



**HAL**  
open science

## Uniform entropy scalings of filtrations

Stéphane Laurent

► **To cite this version:**

| Stéphane Laurent. Uniform entropy scalings of filtrations. 2016. hal-01006337v5

**HAL Id: hal-01006337**

**<https://hal.science/hal-01006337v5>**

Preprint submitted on 16 Aug 2016

**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.

# Uniform entropy scalings of filtrations

Stéphane Laurent

August 16, 2016

## Abstract

We study Vershik and Gorbulsky's notion of entropy scalings for filtrations in the particular case when the scaling is not  $\epsilon$ -dependent, and is then termed as *uniform scaling*. Among our main results, we prove that the scaled entropy of the filtration generated by the Vershik progressive predictions of a random variable is equal to the scaled entropy of this random variable. Standardness of a filtration is the case when the scaled entropy with a constant scaling is zero, thus our results generalize some known results about standardness. As a case-study we consider a family of next-jump time filtrations. We also provide some results about the entropy of poly-adic filtrations, rephrasing or generalizing some old results.

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Vershik's standardness criterion</b>	<b>2</b>
2.1	Vershik's standardness criterion . . . . .	3
2.2	Properties to be generalized later . . . . .	4
2.3	Vershik's standardness criterion in practice . . . . .	5
<b>3</b>	<b>The next-jump time filtrations</b>	<b>7</b>
3.1	Next-jump time process as a random walk on a Bratteli graph . . . . .	7
3.2	Standardness of $\mathcal{F}$ using Vershik's criterion . . . . .	9
3.3	Iterated Kantorovich distances . . . . .	11
<b>4</b>	<b>The uniformly scaled entropy</b>	<b>12</b>
4.1	Definitions . . . . .	13
4.2	Main properties and main theorem . . . . .	16
4.3	Extensions that do not increase entropy . . . . .	20
<b>5</b>	<b>Entropy of next-jump time filtrations</b>	<b>22</b>
<b>6</b>	<b>Entropy of poly-adic filtrations</b>	<b>23</b>
6.1	The $\pi_n X$ in poly-adic filtrations and the exponential entropy . . . . .	25
6.2	Gorbulsky's theorem . . . . .	25
6.3	$(\nabla)$ -adic filtrations have zero exponential entropy . . . . .	29
6.4	$(\neg\Delta)$ -adic filtrations have zero exponential entropy . . . . .	30
6.5	$(\Delta)$ -adic split-word filtrations . . . . .	31
6.6	Dyadic filtrations of unordered pairs . . . . .	33

# 1 Introduction

This is the first paper about Vershik & Gorbulsky's theory of the entropy of filtrations written in the probabilistic language. It deals with the *scaled entropy* introduced in [21]. Our results focus on the case of *uniform entropy scalings*, and, because standardness is equivalent to zero entropy with a constant entropy scaling, they generalize the main properties about standardness.

In Section 2 we recall the definition of Vershik's standardness criterion. We use this criterion in Section 3 to give a new proof of the standardness criterion for the family of next-jump time filtrations studied in [13] (where I-cosiness was used to derive this criterion). In Section 4 we introduce the scaled entropy with uniform scalings. In Section 5 we pursue the work of Section 3 by studying uniform entropy scalings for the next-jump time filtrations; in fact, we just use our results to show that this problem comes down to the scaled entropy of a discrete measure which is studied in [15]. Section 6 deals with the exponential entropy for poly-adic filtrations. In this section, we mainly rephrase some old theorems by Vershik and, with a slight generalization, a theorem by Gorbulsky about the coincidence between the scaled entropy and the exponential entropy.

The relevance of the results provided by this paper is twofold. First, the general properties about standardness become particular cases of the general properties about the uniformly scaled entropy. Among these properties, Theorem 4.11 is one of the main results. It states that the scaled entropy of the filtration generated by the Vershik progressive predictions of a random variable equals the scaled entropy of this random variable, thereby considerably reducing the task of calculating the entropy of a filtration. The power of this theorem is well illustrated on the calculation of the scaled entropy for the next-jump time filtrations. Second, our results provide new knowledge about the known examples of non-standard filtrations. The results of Section 6 providing the scaled entropy of some non-standard split-word filtrations with the exponential scaling, are not new. When this scaled entropy has not the same value for two such filtrations, one can conclude that these two filtrations are not isomorphic. But thanks to our results of Section 4.3, we learn something more, namely that it is not possible to embed the filtration having the smallest entropy in, for example, an independent enlargement of the other one.

## 2 Vershik's standardness criterion

In the probabilistic literature, *standardness* of a filtration  $\mathcal{F} = (\mathcal{F}_n)_{n \leq 0}$  in discrete negative time is usually defined as the possibility to embed  $\mathcal{F}$  in the filtration generated by a sequence of independent random variables (see [4, 10, 11, 12]). As long as the final  $\sigma$ -field  $\mathcal{F}_0$  is essentially separable, standardness is known to be equivalent to Vershik's standardness criterion. In the present paper, we say that a filtration is Vershikian if it satisfies Vershik's standardness criterion, and we say that a filtration is standard if it is Vershikian and its final  $\sigma$ -field is essentially separable.

In this section we recall the statement of Vershik's standardness criterion and we state its main properties which are proved in [11]. In Section 4 we will see that these properties are particular cases of our results about the scaled entropy.

## 2.1 Vershik's standardness criterion

The *Kantorovich* distance plays a major role in the statement of Vershik's standardness criterion, as well as in the definition of the entropy. Given a separable metric space  $(E, \rho)$ , the Kantorovich distance  $\rho'(\mu, \nu)$  between two probability measures  $\mu$  and  $\nu$  is defined by

$$\rho'(\mu, \nu) = \inf_{\Lambda \in \mathcal{J}(\mu, \nu)} \iint \rho(x, y) d\Lambda(x, y),$$

where  $\mathcal{J}(\mu, \nu)$  is the set of joinings of  $\mu$  and  $\nu$ , that is, the set of probabilities on  $E \times E$  whose first and second marginal measures are  $\mu$  and  $\nu$  respectively. In general,  $\rho'(\mu, \nu)$  is possibly infinite, but  $\rho'$  defines a distance on the space  $E'$  of integrable probability measures on  $(E, \rho)$ , when saying that a probability measure  $\mu$  on  $(E, \rho)$  is integrable if the random variables  $X \sim \mu$  satisfy  $\mathbb{E}[\rho(X, x)] < \infty$  for some (  $\iff$  for every) point  $x \in E$ , and such a random variable  $X$  is also said to be integrable. When  $E$  is compact then every  $E$ -valued random variable is integrable. In general, the topology induced by  $\rho'$  on  $E'$  is finer than the topology of weak convergence, but they coincide when  $(E, \rho)$  is compact, and  $(E', \rho')$  is itself compact in this case. We mainly use the fact that the metric space  $(E', \rho')$  is complete and separable whenever  $(E, \rho)$  is (see e.g. [2]).

In order to state Vershik's standardness criterion, one has to introduce the *Vershik progressive predictions*  $\pi_n X$  of a random variable  $X$  (corresponding to the so-called *universal projectors*, or *tower of measures*, in [17] and [20]) and the iterated Kantorovich distance  $\rho^{(n)}$  on the state space  $E^{(n)}$  of  $\pi_n X$ . Let  $(E, \rho)$  be a Polish metric space. For a  $\sigma$ -field  $\mathcal{B}$  we denote by  $L^1(\mathcal{B}; E)$  the space of integrable  $E$ -valued  $\mathcal{B}$ -measurable random variables. Let  $\mathcal{F}$  be a filtration, and  $X \in L^1(\mathcal{F}_0; E)$ . The Vershik progressive predictions  $\pi_n X$  of  $X$  with respect to  $\mathcal{F}$  are recursively defined as follows: we put  $\pi_0 X = X$ , and  $\pi_{n-1} X = \mathcal{L}(\pi_n X | \mathcal{F}_{n-1})$  (the conditional law of  $\pi_n X$  given  $\mathcal{F}_{n-1}$ ). Since  $X$  is integrable, for any  $x \in E$  the conditional expectation  $\mathbb{E}[\rho(X, x) | \mathcal{F}_{-1}]$  is finite, therefore  $\rho'(\mathcal{L}(X | \mathcal{F}_{-1}), \delta_x) < \infty$  and thus the conditional law  $\mathcal{L}(X | \mathcal{F}_{-1}) = \pi_{-1} X$  is integrable. Thus, by a recursive reasoning, the  $n$ -th progressive prediction  $\pi_n X$  is a random variable taking its values in the Polish space  $E^{(n)}$  recursively defined by  $E^{(0)} = E$  and  $E^{(n-1)} = (E^{(n)})'$ , denoting as before by  $E'$  the space of integrable probability measures on any separable metric space  $E$ . Note that  $(\pi_n X)_{n \leq 0}$  is a Markov process. The state space  $E^{(n)}$  of  $\pi_n X$  is Polish when endowed with the distance  $\rho^{(n)}$  obtained by iterating  $|n|$  times the construction of the Kantorovich distance starting with  $\rho$ : we recursively define  $\rho^{(n)}$  by putting  $\rho^{(0)} = \rho$  and by defining  $\rho^{(n-1)} = (\rho^{(n)})'$  as the Kantorovich distance issued from  $\rho^{(n)}$ .

Finally, in order to state Vershik's standardness criterion, one introduces the *dispersion*  $\text{disp } X$  of (the law of) an integrable random variable  $X$  in a Polish metric space  $(E, \rho)$ . It is defined as the expectation of  $\rho(X', X'')$  where  $X'$  and  $X''$  are two independent copies of  $X$ , that is, two independent random variables defined on the same probability space and having the same law as  $X$ . Now, Vershik's standardness criterion is defined as follows. Let  $\mathcal{F}$  be a filtration, let  $E$  be a Polish metric space and  $X \in L^1(\mathcal{F}_0; E)$ . We say that the random variable  $X$  satisfies the *Vershik property*, or, for short, that  $X$  is *Vershikian* (with respect to  $\mathcal{F}$ ) if  $\text{disp } \pi_n X \rightarrow 0$  as  $n$  goes to  $-\infty$ . Then we extend this definition to  $\sigma$ -fields  $\mathcal{E}_0 \subset \mathcal{F}_0$  and to the whole filtration as follows: we say that a  $\sigma$ -field  $\mathcal{E}_0 \subset \mathcal{F}_0$  is *Vershikian* if each random variable  $X \in L^1(\mathcal{E}_0; [0, 1])$  is Vershikian, and we say that the filtration  $\mathcal{F}$  is *Vershikian*, or that  $\mathcal{F}$  satisfies *Vershik's standardness*

criterion, if the final  $\sigma$ -field  $\mathcal{F}_0$  is Vershikian.

When  $\mathcal{F}$  is immersed in a bigger filtration  $\mathcal{G}$  and  $X \in L^1(\mathcal{F}_0; E) \subset L^1(\mathcal{G}_0; E)$ , it is important to note that the iterated conditional law  $\pi_n X$  is the same considering either  $\mathcal{F}$  or  $\mathcal{G}$  as the underlying filtration. We refer to [4] or [10] for details about the immersion property. Consequently, in such a situation, there is no ambiguity in considering the Vershik property without specifying the underlying filtration.

Later, we will use the two following lemmas about the Vershik progressive predictions. The proof of the first one is straightforward from the definitions and we leave it to the reader.

**Lemma 2.1.** *For any Polish space  $(E, \rho)$  and  $X, Y \in L^1(\mathcal{F}_0; E)$ , the stochastic process  $(\rho^{(n)}(\pi_n X, \pi_n Y))_{n \leq 0}$  is a submartingale. In particular the expectation  $\mathbb{E}[\rho^{(n)}(\pi_n X, \pi_n Y)]$  is increasing with  $n$ .*

**Lemma 2.2.** *Let  $\mathcal{F}$  be a filtration. Let  $(E, \rho)$  and  $(\tilde{E}, \tilde{\rho})$  be two Polish metric spaces, and  $f: E \rightarrow \tilde{E}$  a measurable function. If  $X$  is a  $\mathcal{F}_0$ -measurable random variable taking its values in  $E$ , then  $\pi_n(f(X)) = f^n(\pi_n X)$  for some measurable function  $f^n: E^{(n)} \rightarrow \tilde{E}^{(n)}$ , and the two following properties hold:*

- $f^n$  is  $K$ -Lipschitz if  $f$  is  $K$ -Lipschitz;
- $f^n$  is an isometry if  $f$  is an isometry.

*Proof.* The function  $f^n$  is inductively defined by  $f^0 = f$  and  $f^{n-1}(\mu) = f^n \star \mu$  (image measure). It is elementary to check the three claims asserted in the lemma.  $\square$

## 2.2 Properties to be generalized later

Throughout this article, we denote by  $V(X)$  the Vershik property for an integrable random variable  $X$ , when an underlying ambient filtration  $\mathcal{F}$  is understood. We also denote by  $V(\mathcal{E}_0)$  the Vershik property for a  $\sigma$ -field  $\mathcal{E}_0 \subset \mathcal{F}_0$ . We will see in Section 4 that  $V(X)$  can be equivalently stated as  $h_c(X) = 0$  where  $h_c$  is the scaled entropy of  $X$  with a constant scaling function  $c$ . Then our results in Section 4 about the uniformly scaled entropy generalize the following propositions and theorem which are provided in [11].

**Proposition 2.3.** *Let  $\mathcal{F}$  be a filtration,  $n_0 \leq 0$  be an integer, and denote by  $\mathcal{F}^{n_0] = (\mathcal{F}_{n_0+n})_{n \leq 0}$  the filtration  $\mathcal{F}$  truncated at  $n_0$ . Then  $\mathcal{F}^{n_0]}$  is Vershikian if and only if  $\mathcal{F}$  is Vershikian.*

**Proposition 2.4.** *a) If  $(\mathcal{B}_k)_{k \geq 1}$  is an increasing sequence of sub- $\sigma$ -fields of  $\mathcal{F}_0$  then*

$$[\forall k \geq 1, V(\mathcal{B}_k)] \implies V\left(\bigvee_{k \geq 1} \mathcal{B}_k\right).$$

*b) For any Polish metric space  $(E, \rho)$  and  $X \in L^1(\mathcal{F}_0; E)$ ,*

$$V(X) \iff V(\sigma(X)).$$

**Theorem 2.5.** *For any  $X \in L^1(\mathcal{F}_0; E)$ , the filtration  $\mathcal{F}^X$  generated by the Markov process  $(\pi_n X)_{n \leq 0}$  is standard if and only if the random variable  $X$  satisfies the Vershik property.*

Proposition 2.3 is a consequence of Corollary 4.18. Proposition 2.4 is a consequence of Proposition 4.15 and Proposition 4.16. Theorem 2.5 is a consequence of Theorem 4.11.

We will use the two propositions and the theorem above in Section 3, before proving their generalization, and this is why we state them here. But we will also provide generalizations of two other results: Theorem 4.19 generalizes the fact that a parametric extension of a Vershikian filtration is still a Vershikian filtration, and Theorem 4.22 generalizes the fact that the independent product of two Vershikian filtrations is a Vershikian filtration.

### 2.3 Vershik's standardness criterion in practice

Vershik's standardness criterion may appear puzzling and complicated at first glance: calculating the progressive predictions  $\pi_n X$  and the iterated Kantorovich distance  $\rho^{(n)}$  on the strange state space of  $\pi_n X$  do not appear easily practicable.

First note that  $V(X)$  does not depend on the choice of the Polish space  $E$  in which  $X$  takes its values: this stems from the second claim of Proposition 2.4. Also note the importance of Theorem 2.5:  $V(X)$  is equivalent to standardness of the filtration  $\mathcal{F}^X$  generated by the Markov process  $(\pi_n X)_{n \leq 0}$ . Thus, if we intend to show that standardness of  $\mathcal{F}$  holds true, our task is reduced to only show  $V(X)$  if we find  $X$  such that  $\mathcal{F}^X = \mathcal{F}$ .

Observe that any filtration  $\mathcal{F}$  having an essentially separable final  $\sigma$ -field  $\mathcal{F}_0$  can always be generated by a Markov process  $(X_n)_{n \leq 0}$ : just take for  $X_n$  any random variable generating the  $\sigma$ -field  $\mathcal{F}_n$  for every  $n \leq 0$ . Vershik's standardness criterion can be rephrased to a more practical criterion by considering such a Markov process  $(X_n)_{n \leq 0}$ , as we explain below and summarize in Lemma 2.6; but practicality of the rephrased criterion depends on the choice of the generating Markov process. Firstly, thanks to this lemma, the strange state spaces of Vershik's progressive predictions  $\pi_n X$  can be avoided when  $X$  is one of the random variable  $X_k$  of the Markov process  $(X_n)_{n \leq 0}$ . Let us explain this claim for  $X = X_0$ , which is enough to understand. Denote by  $A_n$  the state space of  $X_n$  for every  $n \leq 0$ . Starting with a compact metric  $\rho_0$  on  $A_0$ , we recursively define a pseudometric  $\rho_n$  on the state space of  $X_n$  by setting

$$\rho_n(x_n, x'_n) = (\rho_{n+1})'(\mathcal{L}(X_{n+1} | X_n = x_n), \mathcal{L}(X_{n+1} | X_n = x'_n))$$

where  $(\rho_{n+1})'$  is the Kantorovich pseudometric derived from  $\rho_{n+1}$ . The  $\rho_n$  are more friendly than the  $\rho^{(n)}$  appearing in Vershik's standardness criterion, and Lemma 2.6 states that there are some maps  $\psi_n: A_n \rightarrow A_0^{(n)}$  such that  $\pi_n X_0 = \psi_n(X_n)$  and

$$\rho^{(n)}(\psi_n(x_n), \psi_n(x'_n)) = \rho_n(x_n, x'_n)$$

for every  $x_n, x'_n \in A_n$ . Thus, in order for the Vershik property  $V(X_0)$  to hold true, it suffices that  $\rho_n(X'_n, X''_n) \rightarrow 0$  in  $L^1$  where  $X'_n$  and  $X''_n$  are two independent copies of  $X_n$ . Moreover, Lemma 2.6 states that  $\mathcal{F}^{X_0} = \mathcal{F}$  under the identifiability condition

$$\forall n \leq 0, \forall x_n, x'_n \in A_n, \quad [x_n \neq x'_n] \implies [\mathcal{L}(X_{n+1} | X_n = x_n) \neq \mathcal{L}(X_{n+1} | X_n = x'_n)] \quad (\star)$$

and then, by Theorem 2.5 standardness of  $\mathcal{F}$  is equivalent to  $V(X_0)$  under this condition.

**Lemma 2.6.** *Let  $\mathcal{F}$  be the filtration generated by a Markov process  $(X_n)_{n \leq 0}$ . Denote by  $A_n$  the state space of  $X_n$  for every  $n \leq 0$  and assume that  $A_0$  is a compact metric space under some metric  $\rho_0$ . Consider the pseudometrics  $\rho_n$  introduced above and the iterated Kantorovich metrics  $\rho^{(n)}$  appearing in Vershik's standardness criterion.*

1) There are some maps  $\psi_n: A_n \rightarrow A_0^{(n)}$  such that  $\pi_n X_0 = \psi_n(X_n)$  and

$$\rho^{(n)}(\psi_n(x_n), \psi_n(x'_n)) = \rho_n(x_n, x'_n)$$

for every  $x_n, x'_n \in A_n$  and every  $n \leq 0$ .

2) The Vershik property  $V(X_0)$  is equivalent to  $\mathbb{E}[\rho_n(X'_n, X''_n)] \rightarrow 0$  where  $X'_n$  and  $X''_n$  are two independent copies of  $X_n$ .

3) Under the identifiability condition  $(\star)$ , the  $\rho_n$  are metrics and the  $\psi_n$  are some isometries. Consequently  $\mathcal{F}$  is generated by the process  $(\pi_n X_0)_{n \leq 0}$ .

4) Under the identifiability condition  $(\star)$ , the Vershik property  $V(X_0)$  is equivalent to standardness of  $\mathcal{F}$ .

*Proof.* Obviously  $\pi_0 X_0$  is a  $\sigma(X_0)$ -measurable random variable, and  $\pi_n X_0 = \mathcal{L}(\pi_{n+1} X_0 | \mathcal{F}_n)$  for  $n < 0$  is a  $\sigma(X_n)$ -measurable random variable by the Markov property. Therefore, for each  $n \leq 0$ , the Doob-Dynkin lemma provides a measurable function  $\psi_n$  for which  $\pi_n X_0 = \psi_n(X_n)$ , and  $\psi_0$  is nothing but the identity map. The equality in 1), relating  $\rho^{(n)}$  and  $\rho_n$ , is obviously true for  $n = 0$ . Assuming  $\rho^{(n+1)}(\psi_{n+1}(x_{n+1}), \psi_{n+1}(x'_{n+1})) = \rho_{n+1}(x_{n+1}, x'_{n+1})$ , then the Kantorovich distance  $\rho_n(x_n, x'_n)$  is given by

$$\rho_n(x_n, x'_n) = \inf_{\Lambda_{x_n, x'_n}} \int \rho^{(n+1)}(\psi_{n+1}(x_{n+1}), \psi_{n+1}(x'_{n+1})) d\Lambda_{x_n, x'_n}(x_{n+1}, x'_{n+1}),$$

where the infimum is taken over all joinings  $\Lambda_{x_n, x'_n}$  of  $\mathcal{L}(X_{n+1} | X_n = x_n)$  and  $\mathcal{L}(X_{n+1} | X_n = x'_n)$ , and then  $\rho_n(x_n, x'_n)$  is also given by

$$\rho_n(x_n, x'_n) = \inf_{\Theta_{x_n, x'_n}} \int \rho^{(n+1)}(y_{n+1}, y'_{n+1}) d\Theta_{x_n, x'_n}(y_{n+1}, y'_{n+1}),$$

where the infimum is taken over all joinings  $\Theta_{x_n, x'_n}$  of  $\mathcal{L}(\pi_{n+1} X_0 | X_n = x_n) = \psi_n(x_n)$  and  $\mathcal{L}(\pi_{n+1} X_0 | X_n = x'_n) = \psi_n(x'_n)$ , thereby showing  $\rho^{(n)}(\psi_n(x_n), \psi_n(x'_n)) = \rho_n(x_n, x'_n)$ . That shows 1), and 2) obviously follows.

The claim about the  $\rho_n$  in 3) is recursively shown too. It suffices to show that every  $\psi_n$  is injective. Assuming that  $\psi_{n+1}$  is injective and assuming  $\mathcal{L}(X_{n+1} | X_n = x_n) \neq \mathcal{L}(X_{n+1} | X_n = x'_n)$ , then, obviously,

$$\mathcal{L}(\psi_{n+1}(X_{n+1}) | X_n = x_n) \neq \mathcal{L}(\psi_{n+1}(X_{n+1}) | X_n = x'_n),$$

that is,  $\psi_n(x_n) \neq \psi_n(x'_n)$ , thereby showing 3). Finally, claim 4) stems from Theorem 2.5.  $\square$

Obviously we can similarly state Lemma 2.6 for  $X_k$  instead of  $X_0$ , for any  $k \leq 0$ . When the identifiability condition  $(\star)$  does not hold, one can apply proposition 6.2 of [11], which claims that, in order to prove standardness of  $\mathcal{F}$ , it is sufficient to check that  $V(X_k)$  holds true for every  $k \leq 0$ . .

### 3 The next-jump time filtrations

In Section 5 we will study the scaled entropy of the next-jump time filtrations which are introduced in this section. Standardness of these filtrations has been characterized in [13] with the help of the I-cosiness criterion. In this section we provide a new proof of this characterization with the help of Vershik’s standardness criterion (Section 2.1). More precisely, we will be in the context of Lemma 2.6 and the identifiability condition  $(\star)$  will be fulfilled, and thus our main task will be to derive the metrics  $\rho_n$  of this lemma. This will be achieved in Section 3.2, after we introduce the next-jump filtrations in Section 3.1 as the filtrations generated by some random walks on the vertices of a Bratteli graph (shown on Figure 1).

#### 3.1 Next-jump time process as a random walk on a Bratteli graph

Our presentation of the next-jump time filtrations differs from the one given in [13]. Here we define these filtrations as those generated by a Markov process on the vertices of a Bratteli graph.

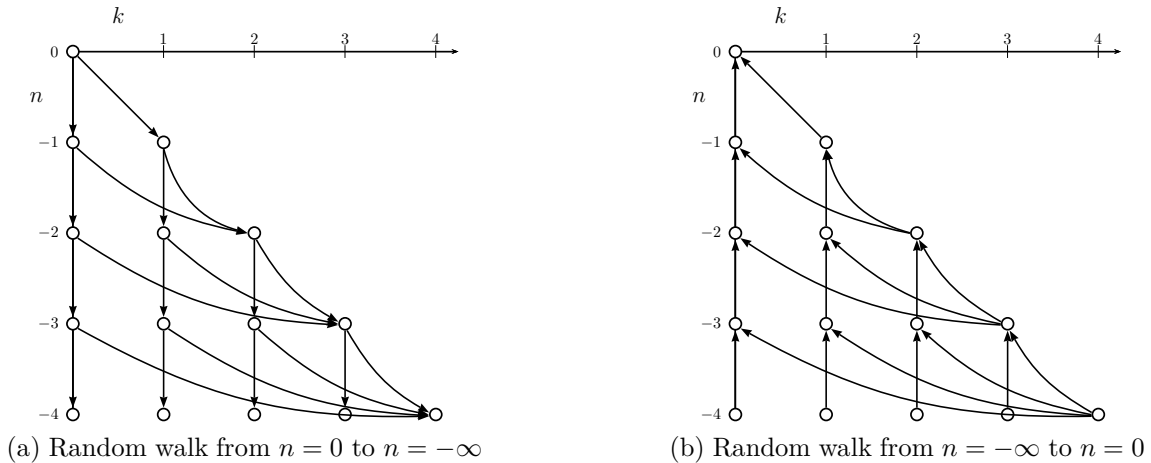


Figure 1: Next-jump time process as a random walk.

Let  $B$  be the  $(-\mathbb{N})$ -graded Bratteli graph shown on Figure 1. At each level  $n$ , there are  $|n| + 1$  vertices labeled by  $k \in \{0, \dots, |n|\}$ , and the vertex labeled by  $k$  is connected to the two vertices at level  $n - 1$  labeled by  $k$  and  $|n| + 1$ . A path on  $B$  is a sequence  $(\gamma_n)_{n \leq 0}$  consisting of edges  $\gamma_n$  such that  $\gamma_n$  connects a vertex at level  $n$  to a vertex at level  $n - 1$  for every  $n \leq 0$ . The set of paths is denoted by  $\Gamma_B$ . When a path is taken at random in  $\Gamma_B$  we denote by  $V_n$  the label of the selected vertex at level  $n$  (thus  $V_0 = 0$ ) and we are interested in the filtration  $\mathcal{F}$  generated by the process  $(V_n)_{n \leq 0}$ . Since this causes no possible confusion, we identify a vertex to its label. We study the case when the process  $(V_0, V_{-1}, \dots)$  is the Markov chain whose transition distributions are defined from a given  $[0, 1]$ -valued sequence  $(p_n)_{n \leq 0}$  satisfying  $p_0 = 1$ , by

$$\mathcal{L}(V_n | V_{n+1} = k) = (1 - p_n)\delta_k + p_n\delta_{|n|},$$

that is to say, given  $V_{n+1}$ , the vertex  $V_n$  is one of the two vertices connected to  $V_{n+1}$  and equals the extreme vertex  $|n|$  with probability  $p_n$ .



In other words, if we consider that the set of paths  $\Gamma_B$  is  $\{0, 1\}^{-\mathbb{N}}$  by labelling the edges connecting a vertex  $v_n$  at level  $n$  to the vertex  $v_{n-1}$  at level  $n-1$  by 0 if  $v_{n-1}$  and  $v_n$  have the same label and by 1 if  $v_{n-1}$  is labeled by  $|n|+1$ , then we are interested in the case when the paths are taken at random according to the independent product measure  $\otimes_{n \leq -1} (1-p_n, p_n)$  by denoting by  $(1-p, p)$  the Bernoulli probability measure with probability of success  $p$ .

The time-directed process  $(V_n)_{n \leq 0}$  is Markovian too. The next-jump time process  $(Z_n)_{n \leq 0}$  defined in [13] is obtained from  $V_n$  by putting  $Z_0 = 0$  and  $Z_n = -V_{n+1}$  for  $n \leq -1$ . Hence the filtration  $\mathcal{F}$  generated by the Markov process  $(V_n)_{n \leq 0} = (Z_{n-1})_{n \leq 0}$  shares the same standardness status as the one studied in [13] because standardness is an asymptotic property (Proposition 2.3).

It is easy to see that  $\Pr(V_n = |n|) = p_n$ . We will say that the  $p_n$  are the *jumping probabilities* because one also has  $p_n = \Pr(V_{n+1} \neq V_n)$  for every  $n < 0$ . It is shown in [13] that

$$\Pr(V_n = |k|) = (1-p_n) \cdots (1-p_{k-1})p_k \quad \text{if } 0 \leq |k| < |n|,$$

and the transitions kernels  $P_n(v, \cdot)$  from  $n-1$  to  $n$  are given by

$$P_n(v, \cdot) := \mathcal{L}(V_n | V_{n-1} = v) = \begin{cases} \delta_v & \text{if } 0 \leq v < |n| + 1 \\ \mathcal{L}(V_n) & \text{if } v = |n| + 1 \end{cases}. \quad (3.1)$$

Obviously the identifiability condition  $(\star)$  defined in Section 2.3 cannot hold for  $(V_n)_{n \leq 0}$  because  $V_0 = 0$  is degenerate. But we will see in Lemma 3.3 that this condition holds for the process truncated at  $-1$  when  $p_{-1} \in ]0, 1[$  and  $p_n < 1$  for every  $n \leq -2$ .

An important particular case is the one when  $p_n = (|n|+1)^{-1}$ . In this case,  $V_n$  has the uniform distribution on  $\{0, \dots, |n|\}$  for every  $n \leq 0$  and the filtration  $\mathcal{F}$  generated by  $(V_n)_{n \leq 0}$  is Kolmogorovian and not standard in this case. This results from the standardness criterion provided by Theorem 3.7, which was proved in [13] with the help of the I-cosiness criterion, and which is proved in the present paper with the help of Vershik's criterion.

When  $p_n < 1$ , the law of  $V_{n+1}$  is the law of  $V_n$  conditioned on  $\{0, \dots, |n|-1\}$ . Thus, when  $p_n < 1$  for every  $n \leq -1$ , the law of  $V_n$  can be represented as the truncation of a measure  $\mu$  on  $\mathbb{N}$ . For example,  $\mu$  is the counting measure in the uniform case  $p_n = (|n|+1)^{-1}$ . When the  $p_n$  are given, this measure is given by

$$\mu(-n) = \frac{p_n}{\prod_{k=n}^{-1} (1-p_k)} = \frac{\Pr(V_n = n)}{\Pr(V_n = 0)}$$

for every  $n \leq 0$ , and

$$\mu(\{0, \dots, -n\}) = \frac{1}{\prod_{k=n}^{-1} (1-p_k)} = \frac{1}{\Pr(V_n = 0)},$$

thus  $\mu$  is normalizable if and only if  $\sum p_n < \infty$ . In this case the law of  $V_n$  goes to the normalized version of  $\mu$  and  $\mathcal{F}$  is not Kolmogorovian, and in the other case  $V_n$  goes to  $\infty$  and  $\mathcal{F}$  is Kolmogorovian. This is due to the following proposition about the tail  $\sigma$ -field  $\mathcal{F}_{-\infty}$ , which is a rewriting of proposition 3.1 in [13], to which we refer for a detailed proof.

**Proposition 3.1.** *The sequence  $(V_n)_{n \leq 0}$  goes to a random variable  $V_{-\infty}$  when  $n$  goes to  $-\infty$ , and the tail  $\sigma$ -field  $\mathcal{F}_{-\infty}$  is generated by  $V_{-\infty}$ . There are three possible situations:*

1) if  $\sum p_n = \infty$  then  $V_{-\infty} = +\infty$  almost surely, therefore  $\mathcal{F}$  is Kolmogorovian;

2) if  $\sum p_k < \infty$  then

(a) either  $V_{-\infty}$  is not degenerate, therefore  $\mathcal{F}$  is not Kolmogorovian,

(b) or we are in the following case

$$p_{n_0} = 1 \text{ and } p_n = 0 \text{ for every } n < n_0 \text{ for some } n_0 \leq 0 \quad (*)$$

and then  $V_{-\infty} = |n_0|$  almost surely, therefore  $\mathcal{F}$  is Kolmogorovian and even standard.

Thus  $\mathcal{F}$  is Kolmogorovian if and only if  $\sum p_n = \infty$  or in case (\*). Standardness of  $\mathcal{F}$  in case (\*) elementarily holds true because  $\mathcal{F}_m = \{\emptyset, \Omega\}$  for every  $m \leq n_0$ .

## 3.2 Standardness of $\mathcal{F}$ using Vershik's criterion

Throughout this section, we denote by  $(V_n)_{n \leq 0}$  the next-jump time process with jumping probabilities  $(p_n)_{n \leq 0}$  and we denote by  $\mathcal{F}$  the filtration it generates. Discarding the elementary case (\*), it is shown in [13] with the help of the I-cosiness criterion that  $\mathcal{F}$  is standard (Vershikian) if and only if  $\sum p_n^2 = \infty$ . In this section we derive again this result by using Vershik's standardness criterion. More precisely we will use the version of Vershik's standardness criterion given by Lemma 2.6. We firstly treat a particular case in lemma below.

**Lemma 3.2.** *If  $p_n = 1$  for infinitely many  $n$ , then  $\mathcal{F}$  is standard.*

*Proof.* For every integer  $k \leq 0$ , define the random vector  $X_k = (V_k, \dots, V_0)$  and denote by  $\mathcal{B}_k = \sigma(V_k, \dots, V_0)$  the  $\sigma$ -field it generates. By the Markov property, the  $n$ -th progressive prediction  $\pi_n X_k$  of  $X_k$  is measurable with respect to  $\sigma(V_n)$  for every  $n \leq k$ , and  $V_n = |n|$  almost surely when  $p_n = 1$ , therefore  $\pi_n X_k$  is a degenerate random variable too, and  $\text{disp}(\pi_n X_k) = 0$ . Consequently,  $\mathcal{F}$  satisfies Vershik's standardness criterion by proposition 2.4(a).  $\square$

We also know by Proposition 3.1 that  $\mathcal{F}$  is standard in the case when  $p_n = 0$  for every  $n < 0$ . Then the following lemma will allow us to restrict our standardness study to the case when the identifiability condition ( $\star$ ) of Section 2.3 holds.

**Lemma 3.3.** 1) *Let  $(X_n)_{n \leq 0} = (V_{n-1})_{n \leq 0}$ . The identifiability condition ( $\star$ ) holds when*

$$p_{-1} \in ]0, 1[ \quad \text{and} \quad p_n < 1 \quad \text{for all } n < 0. \quad (3.2)$$

*In this case,  $\mathcal{F}$  is generated by the process  $(\pi_n V_{-1})_{n \leq 0}$ , and even more precisely,  $\sigma(\pi_n V_{-1}) = \sigma(V_n)$  for every  $n < 0$ .*

2) *If  $p_{n_0} = 1$  for some  $n_0 < 0$ , then the process  $(V_{n_0+n} - |n_0|)_{n \leq 0}$  is the next-jump time process with jumping probabilities  $(p_{n_0+n})_{n \leq 0}$ .*

3) *If  $p_{-1} = 0$ , then the process  $(W_{n-1})_{n \leq 0}$  defined by*

$$W_n = \begin{cases} 0 & \text{if } V_{n-1} = 0 \\ V_{n-1} - 1 & \text{if } V_{n-1} > 0 \end{cases} \quad \text{for } n \leq -1.$$

has the same distribution as  $(V_{n-1})_{n \leq 0}$  where  $(V_n)_{n \leq 0}$  is the next-jump time process with jumping probabilities  $(p'_n)_{n \leq 0}$  given by  $p'_n = p_{n-1}$  for every  $n < 0$ .

*Proof.* For  $v \neq v'$  in the state space of  $V_{n-1}$ , the conditional distributions  $\mathcal{L}(V_n | V_{n-1} = v)$  and  $\mathcal{L}(V_n | V_{n-1} = v')$  have different supports under (3.2), hence the first point follows. The equality  $\sigma(\pi_n V_{-1}) = \sigma(V_n)$  under condition  $(\star)$  is provided by Lemma 2.6. Checking the second and third points do not pose any difficulty.  $\square$

Thus, since standardness is an asymptotic property at  $n = -\infty$  (Proposition 2.3), we will focus on the case when (3.2) holds, and this will allow us to use Lemma 2.6. In Lemma 3.4 we summarize the way we are going. Hereafter we denote by  $\mathbb{V}_n = \{0, \dots, |n|\}$  the state space of  $V_n$  and consider on  $\mathbb{V}_n$  the  $n$ -th iterated Kantorovich metric  $\rho_n$  starting with the discrete 0–1 metric  $\rho_{-1}$  on  $A_{-1} = \{0, 1\}$ . That is,

$$\rho_n(v_n, v'_n) = \inf_{\Lambda_{v_n, v'_n}} \int \rho_{n+1} d\Lambda_{v_n, v'_n}$$

for every  $n \leq -2$ , where  $\Lambda_{v_n, v'_n}$  is a joining of the conditional laws  $\mathcal{L}(V_{n+1} | V_n = v_n) = P_{n+1}(v_n, \cdot)$  and  $\mathcal{L}(V_{n+1} | V_n = v'_n) = P_{n+1}(v'_n, \cdot)$ . Hereafter we denote by  $d_n$  the dispersion of  $V_n$  under  $\rho_n$ , defined by  $d_n = \mathbb{E}[\rho_n(V'_n, V''_n)]$  for two independent copies  $V'_n$  and  $V''_n$ .

**Lemma 3.4.** *Under the identifiability condition (3.2), the filtration  $\mathcal{F}$  is standard if and only if the Vershik property  $V(X)$  holds for  $X = V_{-1}$ . Moreover, this property is equivalent to  $d_n \rightarrow 0$ .*

*Proof.* Consequence of Lemma 2.6 and Lemma 3.3.  $\square$

In lemma below we provide a list of relations about the kernels  $P_n$  of the next-jump time Markov chain and the iterated Kantorovich distances  $\rho_n$ . We denote by  $P_n(v, f)$  the expectation of a function  $f$  under the probability measure  $P_n(v, \cdot)$ . Recall that  $P_{n+1}(|n|, \cdot)$  which occurs several times in the lemma is equal to the law of  $V_{n+1}$ . We use  $P_{n+1}(|n|, \cdot)$  and not  $\mathcal{L}(V_{n+1})$  in the lemma to emphasize that the derivation of the  $\rho_n$  only depends on the kernels  $P_n$  by nature.

**Lemma 3.5.** *Let  $x \geq 0$  and  $x' \geq 0$  be integer numbers.*

- 1) *If  $n \leq -1$  and  $x, x' \leq |n| - 1$ , then  $\rho_n(x, x') = \rho_{n+1}(x, x')$ .*
- 2) *If  $n \leq -2$  and  $x' \leq |n| - 1$ , then  $\rho_n(|n|, x') = P_{n+1}(|n|, \rho_{n+1}(\cdot, x'))$ .*
- 3) *If  $n \leq -3$  and  $x' \leq |n| - 2$ , then  $\rho_n(|n|, x') = \rho_{n+1}(|n+1|, x')$ .*
- 4) *If  $n \leq -1$ , then  $\rho_{n-1}(|n-1|, |n|) = (1 - p_n)P_{n+1}(|n|, \rho_n(|n|, \cdot))$ .*
- 5) *If  $n \leq -2$ , then  $P_n(|n-1|, \rho_{n-1}(|n-1|, \cdot)) = (1 - p_n^2)P_{n+1}(|n|, \rho_n(|n|, \cdot))$ .*
- 6) *For every  $n \leq -1$ ,  $P_n(|n-1|, \rho_{n-1}(|n-1|, \cdot)) = 2p_{-1}(1 - p_{-1}) \prod_{m=n}^{-2} (1 - p_m^2)$ .*

*Proof.* 1) and 2) are easily get from the expression of  $\mathcal{L}(V_{n+1} | V_n = v)$  given in Section 3.1. One obtains 3) as a consequence of 1) and 2) by using the relation

$$\Pr(V_n = k | V_{n-1} = |n-1|) = (1 - p_n) \Pr(V_{n+1} = k | V_n = |n|) \quad (3.3)$$

valid for  $0 \leq k < |n|$  and  $n \leq -2$ . One gets 4) by using 2) and (3.3). Finally, 5) is derived from 3), 4) and (3.3), and one obtains 6) by calculating the right member of 5) for  $n = -2$  and then by applying 5) recursively.  $\square$

**Lemma 3.6.** *The dispersion of  $V_n$  under  $\rho_n$  is given by  $d_n = 2p_{-1}(1-p_{-1}) \prod_{m=n}^{-2} (1-p_m^2)$  for every  $n \leq -1$ .*

*Proof.* Because of  $\mathcal{L}(V_{n+1}) = \mathcal{L}(V_{n+1} | V_n = |n|)$  we get  $d_{n+1} = \mathbb{E}[\rho_n(|n|, V_{n+1}) | V_n = |n|]$  for every  $n \leq -2$  by equality 2) of Lemma 3.5, and then the assertion of the lemma is nothing but equality 6) of Lemma 3.5.  $\square$

**Theorem 3.7.** *The filtration  $\mathcal{F}$  is standard if and only if  $\sum p_n^2 = \infty$  or in case (\*).*

*Proof.* Case (\*) is treated in Proposition 3.1. Under the identifiability condition (3.2), we know that  $\mathcal{F}$  is standard if and only if  $\prod_{n=-\infty}^{-2} (1-p_n^2) = 0$  by Lemma 3.4 and by Lemma 3.6. If the identifiability condition (3.2) does not hold, there are two cases to be treated: either  $p_n = 1$  for infinitely many  $n$  or  $p_n = 1$  for finitely many values of  $n$ . In the first case, use Lemma 3.2. In the second case, the problem can be reduced to the case of the identifiability condition by using assertion 2) of Lemma 3.3 and the fact that standardness is an asymptotic property (Proposition 2.3).  $\square$

### 3.3 Iterated Kantorovich distances

Denote by  $\mathbb{V}_n = \{0, \dots, |n|\}$  the set of vertices at each level  $n$ . The pseudometric spaces  $(\mathbb{V}_n, \rho_n)$  are easily derived from relations 1), 3), 4) and 6) given in Lemma 3.5. Note that 1) means that the canonical embedding  $(\mathbb{V}_n, \rho_n) \rightarrow (\mathbb{V}_{n-1}, \rho_{n-1})$  is an isometry, and this is a very particular situation (we mean this is not a general fact about the intrinsic pseudometrics on Bratteli graphs). The pseudometrics  $\rho_n$  are shown on Table 1.

$k \backslash k'$	0	1	2	3	4	5
0	0	1	$p_{-1}$	$p_{-1}$	$p_{-1}$	$p_{-1}$
1	1	0	$1-p_{-1}$	$1-p_{-1}$	$1-p_{-1}$	$1-p_{-1}$
2	$p_{-1}$	$1-p_{-1}$	0	$(1-p_{-2})d_{-1}$	$(1-p_{-2})d_{-1}$	$(1-p_{-2})d_{-1}$
3	$p_{-1}$	$1-p_{-1}$	$(1-p_{-2})d_{-1}$	0	$(1-p_{-3})d_{-2}$	$(1-p_{-3})d_{-2}$
4	$p_{-1}$	$1-p_{-1}$	$(1-p_{-2})d_{-1}$	$(1-p_{-3})d_{-2}$	0	$(1-p_{-4})d_{-3}$
5	$p_{-1}$	$1-p_{-1}$	$(1-p_{-2})d_{-1}$	$(1-p_{-3})d_{-2}$	$(1-p_{-4})d_{-3}$	0

Table 1: Intrinsic metrics  $\rho_n(k, k')$  for  $n = -1, -2, -3, -4, -5$ .

This table is easily filled by successively and iteratively using the following equalities for  $n \leq -2$ :

$$\begin{cases} \rho_n(0, x) = \begin{cases} 1 & \text{if } x = 1 \\ p_{-1} & \text{otherwise} \end{cases} \\ \rho_n(x, x') = \rho_{n+1}(x, x') & \text{for } x, x' < |n| \\ \rho_n(|n|, x) = \begin{cases} \rho_{n+1}(|n+1|, x) & \text{if } x < |n+1| \\ (1-p_{n+1})d_{n+2} & \text{if } x = |n+1| \end{cases} \end{cases}$$

where the expression of  $d_n$  is given in Lemma 3.6 for every  $n \leq -1$  and we set in addition  $d_0 = 1$ . It follows that the distance  $\rho_n(v_n, v'_n)$  between two vertices  $v_n$  and  $v'_n$  at some

level  $n \leq -2$  is explicitly given when  $v_n < v'_n$  by

$$\rho_n(v_n, v'_n) = \begin{cases} 1 & \text{if } v_n = 0 \text{ and } v'_n = 1 \\ p_{-1} & \text{if } v_n = 0 \text{ and } v'_n > 1 \\ (1 - p_{-v_n})d_{-v_n+1} & \text{if } v_n > 0 \end{cases}$$

The  $\rho_n$  are metrics under the identifiability condition (3.2). The space  $(\mathbb{V}_n, \rho_n)$  is an ultrametric space represented by the dendrogram shown in Figure 2 (numerically, this figure shows the case  $p_n \equiv \frac{1}{2}$  for  $n < 0$ ).

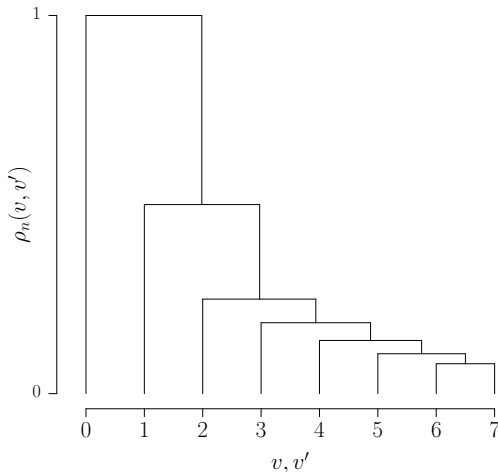


Figure 2: The space  $(\mathbb{V}_n, \rho_n)$ .

## 4 The uniformly scaled entropy

In this section we introduce the scaled entropy of filtrations by following Vershik and Gorbulsky [21], except that we use the probabilistic language and we restrict our attention to scalings which are not  $\epsilon$ -dependent (this is why we term them as *uniform scalings*). Theorem 4.11, our most significant result, generalizes Theorem 2.5.

The definition of the Vershik property  $V(X)$  stated (in Section 2) by  $\text{disp } \pi_n X \rightarrow 0$  can be equivalently stated by:

$$\forall \epsilon > 0, \exists n \leq 0, \exists \mu \in E^{(n)}, \quad \mathbb{E}[\rho_n(\pi_n X, \mu)] < \epsilon. \quad (4.1)$$

In other words, the Vershik progressive prediction  $\pi_n X$  can be approximated by a single value with probability as high as desired when  $n \rightarrow -\infty$ . When this property fails, it is natural to wonder about an optimal asymptotic approximation of  $\pi_n X$ . Roughly speaking, the scaled entropy compares the growth of the minimal entropy approximation of  $\pi_n X$  with the given scaling. Its evident interest is its ability to distinguish locally isomorphic non-standard filtrations.

The definition of the scaled entropy of a filtration relies on a choice of a measure of entropy  $H(\mu)$  for discrete probability distributions  $\mu$ . Common choices include the well-known Shannon entropy

$$H(\mu) = - \sum \mu_i \log \mu_i,$$

and the max-entropy (or Rényi entropy of order 0)

$$H(\mu) = \log \#\{\mu_i \mid \mu_i \neq 0\}.$$

We will state the definition of the scaled entropy for a measure of entropy as defined below. The three conditions of this definition are fulfilled for the Shannon entropy and for the Rényi entropy of any order.

**Definition 4.1.** An application  $H$  associating a quantity  $H(\theta) \in [0, +\infty]$  to a discrete probability measure  $\theta$  is said to be a *measure of entropy* if it satisfies the following conditions.

1.  $H$  is *decisive*:  $H(\theta) = 0$  if  $\theta$  is concentrated on one point.
2.  $H$  is *relative to the maximum probability*:  $H(\theta) \geq C(1 - \theta_{\max})$  for a certain constant  $C > 0$ , where  $\theta_{\max}$  is the highest probability mass of  $\theta$ .
3.  $H$  is *increasing*:  $H(\theta') \leq H(\theta)$  when  $\theta'$  is the image of  $\theta$  under some map.

We simply say that  $H(\theta)$  is the entropy of  $\theta$ , and we denote by  $H(X)$  the entropy of the law of a discrete random variable  $X$ .

## 4.1 Definitions

The definition of the scaled entropy of a filtration  $\mathcal{F}$  has something similar to the definition of Vershik's standardness criterion: one begins by defining the scaled entropy of a  $\mathcal{F}_0$ -measurable random variable, then for a  $\sigma$ -field  $\mathcal{B} \subset \mathcal{F}_0$ , and finally for the filtration  $\mathcal{F}$ . It mainly involves the  $\epsilon$ -entropy, defined below, of the Vershik progressive predictions  $\pi_n X$  (introduced in Section 2). We say that a random variable is *simple* when it takes only finitely many values.

**Definition 4.2.** Let  $Y$  be an integrable random variable taking its values in a Polish metric space  $(E, \rho)$ . The  $\epsilon$ -entropy of  $Y$  is

$$H^\epsilon(Y) = \inf \left\{ H(S) \mid \mathbb{E}[\rho(Y, S)] < \epsilon \right\}$$

where the infimum is taken over all simple  $\sigma(Y)$ -measurable random variables  $S$  taking values in  $E$ .

The *entropy scaling*  $c: (-\mathbb{N}) \rightarrow (0, \infty)$  in the definition below (for short, the *scaling*) is termed as a *uniform scaling* because Vershik and Gorbulsky more generally allow  $\epsilon$ -dependent scalings  $c \mapsto c(\epsilon, n)$ . We define two scaled entropies in this definition, namely the lower scaled entropy and the upper scaled entropy. The scaled entropy defined by Vershik and Gorbulsky, in the case when the scaling does not depend on  $\epsilon$ , is the upper scaled entropy. We think that the lower scaled entropy deserves to be defined in addition to the upper scaled entropy because each property we will give about the scaled entropy holds for both of them, and is proved in the same way.

**Definition 4.3.** Let  $\mathcal{F}$  be a filtration. In 1) and 2) below we consider an integrable  $\mathcal{F}_0$ -measurable random variable  $X$  taking its values in a Polish metric space  $(E, \rho)$ .

- 1) The  $\epsilon$ -entropy of  $X$  (with respect to  $\mathcal{F}$ ) at time  $n$  is  $H_n^\epsilon(X; \mathcal{F}) = H^\epsilon(\pi_n X)$ , shorter denoted by  $H_n^\epsilon(X)$  when  $\mathcal{F}$  is understood, where the  $n$ -th Vershik prediction  $\pi_n X$  is considered as a random variable taking its values in the Polish space  $E^{(n)}$  metrized by the  $n$ -th iterated Kantorovich metric  $\rho^{(n)}$  (Section 2.1).

In the sequel we consider a bounded below function  $c: (-\mathbb{N}) \rightarrow (0, \infty)$ , termed as a *uniform entropy scaling*, or, for short, a *uniform scaling*, or an *entropy scaling* or a *scaling*.

2) The limits in  $[0, \infty]$

$$h_c^-(X; \mathcal{F}) = \lim_{\epsilon \rightarrow 0} \liminf_{n \rightarrow -\infty} \frac{H_n^\epsilon(X)}{c(n)} \quad \text{and} \quad h_c^+(X; \mathcal{F}) = \lim_{\epsilon \rightarrow 0} \limsup_{n \rightarrow -\infty} \frac{H_n^\epsilon(X)}{c(n)}$$

are respectively called the *lower  $c$ -scaled entropy* of  $X$  and the *upper  $c$ -scaled entropy* of  $X$ , or, for short, the (lower/upper) *scaled entropy* of  $X$  when the scaling  $c$  is understood. Note that these limits exist because  $H_n^\epsilon(X)$  increases as  $\epsilon$  decreases. We shorter denote  $h_c^-(X; \mathcal{F})$  by  $h_c^-(X)$  and  $h_c^+(X; \mathcal{F})$  by  $h_c^+(X)$  when  $\mathcal{F}$  is understood without ambiguity.

3) For a  $\sigma$ -field  $\mathcal{B} \subset \mathcal{F}_0$ , the lower and upper  $c$ -scaled entropies of  $\mathcal{B}$  with respect to  $\mathcal{F}$  are defined as

$$h_c^-(\mathcal{B}; \mathcal{F}) = \sup_X h_c^-(X; \mathcal{F}) \quad \text{and} \quad h_c^+(\mathcal{B}; \mathcal{F}) = \sup_X h_c^+(X; \mathcal{F})$$

where the supremum is taken over all  $\mathcal{B}$ -measurable random variables  $X$  taking their values in the interval  $[0, 1]$  equipped with its usual metric. We shorter denote  $h_c^-(\mathcal{B}; \mathcal{F})$  by  $h_c^-(\mathcal{B})$  and  $h_c^+(\mathcal{B}; \mathcal{F})$  by  $h_c^+(\mathcal{B})$  when  $\mathcal{F}$  is understood without ambiguity.

4) The lower and upper  $c$ -scaled entropies of the filtration  $\mathcal{F}$  are defined as

$$h_c^-(\mathcal{F}) = h_c^-(\mathcal{F}_0; \mathcal{F}) \quad \text{and} \quad h_c^+(\mathcal{F}) = h_c^+(\mathcal{F}_0; \mathcal{F}).$$

In the sequel, we will use the notation  $h_c$  as a substitute for either  $h_c^-$  or  $h_c^+$ . When we write a statement about  $h_c$ , that means this statement holds for both  $h_c^-$  and  $h_c^+$ .

As a first remark, note that, obviously, the scaled entropy  $h_c(\mathcal{F})$  only depends on  $\mathcal{F}$  up to isomorphism. Also note that  $h_c(X) = h_{c'}(X)$  when the two scalings  $c$  and  $c'$  are equivalent at  $-\infty$ .

When the Vershik property holds for the random variable  $X$ , it is clear that  $h_c^-(X) = h_c^+(X) = 0$  for the scaling  $c(n) \equiv 1$  and consequently for any scaling  $c$ , in view of the statement (4.1) of the Vershik property and because of the decisiveness property of  $H$  (Definition 4.1). Therefore  $h_c^-(\mathcal{F}) = h_c^+(\mathcal{F}) = 0$  for any scaling  $c$  when  $\mathcal{F}$  is a Vershikian filtration. The converse is true, by virtue of the proposition below.

**Proposition 4.4.** *Let  $X$  be a  $\mathcal{F}_0$ -measurable random variable taking its values in a bounded Polish metric space  $(E, \rho)$  and such that  $h_c^-(X) = 0$  for the scaling  $c(n) \equiv 1$ . Then  $X$  satisfies the Vershik standardness property (hence  $h_c^+(X) = 0$  as well).*

*Proof.* By the assumption  $h_c^-(X) = 0$ , one has  $\liminf_{n \rightarrow -\infty} H^\epsilon(\pi_n X) = 0$  for every  $\epsilon > 0$ . Let  $\delta > 0$ . Take  $n \leq 0$  such that  $H^\delta(\pi_n X) < \delta$ . There exists a simple  $E^{(n)}$ -valued random variable  $S$  such that  $\mathbb{E}[\rho_n(\pi_n X, S)] < \delta$  and  $H(S) < \delta$ . Let  $p$  be the highest probability mass of  $S$  and  $s \in E^{(n)}$  such that  $p = \Pr(S = s)$ . Then  $p > 1 - \delta/C$ , where  $C$  is the constant of condition 2 in Definition 4.1. Therefore  $\mathbb{E}[\rho_n(S, s)] < \text{diam}(E^{(n)})\delta/C < \text{diam}(E)\delta/C$ . Finally,

$$\mathbb{E}[\rho_n(\pi_n X, s)] \leq \mathbb{E}[\rho_n(\pi_n X, S)] + \mathbb{E}[\rho_n(S, s)] \leq \left(1 + \frac{\text{diam}(E)}{C}\right) \delta.$$

Thus we get the statement (4.1) of the Vershik property for  $X$ . □

While the previous proposition is stated for a bounded metric space  $E$ , it is also true for an integrable random variable when  $E$  is unbounded. This stems from the fact mentioned in the following remark.

**Remark 4.5.** The notations  $H_n^\epsilon(X)$  and  $h_c(X)$  do not show the dependence on the metric  $\rho$  on the state space of  $X$ . But this is not important in view of Proposition 4.16 we will see later, which shows that  $h_c(X) = h_c(\sigma(X))$ . Thus, we can replace  $\rho$  with  $\rho \wedge 1$  in the definition of the scaled entropy  $h_c(X)$ , without altering its value. This allows to define  $h_c(X)$  when  $X$  is non-integrable.

**Remark 4.6.** Using the previous remark, it is not difficult to see that we do not alter the value of  $h_c(X)$  if we replace the definition of the  $\epsilon$ -entropy  $H^\epsilon(Y)$  with

$$\inf \left\{ H(S) \mid \mathbb{P}(\rho(Y, S) > \epsilon) < \epsilon \right\},$$

and this  $\epsilon$ -entropy allows to define  $h_c(X)$  when  $X$  is non-integrable.

**Remark 4.7.** Using, as allowed by Remark 4.5, the bounded distance  $\rho \wedge 1$  instead of  $\rho$ , it is easy to prove that we do not alter the value of  $h_c(X)$  if we modify the definition of  $\epsilon$ -entropy  $H^\epsilon(Y)$  by taking the infimum over all discrete random variables  $S$  instead of all simple random variables.

**Remark 4.8.** As already mentioned in the definition, the  $\epsilon$ -entropy  $H_n^\epsilon(X)$  is relative to the underlying filtration  $\mathcal{F}$ . It is important to note that it actually only depends on the filtration  $\mathcal{F}^X$  generated by the Markov process  $(\pi_n X)_{n \leq 0}$  of the Vershik progressive predictions of  $X$ . Indeed, it is easy to see that the value of  $H_n^\epsilon(X)$  is the same whether we consider  $\mathcal{F}$  as the underlying filtration or any filtration  $\mathcal{E}$  immersed in  $\mathcal{F}$  such that  $X$  is measurable with respect to the final  $\sigma$ -field  $\mathcal{E}_0$  of  $\mathcal{E}$ , and the filtration  $\mathcal{F}^X$  is the smallest such filtration (see [4]).

From the previous remark, it is easy to see that  $h_c(\mathcal{F}) \leq h_c(\mathcal{G})$  when the filtration  $\mathcal{F}$  is immersed in the filtration  $\mathcal{G}$ .

The above definition of  $H^\epsilon(\pi_n X)$  is appropriate for deriving the general properties we will give. But for the calculation of  $h_c(X)$  on a case-study, especially when we seek a lower bound of  $h_c(X)$ , it is generally better to use the alternative definition of  $H^\epsilon(\pi_n X)$  given in the following lemma for the case when  $H$  is the Shannon entropy, and in the next lemma for the case when  $H$  is the max-entropy.

**Lemma 4.9.** *Assume  $H$  is the Shannon entropy. In the definition of  $h_c(X; \mathcal{F})$ , one can replace the  $\epsilon$ -entropy  $H^\epsilon(\pi_n X)$  with  $\inf -\sum \mu_n(P_j) \log \mu_n(P_j)$  where  $\mu_n$  is the law of  $\pi_n X$  and the infimum runs over all finite partitions  $\{P_j\}$  of the state space of  $\pi_n X$  having form  $\{A_i, C\}$  where  $\mu_n(C) < \epsilon$  and  $A_i \subset B(\Gamma_i, \epsilon)$ , where  $B(\Gamma_i, \epsilon)$  is the ball of center  $\Gamma_i$  in  $(E^{(n)}, \rho^{(n)})$ .*

*Proof.* Denote by  $H_0^\epsilon(\pi_n X)$  this value. We compare it to the  $\epsilon$ -entropy  $H^\epsilon(\pi_n X)$  as given in Remark 4.6. If  $\{P_j\} = \{A_i, C\}$  is a partition such as the ones described in the lemma, define the function  $f$  by  $f(x) = \operatorname{argmin}_{\Gamma_i} \rho^{(n)}(x, \Gamma_i)$ . Then  $\rho^{(n)}(x, f(x)) < \epsilon$  on  $\cup A_i$ , and that shows  $H_0^\epsilon(\pi_n X) \geq H^\epsilon(\pi_n X)$ . Conversely, take a function  $f$  taking only finitely many values and such that the  $\mu_n$ -measure of the set  $F := \{x \in E \mid \rho^{(n)}(x, f(x)) < \epsilon\}$  is greater than  $1 - \epsilon$ . Let  $\{\Gamma_i\}$  be an enumeration of  $f(F)$ . Set  $A_i = f^{-1}(\Gamma_i) \cap F$ . Then  $\{A_i\}$  is a partition of  $F$  and  $A_i \subset B(\Gamma_i, \epsilon)$ . Moreover, since the law of  $f(\pi_n X)$  equals the weighted average  $\mu_n(F) f(\mu_n(\cdot | F)) + \mu_n(F^c) f(\mu_n(\cdot | F^c))$ , the concavity of the Shannon



entropy yields

$$\begin{aligned} H(f(\pi_n X)) &\geq \mu_n(F) H\left(f\left(\mu_n(\cdot | F)\right)\right) = -\sum \mu_n(A_i) \log \mu_n(A_i) + \mu_n(F) \log \mu_n(F) \\ &= -\sum \mu_n(P_j) \log \mu_n(P_j) - h(\mu_n(F)) \end{aligned}$$

where  $\{P_j\} = \{A_i, F^c\}$  and  $h$  is the binary entropy function defined by  $h(\epsilon) = h(1 - \epsilon) = \epsilon \log \frac{1}{\epsilon} + (1 - \epsilon) \log \frac{1}{1 - \epsilon}$ . The function  $\epsilon \mapsto h(\epsilon)$  is increasing for  $\epsilon < 1/2$ . Hence, for  $\epsilon$  small enough, one has  $h(\mu_n(F)) \leq h(\epsilon)$  and the above inequality shows that  $H^\epsilon(\pi_n X) \geq H_0^\epsilon(\pi_n X) - h(\epsilon)$ . Finally,

$$H_0^\epsilon(\pi_n X) - h(\epsilon) \leq H^\epsilon(\pi_n X) \leq H_0^\epsilon(\pi_n X)$$

and the lemma follows because of  $h(0^+) = 0$ .  $\square$

**Lemma 4.10.** *Assume  $H$  is the max-entropy. Let  $X$  a random variable taking its values in a Polish space  $(E, \rho)$ . Then the  $\epsilon$ -entropy  $H^\epsilon(X)$  as given in Remark 4.6 equals the minimal log-number of  $\epsilon$ -balls in  $E$  such that  $X$  falls in the union of these balls with probability higher than  $1 - \epsilon$ :*

$$H^\epsilon(X) = \min \left\{ \log \#\{x_i\} \mid x_i \in E, \Pr(X \in \cup B(x_i, \epsilon)) > 1 - \epsilon \right\}.$$

*Proof.* Denote by  $H_0^\epsilon(X)$  this value. For given balls  $B(x_i, \epsilon)$ , define the function  $f$  by  $f(x) = \operatorname{argmin}_{x_i} \rho(x, x_i)$ . Then  $\Pr(\rho(X, f(X)) < \epsilon) = \Pr(X \in \cup B(x_i, \epsilon))$ , and that shows that  $H_0^\epsilon(X) \geq H^\epsilon(X)$ . Conversely, take a function  $g: E \rightarrow E$  taking only finitely many values and such that  $\Pr(\rho(X, g(X)) > \epsilon) < \epsilon$ . Let  $\{x_i\}$  be an enumeration of the support of  $g$ . Define, as before,  $f(x) = \operatorname{argmin}_{x_i} \rho(x, x_i)$ . Then  $\rho(X, f(X)) \leq \rho(X, g(X))$ , hence  $\Pr(\rho(X, f(X)) < \epsilon) \geq \Pr(\rho(X, g(X)) < \epsilon)$ . That shows that  $H_0^\epsilon(X) \leq H^\epsilon(X)$ .  $\square$

## 4.2 Main properties and main theorem

This section is devoted to prove the main theorem of this paper: Theorem 4.11, stated below, which generalizes Theorem 2.5. It will be applied in Section 5 to the study the scaled entropy of the next-jump time filtrations. In this theorem and all other results of this section, a measure of entropy  $H$  as defined in Definition 4.1 is understood. We also provide two important properties of the scaled property: the heredity to the generated  $\sigma$ -field, that is to say, the equality  $h_c(X) = h_c(\sigma(X))$  (Proposition 4.16), and the left-continuity of  $\mathcal{B} \mapsto h_c(\mathcal{B})$  (Proposition 4.15). These two properties generalize the two claims of Proposition 2.4.

**Theorem 4.11.** *Let  $\mathcal{F}$  be a filtration,  $X \in L^1(\mathcal{F}_0; E)$  where  $E$  is a Polish space, and  $c: (-\mathbb{N}) \rightarrow (0, \infty)$  a uniform entropy scaling. Then  $h_c(X; \mathcal{F}) = h_c(\mathcal{F}^X)$ , where  $\mathcal{F}^X$  is the filtration generated by the Markov process  $(\pi_n X)_{n \leq 0}$ .*

Note that  $h_c(\mathcal{F}^X)$  is the entropy of the filtration  $\mathcal{F}^X$  as well as the entropy of the  $\sigma$ -field  $\sigma(\pi_n X; n \leq 0)$  when we consider  $\mathcal{F}$  as the underlying filtration (see Remark 4.8). Theorem 4.11 has a straightforward useful consequence when the  $\pi_n X$  are discrete random variables with finite entropy: it gives the upper bound  $h_c(\mathcal{F}^X) \leq 1$  for any scaling  $c(n) \sim H(\pi_n X)$ , because of the increasing property of  $H$ .

This theorem will be derived from the two following lemmas and the two main propositions of this section (heredity to the generated  $\sigma$ -field and left-continuity).

**Lemma 4.12.** *Let  $\mathcal{F}$  be a filtration. Let  $(E, \rho)$  and  $(\tilde{E}, \tilde{\rho})$  be two Polish metric spaces, and  $f: E \rightarrow \tilde{E}$  an isometry. Then  $h_c(X) = h_c(f(X))$  for any random variable  $X \in L^1(\mathcal{F}_0; E)$ .*

*Proof.* This is a straightforward consequence of Lemma 2.2.  $\square$

The key lemma is the following one.

**Lemma 4.13.** *Let  $\mathcal{F}$  be a filtration,  $X \in L^1(\mathcal{F}_0; E)$  where  $E$  is a Polish space metrized by a distance  $\rho$ , and set  $W^n = (\pi_n X, \dots, \pi_{-1} X, X)$  for some  $n \leq 0$ . Consider the metric  $\tilde{\rho}_n = \frac{1}{|n|+1} \sum_{k=n}^{k=0} \rho^{(k)}$  on the state space of  $W^n$ . Then  $\pi_n W^n = \phi(\pi_n X)$  where  $\phi$  is an isometry.*

*Proof.* For the proof we consider the distance  $\tilde{\rho}_n = \sum_{k=n}^{k=0} \rho^{(k)}$  instead of  $\tilde{\rho}_n$  on the state space of  $W_n$ . For each  $n \leq 0$  and  $k \in \{n, \dots, 0\}$ , one has  $\pi_k W_n = g_k^n(\pi_n X, \dots, \pi_k X)$  for some functions  $g_k^n$  related by the fact that  $g_{k-1}^n(\mu_n, \dots, \mu_{k-1})$  is the distribution of  $g_k^n(\mu_n, \dots, \mu_{k-1}, M_k)$  where  $M_k \sim \mu_{k-1}$ . Therefore

$$\begin{aligned} & \tilde{\rho}_n^{(k-1)}\left(g_{k-1}^n(\mu_n, \dots, \mu_{k-1}), g_{k-1}^n(\mu'_n, \dots, \mu'_{k-1})\right) \\ &= \inf_{(M_k, M'_k)} \mathbb{E} \left[ \tilde{\rho}_n^{(k)}\left(g_k^n(\mu_n, \dots, \mu_{k-1}, M_k), g_k^n(\mu'_n, \dots, \mu'_{k-1}, M'_k)\right) \right] \end{aligned} \quad (\#)$$

where the infimum is take over all joinings  $(M_k, M'_k)$  of  $\mu_{k-1}$  and  $\mu'_{k-1}$ . Using this relation, the equality

$$\tilde{\rho}_n^{(k)}\left(g_k^n(\mu_n, \dots, \mu_k), g_k^n(\mu'_n, \dots, \mu'_k)\right) = \rho^{(n)}(\mu_n, \mu'_n) + \tilde{\rho}_{n+1}^{(k)}\left(g_k^{n+1}(\mu_{n+1}, \dots, \mu_k), g_k^{n+1}(\mu'_{n+1}, \dots, \mu'_k)\right)$$

is easy to derive. Indeed, denoting by  $H(n, k)$  this equality, then  $H(n, 0)$  is nothing but the equality  $\tilde{\rho}_n = \rho^{(n)} + \tilde{\rho}_{n+1}$  and the implication from  $H(n, k)$  to  $H(n, k-1)$  is easy to derive from relation (#).

Now, by (#),

$$\tilde{\rho}_n^{(n)}\left(g_n^n(\mu_n), g_n^n(\mu'_n)\right) = \inf_{(M_{n+1}, M'_{n+1})} \mathbb{E} \left[ \tilde{\rho}_n^{(n+1)}\left(g_{n+1}^n(\mu_n, M_{n+1}), g_{n+1}^n(\mu'_n, M'_{n+1})\right) \right]$$

where the infimum is taken over all joinings  $(M_{n+1}, M'_{n+1})$  of  $\mu_n$  and  $\mu'_n$ . Hence, by relation  $H(n, n+1)$

$$\tilde{\rho}_n^{(n)}\left(g_n^n(\mu_n), g_n^n(\mu'_n)\right) = \rho^{(n)}(\mu_n, \mu'_n) + \inf_{(M_{n+1}, M'_{n+1})} \mathbb{E} \left[ \tilde{\rho}_{n+1}^{(n+1)}\left(g_{n+1}^{n+1}(M_{n+1}), g_{n+1}^{n+1}(M'_{n+1})\right) \right],$$

and using this equality we can prove by recursion (starting at  $n = 0$ ) the equality

$$\tilde{\rho}_n^{(n)}\left(g_n^n(\mu_n), g_n^n(\mu'_n)\right) = (|n| + 1) \rho^{(n)}(\mu_n, \mu'_n)$$

which is obviously equivalent to the statement of the lemma.  $\square$

It is interesting to note that Theorem 2.5 is an easy corollary of the previous lemma, Lemma 2.2, and Proposition 2.4. Indeed, this provides a new proof of this theorem, cleaner than the one given in [11].

The following lemma is a continuity-like property of  $X \mapsto h_c(X; \mathcal{F})$ . A cleaner continuity property of the scaled entropy, relying on this lemma, is the content of Proposition 4.15.

**Lemma 4.14.** *Let  $\mathcal{F}$  be a filtration, and  $E$  a Polish space. Let  $(X_k)_{k \geq 1}$  be a sequence in  $L^1(\mathcal{F}_0; E)$  such that  $X_k \rightarrow X$  in  $L^1$  for some random variable  $X \in L^1(\mathcal{F}_0; E)$ , and such that  $\sigma(X_k) \subset \sigma(X)$  for every  $k \geq 1$ . If, for a given scaling  $c: (-\mathbb{N}) \rightarrow ]0, \infty[$ , there exists  $\ell \geq 0$  such that  $h_c(X_k, \mathcal{F}) \leq \ell$  for every  $k$  sufficiently large, then  $h_c(X; \mathcal{F}) \leq \ell$ .*

*Proof.* We denote by  $k(\epsilon)$  an integer such that  $\mathbb{E}[\rho(X_k, X)] \leq \epsilon$  for every  $k \geq k(\epsilon)$ . Hence, for  $k \geq k(\epsilon)$ , the inequality  $H_n^\epsilon(X_k) \geq H_n^{2\epsilon}(X)$  holds for every  $n$  by definition of  $H_n^\epsilon(\cdot)$  and Lemma 2.1.

We write the proof for the upper scaled entropy. The proof for the lower scaled entropy is similar. Set  $a = h_c(X; \mathcal{F})$ . We firstly check that  $a < \infty$ . Assuming  $a = \infty$ , there exists  $\epsilon_0 > 0$  such that  $\limsup_{n \rightarrow -\infty} \frac{H_n^{2\epsilon_0}(X)}{c(n)} > \ell + 1$ . Therefore one can take  $k_0 \geq k(\epsilon_0)$  sufficiently large in order that  $h_c(X_{k_0}, \mathcal{F}) \leq \ell$  and such that  $\limsup_{n \rightarrow -\infty} \frac{H_n^{\epsilon_0}(X_{k_0})}{c(n)} > \ell + 1$ . But  $\epsilon \mapsto H_n^\epsilon(X_{k_0})$  is decreasing, therefore the inequality  $\limsup_{n \rightarrow -\infty} \frac{H_n^\epsilon(X_{k_0})}{c(n)} > \ell + 1$  holds for every  $\epsilon \leq \epsilon_0$ , a contradiction of the assumption of the lemma.

Knowing now that  $a < \infty$ , we check that  $\ell \geq a$ . Given  $\delta > 0$ , there exists  $\epsilon_0 > 0$  such that  $\limsup_{n \rightarrow -\infty} \frac{H_n^{2\epsilon_0}(X)}{c(n)} > a - \delta$ . Taking  $k \geq k(\epsilon_0)$ , one gets  $\limsup_{n \rightarrow -\infty} \frac{H_n^\epsilon(X_k)}{c(n)} > a - \delta$  for every  $\epsilon \leq \epsilon_0$  because  $\epsilon \mapsto H_n^\epsilon(X_k)$  is decreasing. Taking  $k$  sufficiently large in order that  $h_c(X_k, \mathcal{F}) \leq \ell$ , one finally gets  $\ell \geq a$ .  $\square$

The property given in the following proposition will be called the *left-continuity* of the scaled entropy.

**Proposition 4.15** (Left-continuity). *Let  $\mathcal{F}$  be a filtration,  $\mathcal{B} \subset \mathcal{F}_0$  a  $\sigma$ -field, and  $(\mathcal{B}_k)_{k \geq 1}$  an increasing sequence of  $\sigma$ -fields such that  $\mathcal{B}_k \nearrow \mathcal{B}$ . Then  $h_c(\mathcal{B}) = \lim h_c(\mathcal{B}_k)$  for any scaling  $c$ .*

*Proof.* The inequality  $h_c(\mathcal{B}) \geq h_c(\mathcal{B}_k)$  is an obvious consequence of the inclusion  $\mathcal{B}_k \subset \mathcal{B}$ . To show the converse inequality, consider an integrable  $\mathcal{B}$ -measurable random variable  $X$  taking its values in  $[0, 1]$ . Then  $X = \lim X_k$  in  $L^1$  where  $X_k$  is a  $\mathcal{B}_k$ -measurable random variable. By Lemma 4.14,  $h_c(X) \leq \sup h_c(X_k)$ , therefore  $h_c(X) \leq \sup h_c(\mathcal{B}_k)$ . But  $h_c(\mathcal{B}_k)$  is increasing in  $k$ , hence  $\sup h_c(\mathcal{B}_k) = \lim h_c(\mathcal{B}_k)$ . That shows that  $h_c(\mathcal{B}) \leq \lim h_c(\mathcal{B}_k)$ .  $\square$

**Proposition 4.16** (Hereditiy to the generated  $\sigma$ -field). *Let  $\mathcal{F}$  be a filtration and  $X \in L^1(\mathcal{F}_0; E)$  where  $E$  is a Polish space. Then  $h_c(\sigma(X)) = h_c(X)$  for any scaling  $c: (-\mathbb{N}) \rightarrow ]0, \infty[$ .*

*Proof.* If  $Y = f(X)$  for some Lipschitz function  $f$ , then it is easy to check that  $h_c(Y) \leq h_c(X)$  with the help of Lemma 2.2. The set of random variables  $f(X)$ ,  $f: E \rightarrow \mathbb{R}^m$  Lipschitzian, is dense in  $L^1(\sigma(X); \mathbb{R}^m)$ . Indeed, this is the content of lemma 2.15 in [10] when  $m = 1$ , and the case  $m \geq 2$  obviously follows from the case  $m = 1$ . Therefore, by Lemma 4.14, we know that  $h_c(Y) \leq h_c(X)$  for every random variable  $Y \in L^1(\sigma(X); \mathbb{R}^m)$ . The case  $m = 1$  yields the inequality  $h_c(\sigma(X)) \leq h_c(X)$ . It remains to show the converse inequality.

Firstly, consider a random variable  $U$  taking its values in  $[0, 1]$  and such that  $\sigma(X) = \sigma(U)$ . From what we have seen above, we know that  $h_c(Y) \leq h_c(U)$ , hence  $h_c(Y) \leq h_c(\sigma(X))$ , for every random variable  $Y \in L^1(\sigma(X); \mathbb{R}^m)$ .

Now, consider a sequence  $(X_k)_{k \geq 1}$  in  $L^1(\sigma(X); E)$  such that  $X_k \rightarrow X$  in  $L^1$  and each  $X_k$  takes only finitely many values. We know that  $h_c(X) \leq \sup h_c(X_k)$  by Lemma 4.14. For a given  $k \geq 1$ , denote by  $F$  the finite subset of  $E$  in which  $X_k$  takes its values. Note that  $h_c(X_k)$  is the same either we look at  $X_k$  as a  $E$ -valued random variable or a  $F$ -valued random variable. Considering the sup-norm on  $\mathbb{R}^m$ , where  $m = \#F$ , it is well-known that there exists a distance-preserving function  $f: F \rightarrow \mathbb{R}^m$ . By Lemma 4.12,  $h_c(X_k) = h_c(f(X_k))$ . But we have seen that  $h_c(f(X_k)) \leq h_c(\sigma(X))$ , therefore we finally get the inequality  $h_c(X) \leq h_c(\sigma(X))$ .  $\square$

Now we can quickly prove Theorem 4.11.

*Proof of Theorem 4.11.* Let  $\mathcal{B}_n = \sigma(\pi_n X, \dots, \pi_{-1} X, X)$ . By Proposition 4.16 (heredity to the generated  $\sigma$ -field), Lemma 4.13 and Lemma 4.12,  $h_c(\mathcal{B}_n; \mathcal{F}^X) = h_c(\pi_n X; \mathcal{F}^X)$ . It is not difficult to see that  $h_c(\pi_n X; \mathcal{F}^X) = h_c(X; \mathcal{F}^X)$ . Thus,  $h_c(\mathcal{B}_n; \mathcal{F}^X) = h_c(X; \mathcal{F}^X) = h_c(X; \mathcal{F})$ . But  $h_c(\mathcal{B}_n; \mathcal{F}^X) \rightarrow h_c(\mathcal{F}^X)$  by Proposition 4.15, then the theorem follows.  $\square$

We provide the following corollary as an easy consequence of Proposition 4.15 and Proposition 4.16.

**Corollary 4.17.** *When the final  $\sigma$ -field  $\mathcal{F}_0$  of a filtration  $\mathcal{F}$  is essentially separable, then  $h_c(\mathcal{F}) = \sup_X h_c(X)$  where the supremum is taken over all simple  $\mathcal{F}_0$ -measurable random variables  $X$ .*

*Proof.* This supremum is obviously lower than  $h_c(\mathcal{F})$ . To show the converse inequality, take an increasing sequence  $(\mathcal{B}_k)_{k \geq 1}$  of finitely generated sub- $\sigma$ -fields of  $\mathcal{F}_0$  such that  $\mathcal{B}_k \nearrow \mathcal{F}_0$ , and take a simple random variable  $X_k$  generating  $\mathcal{B}_k$  for each  $k \geq 1$ . By Proposition 4.15 and Proposition 4.16,

$$h_c(\mathcal{F}) = \lim_{k \rightarrow \infty} h_c(\mathcal{B}_k) = \lim_{k \rightarrow \infty} h_c(X_k),$$

thereby showing the desired converse inequality.  $\square$

The next corollary is a consequence of Lemma 4.13 and Proposition 4.16. It generalizes Proposition 2.3 (standardness is an asymptotic property).

**Corollary 4.18** (Entropy is an asymptotic quantity). *Let  $\mathcal{F}$  be a filtration,  $n_0 \leq 0$  be an integer, and denote by  $\mathcal{F}^{n_0] = (\mathcal{F}_{n_0+n})_{n \leq 0}$  the filtration  $\mathcal{F}$  truncated at  $n_0$ . Let  $c: (-\mathbb{N}) \rightarrow ]0, \infty[$  be a scaling and denote  $c^{n_0] = (c_{n_0+n})_{n \leq 0}$  its truncation at  $n_0$ . Then  $h_{c^{n_0]}(\mathcal{F}^{n_0]) = h_c(\mathcal{F})$ .*

*Proof.* It is not difficult to derive the equality

$$H_n^\epsilon(X_{n_0}, \mathcal{F}^{n_0]) = H_{n_0+n}^\epsilon(X_{n_0}; \mathcal{F}) \quad (4.2)$$

for every integrable  $\mathcal{F}_{n_0}$ -measurable random variable  $X_{n_0}$ , every  $n \leq 0$  and every  $\epsilon > 0$ . This provides the inequality  $h_{c^{n_0]}(\mathcal{F}^{n_0]) \leq h_c(\mathcal{F})$ .

Conversely, if  $W^{n_0} = (\pi_{n_0} X_0, \dots, X_0)$  where  $X_0$  is an integrable  $\mathcal{F}_0$ -measurable random variable, then  $h_c(X_0; \mathcal{F}) \leq h_c(W^{n_0}; \mathcal{F})$  by Proposition 4.16 and because  $H$  is increasing (Definition 4.1). But Lemma 4.13 provides the equality

$$H_{n_0+n}^\epsilon(W^{n_0}; \mathcal{F}) = H_{n_0+n}^\epsilon(\pi_{n_0} X_0; \mathcal{F})$$

for every  $n \leq 0$ . Hence equality (4.2) gives

$$H_{n_0+n}^\epsilon(W^{n_0}; \mathcal{F}) = H_n^\epsilon(\pi_{n_0} X_0; \mathcal{F}^{n_0]),$$

therefore  $h_c(W^{n_0}; \mathcal{F}) = h_{c^{n_0]}(\pi_{n_0} X_0; \mathcal{F}^{n_0])$  and finally  $h_c(X_0; \mathcal{F}) \leq h_{c^{n_0]}(\pi_{n_0} X_0; \mathcal{F}^{n_0])$ . This provides the inequality  $h_c(\mathcal{F}) \leq h_{c^{n_0]}(\mathcal{F}^{n_0])$ .  $\square$

### 4.3 Extensions that do not increase entropy

Say that a filtration  $\mathcal{G}$  is an *extension* of a filtration  $\mathcal{F}$  if the filtration  $\mathcal{F}$  is immersible in  $\mathcal{G}$  (see [4] or [10] if needed). When  $\mathcal{G}$  is a parametric extension or an independent enlargement of  $\mathcal{F}$  (defined below), it is known that  $\mathcal{G}$  is Vershikian if  $\mathcal{F}$  is Vershikian. We generalize these two results by showing that in such cases,  $\mathcal{G}$  has the same scaled entropy as  $\mathcal{F}$ . We will apply these two theorems on some examples in Section 6.

The next theorem is a generalization of proposition 6.1 in [11], which says that a parametric extension of a Vershikian filtration is a Vershikian filtration. A *superinnovation* of a filtration  $\mathcal{F}$  is a sequence  $(V_n)_{n \leq 0}$  of random variables such that each  $V_n$  is independent of  $\mathcal{F}_{n-1}$  and satisfies  $\mathcal{F}_n \subset \mathcal{F}_{n-1} \vee \sigma(V_n)$ . It is also called a *parameterization* or a *governing process* in some published papers. Once that is given, the filtration  $\mathcal{F}$  is immersed in the filtration  $\mathcal{G}$  defined by  $\mathcal{G}_n = \mathcal{F}_n \vee \sigma(V_m; m \leq n)$  (this follows from lemma 1.6 in [10]), and  $\mathcal{G}$  is called a *parametric extension* of  $\mathcal{F}$ .

**Theorem 4.19** (Parametric extensions do not increase entropy). *Let  $\mathcal{F}$  be a filtration and  $(V_n)_{n \leq 0}$  be a superinnovation of  $\mathcal{F}$ . Denote by  $\mathcal{G}$  the extension of  $\mathcal{F}$ , defined by  $\mathcal{G}_n = \mathcal{F}_n \vee \sigma(V_m; m \leq n)$ . Then  $h_c(\mathcal{G}) = h_c(\mathcal{F})$  for any scaling  $c$ .*

*Proof.* By the left-continuity of the scaled entropy (Proposition 4.15),  $h_c(\mathcal{G}) = \lim h_c(\mathcal{B}_m)$  where  $\mathcal{B}_m = \mathcal{F}_0 \vee \sigma(V_{m+1}, \dots, V_0)$ . But  $\mathcal{B}_m = \mathcal{F}_m \vee \sigma(V_{m+1}, \dots, V_0)$ , therefore  $h_c(\mathcal{B}_m; \mathcal{G}) \geq h_c(\mathcal{F}_m; \mathcal{G})$ . Conversely, using the same equality, it is not difficult to check that  $\pi_m X$  is measurable with respect to  $\mathcal{F}_m$  for any  $\mathcal{B}_m$ -measurable integrable random variable  $X$ . Since  $h_c(X) = h_c(\pi_m X)$ , one gets  $h_c(\mathcal{B}_m; \mathcal{G}) \leq h_c(\mathcal{F}_m; \mathcal{G})$  and finally  $h_c(\mathcal{B}_m; \mathcal{G}) = h_c(\mathcal{F}_m; \mathcal{G})$ . Now,  $\mathcal{F}$  is immersed in  $\mathcal{G}$ , therefore  $h_c(\mathcal{F}_m; \mathcal{G}) = h_c(\mathcal{F}_m; \mathcal{F})$ . But  $h_c(\mathcal{F}_m; \mathcal{F}) = h_c(\mathcal{F}_0; \mathcal{F})$  because of the equality  $h_c(X) = h_c(\pi_m X)$  holding for every  $\mathcal{F}_0$ -measurable integrable random variable  $X$ .  $\square$

The corollary below is an application of the previous theorem. It generalizes corollary 6.1 of [11]. In the proof, we use the fact that for every Markov process  $(X_n)_{n \leq 0}$ , there exists a superinnovation  $(V_n)_{n \leq 0}$  of  $\mathcal{F}$  satisfying the additional property  $\sigma(X_{n-1}, V_n) \supset \sigma(X_n)$ . This stems from lemma 3.41 in [10]. Note that the filtration  $\mathcal{E}$  in the corollary is immersed in the filtration  $\mathcal{F}_{\phi(\cdot)}$ ; that means that the process  $(X_{\phi(n)})_{n \leq 0}$  is Markovian with respect to  $\mathcal{F}_{\phi(\cdot)}$ .

**Corollary 4.20.** *Let  $(X_n)_{n \leq 0}$  be a Markov process and let  $\mathcal{F}$  be the filtration it generates. Let  $\phi: (-\mathbb{N}) \rightarrow (-\mathbb{N})$  be a strictly increasing map, denote by  $\mathcal{F}_{\phi(\cdot)}$  the filtration  $(\mathcal{F}_{\phi(n)})_{n \leq 0}$ , and denote by  $\mathcal{E}$  the filtration generated by the Markov process  $(X_{\phi(n)})_{n \leq 0}$ . Then, for any scaling  $c$ ,*

$$\begin{aligned} h_c(X_{\phi(n)}; \mathcal{F}) &\xrightarrow[n \rightarrow -\infty]{} h_c(\mathcal{F}), \\ h_c(X_{\phi(n)}; \mathcal{E}) &\xrightarrow[n \rightarrow -\infty]{} h_c(\mathcal{E}) \end{aligned}$$

and  $h_c(\mathcal{F}_{\phi(\cdot)}) = h_c(\mathcal{E})$ .

*Proof.* Let  $(V_n)_{n \leq 0}$  be a superinnovation of  $\mathcal{F}$  such that  $\sigma(X_{n-1}, V_n) \supset \sigma(X_n)$  for every  $n \leq 0$ , and let  $\mathcal{G}$  be the parametric extension of  $\mathcal{F}$  with  $(V_n)_{n \leq 0}$ . By Theorem 4.19, we know that  $h_c(\mathcal{F}) = h_c(\mathcal{G})$ . By Proposition 4.15 (left-continuity of the scaled entropy), we know that  $h_c(\mathcal{G}) = \lim h_c(\mathcal{B}_n)$  where  $\mathcal{B}_m = \sigma(X_{\phi(n)}, V_{\phi(n)+1}, \dots, V_0)$ . Now, by noting that the Vershik prediction  $\pi_{\phi(n)}X$  is  $\sigma(X_{\phi(n)})$ -measurable for every  $\mathcal{B}_n$ -measurable random variable  $X$ , one gets  $h_c(\mathcal{B}_n) = h_c(X_{\phi(n)})$ . More precisely, we should write  $h_c(\mathcal{B}_n; \mathcal{G}) = h_c(X_{\phi(n)}; \mathcal{G})$ , but  $h_c(X_{\phi(n)}; \mathcal{G}) = h_c(X_{\phi(n)}; \mathcal{F})$  because  $\mathcal{F}$  is immersed in  $\mathcal{G}$ . Thus we have shown  $h_c(\mathcal{F}) = \lim h_c(X_{\phi(n)})$ . This result applied to the Markov process  $(X_{\phi(n)})_{n \leq 0}$  and with the identity map instead of  $\phi$  shows that  $h_c(\mathcal{E}) = \lim h_c(X_{\phi(n)}; \mathcal{E})$ , and note that  $h_c(X_{\phi(n)}; \mathcal{E}) = h_c(X_{\phi(n)}; \mathcal{F}_{\phi(\cdot)})$  because  $\mathcal{E}$  is immersed in  $\mathcal{F}_{\phi(\cdot)}$ . To finish, the equality  $h_c(\mathcal{E}) = h_c(\mathcal{F}_{\phi(\cdot)})$  follows from an application of Theorem 4.19 to the filtration  $\mathcal{E}$  and its superinnovation  $\left( (V_{\phi(n-1)+1}, \dots, V_{\phi(n)}) \right)_{n \leq 0}$ .  $\square$

The next theorem is a generalization of lemma 19 in [4], which says that the supremum of two independent Vershikian filtrations is a Vershikian filtration (saying that two filtrations are independent when they are defined on the same probability space and their final  $\sigma$ -fields are independent). To prove it, we use the following lemma, whose proof is given in lemma 19 of [4]. Note that  $\pi_n(R_1, R_2)$  in this lemma is defined with respect to the filtration  $\mathcal{G} = \mathcal{F}^1 \vee \mathcal{F}^2$ . On the other hand, each of the filtrations  $\mathcal{F}^1$  and  $\mathcal{F}^2$  is immersed in  $\mathcal{G}$ , therefore there is no possible ambiguity about the underlying filtration with respect to which  $\pi_n R_1$  and  $\pi_n R_2$  are defined.

**Lemma 4.21.** *Let  $\mathcal{F}^1$  and  $\mathcal{F}^2$  be two independent filtrations. Let  $(E_1, \rho_1)$  and  $(E_2, \rho_2)$  be two Polish metric spaces. We take the metric  $\rho((x_1, x_2), (x'_1, x'_2)) = \rho_1(x_1, x'_1) + \rho_2(x_2, x'_2)$  on the product Polish space  $E = E_1 \times E_2$ . Then there are some Borelian maps  $i_n: E_1^{(n)} \times E_2^{(n)} \rightarrow E^{(n)}$  such that*

$$\rho^{(n)}(i_n u, i_n v) \leq \rho_1^{(n)}(u_1, v_1) + \rho_2^{(n)}(u_2, v_2)$$

for every  $u = (u_1, u_2)$  and  $v = (v_1, v_2)$  in  $E_1^{(n)} \times E_2^{(n)}$ , and they are such that

$$\pi_n(R_1, R_2) = i_n(\pi_n R_1, \pi_n R_2)$$

for every  $\mathcal{F}_0^1$ -measurable random variable  $R_1$  taking its values in  $(E_1, \rho_1)$  and every  $\mathcal{F}_0^2$ -measurable random variable  $R_2$  taking its values in  $(E_2, \rho_2)$ .

**Theorem 4.22** (Independent enlargments do not increase entropy). *Let  $\mathcal{F}^1$  and  $\mathcal{F}^2$  be two independent filtrations. If  $\mathcal{F}^2$  is Vershikian, then the equality  $h_c(\mathcal{F}^1 \vee \mathcal{F}^2) = h_c(\mathcal{F}^1)$  holds for any scaling  $c$ .*

*Proof.* We know that  $h_c(\mathcal{F}^1 \vee \mathcal{F}^2) \geq h_c(\mathcal{F}^1)$  because  $\mathcal{F}_1$  is immersed in  $\mathcal{F}_1 \vee \mathcal{F}_2$ . It remains to show  $h_c(\mathcal{F}^1 \vee \mathcal{F}^2) \leq h_c(\mathcal{F}^1)$ . Let  $R = (R_1, R_2)$  where  $R_1$  and  $R_2$  are random variables as in the previous lemma. Take  $\epsilon > 0$ . Thanks to the Vershik property of  $\mathcal{F}^2$ , for every small enough  $n$  there exists  $s_2 \in E_2^{(n)}$  such that  $\mathbb{E}[\rho_2^{(n)}(\pi_n R_2, s_2)] < \epsilon$ . Take this  $n$  small enough in order to get a simple  $\sigma(\pi_n R_1)$ -measurable random variable  $S_1$  such that  $\mathbb{E}[\rho_1^{(n)}(\pi_n R_1, S_1)] < \epsilon$ . The simple random variable  $S = (S_1, s_2)$  has the same entropy as  $S_1$ , and by the previous lemma,

$$\rho^{(n)}(\pi_n R, i_n S) \leq \rho_1^{(n)}(\pi_n R_1, S_1) + \rho_2^{(n)}(\pi_n R_2, s_2),$$

therefore  $\mathbb{E}[\rho^{(n)}(\pi_n R, i_n S)] < 2\epsilon$ . Since  $H(i_n S) \leq H(S)$ , that shows that  $H^{2\epsilon}(\pi_n R) \leq H^\epsilon(\pi_n R_1)$ . Therefore  $h_c(R; \mathcal{F}^1 \vee \mathcal{F}^2) \leq h_c(R_1; \mathcal{F}_1)$ . Consequently,  $h_c(\mathcal{F}^1 \vee \mathcal{F}^2) \leq h_c(\mathcal{F}^1)$ , and the proof is over.  $\square$

## 5 Entropy of next-jump time filtrations

In this section, we consider, for a given sequence  $(p_n)_{n \leq 0}$  of jumping probabilities, the next-jump time process  $(V_n)_{n \leq 0}$  introduced in Section 3. Its filtration is denoted, as before, by  $\mathcal{F}$ . Using the Shannon entropy as the measure of entropy, we study the scaled entropy of  $\mathcal{F}$  in the Kolmogorovian non-standard case, that is, in view of Proposition 3.1 and Theorem 3.7, the case when  $\sum p_n = \infty$  and  $\sum p_n^2 < \infty$ . It is understood that we consider this situation throughout this section. Two non-standard next-jump time filtrations defined by two distinct jumping probabilities sequences are not locally isomorphic, therefore there is no real interest to compare them with the scaled entropy. But this case-study provides a nice example because it illustrates Theorem 4.11 and because of its simplicity.

We assume without loss of generality that  $p_n < 1$  for every  $n < 0$ . Indeed, as seen in Section 3,  $p_n = 1$  only for finitely many values (Lemma 3.2). Therefore, we can make this assumption by applying point 2) of Lemma 3.3 with the largest integer  $n_0$  such that  $p_{n_0} = 1$ , and by knowing that the scaled entropy is an asymptotic quantity (Corollary 4.18).

Thus, by point 1) of Lemma 3.3, the identifiability condition  $(\star)$  holds for the next-jump time process when one takes  $X_0 = V_{-1}$ , and by point 3) of Lemma 2.6, we know that the filtration  $(\mathcal{F}_n)_{n \leq -1}$  is generated by the process of Vershik progressive predictions  $(\pi_n V_{-1})_{n \leq -1}$ . Therefore the calculation of the scaled entropy  $h_c(\mathcal{F})$  is greatly simplified by Theorem 4.11 because  $h_c(\mathcal{F}) = h_c(V_{-1})$  by virtue of this theorem. Moreover, by point 3) of Lemma 2.6, one can replace the  $\epsilon$ -entropy  $H^\epsilon(\pi_n V_{-1})$  with the  $\epsilon$ -entropy  $H^\epsilon(V_n)$ , considering that  $V_n$  takes its values in the set  $\{0, \dots, |n|\}$  equipped with the iterated Kantorovich metric  $\rho_n$ .

But one can even replace the  $\rho_n$  with the discrete 0–1 metric. Indeed, as seen in Section 3.3, the iterated Kantorovich metrics  $\rho_n$  satisfy  $\rho_n(x, x') \geq \epsilon_0$  for every  $n \leq 0$  and every  $x \neq x'$ , where  $\epsilon_0 > 0$  does not depend on  $n$ ,  $x$  and  $x'$ . Therefore, one can replace  $\mathbb{E}[\rho_n(V_n, S)]$  with  $\Pr(V_n \neq S)$  in the definition of  $H^\epsilon(V_n)$ , because  $h_c(V_{-1})$  pertains on  $H^\epsilon(V_n)$  only for small  $\epsilon$ .

To sum up, the lower and upper scaled entropies of  $\mathcal{F}$  are

$$h_c^-(\mathcal{F}) = \lim_{\epsilon \rightarrow 0} \liminf_{n \rightarrow -\infty} \frac{H^\epsilon(V_n)}{c(n)} \quad \text{and} \quad h_c^+(\mathcal{F}) = \lim_{\epsilon \rightarrow 0} \limsup_{n \rightarrow -\infty} \frac{H^\epsilon(V_n)}{c(n)},$$

where  $V_n$  is viewed as a random variable taking its values in the set  $\{0, \dots, |n|\}$  equipped with the discrete 0–1 metric. Thus, the calculation of  $h_c(\mathcal{F})$  comes down to a quite elementary problem. That does not mean its calculation is an easy task. This problem is investigated in [15], where the right member is denoted by  $h_c(\mu)$ , where  $\mu$  is the measure on  $\mathbb{N}$  we introduced before Proposition 3.1. We noted that this measure characterizes the sequence of jumping probabilities  $(p_n)_{n \leq 0}$ , which corresponds to the *reversed hazard rate* of  $\mu$  in [15]. The next-jump time process  $(V_n)_{n \leq 0}$  is the process  $(X_{-n})_{n \leq 0}$  with the notations of [15].

It is clear that the inequality  $h_c(\mathcal{F}) \leq 1$  always holds for the scaling  $c(n) = H(V_n)$  because  $H^\epsilon(V_n) \leq H(V_n)$ . In [15], it is shown that  $H(V_n) \rightarrow \infty$  in the Kolmogorovian non-standard situation. With this scaling, using the examples studied in [15], one has some examples such that  $h_c(\mathcal{F}) = r$  for any  $r \in [0, 1]$ . For instance,  $h_c(\mathcal{F}) = 1$  in the uniform case  $p_n = (|n| + 1)^{-1}$ . The jumping probabilities of the examples in [15] yielding  $h_c(\mathcal{F}) < 1$  are shown on Figure 3. For the examples giving  $h_c(\mathcal{F}) = 0$  for the scaling  $c(n) = H(V_n)$ , a scaling  $c'$  such that  $h_{c'}(\mathcal{F}) = 1$  is derived in [15].

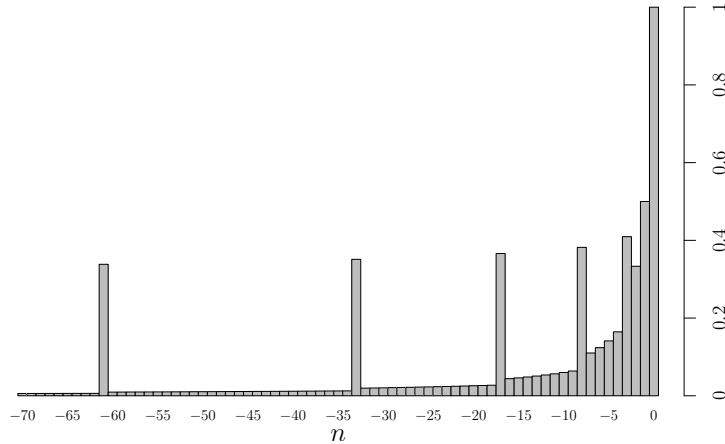


Figure 3: Jumping probabilities yielding  $h_c(\mathcal{F}) < 1$ .

## 6 Entropy of poly-adic filtrations

The pioneering works of Vershik focused on poly-adic filtrations, that is to say, filtrations  $\mathcal{F}$  such that for every  $n \leq 0$ , there exists a random variable  $\eta_n$  uniformly distributed on a finite set, independent of  $\mathcal{F}_{n-1}$ , and such that  $\mathcal{F}_n = \mathcal{F}_{n-1} \vee \sigma(\eta_n)$ . Such a random  $\eta_n$  is called an *innovation* of  $\mathcal{F}$  (at time  $n$ ), and denoting by  $r_n$  the size of the set on which it is uniformly distributed,  $\mathcal{F}$  is said to be  $(r_n)_{n \leq 0}$ -adic (this makes sense because any other innovation of  $\mathcal{F}$  at time  $n$  is uniformly distributed on  $r_n$  values). For such filtrations, Vershik defined the *exponential entropy*, originally in [18]. We will give this definition below.

In spite of the equality  $\mathcal{F}_n = \mathcal{F}_m \vee \sigma(\eta_{m+1}, \dots, \eta_n)$  holding for every  $m < n \leq 0$ , the Kolmogorov property  $\mathcal{F}_{-\infty} = \{\emptyset, \Omega\}$  does not ensure that  $\mathcal{F}_n = \sigma(\eta_m; m \leq n)$ . In other words, it does not ensure that  $\mathcal{F}$  is generated by the process of innovations  $(\eta_n)_{n \leq 0}$ . In fact, standardness of a such a filtration is known to be equivalent to the existence of a process of innovations  $(\eta'_n)_{n \leq 0}$  generating this filtration. This is one of the main results of Vershik's theory of filtrations. The difficult point to prove in this result is the existence of  $(\eta'_n)_{n \leq 0}$  assuming standardness, whereas the converse is easy to prove with the help of Proposition 2.4.

Throughout this section, when a  $(r_n)_{n \leq 0}$ -adic filtration is under consideration, we denote by  $(\ell_n)_{n \leq 0}$  the integer sequence associated to  $(r_n)_{n \leq 0}$  by setting  $\ell_n = \prod_{i=n+1}^0 r_i$  (agreeing with  $\ell_0 = 1$ ).

In Section 6.1, with the help of Lemma 6.3, we will see that the definition below makes sense.



**Definition 6.1.** Let  $\mathcal{F}$  be a  $(r_n)_{n \leq 0}$ -adic filtration and  $X$  a  $\mathcal{F}_0$ -measurable random variable taking only finitely many values.

1. The exponential entropy of  $X$  with respect to  $\mathcal{F}$  is the number

$$h(X; \mathcal{F}) = \lim_{n \rightarrow -\infty} \frac{H(\pi_n X)}{\ell_n} = \inf_{n \leq 0} \frac{H(\pi_n X)}{\ell_n} \in [0, \infty[$$

where  $\ell_n = \prod_{i=n+1}^0 r_i$  and  $H$ , unless something else is said, is the Shannon entropy (in a given logarithmic base).

2. The exponential entropy, of  $\mathcal{F}$  is  $h(\mathcal{F}) = \sup h(X; \mathcal{F}) \in [0, \infty]$  where the supremum is taken over all  $\mathcal{F}_0$ -measurable random variables  $X$  taking only finitely many values.

Note the obvious inequality  $h_c(\mathcal{F}) \leq h(\mathcal{F})$  when  $c$  is the scaling  $c(n) = \ell_n$ .

It is shown in [20] that  $h(\mathcal{F}) = 0$  when  $\mathcal{F}$  is a standard poly-adic filtration<sup>1</sup>. We will not use this result in the present paper. But in the case of the slowness condition  $(\Delta)$  about the poly-adicity sequence  $(r_n)_{n \leq 0}$ , it follows from Theorem 6.4, and in the case of the opposite condition  $(-\Delta)$ , it follows from Theorem 6.14 (see below).

The main results of this section are listed below, where the two conditions  $(\Delta)$  and  $(\nabla)$  about the speed of the poly-adicity sequence  $(r_n)_{n \leq 0}$ , respectively a slowness condition and a fastness condition, are

$$(\Delta): \sum_{n=-\infty}^0 \frac{r_n \log r_n}{\ell_{n-1}} < \infty \quad \text{and} \quad (\nabla): \frac{\log r_n}{\ell_n} \rightarrow \infty.$$

1. Theorem 6.4, whose credit is given to Gorbulsky ([6]), is about the equality  $h_c(\mathcal{F}) = h(\mathcal{F})$  between the exponential entropy and the scaled entropy with scaling  $c(n) = \ell_n$  under the slowness condition  $(\Delta)$ .
2. Theorem 6.10, due to Vershik ([18, 20]), states that poly-adic filtrations have a zero exponential entropy under the fastness condition  $(\nabla)$ , which is stronger than the negation  $(-\Delta)$  of the slowness condition  $(\Delta)$ .
3. Theorem 6.14 includes Theorem 6.10 when the Kolmogorovian assumption holds. It states that Kolmogorovian poly-adic filtrations have a zero exponential entropy when the slowness condition  $(\Delta)$  is not fulfilled. Its proof relies on a highly non-trivial theorem by Heicklen [7], which belongs to ergodic theory rather than probability theory.
4. Theorem 6.15, due to Vershik ([18, 20]), gives the value of the exponential entropy for the split-word filtrations under the slowness condition  $(\Delta)$ .

Gorbulsky showed Theorem 6.4 in the dyadic case  $r_n \equiv 2$  only. Our proof of the generalization to condition  $(\Delta)$  essentially uses the same mathematics. In addition to the points listed above, we will investigate the entropy of some filtrations called *filtrations of unordered pairs* in [22], which are more or less the filtrations  $\mathcal{F}^{X_0}$  generated by the process  $(\pi_n X_0)_{n \leq 0}$  of the Vershik progressive predictions of the final letter  $X_0$  of a split-word process  $(X_n)_{n \leq 0}$ .

---

<sup>1</sup>It is clear that  $h(X; \mathcal{F}) = 0$  when  $X$  is measurable with respect to  $\sigma(\eta_n, \dots, \eta_0)$  for any process of innovations  $(\eta_n)_{n \leq 0}$ . Thus, knowing that standardness of a poly-adic filtration means the existence of a generating process of innovations,  $h(\mathcal{F}) = 0$  follows from a result similar to the left-continuity property of the scaled entropy (Proposition 4.15).

## 6.1 The $\pi_n X$ in poly-adic filtrations and the exponential entropy

For poly-adic filtrations, the Vershik progressive predictions  $\pi_n X$  and the iterated Kantorovich distances  $\rho^{(n)}$  as defined in Section 2 have a convenient representation, the one given in the following lemma which is a consequence of lemma 4.6 and lemma 4.7 in [14]. For our purposes, we only state this result for a countable state space  $A$  equipped with the 0–1 metric. In this lemma and hereafter, it is understood that  $G_n$  is the group of automorphisms of the  $(r_{n+1}, \dots, r_0)$ -ary tree. If needed, the reader is referred to [14] for details about the group  $G_n$  of tree automorphisms and its action on the set of  $\ell_n$ -words  $A^{\ell_n}$ .

**Lemma 6.2.** *Let  $\mathcal{F}$  be a  $(r_n)_{n \leq 0}$ -adic filtration and  $X$  a  $\mathcal{F}_0$ -measurable random variable taking its values in a countable set  $A$ . Then  $\pi_n X$  can be identified to the  $G_n$ -orbit of a random word  $X_n$  on  $A$  having length  $\ell_n$ . Using this identification and starting with the 0–1 metric  $\rho$  on  $A$ , the  $n$ -th iterated Kantorovich metric  $\rho^{(n)}$  on  $A^{(n)}$  is transported to the metric  $\bar{\rho}_n$  on the quotient set  $A^{\ell_n}/G_n$ , given by*

$$\bar{\rho}_n(\Gamma, \Gamma') = \min_{w \in \Gamma, w' \in \Gamma'} \delta_n(w, w')$$

for every pair of orbits  $\Gamma$  and  $\Gamma'$ , where  $\delta_n(w, w')$  is the Hamming distance between the  $\ell_n$ -words  $w$  and  $w'$  (the proportion of positions at which the letters of  $w$  and  $w'$  differ).

We will use this lemma throughout this section. Though we do not provide its proof, it is easy to derive it from the first part of the following lemma.

**Lemma 6.3.** *Let  $\mathcal{F}_{-1}$  be a  $\sigma$ -field,  $\eta_0$  a random variable independent of  $\mathcal{F}_{-1}$  taking its values in a set with finite size  $r_0 \geq 2$ , which we assume to be  $\{1, \dots, r_0\}$  without loss of generality. Define the  $\sigma$ -field  $\mathcal{F}_0 = \mathcal{F}_{-1} \vee \sigma(\eta_0)$ .*

*Let  $X_0$  be a random variable taking its values in a Polish space  $A$ .*

1. *There exist  $r_0$  random variables  $X_{-1}(1), \dots, X_{-1}(r_0)$ , taking their values in  $A$  and measurable with respect to  $\mathcal{F}_{-1}$ , and such that  $X_0 = X_{-1}(\eta_0)$ .*
2. *For such random variables and when  $A$  is finite, one has  $H(X_{-1}) \leq r_0 H(X_0)$ , where  $X_{-1} = (X_{-1}(1), \dots, X_{-1}(r_0))$  and  $H$  is the Shannon entropy.*

*Proof.* For the first point, write  $X_0 = f(F_{-1}, \eta_0)$  for some Borelian function  $f$ , and set  $X_{-1}(i) = f(F_{-1}, i)$ . For the second point, check that the law of  $X_0$  is the average law of the  $X_{-1}(i)$ , hence  $H(X_{-1}) \leq H(X_{-1}(1)) + \dots + H(X_{-1}(r_0)) \leq r_0 H(X_0)$  by concavity of the Shannon entropy.  $\square$

This lemma justifies that the exponential entropy is well defined: the first part shows that  $\pi_n X$  in Definition 6.1 takes only finitely many values, and the second part implies  $H(\pi_{-1} X_0) \leq r_0 H(X_0)$  and by recursively applying this inequality one gets that  $\frac{H(\pi_n X)}{\ell_n}$  is decreasing, thereby justifying the equality  $\lim_{n \rightarrow -\infty} \frac{H(\pi_n X)}{\ell_n} = \inf_{n \leq 0} \frac{H(\pi_n X)}{\ell_n}$ .

## 6.2 Gorbulsky's theorem

In [6], Gorbulsky proved Theorem 6.4 below in the case when  $r_n \equiv 2$ . We show that this result more generally holds for poly-adicity sequences  $(r_n)_{n \leq 0}$  satisfying the slowness

condition

$$(\Delta): \quad \sum_{n=-\infty}^0 \frac{\log r_n!}{\ell_{n-1}} < \infty, \quad \text{equivalent to} \quad \sum_{n=-\infty}^0 \frac{r_n \log r_n}{\ell_{n-1}} < \infty.$$

For example, all bounded sequences  $(r_n)_{n \leq 0}$  satisfy condition  $(\Delta)$ , and it is also fulfilled in the case when  $r_n = |n| + 1$ .

**Theorem 6.4.** *Let  $\mathcal{F}$  be a  $(r_n)_{n \leq 0}$ -adic filtration. Assume that condition  $(\Delta)$  is fulfilled by the poly-adicity sequence  $(r_n)_{n \leq 0}$  and consider the scaling  $c(n) = \ell_n$ . Then the scaled entropy of  $X$  equals its exponential entropy:*

$$h_c(X; \mathcal{F}) = \lim_{n \rightarrow -\infty} \frac{H(\pi_n X)}{\ell_n}$$

for every  $\mathcal{F}_0$ -measurable random variable  $X$  taking only finitely many values. Consequently the scaled entropy  $h_c(\mathcal{F})$  of  $\mathcal{F}$  equals its exponential entropy  $h(\mathcal{F})$ .

In this theorem, it is understood that we use the Shannon entropy as the measure of entropy in the scaled entropy. The last claim of the theorem is derived from Corollary 4.17.

This section is devoted to the proof of this theorem. Only Lemma 6.8 below will be involved in the proof. It is a consequence of Lemma 6.6 and Lemma 6.7, and the result of elementary analysis stated in the following lemma will be used in Lemma 6.6.

**Lemma 6.5.** *Let  $(u_n)_{n \geq 0}$  and  $(v_n)_{n \geq 0}$  be two sequences of positive numbers. Assume that  $v_n \searrow 0$  and  $\sum u_n v_n < \infty$ . Then  $\epsilon \sum_{i=0}^{n(\epsilon)} u_i \rightarrow 0$  when  $\epsilon \rightarrow 0^+$ , where  $n(\epsilon) = \min\{n \mid v_{n+1} < \epsilon\}$ .*

*Proof.* Let  $\delta > 0$ . Take  $M_1$  such that  $\sum_{i=m+1}^n u_i v_i < \delta$  whenever  $n > m \geq M_1$ . Now take  $M_2$  such that  $v_m \sum_{i=0}^{M_1} u_i < \delta$  whenever  $m \geq M_2$ . Set  $N = \max\{M_1, M_2\}$ . If  $\epsilon \leq v_{N+1}$ , then  $n(\epsilon) > N$  and

$$\epsilon \sum_{i=0}^{n(\epsilon)} u_i \leq v_{n(\epsilon)} \sum_{i=0}^{n(\epsilon)} u_i \leq v_{n(\epsilon)} \sum_{i=0}^{M_1} u_i + \sum_{i=M_1+1}^{n(\epsilon)} u_i v_i < 2\delta,$$

thereby showing the desired result.  $\square$

In the next lemmas, we denote by  $|B|$  the number of words contained in a subset  $B \subset A^{\ell_n}$ . Recall that  $\bar{\rho}_n$  in the following lemma is the quotient distance introduced in Lemma 6.2.

**Lemma 6.6.** *Assume that condition  $(\Delta)$  is fulfilled by the poly-adicity sequence  $(r_n)_{n \leq 0}$ . For any pair of orbits  $\Gamma, \Gamma' \in A^{\ell_n}/G_n$ ,*

$$\frac{|\log |\Gamma| - \log |\Gamma'| |}{\ell_n} \leq d(\bar{\rho}_n(\Gamma, \Gamma')),$$

where  $d$  is a function satisfying  $\lim_{\epsilon \rightarrow 0} d(\epsilon) = 0$ .

*Proof.* Consider a word  $w$  of length  $\ell_n$  with its tree structure as shown by Figure 4. We denote by  $\Gamma(w)$  its  $G_n$ -orbit.

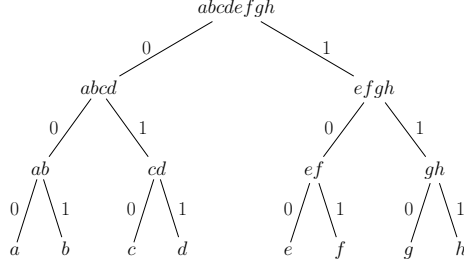


Figure 4: The tree structure of the word  $abcdefgh$ .

The word  $w$  at level  $n$  is the concatenation of the  $r_{n+1}$  words  $w_i$  of length  $\ell_{n+1}$  at level  $n+1$ . If the  $G_{n+1}$ -orbits  $\Gamma(w_i)$  of the subwords  $w_i$  are pairwise distinct, then  $|\Gamma(w)| = r_n! |\Gamma(w_1)| \dots |\Gamma(w_{r_n})|$ . If they are all equal, then  $|\Gamma(w)| = |\Gamma(w_1)| \dots |\Gamma(w_{r_n})|$ . Generally,  $|\Gamma(w)| = M_{n,1} |\Gamma(w_1)| \dots |\Gamma(w_{r_n})|$  where  $M_{n,1}$  is a multinomial coefficient lying between 1 and  $r_{n+1}!$ . Continuing so on, we find

$$\log |\Gamma(w)| = \sum_{j=n}^{-1} \sum_{i=1}^{\ell_n/\ell_j} \log M_{j,i}$$

where  $M_{j,i}$  is a multinomial coefficient lying between 1 and  $r_{j+1}!$ . Writing another word  $w'$  of length  $\ell_n$  in the same way, we get

$$\left| \log |\Gamma(w)| - \log |\Gamma(w')| \right| \leq \sum_{j=n}^{-1} \sum_{i=1}^{\ell_n/\ell_j} \left| \log M_{j,i} - \log M'_{j,i} \right|.$$

Each deviation  $|\log M_{j,i} - \log M'_{j,i}|$  is bounded by  $\log r_{j+1}!$ . If the letters of  $w$  and  $w'$  differ at  $\epsilon \ell_n$  positions, then at each level  $j$  there are at most  $\min\left(\frac{\ell_n}{\ell_j}, \epsilon \ell_n\right)$  non-zero deviations  $|\log M_{j,i} - \log M'_{j,i}|$ . Therefore,

$$\begin{aligned} \frac{\left| \log |\Gamma(w)| - \log |\Gamma(w')| \right|}{\ell_n} &\leq \sum_{j=n}^{k(\epsilon)-1} \frac{\log r_{j+1}!}{\ell_j} + \epsilon \sum_{j=k(\epsilon)}^{-1} \log r_{j+1}! \\ &\leq \sum_{j=n}^{k(\epsilon)-1} \frac{r_{j+1} \log r_{j+1}}{\ell_j} + \epsilon \sum_{j=k(\epsilon)}^{-1} r_{j+1} \log r_{j+1} \end{aligned}$$

where  $k(\epsilon) = \max\{k \mid \ell_{k-1}^{-1} < \epsilon\}$ .

Under the  $(\Delta)$  condition, the first sum in the right member goes to  $\sum_{j=-\infty}^{k(\epsilon)-1} \frac{r_{j+1} \log r_{j+1}}{\ell_j}$  when  $n \rightarrow -\infty$ , and this goes to 0 when  $\epsilon \rightarrow 0$  because  $k(\epsilon)$  goes to  $-\infty$ . The second sum goes to 0 too because of Lemma 6.5.  $\square$

**Lemma 6.7.** *For any subset  $B \subset A^{\ell_n}$ , the log-number of words in an  $\epsilon$ -neighbourhood of  $B$  does not exceed*

$$\log |B| + h(\epsilon) \ell_n + \epsilon \log(\#A - 1) \ell_n$$

where  $h(\epsilon) = \epsilon \log \frac{1}{\epsilon} + (1 - \epsilon) \log \frac{1}{1-\epsilon}$ .

*Proof.* The words in an  $\epsilon$ -ball around a  $\ell_n$ -word  $w$  are obtained by taking  $k = 0, \dots, \lfloor \epsilon \ell_n \rfloor$  positions in  $w$  and changing the letters at these positions. Then the number of such words

is

$$(\#A - 1)^{\lfloor \epsilon \ell_n \rfloor} \times \sum_{k=0}^{\lfloor \epsilon \ell_n \rfloor} \binom{\ell_n}{k}.$$

The lemma follows from the inequality

$$\log \sum_{k=0}^{\lfloor \epsilon \ell_n \rfloor} \binom{\ell_n}{k} \leq h(\epsilon) \ell_n,$$

which is derived from the classical large deviations inequality for independent symmetric Bernoulli variables (corollary 2.20 in [8]).  $\square$

**Lemma 6.8.** *Under condition  $(\Delta)$ , for any orbit  $\Gamma \in A^{\ell_n}/G_n$ , the log-number of orbits in an  $\epsilon$ -neighborhood of  $\Gamma$  (for the  $\bar{\rho}_n$  distance) does not exceed a value  $L_n^\epsilon$  satisfying*

$$\lim_{\epsilon \rightarrow 0} \limsup_{n \rightarrow \infty} \frac{L_n^\epsilon}{\ell_n} = 0.$$

*Proof.* The number of orbits in an  $\epsilon$ -neighbourhood  $V_\epsilon(\Gamma)$  of  $\Gamma$  is less than the number of words in  $V_\epsilon(\Gamma)$  divided by the minimal length of an orbit  $\Gamma'$  in  $V_\epsilon(\Gamma)$ . Let  $L_n^\epsilon$  be this ratio. Applying the two previous lemmas yields the desired result:

$$L_n^\epsilon \leq \left[ \log |\Gamma| + h(\epsilon) \ell_n + \epsilon \log(\#A - 1) \ell_n \right] - \left[ \log |\Gamma| - d(\epsilon) \ell_n \right] = \ell_n \left( h(\epsilon) + \epsilon \log(\#A - 1) + d(\epsilon) \right).$$

$\square$

*Proof of Theorem 6.4.*

In the proof we use notation  $H(E) = -\sum_{x \in E} \mu(x) \log \mu(x)$  where  $\mu$  is the law of  $\pi_n X$  and  $E$  is any set of  $G_n$ -orbits. Note that

$$H(E) \leq \mu(E) \log \#E - \mu(E) \log \mu(E)$$

because  $H(E) = \mu(E) H(\mu') - \mu(E) \log \mu(E)$  where  $\mu' = \mu(\cdot | E)$ .

We use the  $\epsilon$ -entropy as defined in Lemma 4.9. Let  $\{A_i, B\}$  be a partition achieving  $H_n^\epsilon(X)$ , with  $A_i \subset B(\Gamma_i, \epsilon)$  and  $\mu(B) \leq \epsilon$ . One has

$$H_n^\epsilon(X) = -\sum_i \mu(A_i) \log \mu(A_i) - \mu(B) \log(B)$$

and

$$H(\pi_n X) = \sum_i H(A_i) + H(B).$$

Firstly,  $H(B) \leq (\log \#A) \epsilon \ell_n - \epsilon \log \epsilon$ . On the other hand,

$$H(A_i) \leq \mu(A_i) L_n^\epsilon - \mu(A_i) \log \mu(A_i),$$

hence

$$\sum_i H(A_i) \leq H_n^\epsilon(X) + L_n^\epsilon.$$

Thus,

$$\frac{H_n^\epsilon(X)}{\ell_n} \leq \frac{H(\pi_n X)}{\ell_n} \leq \frac{H_n^\epsilon(X)}{\ell_n} + \frac{L_n^\epsilon - \epsilon \log \epsilon}{\ell_n} + (\log \#A) \epsilon,$$

thereby yielding the theorem.  $\square$

The following corollary just emphasizes that the equality  $h_c(X) = h(X)$  of Theorem 6.4 can be stated not only for poly-adic filtrations, but also for filtrations immersed in a poly-adic filtration.

**Corollary 6.9.** *Let  $\mathcal{F}$  be a filtration immersed in a  $(r_n)_{n \leq 0}$ -adic filtration. Consider the scaling  $c(n) = \ell_n$ . If  $(r_n)_{n \leq 0}$  fulfills the  $(\Delta)$  condition, then*

$$h_c(X) = \lim_{n \rightarrow -\infty} \frac{H(\pi_n X)}{\ell_n}$$

for every simple  $\mathcal{F}_0$ -measurable random variable  $X$ .

*Proof.* This is a direct consequence of Theorem 6.4 and the immersion property (Remark 4.8).  $\square$

### 6.3 $(\nabla)$ -adic filtrations have zero exponential entropy

The  $(\nabla)$  condition is stronger than the negation of the  $(\Delta)$  condition, because this condition is the divergence of a certain sequence whereas the  $(\Delta)$  condition is the convergence of the series made up of the same sequence:

$$(\nabla): \quad \frac{\log r_n}{\ell_n} \rightarrow \infty.$$

Vershik's following theorem is proved in [20]. Note that the hypotheses do not require  $\mathcal{F}$  to be Kolmogorovian.

**Theorem 6.10.** *Let  $\mathcal{F}$  be a  $(\nabla)$ -adic filtrations. Then it has zero exponential entropy when we use the max-entropy as the underlying measure of entropy  $H$  (see above Definition 4.1). Consequently it also has zero exponential entropy when we use the Shannon entropy as the underlying measure of entropy, and the same result holds for the scaled entropy with the scaling  $c(n) = \ell_n$ .*

The proof is based on the following combinatorial lemma.

**Lemma 6.11.** *For an alphabet  $A$  having size  $\#A = p$ , the number of orbits  $\chi_n^p$  of the action of  $G_n$  on  $A^{\ell_n}$  is given by  $\chi_0^p = p$  and the recurrence formula*

$$\chi_{n-1}^p = \binom{r_n + \chi_n^p - 1}{r_n},$$

and one has  $\log \chi_n^p = o(\ell_n)$  when condition  $(\nabla)$  holds.

*Proof.* The recurrence formula obviously stems from the fact that an orbit for the action of  $G_{n-1}$  is obtained by choosing a list of  $r_n$  orbits for the action of  $G_n$ , with possible repetitions. Then

$$\log \chi_{n-1}^p = \log[(r_n + \chi_n^p - 1) \cdots (r_n + 1)] - \log[(\chi_n^p - 1)!] \leq \chi_n^p \log r_n$$

by subadditivity of the logarithm. Set  $t_n = \frac{\log \chi_n^p}{\ell_n}$ . Note that  $r_n = \exp(\beta_n \ell_n)$  where  $\beta_n$  is the quantity going to  $\infty$  under the  $(\nabla)$  condition. Thus

$$t_{n-1} \leq \frac{\chi_n^p \log r_n}{\ell_{n-1}} = \frac{\exp(t_n \ell_n) \beta_n \ell_n}{\ell_n \exp(\beta_n \ell_n)} = \beta_n \exp((t_n - \beta_n) \ell_n).$$

The number of orbits cannot exceed the number of words, and this yields the inequality  $t_n \leq \log p$ . Therefore the right member of the last inequality goes to 0 under the  $(\nabla)$  condition.  $\square$

*Proof of Theorem 6.10.* Given any  $\mathcal{F}_0$ -measurable random variable  $X$  taking its values in a finite set  $A$ , one has  $H_0(\pi_n X) \leq \log \chi_n^{\#A}$  with the notations of the previous lemma, where  $H_0$  is the max-entropy. Then the result for the max-entropy follows from this lemma, and the result for the Shannon entropy  $H$  follows because of  $H \leq H_0$ . The result for the scaled entropy follows from the obvious inequality  $h_c(\mathcal{F}) \leq h(\mathcal{F})$ .  $\square$

## 6.4 $(\neg\Delta)$ -adic filtrations have zero exponential entropy

Theorem 6.14 given in this section implies Theorem 6.10 in the case of a Kolmogorovian filtration, but its proof relies on Heicklen's theorem, rephrased below in Theorem 6.13, whose proof is far to be trivial and belongs to ergodic theory rather than probability theory.

Heicklen's theorem deals with the ergodic free actions of the group  $\Gamma = \bigoplus_{n=-\infty}^0 \mathbb{Z}/r_n \mathbb{Z}$  on a Lebesgue space, where  $(r_n)_{n \leq 0}$  is a sequence of integers as before. One can write  $\Gamma = \bigcup_{n=0}^{-\infty} \Gamma_n$  where  $\Gamma_n = \sum_{k=n+1}^0 \mathbb{Z}/r_k \mathbb{Z}$ . Then, when a  $\Gamma$ -action on a Lebesgue space is given, one can associate to it a filtration  $(\mathcal{F}_n)_{n \leq 0}$  on the Lebesgue space by defining  $\mathcal{F}_n$  as the  $\sigma$ -field of  $\Gamma_n$ -invariant sets. This filtration is Kolmogorovian when the  $\Gamma$ -action is ergodic, and it is  $(r_n)_{n \leq 0}$ -adic when the  $\Gamma$ -action is free. Conversely, any Kolmogorovian  $(r_n)_{n \leq 0}$ -adic filtration on a Lebesgue space can be derived in this way from a free ergodic  $\Gamma$ -action. We refer the reader to [7] or [5] for these claims.

In the present paper we never assume that the filtrations  $\mathcal{F}$  are defined on a Lebesgue space. But when the final  $\sigma$ -field  $\mathcal{F}_0$  is essentially separable, then  $\mathcal{F}$  is isomorphic to a Lebesgue space (see [1]). Thus, when restricting oneself to filtrations  $\mathcal{F}$  whose final  $\sigma$ -field  $\mathcal{F}_0$  is essentially separable, one can say that every Kolmogorovian  $(r_n)_{n \leq 0}$ -adic filtration comes from a free ergodic  $\Gamma$ -action in the above way.

The entropy of the filtration  $\mathcal{F}$  corresponding to a  $\Gamma$ -action are related, as shown by the next lemma. The definition of the entropy of a  $\Gamma$ -action is given in the proof of this lemma.

**Lemma 6.12.** *Let  $T$  be a free ergodic  $\Gamma$ -action on a Lebesgue space, and  $\mathcal{F}$  be its corresponding filtration. Then  $h(\mathcal{F}) \leq h(T)$ .*

*Proof.* Let  $X_0$  be a simple  $\mathcal{F}_0$ -measurable taking its values in a finite set  $A$ . Let  $P_0$  be a finite partition of the Lebesgue space generating the same  $\sigma$ -field as  $X_0$ . Define the partition  $P_n = \bigvee_{g \in \Gamma_n} T^g P_0$ . The  $\sigma$ -field generated by the partition  $P_n$  is the same as the  $\sigma$ -field generated by a  $\mathcal{F}_n$ -measurable random word  $X_n$  having  $\ell_n$  letters on the alphabet  $A$ , and which is one of the  $r_n$  subwords of  $X_{n-1}$ , uniformly taken at random, when one splits  $X_{n-1}$  into  $r_n$  subwords each having  $\ell_n$  letters. Therefore, the  $n$ -th Vershik progressive prediction  $\pi_n X_0$  of  $X_0$  is the orbit of  $X_n$  under the action of the group  $G_n$  of tree automorphisms. We refer to [14] for this claim, as we referred for Lemma 6.2. Hence, the exponential entropy  $h(\mathcal{F}, X_0)$  of  $X_0$ ,

$$h(\mathcal{F}, X_0) := \lim_{n \rightarrow -\infty} \frac{H(\pi_n X_0)}{\ell_n},$$

satisfies

$$h(\mathcal{F}, X_0) \leq \lim_{n \rightarrow -\infty} \frac{H(X_n)}{\ell_n} =: h(T, P_0) \leq h(T) := \sup_{P \text{ finite partition}} h(T, P).$$

Therefore  $h(\mathcal{F}) \leq h(T)$ .  $\square$

Heicklen proved the following theorem.

**Theorem 6.13** (Heicklen [7]). *Let  $T$  be a free ergodic  $\Gamma$ -action on a Lebesgue space, and  $\mathcal{F}$  be its corresponding filtration. When the  $(\Delta)$  condition about the sequence  $(r_n)_{n \leq 0}$  does not hold, there exists an action  $S$  with zero entropy and whose corresponding filtration is isomorphic to  $\mathcal{F}$ .*

From this theorem and the previous lemma, one easily gets the following theorem.

**Theorem 6.14.** *Let  $\mathcal{F}$  be a Kolmogorovian  $(\neg\Delta)$ -adic filtration, whose final  $\sigma$ -field  $\mathcal{F}_0$  is essentially separable. Then it has a zero exponential entropy, and consequently it also has a zero scaled entropy  $h_c(\mathcal{F})$  with the scaling  $c(n) = \ell_n$ .*

*Proof.* Up to isomorphism,  $\mathcal{F}$  is a filtration on a Lebesgue space, and it is the filtration corresponding by the way explained above to a free ergodic  $\Gamma$ -action on this Lebesgue space. The result follows from Heicklen's theorem and Lemma 6.12.  $\square$

## 6.5 $(\Delta)$ -adic split-word filtrations

The poly-adic filtrations of the split-word processes with i.i.d. letters were studied in [16], [4], [3], [10]. In the more general case of stationary letters, standardness of these filtrations is closely connected, as shown in [14], to the notion of scale of an automorphism introduced by Vershik in [19]. Theorem 6.15 below, which is a rephrasing of theorem 4.1 in [20], provides the exponential entropy of these filtrations under condition  $(\Delta)$ . The hypotheses of the theorem do not require  $\mathcal{F}$  to be Kolmogorovian.

Given a sequence of integers  $(r_n)_{n \leq 0}$ , setting as before  $\ell_n = r_{n+1} \dots r_0$ , and given an alphabet  $A$ , a  $(r_n)_{n \leq 0}$ -adic split-word process on  $A$  is a Markov process  $(X_n, \eta_n)_{n \leq 0}$  satisfying conditions below for each  $n \leq 0$ , where we denote by  $\mathcal{F}$  the filtration it generates, the following conditions:

- $X_n$  is a random word on  $A$  of length  $\ell_n$ ;
- $\eta_n$  is a random variable uniformly distributed on  $\{1, 2, \dots, r_n\}$  and is independent of  $\mathcal{F}_{n-1}$ , and the word  $X_n$  is the  $\eta_n$ -th letter of  $X_{n-1}$  treated as a  $r_n$ -word on  $A^{\ell_n}$ .

Obviously the filtration  $\mathcal{F}$  generated by  $(X_n, \eta_n)_{n \leq 0}$  is a  $(r_n)_{n \leq 0}$ -adic filtration for which  $(\eta_n)_{n \leq 0}$  is a process of innovations.

For example, one can define such a process by taking a stationary probability measure on  $A^{\mathbb{Z}}$  and then taking for the law of  $X_n$  the projection of this measure on  $\ell_n$  consecutive coordinates. In this case and when  $A$  is countable, by standard ergodic theory, the Kolmogorov entropy of this stationary probability measure can be written

$$\theta_b = \lim_{n \rightarrow -\infty} \frac{H(X_n)}{\ell_n} \in (0, +\infty) \quad (6.1)$$

when we use the Shannon entropy  $H$  with logarithmic base  $b$ . More generally, it follows from Lemma 6.3 that this limit  $\theta_b$  exists for any split-word process.

With the terminology of [9],  $\mathcal{F}$  is an adic filtration on the Bratteli graph shown on Figure 5, called the graph of the ordered pairs by Vershik (in contrast with the graph of unordered pairs that we will see in Section 6.6).



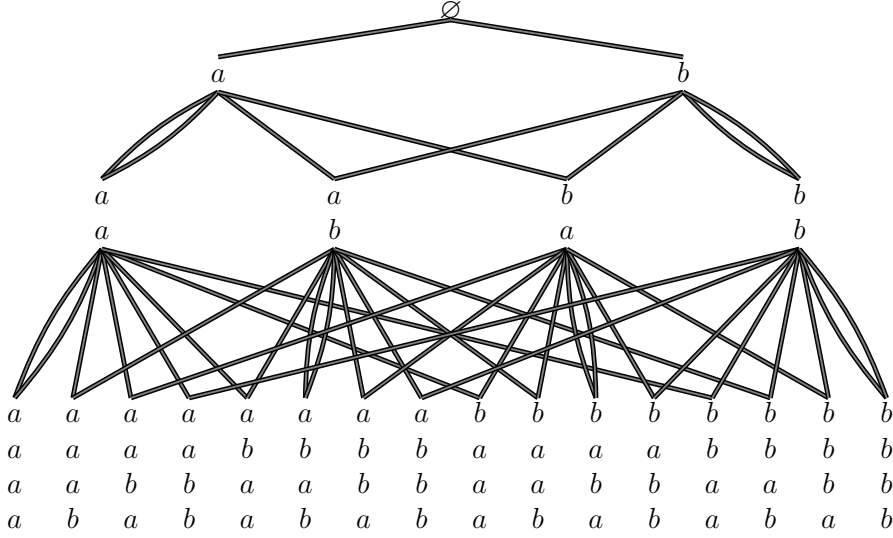


Figure 5: The Brattelli graph of the ordered pairs.

The proof of the theorem involves the cardinal of the group of tree automorphisms  $G_n$  given by

$$\log \#G_n = \ell_n \sum_{m=n+1}^0 \frac{\log r_m!}{\ell_{m-1}}.$$

**Theorem 6.15.** *For the scaling  $c(n) = \ell_n$  and under condition  $(\Delta)$ , the scaled entropy of  $\mathcal{F}$  is  $h_c(\mathcal{F}) = \theta_b$  when we use the Shannon entropy in logarithmic base  $b$  as the underlying measure of entropy. It is also the exponential entropy of  $\mathcal{F}$  because of Theorem 6.4.*

*Proof.* By Theorem 4.19 and Proposition 4.20, we know that  $h_c(\mathcal{F}) = \lim_{m \rightarrow -\infty} h_c(X_m; \mathcal{F})$ . And we know that  $h_c(X_m; \mathcal{F}) = \lim_{n \rightarrow -\infty} \frac{H(\pi_n X_m)}{\ell_n}$  by Theorem 6.4.

We firstly compare  $H(\pi_n X_0)/\ell_n$  with  $H(X_n)/\ell_n$ . Recall that  $\pi_n X_0$  is the orbit of  $X_n$  under the action of the group of tree automorphisms  $G_n$ . Below we denote the group of tree automorphisms by  $G_n(\{r_i\}_{i=n+1}^0)$  to show its dependence on the  $r_n$ . Therefore  $H(\pi_n X_0) = H(X_n) - H(X_n | \pi_n X_0)$ , and the conditional entropy  $H(X_n | \pi_n X_0)$  is less than the logarithm of the length of the orbit  $\pi_n X_0$ , and a fortiori it is less than the logarithm of the number of tree automorphisms. Thus,

$$H(X_n) \geq H(\pi_n X_0) \geq H(X_n) - \log \#G_n(\{r_i\}_{i=n+1}^0).$$

In the same way, for  $n < m$ ,

$$H(X_n) \geq H(\pi_n X_m) \geq H(X_n) - \log \#G_{n-m}(\{r_i\}_{i=n+1}^m),$$

therefore

$$\frac{H(X_n)}{\ell_n} \geq \frac{H(\pi_n X_m)}{\ell_n} \geq \frac{H(X_n)}{\ell_n} - \frac{\log \#G_{n-m}(\{r_i\}_{i=n+1}^m)}{\ell_n} \quad (6.2)$$

$$= \frac{H(X_n)}{\ell_n} - \sum_{k=n}^m \frac{\log r_{k+1}!}{\ell_k}, \quad (6.3)$$

and

$$\theta \geq \lim_{n \rightarrow -\infty} \frac{H(\pi_n X_m)}{\ell_n} \geq \theta - \sum_{k=-\infty}^m \frac{\log r_{k+1}!}{\ell_k}.$$

The  $(\Delta)$  condition being  $\lim_{m \rightarrow -\infty} \sum_{k=-\infty}^m \frac{\log r_{k+1}!}{\ell_k} = 0$ , the proof is over.  $\square$

To illustrate Theorem 4.19 and Theorem 4.22, consider the filtration  $\mathcal{G}$  of the split-word process  $(X_n, \eta_n)_{n \leq 0}$ , with uniform letters on an alphabet  $A = \{a, b, c, d\}$  having four letters. We know that  $h_c(\mathcal{G}) = \log 4$  by the previous theorem. Now take the function  $f: \{a, b, c, d\} \rightarrow \{0, 1\}$  sending  $a$  and  $b$  to 0, and sending  $c$  and  $d$  to 1. Then the process  $(f(X_n), \eta_n)_{n \leq 0}$ , where  $f(X_n)$  denotes the word obtained by applying  $f$  to the letters of  $X_n$ , is the split-word process with uniform letters on the alphabet  $\{0, 1\}$ . The scaled entropy of its filtration  $\mathcal{F}$  is  $h_c(\mathcal{F}) = \log 2$ . The filtration  $\mathcal{F}$  is immersed in  $\mathcal{G}$ . However, Theorem 4.19 shows that  $\mathcal{G}$  is not immersible in a parametric extension of  $\mathcal{F}$  and Theorem 4.22 shows that  $\mathcal{G}$  is not immersible in an independent enlargement of  $\mathcal{F}$  with a standard filtration.

As a side note, let us mention that, for some reasons beyond the scope of this paper, every Kolmogorovian  $(r_n)_{n \leq 0}$ -adic filtration whose final  $\sigma$ -field is essentially separable is generated by a split-word process on a countable alphabet as long as its final  $\sigma$ -field is essentially separable. This follows from the ergodic theoretic fact that such a filtration can always be derived from a free ergodic  $\Gamma$ -action on a Lebesgue space, as said in Section 6.4, and from the existence of a countable generator for such an action.

## 6.6 Dyadic filtrations of unordered pairs

Let  $(X_n, \epsilon_n)_{n \leq 0}$  be a dyadic split-word process and  $\mathcal{F}$  the filtration it generates. Consider the scaling  $c(n) = \ell_n = 2^{|n|}$ . We know the scaling entropy of  $\mathcal{F}$  by Theorem 6.15. It is interesting to wonder about the scaled entropy  $h_c(X_0)$  of the final letter  $X_0$ , that is to say, in view of Theorem 4.11, the scaled entropy  $h_c(\mathcal{F}^{X_0})$  of the filtration  $\mathcal{F}^{X_0}$  generated by the Markov process  $(\pi_n X_0)_{n \leq 0}$ . Here we provide a result for the case of an alphabet  $A = \{a, b\}$  having only two letters.

The Markov process  $(\pi_n X_0)_{n \leq 0}$  can be seen as a random walk on the vertices of the *graph of the unordered pairs* shown on Figure 6 below, and which can be found in [22].

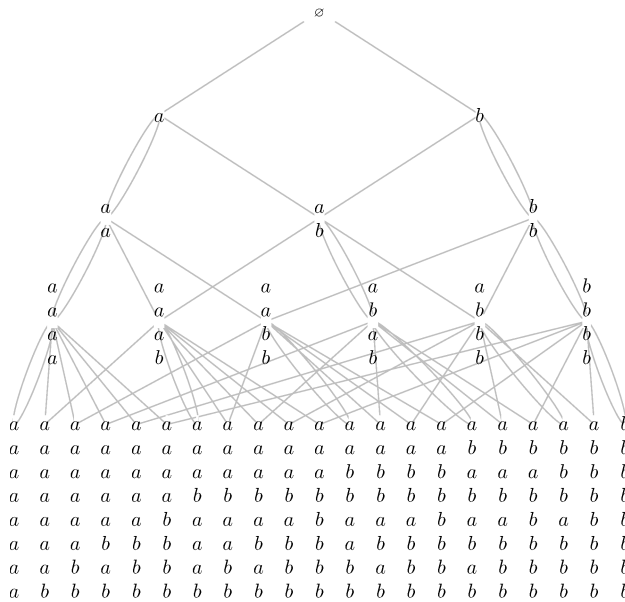


Figure 6: The Bratteli graph of the unordered pairs ([22]).

The filtration  $\mathcal{F}^{X_0}$  generated by  $(\pi_n X_0)_{n \leq 0}$  is not dyadic, because  $\pi_{n+1} X_0$  is deterministic given  $\mathcal{F}_n$  for certain values of  $\pi_n X_0$ . This is shown on Figure 6 by the double edges. Nevertheless, one has  $h_c(X_0) = h(X_0)$  by Corollary 6.9.

Indeed, there is a dyadic superinnovation of  $\mathcal{F}^{X_0}$ , that is to say a sequence  $(\epsilon'_n)_{n \leq 0}$  of independent symmetric Bernoulli random variables which is a superinnovation (see Theorem 4.19) of  $\mathcal{F}^{X_0}$  and which is a process of innovations of the enlarged filtration  $\mathcal{G}^{X_0}$  defined by  $\mathcal{G}_n^{X_0} = \mathcal{F}_n^{X_0} \vee \sigma(\epsilon'_m; m \leq n)$ . To construct such a dyadic superinnovation, consider for every  $n \leq 0$  an arbitrary but fixed order on the set of  $G_n$ -orbits, and set

$$\epsilon'_n = \begin{cases} \epsilon_n & \text{if } \pi_{n-1} X_0 \text{ is symmetric;} \\ 1 & \text{if } \pi_{n-1} X_0 = \{\Gamma_1, \Gamma_2\} \text{ with } \Gamma_1 < \Gamma_2 \text{ and } \pi_n X_0 = \Gamma_1; \\ 2 & \text{if } \pi_{n-1} X_0 = \{\Gamma_1, \Gamma_2\} \text{ with } \Gamma_1 < \Gamma_2 \text{ and } \pi_n X_0 = \Gamma_2. \end{cases}$$

As we previously recalled,  $h_c(X_0) = h_c(\mathcal{F}^{X_0})$  because of Theorem 4.11. Moreover we know that  $h_c(\mathcal{F}^{X_0}) = h_c(\mathcal{G}^{X_0})$  by Theorem 4.19. In fact it can be shown, with the help of Theorem 4.19 and lemma 5.3 in [9], that one can always "drop" the multiple edges when we are interested in the scaling entropy of the filtration associated (in the way explained in [9]) to a Bratteli graph endowed with a central probability measure.

The bounds for  $h_c(X_0)$  we give in Proposition 6.18 are derived from the two following lemmas. The first one, giving the maximal length of a  $G_n$ -orbit, is a copy of lemma 3.6 in [20], to which we refer for the proof. The second one gives an asymptotic equivalent of the number of  $G_n$ -orbits (the number of vertices at level  $n$  of the graph of the unordered pairs). In [20], it is stated in lemma 3.7 but the given value of  $\gamma$  is not correct.

**Lemma 6.16.** *For an alphabet with two letters, the maximal length of a  $G_n$ -orbit is  $2^{\frac{3}{4}2^{|n|-1}}$  for every  $n \leq 2$ .*

**Lemma 6.17.** *For an alphabet with two letters, the base 2 logarithm of the number  $\chi_n$  of  $G_n$ -orbits is equivalent to  $\gamma 2^{|n|}$  where  $0.428 < \gamma < 0.429$ .*

*Proof.* It is easy to see that the number of orbits  $\chi_n$  is given by  $\chi_0 = 2$  and  $\chi_{n-1} = \frac{\chi_n(\chi_n+1)}{2}$  (this is a particular case of Lemma 6.11). By the equality

$$\frac{\log_2 \chi_{n-1}}{2^{|n-1|}} = \frac{\log_2 \chi_n}{2^{|n|}} + \frac{\log_2 \left(1 + \frac{1}{\chi_n}\right) - 1}{2^{|n-1|}},$$

the sequence  $\frac{\log_2 \chi_n}{2^{|n|}}$  is decreasing, and for every  $n < n_0 < 0$ ,

$$\frac{\log_2 \chi_{n_0}}{2^{|n_0|}} - \frac{1}{2^{|n_0|}} \leq \frac{\log_2 \chi_n}{2^{|n|}} \leq \frac{\log_2 \chi_{n_0}}{2^{|n_0|}}.$$

Taking  $n_0 = -11$  gives the bounds on the limit  $\gamma$ . □

**Proposition 6.18.** *For an alphabet with two letters, and taking the scaling  $c(n) = \ell_n = 2^{|n|}$ ,*

$$\theta_2 - \frac{3}{4} \leq h(X_0) \leq \min(\theta_2, \gamma)$$

where  $\gamma$  is given in Lemma 6.17 and  $\theta_2$  is defined in equation (6.1). Here we use the Shannon entropy with logarithmic base 2 as the underlying measure of entropy and  $h(X_0)$  is the exponential entropy (Definition 6.1), but by Theorem 6.4 it is the same as the scaled entropy  $h_c(X_0)$ .

*Proof.* We start, as in the proof of Theorem 6.15, with the equality  $H(\pi_n X_0) = H(X_n) - H(X_n \mid \pi_n X_0)$ . But this time we bound from above the conditional entropy  $H(X_n \mid \pi_n X_0)$  by the logarithm of the maximal length of a  $G_n$ -orbit. In addition we bound from above  $H(\pi_n X)$  by the logarithm of the number of  $G_n$ -orbits. Then, using Lemma 6.16 and Lemma 6.17, we get

$$\frac{H(X_n)}{2^{|n|}} - \frac{3}{4} \leq \frac{H(\pi_n X_0)}{2^{|n|}} \leq \min \left\{ \frac{H(X_n)}{2^{|n|}}, \frac{\log_2 \chi_n}{2^{|n|}} \right\},$$

and then the result follows by taking the limit.  $\square$

For example, this result shows that the filtration  $\mathcal{F}^{X_0}$  in the uniform case  $\theta_2 = 1$  is not isomorphic to the filtration  $\mathcal{F}^{X_0}$  in a case when  $\theta_2 < \frac{1}{4}$ .

As an application of Theorem 4.19 and Theorem 4.22, consider the uniform case  $\theta_2 = 1$ . Then we know that  $h_c(\mathcal{F}) = \theta_2$  by Theorem 6.15, whereas  $h_c(\mathcal{F}^{X_0}) \leq \gamma < \theta_2$  by Proposition 6.18. Thus by Theorem 4.19 we know that  $\mathcal{F}$  is not immersible in a parametric extension of  $\mathcal{F}^{X_0}$  and by Theorem 4.22, we know that  $\mathcal{F}$  is not immersible in an independent enlargement of  $\mathcal{F}^{X_0}$  with a standard filtration.

## References

- [1] Barlow, M., Émery, M., Knight, F., Song, S., Yor, M.: Autour d'un théorème de Tsirelson sur des filtrations browniennes et non-browniennes. Séminaire de Probabilités XXXII (Springer Lectures Notes in Math. 1686, Berlin 1998), 264–305.
- [2] Bolley, F.: Separability and completeness for the Wasserstein distance. Séminaire de Probabilités XLI, Springer Lecture Notes in Mathematics 1934, 371–377 (2008).
- [3] Ceillier, G.: The filtration of the split-words process. Probability Theory and Related Fields 153, Issue 1–2, 269–292 (2012).
- [4] Émery, M., Schachermayer, W.: On Vershik's standardness criterion and Tsirelson's notion of cosiness. Séminaire de Probabilités XXXV, Springer Lectures Notes in Math. 1755 (2001), 265–305.
- [5] Fieldsteel, A., Hasfura-Buenaga, J.R.: Dyadic equivalence to completely positive entropy. Transactions of the American Mathematical Society, Vol 350, Number 3 (1998), 1143–1166.
- [6] Gorbulsky, A.D.: Interrelations between various definitions of the entropy of decreasing sequences of partitions; scaling. J. Math. Sci. (New York) 121, No. 3 (2004), 2319–2325.
- [7] Heicklen, D.: Entropy and  $r$ -equivalence. Ergodic Theory and Dynamical Systems 18, no. 5 (1998), 1139–1157.
- [8] van der Hofstad, R.: Random Graphs and Complex Networks. To appear.
- [9] É. Janvresse, S. Laurent, T. de la Rue: Standardness of monotonic Markov filtrations. arXiv:1501.02166 (2015). To appear in: Markov Processes and Related Fields.

- [10] Laurent, S.: On standardness and I-cosiness. Séminaire de Probabilités XLIII, Springer Lecture Notes in Mathematics 2006, 127–186 (2010).
- [11] Laurent, S.: On Vershikian and I-cosy random variables and filtrations. *Teoriya Veroyatnostei i ee Primeneniya* 55 (2010), 104–132. Also published in: *Theory Probab. Appl.* 55 (2011), 54–76.
- [12] Laurent, S.: Further comments on the representation problem for stationary processes. *Statist. Probab. Lett.* 80 (2010), 592–596.
- [13] Laurent, S.: Standardness and non-standardness of next-jump time filtrations. *Electronic Communications in Probability* 18 (2013), no. 56, 1–11.
- [14] Laurent, S.: 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] Laurent, S.: The reversed discrete hazard rate and the scaled entropy of a discrete measure. Preprint (2016). <https://hal.archives-ouvertes.fr/hal-01251793>
- [16] Smorodinsky, M.: Processes with no standard extension. *Israel Journal of Mathematics* 107, 327–331 (1998).
- [17] Vershik, A.M.: Decreasing sequences of measurable partitions, and their applications. *Dokl. Akad. Nauk SSSR* 193, 748–751 (1970). English translation: *Soviet Math. Dokl.* 11, 1007–1011 (1970).
- [18] Vershik, A.M.: Continuum of pairwise nonisomorphic diadic sequences. *Funktional’nyi Analiz i Ego Prilozheniya* 5:3, 16–18 (1971). English translation: *Functional Analysis and Its Applications* 5:3, 182–184 (1971).
- [19] Vershik, A.M.: Four definitions of the scale of an automorphism. *Funktional’nyi Analiz i Ego Prilozheniya*, 7:3, 1–17 (1973). English translation: *Functional Analysis and Its Applications*, 7:3, 169–181 (1973).
- [20] Vershik, A.M.: The theory of decreasing sequences of measurable partitions (in Russian). *Algebra i Analiz*, 6:4, 1–68 (1994). English translation: *St. Petersburg Mathematical Journal*, 6:4, 705–761 (1995).
- [21] Vershik, A.M, Gorbulsky, A.D.: Scaled entropy of filtrations of  $\sigma$ -fields. *Theory of Probability and its Applications*, **52**, 3, 493–508, (2008).
- [22] A. M. Vershik, The problem of describing central measures on the path spaces of graded graphs. *Funct. Anal. Appl.* 48, No. 4, 26–46 (2014).