq 0}$  is dyadic. But  $(\xi_n)_{n \leq 0}$  is also a sequence of innovations of the larger filtration  $(\mathcal{F}_n^Z)_{n \leq 0}$ , hence  $(\mathcal{F}_n^{W, \xi})_{n \leq 0}$  is immersed in  $(\mathcal{F}_n^Z)_{n \leq 0}$ .

Therefore, to prove that  $(\mathcal{F}_n^Z)_{n \leq 0}$  is not product-type, it is sufficient to prove that  $(\mathcal{F}_n^{W, \xi})_{n \leq 0}$  is not product-type. By Theorem 28, it is sufficient to check that the random variable  $\gamma(0) = \gamma(S_0) = W_0(0)$  does not satisfy the Vershik property. We will work with the Vershik intermediate property, so we introduce the split-word process associated to  $\gamma(0)$  and to the innovations  $(\xi_n)_{n \leq 0}$ . This process is closely related to the nibbled-word process introduced above. Some notations are necessary to spell out this relation.

**Definition 34** Let  $n \leq 0$ . Call  $s_n$  the map from  $F^{|n|}$  to  $I_n$  defined by  $s_n(x_{n+1} \cdots, x_0) = x_{n+1} + \cdots + x_0$ . To every word  $w \in A^{I_n}$ , we associate its extension  $\bar{w} = w \circ s_n \in A^{F^{|n|}}$ .

Figure 2 illustrates the example where  $n = -3$  and  $w$  sends  $-3, -1, 1, 3$  on  $a, b, c, d$  respectively (so  $w$  is identified with the word  $abcd$ ). Each element of  $F^3$  is viewed as a leaf of the binary tree with height 3. The map  $\bar{w}$  sends the elements of  $F^3$  in lexicographic order on  $a, b, b, c, b, c, c, d$ .

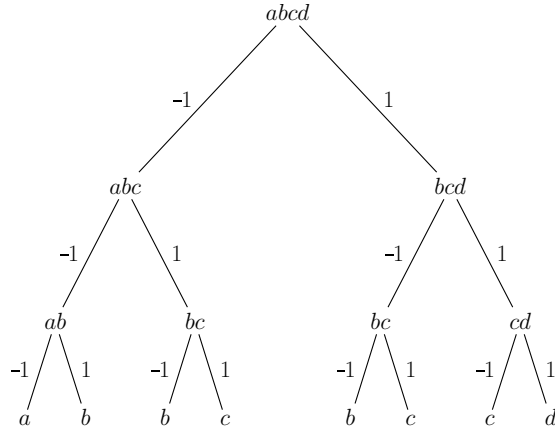


Figure 2:  $\bar{w}: F^3 \rightarrow A$  when  $w = abcd \in A^{I^3}$ .

**Proposition 35** The split-word process associated to the random variable  $\gamma(0)$  and to the innovations  $(\xi_n)_{n \leq 0}$  is  $(\bar{W}_n, \xi_n)_{n \leq 0}$ . This process generates the same filtration as the nibbled-word process  $(W_n, \xi_n)_{n \leq 0}$ .

Proof. For every  $n \leq 0$ , the random map  $\bar{W}_n$  is  $\mathcal{F}_n^Z$ -measurable and

$$\gamma(0) = \gamma(S_n + \xi_{n+1} + \cdots + \xi_0) = W_n(\xi_{n+1} + \cdots + \xi_0) = \bar{W}_n(\xi_{n+1}, \dots, \xi_0),$$

so  $\gamma(0)$  coincides with  $\bar{W}_n(x_{n+1} \cdots, x_0)$  on the event  $\{(\xi_{n+1}, \dots, \xi_0) = (x_{n+1} \cdots, x_0)\}$ . Last,  $\bar{W}_n$  generates the same filtration as  $W_n$  since the map  $w \mapsto w \circ s_n$  from  $A^{I_n}$  to  $A^{F^{|n|}}$  is injective. The result follows.

To negate Vershik's intermediate criterion, we use the metric  $\delta_n$  and the pseudo-metric  $d_n$  on  $A^{F^{|n|}}$  introduced in definition 20. We have to show that  $\text{disp}(\bar{W}_n, d_n)$  does not tend to 0 as  $n$  goes to  $-\infty$ . This leads us to search positive lower bounds for the expectation of  $d_n(\bar{u}, \bar{v})$ , where  $u$  and  $v$  are chosen independently in  $A^{I_n}$  according to the law  $\nu^{\otimes I_n}$

## 5.2 Key-lemmas

Keep in mind the notations introduced in section 3. We begin with the case where  $A$  is countable and  $d$  is the discrete metric on  $A$ . Thus,  $\delta_n$  is the Hamming metric on  $A^{F^{|n|}}$  (normalized to vary between 0 and 1), and  $\delta_n$  is the quotient pseudo-metric by the action of the automorphism group of the binary tree with height  $|n|$ .

To get a positive lower bound of  $\text{disp}(\overline{W}_n, d_n)$  for arbitrary large (negative) integers  $n$ , we will make a recursion. The next lemma will help us to start the recursion. This lemma is stated and its proof is outlined in [7].

Before stating it, we must define the adjoint of a word. To handle with non-negative integers, we denote by  $h$ , or by  $H$  the height of the trees considered.

**Definition 36** *Let  $w \in A^{I_h}$ . Call  $w^r \in A^{I_h}$  the word obtained by reversing the word  $w$ , namely  $w^r(i) = w(-i)$  for every  $i \in I_h$ .*

*If  $w$  can be identified with a word compound with two alternating letters (for example  $ababa\dots$ ), we define its adjoint  $w^* \in A^{I_h}$  as the word obtained from  $w$  by switching these two letters (for example  $babab\dots$ ).*

*Otherwise, we define its adjoint by  $w^* = w^r$ .*

Note that the only possibility for  $w^* \neq w^r$  only when  $w$  is a word with odd length and compound with two alternating letters; in this case,  $w$  is palindromic. Moreover, the reversal map and the adjoint map are involutions and they commute.

**Lemma 37** *Let  $u, v \in A^{I_h}$ . Then  $\bar{u}$  and  $\bar{v}$  belong to the same orbit under the action of  $G_h$  if and only if  $v = u$  or  $v = u^*$ .*

*Proof.* The ‘if’ part will not be used in the sequel and can be proved directly. Indeed,  $\bar{u}^r = \sigma \cdot \bar{u}$  where  $\sigma \in \mathfrak{S}(F)^{T_{h-1}}$  is the map which sends every vertex of  $T_{h-1}$  on  $-\text{Id}$ . Moreover, if  $u = abab\dots$  then for every  $(x_1, \dots, x_h) \in F^h$ ,  $\bar{u}(x_1, \dots, x_h)$  is  $a$  or  $b$  according the number of 1 among  $x_1, \dots, x_h$  is even or odd. The converse holds for  $\bar{u}^*(x_1, \dots, x_h)$ . Therefore,  $\bar{u}^r = \sigma \cdot \bar{u}$  where  $\sigma \in \mathfrak{S}(F)^{T_{h-1}}$  is the map which sends the root  $()$  on  $-\text{Id}$  and every other vertex on  $\text{Id}$ .

We now prove the ‘only if’ part, by recursion on  $h$ . The result is immediate when  $h = 0$  or  $h = 1$ .

Let  $h \geq 2$ . Assume that the result holds at the rank  $h - 1$ . Let  $u, v \in A^{I_h}$  and  $\sigma \in F^{T_{h-1}}$  such that  $\bar{v} = \bar{u} \circ g_\sigma$ .

**1. Case where  $\sigma() = \text{Id}$ .** Call  $\sigma_1$  et  $\sigma_{-1}$  the elements of  $F^{T_{h-2}}$  defined by  $\sigma_1(x) = \sigma(1, x)$  and  $\sigma_{-1} = \sigma(-1, x)$  for every  $x \in T_{h-2}$ . Let  $u_+ = (u(i+1))_{i \in I_{h-1}}$  and  $u_- = (u(i-1))_{i \in I_{h-1}}$  be the words obtained from  $u$  by suppressing the first and the last letter.

Then for every  $(x_1, \dots, x_h) \in F^h$ ,

$$\begin{aligned}
\overline{v_{x_1}}(x_2, \dots, x_h) &= v_{x_1}(x_2 + \dots + x_h) \\
&= v(x_1 + \dots + x_h) \\
&= \overline{v}(x_1, \dots, x_h) \\
&= \overline{u}(x_1, g_{\sigma_{x_1}}(x_2, \dots, x_h)) \\
&= u(x_1 + s_{h-1}(g_{\sigma_{x_1}}(x_2, \dots, x_h))) \\
&= u_{x_1}(s_{h-1}(g_{\sigma_{x_1}}(x_2, \dots, x_h))) \\
&= \overline{u_{x_1}}(g_{\sigma_{x_1}}(x_2, \dots, x_h)).
\end{aligned}$$

Using the recursion hypothesis, we get

$$\begin{cases} \overline{v_-} = \overline{u_-} \circ g_{\sigma_{-1}} \\ \overline{v_+} = \overline{u_+} \circ g_{\sigma_1} \end{cases} \quad \text{so} \quad \begin{cases} v_- = u_- \text{ or } (u_-)^* \\ v_+ = u_+ \text{ or } (u_+)^* \end{cases}.$$

Four cases have to be considered.

1. If  $v_- = u_-$  and  $v_+ = u_+$ , then  $v = u$ .
2. If  $v_- = u_-$  and  $v_+ = (u_+)^*$ , then
  - either  $u_+$  has the form  $ababa\dots$  and  $v_+$  has the form  $babab\dots$ , hence the equality  $v_- = u_-$  entails  $a = b$ , so  $v_+ = u_+$  and  $v = u$ ;
  - or  $(u_+)^* = (u_+)^r$ , hence the equalities  $v_- = u_-$  and  $v_+ = (u_+)^r$  yield that for every  $i \in I_{h-1} \setminus \{h-1\}$ ,  $u_+(i) = u_-(i+2) = v_-(i+2) = v_+(i) = u_+(-i)$ , so  $(u_+)^r = u_r$  and  $v = u$ .
3. If  $v_- = (u_-)^*$  and  $v_+ = u_+$ , we get in the same way  $v = u$ .
4. If  $v_- = (u_-)^*$  and  $v_+ = (u_+)^*$ , then at least one of the three following cases below occurs:
  - $u_+$  has the form  $ababa\dots$  and  $v_+$  has the form  $babab\dots$ , so the equality  $v_- = (u_-)^*$  forces the alternation of the letters  $a$  and  $b$  to occur from the very beginning of the words  $u$  and  $v$ , and  $v = u^*$ .
  - $u_-$  has the form  $ababa\dots$  and  $v_-$  has the form  $babab\dots$ , so we get in the same way that  $v = u^*$ .
  - $v_+ = (u_+)^r$  and  $v_- = (u_-)^r$ , namely for every  $i \in I_{h-1}$ ,  $v(i+1) = u(-i+1)$  and  $v(i-1) = u(-i-1)$ . Hence, for every  $j \in I_{h-2}$ ,  $v(j+2) = u(-j) = v(j-2)$ , so  $v$  has the form  $ababa\dots$ . Therefore,  $u = ababa\dots = v$  if  $h$  is odd, and  $u = babab\dots = v^*$  if  $h$  is even.

**2. Case where  $\sigma() = -\text{Id}$ .** Since  $\overline{v^r} = \overline{u} \circ g_{-\sigma}$  and  $-\sigma() = \text{Id}$ , the first case already proved can be applied to  $u$  et  $v^r$ , so  $v^r = u$  or  $v^r = u^*$ , which yields the desired result.

Hence, in all cases, one has  $v = u$  or  $v = u^*$ . The proof is complete.

The next lemma will provide the recursion step. It is buried in the proof of Hecklen and Hoffman [7] but deserves to be given separately for the sake of clarity. The proof we give is close to Hoffman's proof but we change some constants and use sharpest inequalities to get better bounds.

**Lemma 38** *Let  $C \geq 2$ ,  $\eta > 0$ ,  $\varepsilon = (1 - 3C^{-2})\eta > 0$  and two integers entiers  $h \geq 1$  et  $H \geq C^6 h$ . Set  $D = H - h$ .*

*If  $u, v \in A^{I_H}$  satisfy  $d_H(\bar{u}, \bar{v}) < \varepsilon$ , then there exists  $i, j, k, l$  in  $I_D$  such that  $|i| \leq C(\sqrt{H} - \sqrt{h})$ ,  $|j| \leq C(\sqrt{H} - \sqrt{h})$ ,  $j - i > 2C\sqrt{h}$  and*

$$d_h\left(\overline{u(i + \cdot)}|_{I_h}, \overline{v(k + \cdot)}|_{I_h}\right) < \eta,$$

$$d_h\left(\overline{u(j + \cdot)}|_{I_h}, \overline{v(l + \cdot)}|_{I_h}\right) < \eta.$$

Proof. Fix  $\sigma \in \mathfrak{S}(F)^{T_{H-1}}$  such that  $\delta_H(\bar{u}, \bar{v} \circ g_\sigma) < \varepsilon$ .

Split each  $H$ -uple  $x = (x_1, \dots, x_H) \in F^H$  into  $y = (x_1, \dots, x_D) \in F^D$  and  $z = (x_{D+1}, \dots, x_H) \in F^h$ . Set  $s(y) = x_1 + \dots + x_D$ ,  $s(z) = x_{D+1} + \dots + x_H$  and call  $\sigma_y$  the element of  $\mathfrak{S}(F)^{T_{h-1}}$  defined by  $\sigma_y(z_1, \dots, z_i) = \sigma(y, z_1, \dots, z_i)$  for every  $0 \leq i \leq h - 1$  and  $(z_1, \dots, z_i) \in F^i$ . Then

$$\bar{u}(x) = u(s(y) + s(z)) = \overline{u(s(y) + \cdot)}|_{I_h}(z),$$

and

$$\bar{v}(g_\sigma(x)) = v\left(s(g_\sigma(y)) + s(g_{\sigma_y}(z))\right) = \overline{v\left(s(g_\sigma(y)) + \cdot\right)}|_{I_h}(g_{\sigma_y}(z)).$$

Thus

$$\begin{aligned} \delta_H(\bar{u}, \bar{v} \circ g_\sigma) &= 2^{-D} \sum_{y \in F^D} \delta_h\left(\overline{u(s(y) + \cdot)}|_{I_h}, \overline{v\left(s(g_\sigma(y)) + \cdot\right)}|_{I_h} \circ g_{\sigma_y}\right) \\ &\geq 2^{-D} \sum_{y \in F^D} d_h\left(\overline{u(s(y) + \cdot)}|_{I_h}, \overline{v\left(s(g_\sigma(y)) + \cdot\right)}|_{I_h}\right). \end{aligned}$$

Let

$$E_1 = \left\{y \in F^D : d_h\left(\overline{u(s(y) + \cdot)}|_{I_h}, \overline{v\left(s(g_\sigma(y)) + \cdot\right)}|_{I_h}\right) \geq \eta\right\},$$

$$E_2 = \left\{y \in F^D : |s(y)| > C(\sqrt{H} - \sqrt{h})\right\} \text{ and } E = F^D \setminus (E_1 \cup E_2).$$

Since  $\delta_H(\bar{u}, \bar{v} \circ g_\sigma) < \varepsilon$ , Markov inequality shows that  $\mu^{\otimes D}(E_1) \leq \varepsilon/\eta = 1 - 3C^{-2}$ .

But when  $y$  is chosen according to the probability measure  $\mu^{\otimes D}$ ,  $s(y)$  has expectation 0 and variance  $D$ , so Bienaymé-Chebycheff inequality yields

$$\begin{aligned} \mu^{\otimes d}(E_2) &\leq \frac{H - h}{C^2(\sqrt{H} - \sqrt{h})^2} \\ &= \frac{1}{C^2} \times \frac{\sqrt{H} + \sqrt{h}}{\sqrt{H} - \sqrt{h}} \\ &= \frac{1}{C^2} \times \left(1 + \frac{2}{\sqrt{H/h} - 1}\right) \\ &\leq \frac{9}{7C^2} \text{ since } \sqrt{H/h} \geq C^3 \geq 8. \end{aligned}$$

Hence

$$\mu^{\otimes d}(E) \geq \frac{3}{C^2} - \frac{9}{7C^2} = \frac{12}{7C^2}.$$

But  $s(F^D) = I_D = \{2k - D : k \in \llbracket 0, d \rrbracket\}$  and for every  $k \in \llbracket 0, d \rrbracket$ ,

$$\mu^{\otimes d} \{y \in F^D : s(y) = 2k - D\} = \frac{1}{2^D} \binom{D}{k} \leq \sqrt{\frac{2}{\pi D}},$$

by lemma 51 in section 7. Since  $2C\sqrt{h} \geq 4$ , any interval  $J \subset \mathbf{R}$  with length  $2C\sqrt{h}$  contains at most  $(3/2)C\sqrt{h}$  points of  $I_D$ , so

$$\begin{aligned} \mu^{\otimes d} \{y \in F^D : s(y) \in J\} &\leq \frac{3}{2}C\sqrt{h} \times \sqrt{\frac{2}{\pi D}} \\ &= \frac{3\sqrt{2}}{2\sqrt{\pi}} \times \frac{C}{\sqrt{C^6 - 1}} \\ &\leq \frac{4\sqrt{2}}{\sqrt{7\pi}} \times \frac{1}{C^2} \text{ since } C^6 - 1 \geq \frac{63}{64}C^6 \\ &< \frac{12}{7C^2}. \end{aligned}$$

Since  $\mu^{\otimes d}(E) \geq 12/(7C^2)$ , we deduce that  $\max s(E) - \min s(E) > 2C\sqrt{h}$ . Choosing  $y_1$  and  $y_2$  in  $E$  achieving the minimum and the maximum of  $s$  over  $E$  and setting  $i = s(y_1)$ ,  $j = s(y_2)$ ,  $k = s(g_\sigma(y_1))$ ,  $l = s(g_\sigma(y_2))$  yields the result.

We now introduce some notations to continue the proof.

**Notations 39** For every  $H \in \mathbf{Z}_+$  and  $C > 0$ , we define the  $C$ -middle of  $I_H$  by

$$I_{H,C} = I_H \cap [-C\sqrt{H}, C\sqrt{H}].$$

For every  $u \in A^{I_H}$ ,  $C > 0$  and  $\varepsilon > 0$ , let

$$\Theta_{\varepsilon,C}(u) = \{w \in A^{I_H} : \exists v \in A^{I_H}, w = v \text{ on } I_{H,C} \text{ and } d_H(\bar{u}, \bar{v}) < \varepsilon\}.$$

Last, set

$$p_H(\varepsilon, C) = \max_{u \in A^{I_H}} \mathbf{P}[W_{-H} \in \Theta_{\varepsilon,C}(u)].$$

**Remark 40** The larger is  $C$ , the smaller is the set  $\Theta_{\varepsilon,C}(u)$ . If  $C \geq \sqrt{H}$ ,  $I_{H,C} = I_H$  so  $\Theta_{\varepsilon,C}(u)$  is the set of all  $w \in I_H$  such that  $d_H(\bar{u}, \bar{w}) < \varepsilon$ . Therefore, in all cases,

$$\mathbf{P}[d_H(\bar{u}, \bar{W}_{-H}) < \varepsilon] \leq \mathbf{P}[W_{-H} \in \Theta_{\varepsilon,C}(u)] \leq p_H(\varepsilon, C),$$

so

$$\mathbf{E}[d_H(\bar{u}, \bar{W}_{-H})] \geq \varepsilon \mathbf{P}[d_H(\bar{u}, \bar{W}_{-H}) \geq \varepsilon] \geq \varepsilon(1 - p_H(\varepsilon, C)).$$

Since this inequality holds for every  $u \in A^{I_H}$ , we get

$$\text{disp}(\bar{W}_{-H}, d_H) \geq \varepsilon(1 - p_H(\varepsilon, C)).$$

The remark above explains the interest to bound  $\varepsilon(1 - p_H(\varepsilon, C))$  away from 0 to negate Vershik's intermediate criterion. The last lemma provides the inequality below.

**Corollary 41** Take, like in the previous lemma,  $C \geq 2$ ,  $\eta > 0$ ,  $\varepsilon = (1 - 3C^{-2})\eta > 0$  and two integers  $h \geq 1$  and  $H \geq C^6 h$ . Then

$$p_H(\varepsilon, C) \leq C^2 H^3 p_h(\eta, C)^2 / 2.$$

Proof. Let  $X = (X_k)_{k \in I_H}$  be a random word whose letters are chosen independently and according to the law  $\nu$ .

On the event  $\{X \in \Theta_{\varepsilon, C}(u)\}$ , there exists some  $v \in A^{I_H}$  such that  $X$  coincides with  $v$  on  $I_{H, C}$  and  $d_H(\bar{u}, \bar{v}) < \varepsilon$ . Let  $D = H - h$ . The last lemma provides the existence of  $i, j, k, l$  in  $I_D$  such that  $|i| \leq C(\sqrt{H} - \sqrt{h})$ ,  $|j| \leq C(\sqrt{H} - \sqrt{h})$ ,  $j - i > 2C\sqrt{h}$  and

$$\begin{aligned} d_h(\overline{v(i+\cdot)}|_{I_h}, \overline{u(k+\cdot)}|_{I_h}) &< \eta, \\ d_h(\overline{v(j+\cdot)}|_{I_h}, \overline{u(l+\cdot)}|_{I_h}) &< \eta. \end{aligned}$$

The inequalities satisfied by  $i$  and  $j$  entail

$$i + I_{h, C} \subset I_{H, C}, \quad j + I_{h, C} \subset I_{H, C}, \quad \text{and} \quad (i + I_{h, C}) \cap (j + I_{h, C}) = \emptyset.$$

Therefore, the random variables  $X(i+\cdot)|_{I_h}$  and  $X(j+\cdot)|_{I_h}$  coincide on  $I_{h, C}$  with  $v(i+\cdot)|_{I_h}$  et  $v(j+\cdot)|_{I_h}$  and they are independent.

This shows that this event  $\{X \in \Theta_{\varepsilon, C}(u)\}$  is contained in the union of the events

$$\{X(i+\cdot)|_{I_h} \in \Theta_{\eta, C}(u(k+\cdot)|_{I_h}); X(j+\cdot)|_{I_h} \in \Theta_{\eta, C}(u(l+\cdot)|_{I_h})\}$$

over all  $(i, j, k, l)$  satisfying the conditions above. Each one of these events has probability  $\leq p_h(\eta, C)^2$ , and since  $C(\sqrt{H} - \sqrt{h}) \leq C\sqrt{H} - 1$ , the number of 4-uples  $(i, j, k, l)$  considered is bounded above by  $(C\sqrt{H})^2/2 \times (D+1)^2 \leq C^2 H^3/2$ . The result follows.

### 5.3 End of the proof

We begin with the case where the alphabet  $(A, d)$  is countable and endowed with the discrete metric.

To prove theorem 4, we show that  $\gamma(0)$  does not satisfy Vershik's intermediate criterion. By remark 40, it suffices to find sequences  $(C_k)_{k \geq 0}$ ,  $(H_k)_{k \geq 0}$ , and  $(\varepsilon_k)_{k \geq 0}$  tending respectively to  $+\infty$ ,  $+\infty$  and  $\varepsilon_\infty > 0$ , such that the probabilities  $p_{H_k}(\varepsilon_k, C_k)$  tend to 0. Lemma 37 and corollary 41 enable us to do that.

**Lemma 42** *Let  $q$  be the mass of the heaviest atom of  $\nu$  (so  $0 < q < 1$ ). Define the sequences  $(C_k)_{k \geq 0}$ ,  $(H_k)_{k \geq 0}$ ,  $(\varepsilon_k)_{k \geq 0}$  and  $(\alpha_k)_{k \geq 0}$  by  $C_0 = 2$ ,  $H_0$  equals the square of some even positive integer,  $\varepsilon_0 = 2^{-H_0}$ ,  $\alpha_0 = 2q^{C_0\sqrt{H_0}+1}$  and for every  $k \geq 1$ ,*

$$C_k = k + 1, \quad H_k = C_k^6 H_{k-1}, \quad \varepsilon_k = \varepsilon_{k-1}(1 - 3C_k^{-2}), \quad \alpha_k = C_k^2 H_k^3 \alpha_{k-1}^2/2.$$

Then for every  $k \in \mathbf{N}$ ,

$$p_{H_k}(\varepsilon_k, C_k) \leq \alpha_k.$$

Moreover, the sequence  $(\varepsilon_k)_{k \geq 0}$  has a positive limit and if  $H_0$  is large enough, the sequence  $(\alpha_k)_{k \geq 0}$  tends to 0.

Proof. We make a recursion on the integer  $k$ .

Given  $u, v \in A^{I_{H_0}}$ , the inequality  $d_{H_0}(\bar{u}, \bar{v}) < \varepsilon_0 = 2^{-H_0}$  holds only when  $\bar{u}$  and  $\bar{v}$  belong to a same orbite under the action of  $G_{H_0}$ , namely when  $v = u$  or  $v = u^*$ , thanks to lemma 37. But  $I_{H_0, C_0} = \{2j - C_0\sqrt{H_0} : j \in \llbracket 0, C_0\sqrt{H_0} \rrbracket\}$ , hence for every  $u \in A^{I_{H_0}}$ ,

$$\begin{aligned} \mathbf{P}[X \in \Theta_{\varepsilon_0, C_0}(u)] &\leq \mathbf{P}[X = u \text{ sur } I_{H_0, C_0}] + \mathbf{P}[X = u^* \text{ sur } I_{H_0, C_0}] \\ &\leq 2q^{C_0\sqrt{H_0}+1} = \alpha_0. \end{aligned}$$



Hence  $p_{H_0}(\varepsilon_0, C_0) \leq \alpha_0$

Let  $k \geq 1$ . Assume  $p_{H_{k-1}}(\varepsilon_{k-1}, C_{k-1}) \leq \alpha_{k-1}$ . Then corollary 41 applied to  $C = C_k \geq 2$ ,  $h = H_{k-1}$ ,  $H = C^6 h = H_k$ ,  $\eta = \varepsilon_{k-1}$  and  $\varepsilon = (1 - 3C_k^{-2})\eta = \varepsilon_k$  yields

$$\begin{aligned} p_{H_k}(\varepsilon_k, C_k) &\leq C_k^2 H_k^3 p_{H_{k-1}}(\varepsilon_{k-1}, C_k)^2 / 2 \\ &\leq C_k^2 H_k^3 p_{H_{k-1}}(\varepsilon_{k-1}, C_{k-1})^2 / 2 \text{ since } C_k \geq C_{k-1} \\ &\leq C_k^2 H_k^3 \alpha_{k-1}^2 / 2 = \alpha_k, \end{aligned}$$

which achieves the recursion.

Next,

$$\varepsilon_k = \varepsilon_0 \prod_{i=1}^k \left(1 - \frac{3}{(i+1)^2}\right) \xrightarrow{k \rightarrow \infty} \varepsilon_0 \prod_{n \geq 2} \left(1 - \frac{3}{n^2}\right) = 2^{-H_0} \frac{\sin(\pi\sqrt{3})}{-2\pi\sqrt{3}} > 0.$$

Last, equality  $\log_2 \alpha_k = 2 \log_2 \alpha_{k-1} + 3 \log_2 H_k + 2 \log C_k - 1$ , yields by recursion

$$\begin{aligned} 2^{-k} \log_2 \alpha_k &= \log_2 \alpha_0 + \sum_{i=1}^k 2^{-i} (3 \log_2 H_i + 2 \log_2 C_i - 1), \\ &= (2\sqrt{H_0} + 1) \log_2 q + 3 \log_2 H_0 \\ &\quad + 18 \sum_{i=1}^k 2^{-i} \log_2((i+1)!) + 2 \sum_{i=1}^k 2^{-i} \log_2(i+1). \end{aligned}$$

But

$$\sum_{i=1}^{+\infty} 2^{-i} \log_2((i+1)!) = \sum_{i=1}^{+\infty} 2^{-i} \sum_{j=2}^{i+1} \log_2 j = \sum_{j=2}^{+\infty} \log_2 j \sum_{i=j-1}^{+\infty} 2^{-i} = 4 \sum_{j=2}^{+\infty} 2^{-j} \log_2 j.$$

Hence

$$2^{-k} \log_2 \alpha_k \xrightarrow{k \rightarrow +\infty} (2\sqrt{H_0} + 1) \log_2 q + 3 \log_2 H_0 + 76 \sum_{j=1}^{\infty} 2^{-j} \log_2 j.$$

Since  $\log_2 q < 0$  and

$$76 \sum_{j=1}^{\infty} 2^{-j} \log_2 j < 55, 7 < +\infty,$$

we get a negative limit provided  $H_0$  is large enough, which yields the convergence of  $(\alpha_k)_{k \geq 0}$  to 0. The proof is complete

**Remark 43** When  $q = 1/2$ , one can choose  $H_0 = 44^2 = 1936$ , which is far less than the choice  $H_0 = 40000$  made by Heicklen and Hoffman.

We now deduce the result in the general case, namely when  $(A, d)$  is a separable metric complete space endowed with a non-trivial probability measure  $\nu$ . Recall that we chose a random variable  $Z = ((\xi_k)_{k \leq 0}, (\gamma(s))_{s \in \mathbf{Z}})$  with law  $\mu^{\otimes \mathbf{Z}^-} \otimes \nu^{\otimes \mathbf{Z}}$  and defined 'the' filtration of  $[T, T^{-1}]$  as the natural filtration of the process  $(Z_n)_{n \leq 0}$  defined by

$$Z_n = [T, T^{-1}](Z) = ((\xi_{k+n})_{k \leq 0}, (\gamma(s - \xi_0 - \dots - \xi_{n+1}))_{s \in \mathbf{Z}}).$$

Fix a Borel subset  $B$  of  $A$  such that  $p := \nu(B) \in ]0, 1[$ , set  $Z' = ((\xi_k)_{k \leq 0}, (\mathbf{1}_B(\gamma(s)))_{s \in \mathbf{Z}})$  and call  $T'$  the shift on  $\{0, 1\}^{\mathbf{Z}}$ , endowed with the probability  $\mathcal{B}(1, p)^{\otimes \mathbf{Z}}$ . Then the law of  $Z'$  is  $\mu^{\otimes \mathbf{Z}^-} \otimes \mathcal{B}(1, p)^{\otimes \mathbf{Z}}$ , so ‘the’ filtration of  $[T', T'^{-1}]$  is the natural filtration of the process  $(Z'_n)_{n \leq 0}$  defined by

$$Z'_n = [T', T'^{-1}](Z') = ((\xi_{k+n})_{k \leq 0}, (\mathbf{1}_B(\gamma(s - \xi_0 - \dots - \xi_{n+1})))_{s \in \mathbf{Z}}).$$

The filtration  $(\mathcal{F}_n^{Z'})_{n \leq 0}$  is contained in  $(\mathcal{F}_n^Z)_{n \leq 0}$ , and admits  $(\xi_n)_{n \leq 0}$  as sequence of innovations, like  $(\mathcal{F}_n^Z)_{n \leq 0}$ . Hence  $(\mathcal{F}_n^{Z'})_{n \leq 0}$  is dyadic and immersed in  $(\mathcal{F}_n^Z)_{n \leq 0}$ . We have proved that  $(\mathcal{F}_n^{Z'})_{n \leq 0}$  is not product-type hence  $(\mathcal{F}_n^Z)_{n \leq 0}$  cannot be product-type thanks to corollary 15.

Actually, the conclusion still holds if one replaces  $[T, T^{-1}]$  by any  $[T^{k_1}, T^{k_2}]$  where  $k_1$  and  $k_2$  are two distinct integers. Indeed, the random variables  $(\gamma(s))_{s \in \mathbf{Z}}$  are i.i.d, so this replacement preserves the law of the split-word process associated to the random variable  $\gamma(0)$  up to a renumbering of the sites.

## 5.4 proof of corollary 5

Let  $T$  be an automorphism of a Lebesgue space  $(G, \mathcal{G}, Q)$ , with positive entropy. By Sinai’s factor theorem [6], one can find a measurable partition  $\alpha = \{A_1, A_2\}$  of  $(G, \mathcal{G}, Q)$  into two blocks of positive probability such that the partitions  $(T^{-k}\alpha)_{k \in \mathbf{Z}}$  are independent. For each  $g \in G$ , call  $\varphi(g)$  the only index in  $\{1, 2\}$  such that set  $g \in A_{\varphi(g)}$ , and set

$$\Phi(g) = (\varphi(T^k(g)))_{k \in \mathbf{Z}} \in \{1, 2\}^{\mathbf{Z}}.$$

The sequence  $\Phi(g)$  thus defined is called the  $\alpha$ -name of  $g$ .

By construction, the random variables  $\varphi \circ T^k$ , defined on the probability space  $(G, \mathcal{G}, Q)$ , are independent and have the same distribution. Call  $\nu$  this distribution and  $T_B$  be the shift on  $\{1, 2\}^{\mathbf{Z}}$ . Then  $\Phi(Q) = \nu^{\otimes \mathbf{Z}}$  and  $T_B \circ \Phi = \Phi \circ T$ , so the Bernoulli shift  $(\{1, 2\}^{\mathbf{Z}}, \mathcal{P}(\{1, 2\})^{\otimes \mathbf{Z}}, \nu^{\otimes \mathbf{Z}}, T_B)$  is a factor of the dynamical system  $(G, \mathcal{G}, Q, T)$ .

Choose a random variable  $Z = ((\xi_k)_{k \leq 0}, \gamma)$  with law  $\mu^{\otimes \mathbf{Z}^-} \otimes Q$ . Then ‘the’ filtration associated to  $[T, T^{-1}]$  is the natural filtration of the process  $(Z_n)_{n \leq 0}$  defined by

$$Z_n = [T, T^{-1}](Z) = ((\xi_{k+n})_{k \leq 0}, T^{\xi_0 + \dots + \xi_{n+1}}(\gamma)).$$

In the same way, the law of the random variable  $Z' := ((\xi_k)_{k \leq 0}, \Phi(\gamma))$  is  $\mu^{\otimes \mathbf{Z}^-} \otimes \nu^{\otimes \mathbf{Z}}$  so ‘the’ filtration associated to  $[T_B, T_B^{-1}]$  is the natural filtration of the process  $(Z'_n)_{n \leq 0}$  defined by  $Z'_n = [T_B, T_B^{-1}](Z')$ . But the equality  $T_B \circ \Phi = \Phi \circ T$  yields

$$\begin{aligned} Z'_n &= ((\xi_{k+n})_{k \leq 0}, T_B^{\xi_0 + \dots + \xi_{n+1}}(\Phi(\gamma))) \\ &= ((\xi_{k+n})_{k \leq 0}, \Phi(T^{\xi_0 + \dots + \xi_{n+1}}(\gamma))) \end{aligned}$$

Therefore, the filtration  $(\mathcal{F}_n^{Z'})_{n \leq 0}$  is contained in  $(\mathcal{F}_n^Z)_{n \leq 0}$ , and admits  $(\xi_n)_{n \leq 0}$  as sequence of innovations, like  $(\mathcal{F}_n^Z)_{n \leq 0}$ . Hence  $(\mathcal{F}_n^{Z'})_{n \leq 0}$  is dyadic and immersed in  $(\mathcal{F}_n^Z)_{n \leq 0}$ . We have proved that  $(\mathcal{F}_n^{Z'})_{n \leq 0}$  is not product-type hence  $(\mathcal{F}_n^Z)_{n \leq 0}$  cannot be product-type, thanks to corollary 15.

## 6 Proof of theorem 6

We now present a slight variant of Hoffman's example of an automorphism  $T$  of Lebesgue probability space such that  $[T, \text{Id}]$  is not standard. We modify some numerical values, to be more 'parcimonious' and we detail the proof, following Hoffman's strategy.

### 6.1 Construction of null-entropy shifts

Let  $A$  be a finite alphabet with size  $\geq 2$ . In the whole section,  $T$  denotes the bilateral shift  $(x_k)_{k \in \mathbf{Z}} \mapsto (x_{k+1})_{k \in \mathbf{Z}}$  on  $A^{\mathbf{Z}}$ . The purpose of this subsection is to construct a shift-invariant probability measure  $Q$  on  $A^{\mathbf{Z}}$  such that  $T$  is ergodic and has null entropy under  $Q$ .

The 'cut-and-stack' procedure provides a lot of such probability measures. A general treatment of this kind of construction can be found in [20]. Yet, we restrict ourselves to a particular subclass which includes Hoffman's example and enables a more elementary definition.

We fix two sequences  $(N_n)_{n \geq 1}$  and  $(\ell_n)_{n \geq 1}$  of positive integers (tending to infinity) such that  $N_n \geq 2$  and  $N_{n-1}\ell_{n-1}$  divides  $\ell_n$  for every  $n \geq 2$ . For every  $n \geq 1$ , we define a family of distinct elements  $B_{n,0}, \dots, B_{n,N_n-1}$  of  $A^{\ell_n}$ , called the  $n$ -blocks, as follows <sup>4</sup>.

The 1-blocks are the elements of  $A$ , so  $\ell(1) = 1$  and  $N(1) = |A|$ . When  $n \geq 2$ , each block  $B_{n,i}$  is obtained as a concatenation of  $(n-1)$ -blocks in such a way that the number of occurrences of  $B_{n-1,j}$  in  $B_{n,i}$  does not depend on  $i$  and  $j$ : this is the major simplification with regard to the general 'cut-and-stack' constructions, and that is why we assume that  $N_{n-1}\ell_{n-1}$  divides  $\ell_n$ .

The admissible concatenations, and the sequences  $(N_n)_{n \geq 1}$  and  $(\ell_n)_{n \geq 1}$  will be specified later. For now, we explain how to derive a shift-invariant probability measure from this block structure. Informally, the typical sample paths under the probability measure  $Q$  are for each  $n \geq 1$  infinite concatenations of  $n$ -blocks. To construct the probability measure  $Q$ , we construct a compatible family of finite-dimensional marginals.

**Proposition 44** *For every integer  $d \geq 0$ , every word  $w \in A^d$  and every word  $B = (b_0, \dots, b_{\ell-1})$  with length  $\ell \geq d$ , set*

$$N(w, B) = \sum_{k=0}^{\ell-d} |\{k \in \llbracket 0, \ell-d \rrbracket : w = (b_k, \dots, b_{k+d-1})\}| \quad \text{and} \quad p(w|B) = \frac{N(w, B)}{\ell - d + 1},$$

so  $N(w, B)$  and  $p(w|B)$  are respectively the number of occurrences and the frequency of occurrence of  $w$  among the subwords of  $B$  with length  $d$ . Then

1. Let  $(i_n)_{n \geq 1}$  a sequence such that  $i_n \in \llbracket 0, N_n - 1 \rrbracket$  for every  $n \geq 1$ . Then  $p(w|B_{n,i_n})$  has a limit  $p_d(w)$  as  $n$  goes to infinity, and this limit  $p_d(w)$  does not depend on the choice of  $(i_n)_{n \geq 1}$ .
2. The maps  $p_d : A^d \rightarrow [0, 1]$  thus defined are the marginals of some shift-invariant probability measure  $Q$  on  $A^{\mathbf{Z}}$ .

---

<sup>4</sup>We index the  $n$ -blocks by  $\llbracket 0, N_n - 1 \rrbracket$  instead of  $\llbracket 0, N_n \rrbracket$  to handle simpler formulas. For the same reason, we view each word with length  $\ell$  as a map from  $\llbracket 0, \ell - 1 \rrbracket$  to  $A$ .

Proof. For every  $n \geq 2$ ,  $\ell_n \geq N_{n-1}\ell_{n-1} \geq 2\ell_{n-1}$ . A recursion yields  $\ell_n \geq 2^{n-1}$ , so  $\ell_n$  goes to infinity and the series  $\sum_n 1/\ell_n$  converges.

1. Fix a word  $w$  with length  $d$ . Let  $n \geq 2$  such that  $\ell_{n-1} \geq d$  and  $i \in \llbracket 0, N_n - 1 \rrbracket$ . Let  $M_n$  the integer such that  $\ell_n = M_n N_{n-1} \ell_{n-1}$ . Then by construction, the block  $B_{n,i}$  is a concatenation of  $(n-1)$ -blocks in which each  $B_{n-1,j}$  is involved  $M_n$  times.

We obtain the subwords of  $B_{n,i}$  with length  $d$  by choosing  $k \in \llbracket 0, \ell_n - d \rrbracket$  and by looking at the letters at positions  $k+1, \dots, k+d$ . For most of these  $k$ , the interval  $\llbracket k+1, k+d \rrbracket$  is entirely contained in some subinterval  $\llbracket q\ell_{n-1}, (q+1)\ell_{n-1} - d \rrbracket$  with  $q \in \llbracket 0, M_n N_{n-1} \rrbracket$ . The restriction of  $B_{n,i}$  to these subintervals are precisely the  $(n-1)$ -blocks, each block  $B_{n-1,j}$  occurring  $M_n$  times. Since there are exactly  $M_n N_{n-1} (d-1)$  remaining  $k$ , we get

$$0 \leq N(w, B_{n,i}) - M_n \sum_{j=1}^{N_{n-1}} N(w, B_{n-1,j}) \leq M_n N_{n-1} (d-1).$$

Dividing by  $\ell_n = M_n N_{n-1} \ell_{n-1}$  yields

$$0 \leq \frac{N(w, B_{n,i})}{\ell_n} - \frac{1}{N_{n-1}} \sum_{j=1}^{N_{n-1}} \frac{N(w, B_{n-1,j})}{\ell_{n-1}} \leq \frac{d-1}{\ell_{n-1}}. \quad (1)$$

The same inequality holds if  $N(w, B_{n,i})/\ell_n$  is replaced by its mean value over all  $i \in \llbracket 0, N_n - 1 \rrbracket$ . Hence, the convergence of these means follows from the convergence of the series  $\sum_n 1/\ell_{n-1}$ . Using equality 1 again together with the convergence  $(\ell_n - d + 1)/\ell_n \rightarrow 1$  yields item 1.

2. Let  $d \geq 0$  and  $w \in A^d$ . Then

$$p_d(w) = \lim_{n \rightarrow +\infty} p(w|B_{n,1}) \geq 0.$$

In particular,  $p_0() = 1$ , where  $() \in A^0$  denotes the empty word. Moreover,

$$N(w, B_{n,1}) = \sum_{a \in A} N(wa, B_{n,1}) + \mathbf{1}_{\{B_{n,1} \text{ ends with } w\}}.$$

Dividing by  $\ell_n$  and letting  $n$  go to infinity yields

$$p_d(w) = \sum_{a \in A} p_{d+1}(wa).$$

In the same way, we get

$$p_d(w) = \sum_{a \in A} p_{d+1}(aw).$$

Therefore  $(p_d)_{d \geq 0}$  is a sequence of probability measures on the products  $(A^d)_{d \geq 0}$  such that each  $p_d$  is image of  $p_{d+1}$  by the projection on the first  $d$ , or on the last  $d$  components. Item 2 follows, by Kolmogorov extension theorem.

We now construct a stationary process  $\gamma = (\gamma(k))_{k \in \mathbf{Z}}$  with law  $Q$ . To do so, we index the letters of the  $n$ -blocks by the set  $\llbracket 0, \ell_n - 1 \rrbracket$ . Recall that every  $n$ -block is a concatenation of  $(n-1)$ -blocks in which each one of the  $N_{n-1}$  different  $(n-1)$ -blocks occurs exactly  $M_n$  times. Hence, the beginning of the  $(n-1)$ -blocks in any  $n$ -block are the positions  $q\ell_{n-1}$  with  $q \in \llbracket 0, M_n N_{n-1} - 1 \rrbracket$ .

**Proposition 45** *Let  $(I_n)_{n \geq 1}$  be a sequence of independent uniform random variables taking values in the sets  $(\llbracket 0, N_n - 1 \rrbracket)_{n \geq 1}$ , defined on some large enough large enough probability space  $(\Omega, \mathcal{A}, \mathbf{P})$ . Then*

1. *One can construct a sequence  $(U_n)_{n \geq 1}$  of uniform random variables taking values in the sets  $(\llbracket 0, \ell_n - 1 \rrbracket)_{n \geq 1}$  such that*
  - (a) *for all  $n \geq 1$ ,  $U_n$  is independent of  $(I_m)_{m \geq n}$ ;*
  - (b) *the sequence of random intervals  $(\llbracket -U_n, \ell_n - 1 - U_n \rrbracket)_{n \geq 1}$  is increasing;*
  - (c) *for all  $n \geq 2$  and  $k \in \llbracket -U_{n-1}, \ell_{n-1} - 1 - U_{n-1} \rrbracket$ ,*

$$B_{n, I_n}(U_n + k) = B_{n-1, I_{n-1}}(U_{n-1} + k).$$

2. *The intervals  $\llbracket -U_{n-1}, \ell_{n-1} - 1 - U_{n-1} \rrbracket$  cover  $\mathbf{Z}$  almost surely, so one can define a process  $\gamma = (\gamma(k))_{k \in \mathbf{Z}}$  by  $\gamma(k) = B_{n, I_n}(U_n + k)$  whenever  $k \in \llbracket -U_n, \ell_n - 1 - U_n \rrbracket$ .*
3. *The law of the process thus defined is  $Q$ .*

**Proof.** In the statement above, saying that the probability space  $(\Omega, \mathcal{A}, \mathbf{P})$  is large enough means that one can define a uniform random variable with values in  $[0, 1]$  which is independent on the sequence  $(I_n)_{n \geq 1}$ .

1. We construct the sequence  $(U_n)_{n \geq 1}$  recursively.

First, we set  $U_1 = 0$ . Since  $\ell_1 = 1$ , the random variable  $U_1$  is uniform on the set  $\llbracket 0, \ell_1 - 1 \rrbracket$ .

Let  $n \geq 2$ . Assume that  $U_1, \dots, U_{n-1}$  are constructed, that  $U_{n-1}$  is uniform on  $\llbracket 0, \ell_{n-1} - 1 \rrbracket$  and  $U_{n-1}$  is independent of  $(I_m)_{m \geq n}$ . Conditionally on  $(U_1, \dots, U_{n-1})$  and on the whole sequence  $(I_m)_{m \geq 1}$ , choose  $D_n$  uniformly among the  $M_n$  beginnings of the blocks  $B_{n-1, I_{n-1}}$  in the block  $B_{n, I_n}$ . Then for every  $k \in \llbracket 0, \ell_{n-1} \rrbracket$ ,  $B_{n, I_n}(D_n + k) = B_{n-1, I_{n-1}}(k)$ . Moreover, the random variable  $D_n$  is uniform on  $\{q\ell_{n-1} : q \in \llbracket 0, M_n N_n - 1 \rrbracket\}$  and independent on  $(U_{n-1}, (I_m)_{m \geq n})$ , so  $D_n$ ,  $U_{n-1}$  and  $(I_m)_{m \geq n}$  are independent by the recursion hypothesis. Hence the random variable  $U_n := U_{n-1} + D_n$  is uniform on  $\llbracket 0, \ell_n - 1 \rrbracket$  and independent of  $(I_m)_{m \geq n}$ . Moreover,  $U_{n-1} \leq U_n \leq U_{n-1} + (M_n N_n - 1)\ell_{n-1} = U_{n-1} + \ell_n - \ell_{n-1}$ . Item 1 follows.

2. Fix  $k \in \mathbf{Z}$ . For every large enough  $n$ ,  $\ell_n \geq |k|$ , so

$$\mathbf{P}[k \notin \llbracket -U_n, \ell_n - 1 - U_n \rrbracket] = \left\{ \begin{array}{ll} \mathbf{P}[U_n < -k] & \text{if } k \leq 0 \\ \mathbf{P}[\ell_n - 1 - U_n < k] & \text{if } k \geq 0 \end{array} \right\} = \frac{|k|}{\ell_n}.$$

Since  $\ell_n \rightarrow +\infty$ , item 2 follows.

3. For every  $n \leq 0$ , consider the process  $\gamma^{(n)} = (\gamma^{(n)}(k))_{k \in \mathbf{Z}}$  given by

$$\gamma^{(n)}(k) = B_{n, I_n}((U_n + k) \bmod \ell_n),$$

where  $(U_n + k) \bmod \ell_n$  denotes the remainder of  $U_n + k$  modulo  $\ell_n$ . The process  $\gamma^{(n)}$  thus defined is  $\ell_n$ -periodic and stationary, since the process  $((U_n + k) \bmod \ell_n)_{k \in \mathbf{Z}}$  is stationary and independent of  $I_n$ . But the process  $\gamma^{(n)}$  coincides with  $\gamma$  on the random interval  $\llbracket -U_n, \ell_n - 1 - U_n \rrbracket$ . Since the intervals  $\llbracket -U_n, \ell_n - 1 - U_n \rrbracket$  increase to  $\mathbf{Z}$  almost surely,  $\gamma$  is the almost sure limit of the processes  $\gamma^{(n)}$ , so

it is stationary. Hence we only need to check that for every  $d \geq 0$ , the law of  $\gamma_{0:d-1} := (\gamma(0), \dots, \gamma(d-1))$  is  $p_d$ . Let  $w \in A^d$ . Using twice that for every  $n \geq 0$ ,  $U_n$  is independent of  $I_n$ , we get

$$\begin{aligned}
\mathbf{P}[\gamma_{0:d-1} = w] &= \lim_n \mathbf{P}[(\gamma_0, \dots, \gamma_{d-1}) = w ; d-1 \leq \ell_n - 1 - U_n] \\
&= \lim_n \mathbf{P}[(B_{n,I_n}(U_n), \dots, B_{n,I_n}(U_n + d - 1)) = w ; U_n \leq \ell_n - d] \\
&= \lim_n \frac{1}{\ell_n} \sum_{u=0}^{\ell_n-d} \mathbf{P}[(B_{n,I_n}(u), \dots, B_{n,I_n}(u + d - 1)) = w] \\
&= \lim_n \frac{1}{\ell_n - d + 1} \sum_{u=0}^{\ell_n-d} \mathbf{P}[(B_{n,I_n}(u), \dots, B_{n,I_n}(u + d - 1)) = w] \\
&= \lim_n \mathbf{E}[p(w|B_{n,I_n})] \\
&= p_d(w).
\end{aligned}$$

Item 3 follows.

We now deduce some properties on the shift  $T$  under the probability measure  $Q$ .

**Proposition 46** (*Properties of the shift  $T$  under  $Q$* )

1. For every  $d \geq 0$ ,  $w \in A^d$ , and  $Q$ -almost every  $(x_k)_{k \in \mathbf{Z}} \in A^{\mathbf{Z}}$ ,

$$\lim_{L \rightarrow +\infty} p(w|(x_0, \dots, x_{L-1})) = p_d(w).$$

2.  $T$  is ergodic under  $Q$ .

3. If  $\ell_n^{-1} \log_2 N_n \rightarrow 0$  as  $n \rightarrow +\infty$ , then  $T$  has null entropy under  $Q$ .

*Proof.* 1. Since  $Q$  is the law of the process  $\gamma$  constructed in proposition 45, it gives full measure to the set of sample paths  $(x_k)_{k \in \mathbf{Z}}$  such that for every  $n \geq 0$ , there exists  $u_n \in \llbracket 0, \ell_n - 1 \rrbracket$  such that for every slice  $(x_{-u_n+q\ell_n}, \dots, x_{-u_n+q\ell_n+\ell_n-1})$  is an  $n$ -block. Roughly speaking, for each  $n \geq 1$ , every typical sample path can be obtained for as a randomly shifted infinite concatenation of  $n$ -blocks.

Fix such a path and  $\varepsilon > 0$ . Provided  $n \geq 0$  is large enough, one has

$$\forall i \in \llbracket 0, N_n - 1 \rrbracket, \quad |p(w|B_{n,i}) - p_d(w)| \leq \varepsilon/3,$$

and more generally,

$$\forall i_1, \dots, i_n \in \llbracket 0, N_n - 1 \rrbracket, \quad |p(w|B_{n,i_1} \dots B_{n,i_m}) - p_d(w)| \leq 2\varepsilon/3,$$

since the edge effect at the boundaries of the  $n$ -blocks is small if  $\ell_n$  is large with regard to  $d$ . Given  $L \geq 2\ell_n$ , the word  $(x_0, \dots, x_{L-1})$  can be split into  $m(L) := \lfloor L/\ell_n \rfloor - 1$   $n$ -blocks, namely  $(x_{-u_n+q\ell_n}, \dots, x_{-u_n+q\ell_n+\ell_n-1})$  with  $q \in \llbracket 1, m(L) \rrbracket$ , plus two pieces of  $n$ -blocks. If  $L$  is large enough with regard to  $\ell_n$ , the effect of these two pieces is small, so  $|p(w|(x_0, \dots, x_{L-1})) - p_d(w)| \leq \varepsilon$ . Item 1 follows.

2. Call  $(X_k)_{k \in \mathbf{Z}}$  the coordinate process on  $A^{\mathbf{Z}}$  and  $\mathcal{I}$  the  $\sigma$ -field of all  $T$ -invariant subsets of  $A^{\mathbf{Z}}$ . Let  $d \geq 0$  and  $w \in A^d$ . Birkhoff ergodic theorem and item 1 yield the almost sure equalities

$$\begin{aligned}
Q[(X_0, \dots, X_{d-1}) = w | \mathcal{I}] &= \lim_{L \rightarrow +\infty} \frac{1}{L} \sum_{k=0}^{L-1} \mathbf{1}_{\{(X_0, \dots, X_{d-1}) = w\}} \circ T^k \\
&= \lim_{L \rightarrow +\infty} \frac{1}{L} \sum_{k=0}^{L-1} \mathbf{1}_{\{(X_k, \dots, X_{k+d-1}) = w\}} \\
&= \lim_{L \rightarrow +\infty} p(w | (X_0, \dots, X_{L+d-2})) \\
&= p_d(w) = Q[(X_0, \dots, X_{d-1}) = w].
\end{aligned}$$

Since  $T$  preserves  $Q$  and  $A$  is finite, the almost sure equality  $Q(\Gamma | \mathcal{I}) = Q(\Gamma)$  holds for every subset  $\Gamma$  of  $A^{\mathbf{Z}}$  depending on finitely many coordinates, and therefore for every measurable subset of  $A^{\mathbf{Z}}$ . The ergodicity of  $T$  under  $Q$  follows.

3. The entropy of  $T$  is

$$h(T) = \lim_{L \rightarrow +\infty} H(X_0, \dots, X_{L-1})/L = \lim_{n \rightarrow +\infty} H(X_0, \dots, X_{\ell_n})/(\ell_n + 1).$$

Almost surely, the string  $(X_0, \dots, X_{\ell_n})$  is the concatenation of the last  $k$  letters of some  $n$ -block and the first  $\ell_n - k + 1$  letters of some  $n$  block, with  $k \in \llbracket 1, \ell_n \rrbracket$ . Hence, the number of possible values of  $(X_0, \dots, X_{\ell_n})$  is at most  $\ell_n \times N_n^2$ , so  $H(X_0, \dots, X_{\ell_n}) \leq \log_2 \ell_n + 2 \log_2 N_n$ . Item 4 follows.

We now give a precise description of the block structure in a slight variant of Hoffman's example.

The 1-blocks are the elements of  $A$ . By assumption  $|A| \geq 2$ . Once the  $(n-1)$ -blocks  $B_{n-1,0}, \dots, B_{n-1, N_{n-1}-1}$  are constructed<sup>5</sup>, the  $n$ -blocks are defined by

$$\forall i \in \llbracket 0, n^4 - 1 \rrbracket, \quad B_{n,i} = ((B_{n-1,1})^{n^{5i}} \dots (B_{n-1, N_{n-1}})^{n^{5i}})^{n^{5(n^4-1-i)}},$$

so the length of  $B_{n,i}$  is  $\ell_n = n^{5(n^4-i-1)} \times N_{n-1} \times n^{5i} \times \ell_{n-1} = n^{5(n^4-1)} N_{n-1} \ell_{n-1}$ , each block  $B_{n-1,j}$  occurs exactly  $n^{4(n^2-1)}$  times in  $B_{n,i}$ , and the number of different  $n$ -blocks is  $N_n = n^4$ .

The successive repetitions of a same  $(n-1)$ -block inside an  $n$ -block form what we call an  $n$ -region. The length of the of the  $n$ -regions in  $B_{n,i}$ , namely  $r_{n,i} = n^{5i} \ell_{n-1}$ , depends highly on  $i$ . For every integers  $k \geq 0$  and  $\ell \geq 1$ , denote by  $\lfloor k/\ell \rfloor$  and  $k \bmod \ell$  the quotient and the remainder of  $k$  modulo  $\ell$ . Then by construction,

$$\forall i \in \llbracket 0, n^2 - 1 \rrbracket, \quad \forall k \in \llbracket 0, \ell_n - 1 \rrbracket, \quad B_{n,i}(k) = B_{n-1, (\lfloor k/r_{n,i} \rfloor \bmod N_{n-1})} (k \bmod \ell_{n-1}).$$

By proposition 46,  $T$  has null entropy under  $Q$  since  $\ell_n^{-1} \log_2 N_n \rightarrow 0$  as  $n \rightarrow +\infty$ . Moreover,  $T$  is ergodic under  $Q$ , so  $[T, \text{Id}]$  is exact by theorem 1. The filtration associated to  $[T, \text{Id}]$  is dyadic and Kolmogorovian and we want to prove that it is not product-type.

<sup>5</sup>We index the  $n$ -blocks by  $\llbracket 0, N_n - 1 \rrbracket$  instead of  $\llbracket 1, N_n \rrbracket$  to handle simpler formulas. For the same reason, we view each word with length  $\ell$  as a map from  $\llbracket 0, \ell - 1 \rrbracket$  to  $A$ .

## 6.2 Non-standardness of $[T, \text{Id}]$

The description of the associated to  $[T, \text{Id}]$  is close to the description of the ordinary  $[T, T^{-1}]$  filtration that we made in the previous section: we take two independent random variables  $(\xi_n)_{n \in \mathbf{Z}}$  and  $\gamma = (\gamma(s))_{s \in \mathbf{Z}}$  with values in  $F^{\mathbf{Z}}$  and  $A^{\mathbf{Z}}$ , defined on a same probability space  $(\Omega, \mathcal{A}, \mathbf{P})$  with respective law  $\mu^{\otimes \mathbf{Z}}$  and  $Q$ . But this time  $F = \{0, 1\}$ , so  $\mu$  is the uniform law on  $\{0, 1\}$ , and  $(\gamma(s))_{s \in \mathbf{Z}}$  is no more an i.i.d. sequence.

Set  $S_n = -\xi_{n+1} - \dots - \xi_0$  for every  $n \leq 0$  and let  $Z := ((\xi_n)_{n \leq 0}, (\gamma(s))_{s \in \mathbf{Z}})$ . Then ‘the’ filtration associated to  $[T, \text{Id}]$  is the natural filtration  $(\mathcal{F}_n^Z)_{n \leq 0}$  of the process  $(Z_n)_{n \leq 0}$  defined by

$$Z_n = [T, \text{Id}](Z) = ((\xi_{k+n})_{k \leq 0}, (\gamma(S_n + s))_{s \in \mathbf{Z}}).$$

By construction,  $(\xi_n)_{n \leq 0}$  is a sequence of innovations of the filtration  $(\mathcal{F}_n^Z)_{n \leq 0}$ .

By theorem 28, it is sufficient to check that the random variable  $\gamma(0)$  does not satisfy the Vershik intermediate property. To do this, it is convenient to introduce the nibbled-word process  $(W_n, \xi_n)_{n \leq 0}$  by

$$\forall n \leq 0, \quad W_n = (\gamma(S_n + i))_{i \in \llbracket 0, |n| \rrbracket}.$$

The set  $\llbracket 0, |n| \rrbracket$  is exactly the set of all possible values of the sum  $x_{n+1} + \dots + x_0$  when  $x_{n+1}, \dots, x_0$  range over  $F = \{0, 1\}$ , we can define a map  $s_n$  from  $F^{|n|}$  to  $\llbracket 0, |n| \rrbracket$  by  $s_n(x_{n+1}, \dots, x_0) = x_{n+1} + \dots + x_0$ . To every word  $w \in A^{|n|}$ , we associate its extension  $\bar{w} = w \circ s_n \in A^{F^{|n|}}$ . As in section 5, we check the split-word process associated to  $\gamma(0)$  and to the innovations  $(\xi_n)_{n \leq 0}$  is  $(\overline{W}_n, \xi_n)_{n \leq 0}$ .

For every  $n \geq 2$ , let

$$h_n := (\ell_n / N_n)^2 = (r_{n, N(n)-1})^2 = (\ell_{n-1} n^{5(n^4-1)})^2.$$

Note that  $h_n$  is even. We define a decreasing sequence  $(\varepsilon_n)_{n \geq 2}$  of positive real numbers by  $\varepsilon_2 = 2^{-h(2)}$  and for every  $n \geq 3$ ,

$$\varepsilon_n = \varepsilon_{n-1} \left( 1 - \frac{3}{(n-1)^3} - \frac{4}{(n-1)^4} - \frac{1}{n^2} \right).$$

This sequence has a positive limit  $\varepsilon_\infty$ .

To negate the Vershik intermediate property, we have to show that  $\text{disp}(\overline{W}_{-h_n}, d_{h_n})$  is bounded away from 0. This will follow from the next lemma.

**Lemma 47** *Let  $n \geq 2$ . Let  $w'$  and  $w''$  be two words in  $A^{\llbracket 0, h_n \rrbracket}$  whose restrictions to the middle interval  $M_n := \llbracket (h_n/2) - n\sqrt{h_n}, (h_n/2) + n\sqrt{h_n} \rrbracket$  are entirely contained in two different  $n$ -blocks, namely  $B_{n, i'}$  and  $B_{n, i''}$ , with  $i' \neq i''$ . Then  $d_{h_n}(\overline{w'}, \overline{w''}) \geq \varepsilon_n$ .*

Before proving lemma 47, let us deduce that the random variable  $\gamma(0)$  does not satisfy the Vershik intermediate property. Fix  $n \geq 2$ . For every  $(x_{-h_n+1}, \dots, x_0) \in F^{h_n}$ ,

$$\overline{W}_{-h_n}(x_{-h_n+1}, \dots, x_0) = W_n(x_{-h_n+1} + \dots + x_0) = \gamma(S_{-h_n} + x_{-h_n+1} + \dots + x_0).$$

Let  $\gamma'$  and  $\gamma''$  be two independent copies of  $\gamma$  defined on some probability space  $(\overline{\Omega}, \overline{\mathcal{A}}, \overline{\mathbf{P}})$ . The shifted process  $\gamma(S_{-h_n} + h_n/2 + \cdot)$  has the same law as  $\gamma$ , so one gets two independent



copies of  $W_{-h_n}$  by setting  $W'_{-h_n}(i) = \gamma'(i - h_n/2)$  and  $W''_{-h_n}(i) = \gamma''(i - h_n/2)$ . The interest of this translation is that when  $n$  is large, the binomial law with parameters  $h_n$  and  $1/2$  gives probability close to 1 to the interval  $\llbracket h_n/2 - n\sqrt{h_n}, h_n/2 + n\sqrt{h_n} \rrbracket$ , so most of the values

$$\overline{W'_{-h_n}}(x_{-h_n+1}, \dots, x_0) = W'_{-h_n}(x_{-h_n+1} + \dots + x_0) = \gamma'_{x_{-h_n+1} + \dots + x_0 - h_n/2}$$

are provided by the restriction of  $\gamma'$  to the interval  $\llbracket -n\sqrt{h_n}, n\sqrt{h_n} \rrbracket$ .

Following the construction of proposition 45, one may assume that the processes  $\gamma'$  and  $\gamma''$  derive from two independent copies  $(I'_n, U'_n)_{n \geq 1}$  and  $(I''_n, U''_n)_{n \geq 1}$  of the sequence  $(I_n, U_n)_{n \geq 1}$ . Then the restriction of  $\gamma'$  to the interval  $\llbracket -U'_n, \ell_n - 1 - U'_n \rrbracket$  is a time-translation of the  $n$ -block  $B_{n, I'_n}$ , and a similar statement holds for  $\gamma''$ .

Therefore, when the three following conditions hold

- $I'_n \neq I''_n$ ,
- $\llbracket -n\sqrt{h_n}, n\sqrt{h_n} \rrbracket \subset \llbracket -U'_n, \ell_n - 1 - U'_n \rrbracket$ ,
- $\llbracket -n\sqrt{h_n}, n\sqrt{h_n} \rrbracket \subset \llbracket -U''_n, \ell_n - 1 - U''_n \rrbracket$ ,

lemma 47 applies, so  $d_{h_n}(\overline{W'_{-h_n}}, \overline{W''_{-h_n}}) \geq \varepsilon_n$ . But by independence of the random variables  $I'_n, U'_n, I''_n, U''_n$  the probability that these three conditions hold is

$$\left(1 - \frac{1}{N(n)}\right) \times \left(1 - \frac{2n\sqrt{h_n}}{\ell_n}\right)^2 = \left(1 - \frac{1}{n^4}\right) \left(1 - \frac{2}{n^3}\right).$$

Hence

$$\text{disp}(\overline{W_{-h_n}}, d_{h_n}) = \mathbf{E}[d_{h_n}(\overline{W'_{-h_n}}, \overline{W''_{-h_n}})] \geq \varepsilon_n \left(1 - \frac{1}{n^4}\right) \left(1 - \frac{2}{n^3}\right),$$

which remains bounded away from 0. Theorem 6 follows.

We now prove lemma 47.

Proof. We argue by recursion. To make the notations lighter, we will introduce a symbol  $\mathbb{M}$  to denote arithmetic means: given any non-empty finite set  $E$ ,

$$\mathbb{M}_{x \in E} \text{ stands for } \frac{1}{|E|} \sum_{x \in E}.$$

First, assume that  $n = 2$ . The assumption  $i' \neq i''$  and the construction of the 2-blocks prevent  $w''$  to be equal to  $w'$  or to its adjoint  $w'^*$  (which is the reversed word  $w'^r$  since  $w'$  is not 2-periodic). By lemma 37, their extensions  $w'$  and  $w''$  belong to two different orbit modulo the action of the automorphism group of the binary tree with height  $h_2$ , so  $d_{h_2}(\overline{w'}, \overline{w''}) \geq 2^{-h_2} = \varepsilon_2$ .

Now, let  $n \geq 3$ . Assume that the implication is established at level  $n - 1$ . Let  $i' < i''$  in  $\llbracket 0, N_n - 1 \rrbracket$ , and take two subwords  $w'$  of  $B_{n, i'}$  and  $w''$  of  $B_{n, i''}$  having length  $h_n + 1$ : there exist two integers  $u'$  and  $u''$  in  $\llbracket 0, \ell_n - h_n - 1 \rrbracket$ , such that for every  $s \in \llbracket 0, h_n \rrbracket$ ,  $w'(s) = B_{n, i'}(u' + s)$  and  $w''(s) = B_{n, i''}(u'' + s)$ .

Remind the notations of subsection 3.3. For every non-negative integer  $h$ , we view the set  $T_h = \bigcup_{i \in \llbracket 0, h \rrbracket} F^i$  as the binary tree with height  $h$ . To each  $\sigma \in \mathfrak{S}(F)^{T_{h_n-1}}$ , we associate the automorphism of  $T_{h_n}$  given by

$$g_\sigma(x_{-h_n+1}, \dots, x_0) = (x_{-h_n+1}^\sigma, \dots, x_0^\sigma),$$

where

$$\forall t \in \llbracket -h_n + 1, 0 \rrbracket, \quad x_t^\sigma := \sigma(x_{-h_n+1}, \dots, x_{t-1})(x_t).$$

Given  $\sigma \in \mathfrak{S}(F)^{T_{h_n-1}}$ , we have to prove that

$$\delta_{h_n}(\overline{w'}, \overline{w''} \circ g_\sigma) \geq \varepsilon_n.$$

Split each  $h_n$ -uple  $(x_{-h_n+1}, \dots, x_0) \in F^{h_n}$  into  $y = (x_{-h_n+1}, \dots, x_{-h_{n-1}}) \in F^{h_n-h_{n-1}}$  and  $z = (x_{-h_{n-1}+1}, \dots, x_0) \in F^{h_n}$ . Call  $\sigma_y$  the element of  $\mathfrak{S}(F)^{T_{h_n-1}}$  defined by  $\sigma_y(z_1, \dots, z_i) = \sigma(y, z_1, \dots, z_i)$  for every  $0 \leq i \leq h_n - 1$  and  $(z_1, \dots, z_i) \in F^i$ . Set  $s(y) = x_{-h_n+1} + \dots + x_{-h_{n-1}}$ ,  $s(z) = x_{-h_{n-1}+1} + \dots + x_0$ . Then

$$\begin{aligned} \delta_{h_n}(\overline{w'}, \overline{w''} \circ g_\sigma) &= \prod_{y \in F^{h_n-h_{n-1}}} \prod_{z \in F^{h_n-1}} \mathbf{1}_{\{\overline{w'}(yz) \neq \overline{w''} \circ g_\sigma(yz)\}} \\ &= \prod_{y \in F^{h_n-h_{n-1}}} \delta_{h_{n-1}}(\overline{w'}(y, \dots), \overline{w''} \circ g_\sigma(y, \dots)). \end{aligned}$$

But for every  $y \in F^{h_n-h_{n-1}}$  and  $z \in F^{h_n}$ ,

$$\overline{w'}(y, z) = w'(s(yz)) = w'(s(y) + s(z)),$$

$$\overline{w''}(g_\sigma(y, z)) = \overline{w''}(g_\sigma(y)g_{\sigma_y}(z)) = w''(s(g_\sigma(y)) + s(g_{\sigma_y}(z))),$$

so

$$\overline{w'}(y, \dots) = \overline{(w'(s(y) + \cdot))|_{\llbracket 0, h_{n-1} \rrbracket}},$$

$$(\overline{w''} \circ g_\sigma)(y, \dots) = \overline{(w''(s(g_\sigma(y)) + \cdot))|_{\llbracket 0, h_{n-1} \rrbracket}} \circ g_{\sigma_y}.$$

Hence

$$\begin{aligned} \delta_{h_n}(\overline{w'}, \overline{w''} \circ g_\sigma) &= \prod_{y \in F^{h_n-h_{n-1}}} \delta_{h_{n-1}}(\overline{(w'(s(y) + \cdot))|_{\llbracket 0, h_{n-1} \rrbracket}}, \overline{(w''(s(y^\sigma) + \cdot))|_{\llbracket 0, h_{n-1} \rrbracket}} \circ g_{\sigma_y}) \\ &\geq \prod_{y \in F^{h_n-h_{n-1}}} d_{h_{n-1}}(\overline{(w'(s(y) + \cdot))|_{\llbracket 0, h_{n-1} \rrbracket}}, \overline{(w''(s(y^\sigma) + \cdot))|_{\llbracket 0, h_{n-1} \rrbracket}}). \end{aligned}$$

To apply the recursion hypothesis, we look at the restrictions of  $w'(s(y) + \cdot)$  and  $w''(s(y^\sigma) + \cdot)$  to the interval  $M_{n-1}$ . For every  $k \in M_{n-1}$ , one has

$$\begin{aligned} w'(s(y) + k) &= B_{n, i'}(u' + s(y) + k) \\ &= B_{n-1, (\lfloor (u' + s(y) + k) / r_{n, i'} \rfloor \bmod N_{n-1})}((u' + s(y) + k) \bmod \ell_{n-1}), \end{aligned}$$

and  $w''(s(y^\sigma) + k)$  is given by a similar formula. Set

$$\begin{aligned} J'(y) &:= \lfloor (u' + s(y) + h_{n-1}/2) / r_{n, i'} \rfloor \bmod N_{n-1}, \\ J''(y^\sigma) &:= \lfloor (u'' + s(y^\sigma) + h_{n-1}/2) / r_{n, i''} \rfloor \bmod N_{n-1}, \\ K'(y) &:= (u' + s(y) + h_{n-1}/2) \bmod \ell_{n-1}, \\ K''(y^\sigma) &:= (u'' + s(y^\sigma) + h_{n-1}/2) \bmod \ell_{n-1}, \\ \Lambda_{n-1} &:= \llbracket (n-1)\sqrt{h_{n-1}}, \ell_{n-1} - 1 - (n-1)\sqrt{h_{n-1}} \rrbracket. \end{aligned}$$

If  $K'(y)$  and  $K''(y^\sigma)$  belong to the interval  $\Lambda_{n-1}$ , the restrictions of  $w'(s(y) + \cdot)$  and  $w''(s(y^\sigma) + \cdot)$  to the interval  $M_{n-1}$  are entirely contained respectively in  $B_{n-1, J'(y)}$  and

$B_{n-1, J''(y^\sigma)}$ . Therefore, since the random variables  $Y := (\xi_{-h_n+1}, \dots, \xi_{-h_{n-1}})$  and  $Y^\sigma := (\xi_{-h_n+1}^\sigma, \dots, \xi_{-h_{n-1}}^\sigma)$  are uniform on  $F^{h_n-h_{n-1}}$ , the recursion hypothesis yields

$$\begin{aligned} \delta_{h_n}(\overline{w'}, \overline{w''} \circ g_\sigma) &\geq \bigvee_{y \in F^{h_n-h_{n-1}}} \varepsilon_{n-1} \mathbf{1}_{\{K(y) \in \Lambda_{n-1}; K''(y^\sigma) \in \Lambda_{n-1}; J'(y) \neq J''(y^\sigma)\}} \\ &= \varepsilon_{n-1} \mathbf{P}[K(Y) \in \Lambda_{n-1}; K''(Y^\sigma) \in \Lambda_{n-1}; J'(Y) \neq J''(Y^\sigma)]. \end{aligned}$$

Thus

$$\delta_{h_n}(\overline{w'}, \overline{w''} \circ g_\sigma) \geq \varepsilon_{n-1} \left( \mathbf{P}[J'(Y) \neq J''(Y^\sigma)] - 2\mathbf{P}[K(Y) \notin \Lambda_{n-1}] \right). \quad (2)$$

To bound above  $\mathbf{P}[K(Y) \notin \Lambda_{n-1}]$ , we note that  $s(Y)$  has a binomial distribution with parameters  $h_n - h_{n-1}$  and  $1/2$ . By lemma 51, for all  $k \in \llbracket 0, \ell_{n-1} \rrbracket$ ,

$$\begin{aligned} \mathbf{P}[K(Y) = k] &= \sum_{q \in \mathbf{Z}} \mathbf{P}\left[s(Y) = \ell_{n-1}q + k - u' - \frac{h_{n-1}}{2}\right] \\ &\leq \frac{1}{\ell_{n-1}} + \frac{1}{\sqrt{h_n - h_{n-1}}} \\ &= \frac{1}{\ell_{n-1}} + \frac{1}{\ell_{n-1} \sqrt{n^5(n^4-1) - N_{n-1}^{-2}}} \\ &\leq \frac{3}{2\ell_{n-1}}. \end{aligned}$$

Thus

$$\mathbf{P}[K(Y) \notin \Lambda_{n-1}] \leq \frac{3}{2\ell_{n-1}} \times 2(n-1)\sqrt{h_{n-1}} = \frac{3(n-1)}{N_{n-1}} = \frac{3}{(n-1)^3}. \quad (3)$$

We now want to bound above  $\mathbf{P}[J'(Y) = J''(Y^\sigma)]$ . To do this, we set  $D = n^8 r_{n,i'}^2$ , so  $n^4 r_{n,i'} = \sqrt{D} \leq n^{-1} r_{n,i''}$  and  $D \leq n^{-2} (r_{n, N_{n-1}})^2 \leq h_n/9 \leq h_n - h_{n-1}$ , and we split  $Y$  and  $Y^\sigma$  into two independent parts, namely

$$\begin{aligned} Y_1 &:= (\xi_{-h_n+1}, \dots, \xi_{-h_{n-1}-D}) \quad \text{and} \quad Y_2 := (\xi_{-h_{n-1}-D+1}, \dots, \xi_{-h_{n-1}}), \\ (Y^\sigma)_1 &:= (\xi_{-h_n+1}^\sigma, \dots, \xi_{-h_{n-1}-D}^\sigma) \quad \text{and} \quad (Y^\sigma)_2 := (\xi_{-h_{n-1}-D+1}^\sigma, \dots, \xi_{-h_{n-1}}^\sigma), \end{aligned}$$

Then we show that the law of  $J'(Y)$  given  $Y_1$  is spread out on the whole interval  $\llbracket 0, N_{n-1} - 1 \rrbracket$  whereas the law of  $J''(Y^\sigma)$  given  $Y_1$  is mainly concentrated on at most two points.

On the one hand, one checks that

$$J'(Y) = \left\lfloor \frac{(u' + s(Y) + h_{n-1}/2) \bmod r_{n,i'} N_{n-1}}{r_{n,i'}} \right\rfloor$$

Using the equality  $s(Y) = s(Y_1) + s(Y_2)$ , that  $s(Y_2)$  has a binomial distribution with parameters  $D$  and  $1/2$ , and lemma 51 again, one gets that for all  $k \in \llbracket 0, r_{n,i'} N_{n-1} - 1 \rrbracket$ ,

$$\begin{aligned} \mathbf{P}[(u' + s(Y) + h_{n-1}/2) \bmod r_{n,i'} N_{n-1} = k | \sigma(Y_1)] &\leq \frac{1}{r_{n,i'} N_{n-1}} + \frac{1}{\sqrt{D}} \\ &\leq \frac{1}{r_{n,i'} N_{n-1}} + \frac{1}{n^4 r_{n,i'}} \\ &\leq \frac{2}{r_{n,i'} N_{n-1}} \end{aligned}$$

Hence, for every  $j \in \llbracket 0, N_{n-1} - 1 \rrbracket$ ,

$$\mathbf{P}[J'(Y) = j | \sigma(Y_1)] \leq \frac{2}{N_{n-1}}.$$

On the other hand,  $s(Y^\sigma) = s(Y_1^\sigma) + s(Y_2^\sigma)$  and  $Y_1^\sigma$  is a function of  $Y_1$ , whereas  $Y_2^\sigma$  is independent of  $Y_1$  and has the same law as  $Y_2$ . Hence

$$\text{Var}\left(u' + s(Y^\sigma) + \frac{h_{n-1}}{2} \mid \sigma(Y_1)\right) = \text{Var}(s(Y_2^\sigma)) = \frac{D}{4}.$$

Set

$$M = r_{n,i''}^{-1} \mathbf{E}\left[u' + s(Y^\sigma) + \frac{h_{n-1}}{2} \mid \sigma(Y_1)\right], \quad M_- = \left\lfloor M - \frac{1}{2} \right\rfloor, \quad M_+ = \left\lceil M + \frac{1}{2} \right\rceil.$$

Then Bienaymé-Chebycheff inequality yields

$$\mathbf{P}\left[\left|u' + s(Y^\sigma) + \frac{h_{n-1}}{2} - r_{n,i''} M\right| \geq \frac{r_{n,i''}}{2} \mid \sigma(Y_1)\right] \leq \frac{D}{r_{n,i''}^2} \leq \frac{1}{n^2},$$

so

$$\begin{aligned} 1 - \frac{1}{n^2} &\leq \mathbf{P}\left[\left|\frac{u' + s(Y^\sigma) + \frac{h_{n-1}}{2}}{r_{n,i''}} - M\right| < \frac{1}{2} \mid \sigma(Y_1)\right] \\ &\leq \mathbf{P}\left[\left[\frac{u' + s(Y^\sigma) + \frac{h_{n-1}}{2}}{r_{n,i''}}\right] \in \{M_-; M_+\} \mid \sigma(Y_1)\right] \\ &\leq \mathbf{P}\left[J''(Y^\sigma) \in \{M_- \bmod N_{n-1}; M_+ \bmod N_{n-1}\} \mid \sigma(Y_1)\right]. \end{aligned}$$

Comparing the conditional laws of  $J'(Y)$  and  $J''(Y^\sigma)$  given  $Y_1$  yields

$$\mathbf{P}[J'(Y) = J''(Y^\sigma)] \leq \frac{4}{N_{n-1}} + \frac{1}{n^2} = \frac{4}{(n-1)^4} + \frac{1}{n^2}. \quad (4)$$

Plugging inequalities 3 and 4 into inequality 2 yields

$$\delta_{h_n}(\bar{w}', \bar{w}'' \circ g_\sigma) \geq \varepsilon_{n-1} \left(1 - \frac{3}{(n-1)^3} - \frac{4}{(n-1)^4} - \frac{1}{n^2}\right) = \varepsilon_n.$$

The proof is complete.

## 7 Annex

### 7.1 Useful results on Polish spaces

Fix a non-empty separable complete metric space  $(A, d)$ , endowed with the Borel  $\sigma$ -field. We begin with a lemma abridged from de la Rue's paper on Lebesgue spaces [19].

**Lemma 48** *Fix a countable basis  $(B_n)_{n \geq 1}$  of bounded open sets (for example, the balls whose center lies in some countable dense subset and whose radius is the inverse of a positive integer). Let  $C = \{0, 1\}^\infty$  and  $\Phi : A \rightarrow C$  be the map defined by  $\Phi(x) = (\mathbf{1}_{B_n}(x))_{n \geq 1}$ . Then  $\Phi$  is injective,  $\Phi(A)$  is a Borel subset of the compact set  $C$  and  $\Phi^{-1} : \Phi(A) \rightarrow A$  is a Borel map.*

Proof. First, note that the sets  $(B_n)_{n \geq 1}$  separate the points of  $A$ , so the map  $\Phi$  is injective. Moreover, for every  $y = (y_n)_{n \geq 1} \in C$  and  $n \geq 1$ , set  $B_n^{y_n} = B_n$  if  $y_n = 1$ ,  $B_n^{y_n} = B_n^c$  if  $y_n = 0$ . Then

$$\Phi^{-1}(\{y\}) = \bigcap_{n \geq 1} B_n^{y_n}.$$

When  $B$  is a bounded subset in  $(A, d)$ , denote by  $\text{diam}(B)$  its diameter. For every  $n \geq 1$ , set  $I_n = \{m \geq 1 : \overline{B_m} \subset B_n \text{ and } \text{diam}(B_m) \leq \text{diam}(B_n)/2\}$ . Then the set  $\Phi(A)$  is the set of all  $y = (y_n)_{n \geq 1} \in C$  satisfying the three conditions below:

1. For every  $N \geq 1$ ,  $B_1^{y_1} \cap \dots \cap B_N^{y_N} \neq \emptyset$ .
2. There exists  $n \geq 1$  such that  $y_n = 1$ .
3. For every  $n \geq 1$  such that  $y_n = 1$ , there exists  $m \in I_n$  such that  $y_m = 1$ .

Indeed, these conditions are necessary for  $y$  to be in  $\Phi(A)$  since  $(B_n)_{n \geq 1}$  is a countable basis  $(B_n)_{n \geq 1}$  of bounded open sets in the metric space  $(A, d)$ . Conversely, these conditions ensure that the diameter of the non-empty closed subset  $F_N := \overline{B_1^{y_1}} \cap \dots \cap \overline{B_N^{y_N}}$  tends to 0 as  $N$  goes to infinity. Since  $(A, d)$  is complete, and  $(F_N)_{N \geq 1}$  is a non-increasing sequence, its intersection is a single set  $\{x\}$ . To see that  $\Phi(x) = y$ , we have to check that for every  $n \geq 1$ ,  $x$  belongs to  $B_n^{y_n}$ . If  $y_n = 0$ , this is true since  $B_n^{y_n} = B_n^c$  is closed. If  $y_n = 1$ , then  $y_m = 1$  for some  $m \in I_n$ , so  $x \in \overline{B_m} \subset B_n$ . This proves the characterization above, so  $\Phi(A)$  is a Borel subset of  $C$ .

For every closed subset  $F$  in  $(A, d)$ ,  $(\Phi^{-1})^{-1}(F) = \Phi(F)$  is still a Borel subset of  $C$ , since the induced metric space  $(F, d_F)$  is complete and separable, and  $(F \cap B_n)_{n \geq 1}$  is a countable basis of bounded open sets in  $(F, d_F)$ . Hence,  $\Phi^{-1} : \Phi(A) \rightarrow A$  is a Borel map. The proof is complete.

We now state and prove the lemma which legitimates the definition 18.

**Lemma 49** *On a probability space  $(\Omega, \mathcal{A}, \mathbf{P})$ , let  $\mathcal{F}$  be a sub- $\sigma$ -field,  $\xi$  a random variable taking values in a countable set  $F$ , independent of  $\mathcal{F}$ , and  $X$  an  $\mathcal{F} \vee \sigma(\xi)$ -measurable random variable taking values in  $A$ . Given  $x \in F$ , one can find an  $\mathcal{F}$ -measurable random variable  $W_x$  taking values in  $A$  such that  $X$  and  $W_x$  coincide on the event  $\{\xi = x\}$ . If  $\mathbf{P}[\xi = x] > 0$ , such a random variable is almost surely unique.*

Proof. Let  $\Phi : A \rightarrow C$  be the map defined in lemma 48. We will only use the injectivity of  $\Phi$  and the measurability of  $\Phi^{-1}$ . Denote by  $\Phi_n = \mathbf{1}_{B_n}$  the  $n$ -th component of  $\Phi$ .

We begin with the almost sure uniqueness when  $\mathbf{P}[\xi = x] > 0$ . If  $W_x$  exists, then for every  $n \geq 1$ ,

$$\Phi_n(X) \mathbf{1}_{\{\xi = x\}} = \Phi_n(W_x) \mathbf{1}_{\{\xi = x\}}.$$

Conditioning by  $\mathcal{F}$  yields

$$\mathbf{E}[\Phi_n(X) \mathbf{1}_{\{\xi = x\}} | \mathcal{F}] = \Phi_n(W_x) \mathbf{P}[\xi = x].$$

This formula shows that the random variables  $\Phi(W_x)$  is completely determined (almost surely). By injectivity of  $\Phi$ , the almost sure uniqueness of  $W_x$  follow.

Now, let us prove the existence. First, one checks that  $\mathcal{F} \vee \sigma(\xi)$  is the exactly the set of all events of the form

$$E = \biguplus_{x \in F} (E^x \cap \{\xi = x\}),$$

where  $(E^x)_{x \in F}$  is any family of events in  $\mathcal{F}$ . Observe that if  $E$  is given by this formula then  $E \cap \{\xi = x\} = E^x \cap \{\xi = x\}$  for every  $x \in F$ .

Fix  $x \in F$ . For every  $n \geq 1$ ,  $\{X \in B_n\} \in \mathcal{F} \vee \sigma(\xi)$ , so one can find an event  $E_n^x \in \mathcal{F}$  such that  $\{X \in B_n\} \cap \{\xi = x\} = E_n^x \cap \{\xi = x\}$ . The random variable  $Y^x = (Y_n^x)_{n \geq 1}$  with values in  $C$  and defined by  $Y_n^x = \mathbf{1}_{E_n^x}$  is  $\mathcal{F}$ -measurable.

Fix  $a \in A$ . We can define an  $\mathcal{F}$ -measurable-random variable  $W_x$  with values in  $A$  by

$$\begin{aligned} W_x(\omega) &:= \Phi^{-1}(Y^x(\omega)) \text{ if } Y^x(\omega) \in \Phi(A), \\ W_x(\omega) &:= a \text{ otherwise.} \end{aligned}$$

On the event  $\{\xi = x\}$ , one has  $Y^x = (\mathbf{1}_{E_n^x})_{n \geq 1} = (\mathbf{1}_{\{X \in B_n\}})_{n \geq 1} = \Phi(X) \in \Phi(A)$ , so  $W_x = \Phi^{-1}(\Phi(X)) = X$ . The proof is complete.

## 7.2 Equivalence of Vershik's first level and intermediate properties

In the whole subsection, we fix an  $r$ -adic filtration  $(\mathcal{F}_n)_{n \leq 0}$  and a sequence  $(\xi_n)_{n \leq 0}$  of innovations taking values in a same finite set  $F$  (with size  $r$ ). We fix a Polish metric space  $(A, d)$  and  $X \in L^1(\mathcal{F}_0, A)$ . We call  $(W_n)_{n \leq 0}$  the split-word process associated to the random variable  $X$ , the filtration  $(\mathcal{F}_n)_{n \leq 0}$  and the sequence of innovations  $(\xi_n)_{n \leq 0}$ .

### Proof of proposition 24

Let  $(W'_n, \xi'_n)_{n \geq 0}$  and  $(W''_n, \xi''_n)_{n \geq 0}$  be two independent copies of  $(W_n, \xi_n)_{n \geq 0}$  defined on a same probability space  $(\bar{\Omega}, \bar{\mathcal{A}}, \bar{\mathbf{P}})$ . Call  $(\mathcal{F}'_n)_{n \leq 0}$  and  $(\mathcal{F}''_n)_{n \leq 0}$  the natural filtrations of these two processes.

Fix  $n \leq 0$ . Then  $W'_n = W'_{n-1}(\xi'_n, \cdot)$  and  $W''_n = W''_{n-1}(\xi''_n, \cdot)$  almost surely. Conditionally on  $\mathcal{F}'_{n-1} \vee \mathcal{F}''_{n-1}$  the random variable  $(\xi'_n, \xi''_n)$  is uniform on  $F^2$  since the  $\sigma$ -fields  $\mathcal{F}'_{n-1}$ ,  $\mathcal{F}''_{n-1}$  and the innovations  $\xi'_n, \xi''_n$  are independent. Hence

$$\mathbf{E}[d_n(W'_n, W''_n) | \mathcal{F}'_{n-1} \vee \mathcal{F}''_{n-1}] = \frac{1}{|F|^2} \sum_{(x'_n, x''_n) \in F^2} d_n(W'_{n-1}(x'_n, \cdot), W'_{n-1}(x''_n, \cdot)).$$

But the uniform law on  $F^2$  can be written as an average of uniform laws on graphs of permutations on  $F$ . Hence

$$\begin{aligned} \mathbf{E}[d_n(W'_n, W''_n) | \mathcal{F}'_{n-1} \vee \mathcal{F}''_{n-1}] &\geq \min_{\varsigma \in \mathfrak{S}(F)} \frac{1}{|F|} \sum_{x \in F} d_n(W'_{n-1}(x, \cdot), W'_{n-1}(\varsigma(x), \cdot)) \\ &= d_{n-1}(W'_{n-1}, W'_{n-1}). \end{aligned}$$

Actually, we showed that the process  $(d_n(W'_n, W''_n))_{n \leq 0}$  is a submartingale in the filtration  $(\mathcal{F}'_n \vee \mathcal{F}''_n)_{n \leq 0}$ . Taking expectation yields  $\text{disp}(W_n, d_n) \geq \text{disp}(W_{n-1}, d_{n-1})$ .

Now, let  $(\eta_n)_{n \leq 0}$  be another sequence of innovations of the filtration  $(\mathcal{F}_n)_{n \leq 0}$ . Call  $(W_n^*, \eta_n)_{n \leq 0}$  the split-word process associated to the random variable  $X$ , the filtration  $(\mathcal{F}_n)_{n \leq 0}$  and the sequence of innovations  $(\eta_n)_{n \leq 0}$ . By lemma 16, for each  $n \leq 0$ , one can find an  $\mathcal{F}_{n-1}$ -measurable random permutation  $\Sigma_n$  on  $F$  such that  $\eta_n = \Sigma_n(\xi_n)$ .

Fix  $n \leq -1$ . Let us check that the words  $W_n$  and  $W_n^*$  are in the same orbit under the action of the group  $G_{|n|}$ . For every  $k \in \llbracket -n + 1, 0 \rrbracket$ , the random variable  $\Sigma_n$  is  $\mathcal{F}_n \vee \sigma(\xi_{n+1}, \dots, \xi_{k-1})$ -measurable. By lemma 49, for each  $(x_{n+1}, \dots, x_{k-1}) \in F^{k-n-1}$ , one can find an  $\mathcal{F}_n$ -measurable random variable  $\Sigma_{(x_{n+1}, \dots, x_{k-1})}$  taking values in  $\mathfrak{S}(F)$  such that  $\Sigma_n$  and  $\Sigma_{(x_{n+1}, \dots, x_{k-1})}$  coincide on the event  $\{(\xi_{n+1}, \dots, \xi_{k-1}) = (x_{n+1}, \dots, x_{k-1})\}$ .

Define an  $\mathcal{F}_n$ -measurable random map  $\Sigma$  from the tree  $T_{|n|} = F^0 \cup \dots \cup F^{|n|-1}$  to  $\mathfrak{S}(F)$  by  $\Sigma(x_{n+1}, \dots, x_{k-1}) = \Sigma_{(x_{n+1}, \dots, x_{k-1})}$ . Then

$$(\eta_{n+1}, \dots, \eta_0) = (\Sigma(\xi_{n+1}), \Sigma(\xi_{n+1})(\xi_{n+2}), \dots, \Sigma(\xi_{n+1}, \dots, \xi_{-1})(\xi_0)),$$

Since  $X = W_n(\xi_{n+1}, \dots, \xi_{-1}) = W_n^*(\eta_{n+1}, \dots, \eta_0)$ , we get  $W_n = W_n^* \circ g_\Sigma$  with the notations of subsection 3.3, so  $W_n$  and  $W_n^*$  are on the same orbit modulo the action of the group  $G_{|n|}$ . Hence, taking two independent copies  $(W'_n, W''_n)$  and  $(W''_n, W''_n^*)$  of  $(W_n, W_n^*)$  and defined on a same probability space, we get

$$\text{disp}(W_n) = \mathbf{E}[d_n(W'_n, W''_n)] = \mathbf{E}[d_n(W''_n, W''_n^*)] = \text{disp}(W_n^*).$$

The proof is complete.  $\square$

### Proof of the equivalence of Vershik's first level and intermediate properties

If  $X$  satisfies Vershik's first level property, then given  $\varepsilon > 0$ , there exist an integer  $n \leq 0$ , some innovations  $\eta_{n+1}, \dots, \eta_0$  at times  $n+1, \dots, 0$  with values in  $F$  and some map  $f : F^{|n|} \rightarrow A$  such that the distance between  $X$  and  $\tilde{X} := f(\eta_{n+1}, \dots, \eta_0)$  in  $L^1(\mathcal{F}_0, A)$  is at most  $\varepsilon$ .

Set  $\eta_k = \xi_k$  for every  $k \leq n$ . Call  $(W_k^*, \eta_k)_{k \leq 0}$  the split-word process associated to the random variables  $X$ , the filtration  $(\mathcal{F}_k)_{k \leq 0}$  and the innovations  $(\eta_k)_{k \leq 0}$ . Since  $(\eta_{n+1}, \dots, \eta_0)$  is uniform on  $F^{|n|}$  and independent of  $\mathcal{F}_n$ ,

$$\begin{aligned} \mathbf{E}[d_n(W_n^*, f)] &\leq \mathbf{E}[\delta_n(W_n^*, f)] \\ &= \mathbf{E}\left[\frac{1}{|F|^n} \sum_{(x_{n+1}, \dots, x_0) \in F^{|n|}} d(W_n^*(x_{n+1}, \dots, x_0), f(x_{n+1}, \dots, x_0))\right] \\ &= \mathbf{E}[d(W_n^*(\eta_{n+1}, \dots, \eta_0), f(\eta_{n+1}, \dots, \eta_0))] \\ &= \mathbf{E}[d(X, \tilde{X})] \\ &\leq \varepsilon. \end{aligned}$$

Thus proposition 24 and the triangle inequality yield  $\text{disp}(W_n, d_n) = \text{disp}(W_n^*, d_n) \leq 2\varepsilon$ , so  $\text{disp}(W_m, d_m) \leq 2\varepsilon$  for every  $m \leq n$ . Hence  $X$  satisfies Vershik's intermediate property.

Conversely, assume that  $X$  satisfies Vershik's intermediate property. Given  $\varepsilon > 0$ , there exists some  $n \leq 0$  such that  $\text{disp}(W_n, d_n) \leq \varepsilon$ . Call  $\pi_n$  the law of  $W_n$ . Since

$$\text{disp}(W_n, d_n) = \int_{A^{F^{|n|}}} \mathbf{E}[d_n(W_n, w)] d\pi_n(w),$$

there exists some  $w \in A^{F^{|n|}}$  such that  $d_n(W_n, w) \leq \varepsilon$ . But  $d_n(W_n, w)$  is the minimum of  $\delta_n(W_n, w \circ g_\sigma)$  over all  $\sigma$  in  $G_{|n|}$ , so there exists some  $\mathcal{F}_n$ -measurable random variable  $\Sigma$  taking values in  $G_{|n|}$  such that  $d_n(W_n, w) = \delta_n(W_n, w \circ g_\Sigma)$ : to get such a random variable, fix an arbitrary order on the finite set  $G_{|n|}$  and take the first  $\sigma$  which achieves the minimum above. Let  $(\eta_{n+1}, \dots, \eta_0) = g_\Sigma(\xi_{n+1}, \dots, \xi_0)$  and  $\tilde{X} := w(\eta_{n+1}, \dots, \eta_0)$ . By lemma 16,  $\eta_{n+1}, \dots, \eta_0$  are innovations at times  $n+1, \dots, 0$  and

$$\mathbf{E}[d(X, \tilde{X})] = \mathbf{E}[d(W_n(\xi_{n+1}, \dots, \xi_0), w \circ g_\Sigma(\xi_{n+1}, \dots, \xi_0))] = \mathbf{E}[d_n(W_n, w)] \leq \varepsilon.$$

Hence  $X$  satisfies Vershik's first level property. The proof is complete.  $\square$

### 7.3 Inequalities involving binomial coefficients

**Lemma 50** *Let  $S_n$  be a binomial random variable with parameters  $n \geq 1$  and  $p \in ]0, 1[$ . For every  $\varepsilon \in ]0, p[$ ,*

$$P[S_n \leq n\varepsilon] \leq f_p(\varepsilon)^n \text{ where } f_p(\varepsilon) = \left(\frac{p}{\varepsilon}\right)^\varepsilon \left(\frac{1-p}{1-\varepsilon}\right)^{1-\varepsilon}.$$

Proof. Let  $x \in ]0, 1[$ . Then Markov's inequality yields

$$P[S_n \leq n\varepsilon] \leq P[x^{S_n} \geq x^{n\varepsilon}] \leq x^{-n\varepsilon} \mathbf{E}[x^{S_n}] = (x^{-\varepsilon}(1-p+px))^n.$$

Choosing

$$x = \frac{\varepsilon}{p} \times \frac{1-p}{1-\varepsilon}$$

to minimize the right-hand side yields the desired inequality.

**Lemma 51** *Let  $D \geq 1$  be an integer.*

1. *The map  $k \mapsto \binom{D}{k}$  increases on  $[0, D/2] \cap \mathbf{Z}$  and decreases on  $[D/2, D] \cap \mathbf{Z}$ . The maximum is achieved when  $k = \lfloor D/2 \rfloor$  and when  $k = \lceil D/2 \rceil$ .*
2. *For every  $k \in \llbracket 0, D \rrbracket$ ,*

$$\frac{1}{2^D} \binom{D}{k} \leq \sqrt{\frac{2}{\pi D}}.$$

3. *Fix  $L \geq 1$ . For every  $r \in \mathbf{Z}$ ,*

$$\left| \sum_{q \in \mathbf{Z}} \frac{1}{2^D} \binom{D}{Lq+r} - \frac{1}{L} \right| \leq \sqrt{\frac{2}{\pi D}} \leq \frac{1}{\sqrt{D}},$$

*with the convention  $\binom{D}{k} = 0$  whenever  $k \in \mathbf{Z} \setminus \llbracket 0, D \rrbracket$ .*

Proof. For every  $k \in \llbracket 0, D-1 \rrbracket$ ,

$$\binom{D}{k+1} / \binom{D}{k} = \frac{D-k}{k+1} \begin{cases} > 1 & \text{if } 2k+1 < D \\ = 1 & \text{if } 2k+1 = D \\ < 1 & \text{if } 2k+1 > D \end{cases}.$$

Distinguishing two cases, according to the parity of  $D$ , yields item 1.

For every integer  $n \geq 1$ , set

$$r_n = \sqrt{n} \frac{1}{2^{2n}} \binom{2n}{n} = \sqrt{n} \frac{1}{2^{2n-1}} \binom{2n-1}{n-1}$$

Since

$$\frac{r_{n+1}}{r_n} = \frac{\sqrt{n+1}}{\sqrt{n}} \times \frac{1}{4} \times \frac{(2n+1)(2n+2)}{(n+1)^2} = \frac{n+1/2}{\sqrt{n(n+1)}} > 1,$$

the sequence  $(r_n)_{n \geq 1}$  is increasing. But Stirling's formula shows that it converges to  $1/\sqrt{\pi}$ . Hence  $r_n \leq 1/\sqrt{\pi}$  for every  $n \geq 1$ . Item 2 follows.



For every  $r \in \mathbf{Z}$ , set

$$S_r = \sum_{q \in \mathbf{Z}} p_{Lq+r} \text{ where } p_k = \frac{1}{2^D} \binom{D}{k}.$$

First, let us prove that for every  $r$  and  $s$ ,

$$|S_r - S_s| \leq \sqrt{\frac{2}{\pi D}}.$$

By symmetry, one needs only to bound above  $S_r - S_s$ . Using the  $L$ -periodicity of the map  $k \mapsto p_k$  and the symmetry  $p_{D-k} = p_k$  for every  $k \in \mathbf{Z}$ , one may assume that  $r \leq s \leq D/2 < r + L \leq s + L$ . In this case, one has

$$\begin{aligned} S_s - S_r &= \sum_{q \leq 0} (p_{Lq+s} - p_{Lq+r}) + \sum_{q \geq 1} (p_{Lq+s} - p_{Lq+L+r}) - p_{L+r} \\ &\geq 0 + 0 - \sqrt{\frac{2}{\pi D}}, \end{aligned}$$

by item 1 and 2. The desired upper bound of  $|S_r - S_s|$  and item 3 follow, by taking the mean over all  $s \in \llbracket 0, L-1 \rrbracket$ , since  $S_0 + \dots + S_{L-1} = 1$ .

## Acknowledgements

Many thanks to Stéphane Laurent, who provided a substantial help to understand Heicklen and Hoffman's proof, who checked and improved the drafting of section 5 and provide all the figures of the paper. Many thanks also to Michel Emery and to Jean Brossard for their careful reading of large parts of this paper. I also thank Jean Paul Thouvenot and the referee for their suggestions which enrich the paper and Thierry de la Rue for stimulating conversations on this subject.

## References

- [1] T. Austin, *Measure concentration and the weak Pinsker property* Publications Mathématiques de l'IHÉS **128**-1, 1–119 (2018).
- [2] G. Ceillier, *The filtration of the split-word process*. Probability Theory and Related Fields **153**, 269–292 (2012).
- [3] I. Cornfeld, S. Fomin, Y. Sinai, *Ergodic theory*. Translated from the Russian by A. B. Sosinskii. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 245 (1982).
- [4] M. Émery, W. Schachermayer, *On Vershik's standardness criterion and Tsirelson's notion of cosiness*. Séminaire de Probabilités XXXV (Springer Lectures Notes in Math. 1755) 265–305, (2001).
- [5] J. Feldman, D. Rudolph, *Standardness of sequences of  $\sigma$ -fields given by certain endomorphisms*. Fundamenta Mathematicae. 157, 175–189 (1998).
- [6] E. Glasner, *Ergodic theory via joinings*. American Mathematical Society (2003).

- [7] D. Heicklen, C Hoffman,  $[T, T^{-1}]$  is not standard. Ergodic Theory and Dynamical Systems **18**, no. 4, 875–878 (1998).
- [8] C. Hoffman, A zero entropy  $T$  such that the  $[T, \text{Id}]$  endomorphism is nonstandard. Proceedings of the American Mathematical Society **128**, no. 1, 183–188 (2000).
- [9] F. den Hollander, J. Steif, Mixing properties of the generalised  $[T, T^{-1}]$ -process, Journal d'Analyse Mathématique **72-1**, 165–202 (1997).
- [10] C. Hoffman, D. Rudolph Uniform endomorphisms which are isomorphic to a Bernoulli shift. Annals of Mathematics, Second Series, **156**, No. 1, pp. 79-101 (2002).
- [11] S. Laurent: *Filtrations à temps discret négatif*. PhD Thesis, Université de Strasbourg, Strasbourg (2004).
- [12] S. Laurent, *On standardness and I-cosiness*. Séminaire de Probabilités XLIII, Springer Lecture Notes in Mathematics 2006, 127–186 (2010).
- [13] S. Laurent: *On Vershikian and I-cosy random variables and filtrations*. Teoriya Veroyatnostei i ee Primeneniya **55**, 104–132 (2010). Also published in: Theory Probab. Appl. **55**, 54–76 (2011).
- [14] S. Laurent, *Vershik's Intermediate Level Standardness Criterion and the Scale of an Automorphism*. Séminaire de Probabilités XLV, Springer Lecture Notes in Mathematics 2078, 123–139 (2013).
- [15] C. Leuridan, *Filtration d'une marche aléatoire stationnaire sur le cercle*. Séminaire de Probabilités XXXVI, Springer Lecture Notes in Mathematics 1801, 335–347 (2002).
- [16] E. Lindenstrauss, Y. Peres, W. Schlag *Bernoulli convolutions and an intermediate value theorem for entropies of  $K$ -partitions* Journal d'Analyse Mathématique **87-1**, 337–367 (2002).
- [17] I. Meilijson, *Mixing properties of a class of skew-products*. Israel J. Math. **19**, 266–270 (1974).
- [18] K. Petersen, *Ergodic theory*. Corrected reprint of the 1983 original. Cambridge University Press (1989).
- [19] T. de la Rue, *Espaces de Lebesgue*. Séminaire de probabilités, **27**, Lecture Notes in Mathematics 1557, 15–21 (1993).
- [20] P.C. Shields, *The ergodic theory of discrete sample paths*. Graduate Studies in Mathematics, 13. American Mathematical Society, (1996).
- [21] M. Smorodinsky, *Processes with no standard extension*. Israël Journal of Mathematics **107**, 327–331 (1998).
- [22] J. E. Steif, *The  $T, T^{-1}$ -process, finitary codings and weak Bernoulli*. Israel Journal of Mathematics, **125**, 29–43 (2001).
- [23] A.M. Vershik, *Decreasing sequences of measurable partitions, and their applications*. Dokl. Akad. Nauk SSSR, **193**, 748–751 (1970). English translation: Soviet mathematics - Doklady, **11**, 1007–1011 (1970).

- [24] A.M. Vershik, *Approximation in measure theory* (in Russian). PhD Thesis, Leningrad University, Leningrad (1973).
- [25] A.M. Vershik, *The theory of decreasing sequences of measurable partitions*. *Algebra i Analiz* **6**:4, 1–68 (1994) (in Russian). English translation: *St. Petersburg Mathematical Journal* **6**:4, 705–761 (1995).
- [26] A.M. Vershik, *Dynamic theory of growth in groups: Entropy, boundaries, examples*. *Russian Mathematical Surveys* **55**-4, 667-733 (2000).