



HAL
open science

Uniform entropy scalings of filtrations

Stéphane Laurent

► **To cite this version:**

| Stéphane Laurent. Uniform entropy scalings of filtrations. 2015. hal-01006337v3

HAL Id: hal-01006337

<https://hal.science/hal-01006337v3>

Preprint submitted on 22 Jun 2015 (v3), last revised 16 Aug 2016 (v5)

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

June 22, 2015

Abstract

We study Vershik and Gorbulsky's notion of entropy scalings for filtrations in the particular case when the scaling is not ϵ -dependent, and is then termed as *uniform scaling*. Our main result states 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 rephrase some old theorems by Vershik and provide a generalization of a theorem by Gorbulsky about the case of poly-adic filtrations.

Contents

1	Introduction	2
2	Vershik's standardness criterion	2
2.1	Vershik's standardness criterion	2
2.2	Properties to be generalized later	3
2.3	Vershik's standardness criterion in practice	3
3	The next-jump time filtrations	5
3.1	Next-jump time process as a random walk on a Bratteli graph	5
3.2	Standardness of \mathcal{F} using Vershik's criterion	7
3.3	Iterated Kantorovich distances	9
4	The uniformly scaled entropy	10
4.1	Definition	10
4.2	Main theorem	12
4.3	Other properties	14
5	Entropy of next-jump time filtrations	15
5.1	Entropy of V_n	15
5.2	Entropy of the filtration	16
6	Entropy of poly-adic filtrations	17
6.1	The $\pi_n X$ in poly-adic filtrations and the exponential entropy	18
6.2	Gorbulsky's theorem	19
6.3	∇ -adic filtrations	21
6.4	Δ -adic split-word filtrations	22
6.5	Dyadic filtrations of unordered pairs	23

1 Introduction

This is the first paper written in the probabilistic language about Vershik & Gorbulsky's theory of scaled entropy introduced in [19]. We focus on the case of *uniform entropy scalings*. Our results contain as particular cases some known results 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 [12] (where I-cosiness was used to derive this standardness 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. Section 6 deals with the exponential entropy for poly-adic filtrations. We 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, and we investigate the entropy of the filtrations of unordered pairs.

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 [3, 9, 10, 11]). As long as the final σ -field \mathcal{F}_0 is essentially separable, standardness is known to be equivalent to Vershik's standardness criterion. In this section we recall the statement of Vershik's standardness criterion and we state some of its elementary properties which are proved in [10]. In Section 4 we will see that these properties are particular cases of some elementary properties 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, ρ) , the Kantorovich distance $\rho'(\mu, \nu)$ between two probability measures μ and ν 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 μ and ν , that is, the set of probabilities on $E \times E$ whose first and second marginal measures are μ and ν respectively. In general $\rho'(\mu, \nu)$ is possibly infinite, but ρ' defines a distance on the space E' of integrable probability measures on (E, ρ) , where a probability measure μ on (E, ρ) is said to be integrable when 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 ρ' on E' is finer than the topology of weak convergence, but they coincide when (E, ρ) is compact, and (E', ρ') is itself compact in this case. We mainly use the fact that the metric space (E', ρ') is complete and separable whenever (E, ρ) is (see e.g. [1]).

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 [15] and [18]) and the iterated Kantorovich distance $\rho^{(n)}$ on the state space $E^{(n)}$ of $\pi_n X$. Let (E, ρ) be a Polish metric space. For a σ -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 ρ : we recursively define $\rho^{(n)}$ by putting $\rho^{(0)} = \rho$ and by defining $\rho^{(n-1)} = (\rho^{(n)})'$ as the Kantorovich distance issued from $\rho^{(n)}$.

The proof of the following lemma is straightforward from the definitions.

Lemma 2.1. *For any Polish space (E, ρ) and $X, Y \in L^1(\mathcal{F}_0; E)$, the 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 .*

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, ρ) . 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 σ -fields $\mathcal{E}_0 \subset \mathcal{F}_0$ and to the whole filtration as follows: we say that a σ -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 σ -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 [3] or [9] 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.

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 σ -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 scaling entropy generalize the following propositions and theorem which are provided in [10].

Proposition 2.2. *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 2.3. *a) If $(\mathcal{B}_k)_{k \geq 1}$ is an increasing sequence of sub- σ -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, ρ) and $X \in L^1(\mathcal{F}_0; E)$,

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

Theorem 2.4. *For any $X \in L^1(\mathcal{F}_0; E)$, the filtration \mathcal{F}^X generated by the Markov process $(\pi_n X)_{n \leq 0}$ satisfies the Vershik property if and only if the random variable X satisfies the Vershik property.*

Proposition 2.2 is a consequence of Proposition 4.15. Proposition 2.3 is a consequence of Proposition 4.13 and Proposition 4.14. Theorem 2.4 is a particular case of Theorem 4.9.

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 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 2.3. Also note the importance of

Theorem 2.4: $V(X)$ 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 σ -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 σ -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 2.5; but practicality of the rephrased criterion depends on the choice of the generating Markov process. Firstly, the strange state spaces of Vershik's progressive predictions $\pi_n X$ can be avoided when X is some random variable X_k . It suffices to explain this for $X = X_0$. Denote by A_n the state space of X_n for every $n \leq 0$. Starting with a compact metric ρ_0 on A_0 , we recursively define a pseudometric ρ_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 ρ_{n+1} . The ρ_n are more friendly than the $\rho^{(n)}$ appearing in Vershik's standardness criterion, and Lemma 2.5 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 2.5 states that $\mathcal{F}^{X_0} = \mathcal{F}$ under the identifiability condition

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

and then, by Theorem 2.4 standardness of \mathcal{F} is equivalent to $V(X_0)$ under this condition

Lemma 2.5. *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 ρ_0 . Consider the pseudometrics ρ_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 ρ_n are metrics and the ψ_n are isometries. Consequently \mathcal{F} is generated by the process $(\pi_n X_0)_{n \leq 0}$, and $V(X_0)$ 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 ψ_n for which $\pi_n X_0 = \psi_n(X_n)$, and ψ_0 is nothing but the identity map. The equality in 1) relating $\rho^{(n)}$ and ρ_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 Λ_{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 Θ_{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 ρ_n in 3) is recursively shown too. It suffices to show that every ψ_n is injective. Assuming that ψ_{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). The last claim of 3), asserting equivalence between $V(X_0)$ and \mathcal{F} , stems from Theorem 2.4. \square

Obviously we can similarly state Lemma 2.5 for X_k instead of X_0 , for any $k \leq 0$. When the identifiability condition (\star) does not hold, then in order to prove standardness of \mathcal{F} , it is sufficient to check that $V(X_k)$ holds true for every $k \leq 0$. This is a consequence of proposition 6.2 in [10].

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 [12] 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 2.5 and the identifiability condition (\star) will be fulfilled, and thus our main task will be to derive the metrics ρ_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 [12]. Here we define these filtrations as those generated by a Markov process on the vertices of a Bratteli graph.

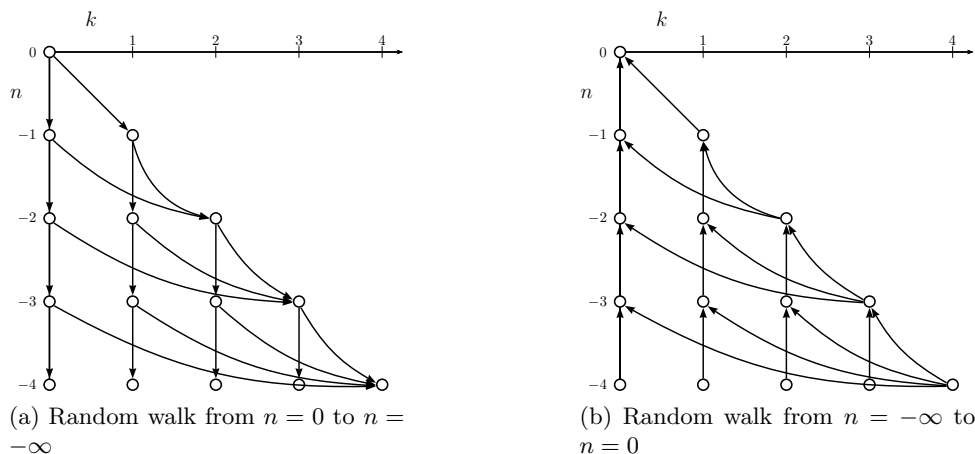


Figure 1: Next-jump time process as a random walk

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 in B is a sequence $(\gamma_n)_{n \leq 0}$ consisting of edges γ_n such that γ_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 Γ_B . When a path is taken at random in Γ_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 $(V_n)_{n \leq 0}$. When 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 Γ_B is $\{0, 1\}^{-\mathbb{N}}$ by labeling 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 $\bigotimes_{n \leq -1} (1 - p_n, p_n)$ 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 [12] 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 [12] because standardness is an asymptotic property (Proposition 2.2).

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 [12] 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 [12] 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 μ on \mathbb{N} . When the p_n are given, this measure is given by

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

The μ -measure of the interval $\{0, \dots, -n\}$ is

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

thus μ is normalizable if and only if $\sum p_n < \infty$. In this case the law of V_n goes to the normalized version of μ and \mathcal{F} is not Kolmogorovian, and in the other case V_n goes to ∞ and \mathcal{F} is Kolmogorovian. This is due to the following proposition about the tail σ -field $\mathcal{F}_{-\infty}$, which is a rewriting of proposition 3.1 in [12], 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 σ -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 [12] 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 2.5. 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 σ -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 2.3(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 than $(V_{n-1})_{n \leq 0}$ 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 2.5. Checking the second and third points do not pose any difficulty. \square

Thus, since standardness is an asymptotic property at $n = -\infty$ (Proposition 2.2), we will focus on the case when (3.2) holds, and this will allow us to use Lemma 2.5. 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 ρ_n starting with the discrete metric ρ_{-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 Λ_{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 also denote by d_n the dispersion of V_n under ρ_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 Vershikian if and only if the Vershik property $V(X)$ holds true for $X = V_{-1}$. Moreover, this property is equivalent to $d_n \rightarrow 0$.*

Proof. Consequence of Lemma 2.5 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 ρ_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 ρ_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 ρ_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. We finally get the statement of the theorem by using Lemma 3.2 and assertion 2) of Lemma 3.3. \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, ρ_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. The pseudometrics ρ_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 ρ_n are metrics under the identifiability condition (3.2). The space (\mathbb{V}_n, ρ_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 $n < 0$).

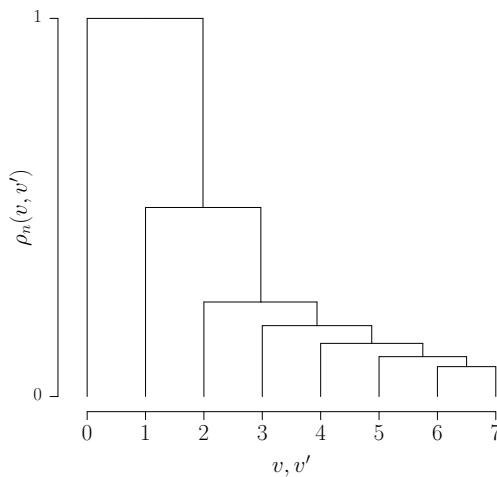


Figure 2: The space (\mathbb{V}_n, ρ_n) .

4 The uniformly scaled entropy

In this section we introduce scaled entropies of filtrations by following Vershik and Gorbulsky [19], except that we use the probabilistic language and we restrict our attention to scalings which are not ϵ -dependent (this is why we term them as *uniform scalings*). Theorem 4.9, our main result, generalizes Theorem 2.4.

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

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

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 some best 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 main 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 μ . Common choices include the Shannon entropy

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

and the min-entropy (or Renyi entropy of order 0)

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

The general properties we will derive about the scaled entropy hold for a measure entropy as defined below.

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

1. $H(\theta) = 0$ if and only if θ is concentrated on one point;
2. H is increasing: $H(\theta') \leq H(\theta)$ when θ' is the image of θ under some map.

When X is a discrete random variable we denote by $H(X)$ the entropy of its law.

4.1 Definition

The definition of the scaled entropy of a filtration \mathcal{F} has something similar to the definition of standardness: we begin by defining the scaled entropy for a \mathcal{F}_0 -measurable random variable, then for a σ -field $\mathcal{B} \subset \mathcal{F}_0$, and finally for the filtration \mathcal{F} . It mainly involves the ϵ -entropy of the Vershik progressive predictions $\pi_n X$ (introduced in Section 2).

Definition 4.2. Let X be an integrable random variable taking its values in a Polish metric space (E, ρ) . The ϵ -entropy of X is

$$H^\epsilon(X) = \inf\{H(F) \mid \mathbb{E}[\rho(X, F)] < \epsilon\}$$

where the infimum is taken over E -valued but discrete $\sigma(X)$ -measurable random variables F .

The *scaling* $c: (-\mathbb{N}) \rightarrow]0, \infty[$ in definition below is termed as *uniform scaling* because Vershik and Gorbulsky more generally allow ϵ -dependent scaling $c \mapsto c(\epsilon, n)$. Thus a uniform scaling is a particular scaling in the sense of Vershik and Gorbulsky's ϵ -dependent general definition, but when it is *proper* in the sense of our definition, then it is also a proper scaling in the sense of Vershik and Gorbulsky.

Definition 4.3. Let \mathcal{F} be a filtration and X an integrable \mathcal{F}_0 -measurable random variable taking its values in a Polish metric space (E, ρ) .

- 1) The ϵ -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 next definitions we consider a nonincreasing function $c: (-\mathbb{N}) \rightarrow]0, \infty[$, termed as *uniform scaling*.

- 2) The limit

$$h_c(X; \mathcal{F}) = \limsup_{\epsilon \rightarrow 0} \limsup_{n \rightarrow 0} \frac{H_n^\epsilon(X)}{c(n)}$$

is called the c -scaled entropy of X . For short, we also denote it by $h_c(X)$ when \mathcal{F} is understood. The uniform scaling c is said to be *proper* for X when $h_c(X; \mathcal{F}) \in]0, \infty[$.

- 3) For a σ -field $\mathcal{B} \subset \mathcal{F}_0$, the c -scaled entropy of \mathcal{B} with respect to \mathcal{F} is defined as

$$h_c(\mathcal{B}; \mathcal{F}) = \sup_X h_c(X),$$

where the supremum is taken over all \mathcal{B} -measurable random variables X taking their values in the interval $[0, 1]$ equipped with the usual metric.

- 4) The c -scaled entropy of \mathcal{F} is defined as

$$h_c(\mathcal{F}) = h_c(\mathcal{F}_0; \mathcal{F}).$$

The uniform scaling c is said to be *proper* for \mathcal{F} when $h_c(\mathcal{F}) \in]0, \infty[$.

Note that the Vershik standardness property for X (Section 2.1) is equivalent to $h_c(X) = 0$ with $c(n) \equiv 1$. Thus, proper uniform scalings do not exist for standard filtrations, and they provide a certain measure of nonstandardness for nonstandard filtrations.

Remark 4.4. The notations $H_n^\epsilon(X)$ and $h_c(X)$ do not show the dependence on the metric ρ on the state space of X . But this is not important in view of Proposition 4.14 which will show that $h_c(X) = h_c(\sigma(X))$. Thus, we could also define $h_c(X)$ when X is non-integrable by replacing ρ with $\rho \wedge 1$.

Remark 4.5. As already mentioned in the definition, the ϵ -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} so long as X is measurable with respect to the final σ -field \mathcal{E}_0 , and \mathcal{F}^X is the smallest such filtration (see [3]).

Remark 4.6. We could replace the ϵ -entropy $H^\epsilon(X)$ with

$$\inf\{H(F) \mid \mathbb{P}(\rho(X, F) > \epsilon) < \epsilon\}$$

without altering the value of $h_c(X; \mathcal{F})$.

The above definition $H^\epsilon(\pi_n X)$ is appropriate for deriving the general properties we will give. But for the derivation of $h_c(X; \mathcal{F})$ on a case-study, it could be better to use the alternative definition of $H^\epsilon(\pi_n X)$ given in the following lemma for the Shannon entropy, and in the next lemma for the min-entropy.

Lemma 4.7. *Assume H is the Shannon entropy. In the definition of $h_c(X; \mathcal{F})$, one can replace the ϵ -entropy $H^\epsilon(\pi_n X)$ with $\inf -\sum \mu_n(P_j) \log \mu_n(P_j)$ where μ_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 $\{A_i, B\}$ where $\mu_n(B) < \epsilon$ and $A_i \subset B(\Gamma_i, \epsilon)$ for some Γ_i .*

Proof. Denote by $H_0^\epsilon(\pi_n X)$ this value. We compare it to the ϵ -entropy $H^\epsilon(\pi_n X)$ as given in Remark 4.6. If $\{P_j\}$ is such a partition, define $f(x) = \operatorname{argmin}_{\Gamma_i} \rho_n(x, \Gamma_i)$. Then $\rho_n(x, f(x)) < \epsilon$ on $\cup A_i$. That shows that $H_0^\epsilon(\pi_n X) \geq H^\epsilon(\pi_n X)$. Conversely, take a function f such that the set $E := \{\rho_n(x, f(x)) < \epsilon\}$ has μ_n -measure greater than $1 - \epsilon$. Let $\{\Gamma_i\}$ be an enumeration of $f(E)$. Set $A_i = f^{-1}(\Gamma_i) \cap E$. Then $\{A_i\}$ is a partition of E and $A_i \subset B(\Gamma_i, \epsilon)$. Now note that

$$\begin{aligned} H(f(\pi_n X)) &\geq \mu_n(E)H(f(\pi_n X) | E) = - \sum \mu_n(A_i) \log \mu_n(A_i) + \mu_n(E) \log \mu_n(E) \\ &= - \sum \mu_n(P_j) \log \mu_n(P_j) - \gamma(\mu_n(E)) \end{aligned}$$

where $\{P_j\} = \{A_i, E^c\}$ and $\gamma(\epsilon) = \epsilon \log \frac{1}{\epsilon} + (1 - \epsilon) \log \frac{1}{1 - \epsilon}$. That shows that $H^\epsilon(\pi_n X) \geq H_0^\epsilon(\pi_n X) - \gamma(\epsilon)$. \square

Lemma 4.8. *Assume H is the min-entropy. Let X a random variable taking its values in a Polish space (A, ρ) . Then the ϵ -entropy $H^\epsilon(X)$ as given in Remark 4.6 equals the minimal log-number of ϵ -balls in A such that X falls in their union with probability higher than $1 - \epsilon$:*

$$\min \{ \log \#\{x_i\} \mid x_i \in A, \Pr(X \in \cup B(x_i, \epsilon)) > 1 - \epsilon \}$$

Proof. Denote by $H_0^\epsilon(X)$ this value. For given x_i , define $f(x) = \operatorname{argmin}_{x_i} \rho(x, x_i)$. Then

$$\Pr(\rho(X, f(X)) < \epsilon) = \Pr(X \in \cup B(x_i, \epsilon)).$$

That shows that $H_0^\epsilon(X) \geq H^\epsilon(X)$. Conversely, take a function g 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 theorem

This section is devoted to prove the main theorem of this paper: Theorem 4.9, which is a deep generalization of Theorem 2.4. It will be used in Section 5 to study the uniform entropy scalings 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.

Theorem 4.9. *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 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 σ -field $\sigma(\pi_n X; n \leq 0)$ when we consider \mathcal{F} as the underlying filtration (see remark 4.5). Theorem 4.9 is particularly useful when the $\pi_n X$ are discrete random variables with finite entropy, because it gives the upper bound $h_c(\mathcal{F}^X) \leq 1$ for any scaling $c(n) \sim H(\pi_n X)$.

Theorem 4.9 will be derived from the following series of lemmas and propositions.

Lemma 4.10. *Let \mathcal{F} be a filtration. If X and Y are two \mathcal{F}_0 -measurable Polish-valued random variables related by $Y = f(X)$ for some measurable function f , then $\pi_n Y = f^n(\pi_n X)$ for some measurable function f^n which is K -Lipschitz if f is K -Lipschitz.*

Proof. See [3]. \square

Theorem 2.4 is an easy corollary of the following lemma and Proposition 2.3, and this provides a new and cleaner proof of Theorem 2.4 than the one given in [10].

Lemma 4.11. *Let \mathcal{F} be a filtration, $X \in L^1(\mathcal{F}_0; E)$ where E is a Polish space metrized by a distance ρ , and set $W^n = (\pi_n X, \dots, \pi_{-1} X, X)$ for some $n \leq 0$. Consider the metric $\bar{\rho}_n = \frac{1}{|n|+1} \sum_{k=0}^n \rho^{(k)}$ on the state space of W^n . Then $\pi_n W^n = \phi(\pi_n X)$ where ϕ is an isometry.*

Proof. For the proof we consider the distance $\tilde{\rho}_n = \sum_{k=n}^{k=0} \rho^{(k)}$ instead of $\bar{\rho}_n$ 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

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

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

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

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 $(\#)$,

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

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

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

and recursively using this equality we finally get

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

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

Lemma 4.12. *Let \mathcal{F} be a filtration, $(X_k)_{k \geq 1}$ be a sequence in $L^1(\mathcal{F}_0; E)$ where E is Polish. If $X_k \rightarrow X$ in L^1 for some random variable $X \in L^1(\mathcal{F}_0; E)$, and if $\sigma(X_k) \subset \sigma(X)$ for every $k \geq 1$, then for every $\epsilon_0 > 0$ there exists k_0 such that $H_n^{\epsilon_0}(X_k) \geq H_n^{2\epsilon_0}(X)$ for every $n \leq 0$ and every $k \geq k_0$.*

Proof. Let $k_0 = k(\epsilon_0)$ such that $\mathbb{E}[\rho(X_k, X)] \leq \epsilon_0$ for every $k \geq k_0$, hence $H_n^{\epsilon_0}(X_k) \geq H_n^{2\epsilon_0}(X)$ for every n by definition of $H_n^\epsilon(\cdot)$ and Lemma 2.1. \square

The following lemma is a continuity-like property of $X \mapsto h_c(X; \mathcal{F})$.

Proposition 4.13. *Let $c: (-\mathbb{N}) \rightarrow]0, \infty[$ be a scaling. If, under the same hypotheses as Lemma 4.12, 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. Put $a = h_c(X; \mathcal{F})$. We firstly check that $a < \infty$. Assuming $a = \infty$, the definition of the superior limit provides ϵ_0 such that $\limsup_{n \rightarrow -\infty} \frac{H_n^{2\epsilon_0}(X)}{c(n)} > \ell + 1$. Therefore one can take k_0 sufficiently large in order that $h_c(X_{k_0}, \mathcal{F}) \leq \ell$ and such that $\limsup_{n \rightarrow -\infty} \frac{H_n^{\epsilon_0}(X_{k_0})}{c(n)} > \ell + 1$ by Lemma 4.12. But $\epsilon \mapsto H_n^\epsilon(X_{k_0})$ is decreasing, therefore 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 $\delta > 0$, the definition of the superior limit provides ϵ_0 such that $\limsup_{n \rightarrow -\infty} \frac{H_n^{2\epsilon_0}(X)}{c(n)} > a - \delta$. Taking k as in Lemma 4.12, one gets $\limsup_{n \rightarrow -\infty} \frac{H_n^\epsilon(X_k)}{c(n)} > a - \delta$ 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

Proposition 4.14. *Let \mathcal{F} be a filtration and $X \in L^1(\mathcal{F}_0; E)$ where E is a Polish space. Then $h_c(\sigma(X); \mathcal{F}) = h_c(X; \mathcal{F})$ 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 $h_c(Y; \mathcal{F}) \leq h_c(X; \mathcal{F})$ with the help of Lemma 4.10. The result follows from Proposition 4.13 and from the density of the set of random variables $f(X)$, f Lipschitzian, in $L^1(\sigma(X), [0, 1])$ (see lemma 2.15 in [9]). \square

Now we can quickly prove Theorem 4.9.

Proof of Theorem 4.9. Let $\mathcal{B}_n = \sigma(\pi_n X, \dots, \pi_{-1} X, X)$. By Lemma 4.11 and Proposition 4.14, $h_c(\mathcal{B}_n; \mathcal{F}) = h_c(X; \mathcal{F})$. Then the theorem follows from Proposition 4.13. \square

4.3 Other properties

Proposition 4.15 below is another corollary of Lemma 4.11, generalizing Proposition 2.2.

Proposition 4.15. *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

$$H_n^\epsilon(X_{n_0}, \mathcal{F}^{n_0]) = H_{n_0+n}^\epsilon(X_{n_0}; \mathcal{F}) \quad (4.1)$$

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 X_0 is an integrable \mathcal{F}_0 -measurable random variable and $W^{n_0} = (\pi_{n_0} X_0, \dots, X_0)$, then $h_c(X_0; \mathcal{F}) \leq h_c(W^{n_0}; \mathcal{F})$ by Proposition 4.14 and because H is increasing (definition 4.1). But Lemma 4.11 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 (4.1) 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

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

Proposition 4.16. *Let \mathcal{F} be a filtration, $\mathcal{B} \subset \mathcal{F}_0$ a σ -field, and $(\mathcal{B}_k)_{k \geq 1}$ an increasing sequence of σ -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. Let X be a \mathcal{B} -measurable integrable random variable. Then $X = \lim X_k$ in L^1 where each X_k is \mathcal{B}_k -measurable. By Proposition 4.13, $h_c(X) \leq \sup h_c(X_k)$, and therefore $h_c(X) \leq \sup h_c(\mathcal{B}_k)$. But H is increasing (definition 4.1), hence $h_c(\mathcal{B}_k)$ is increasing in k and $\sup h_c(\mathcal{B}_k) = \lim h_c(\mathcal{B}_k)$. That shows that $h_c(\mathcal{B}) \leq \lim h_c(\mathcal{B}_k)$, and the reverse inequality obviously stems from the increasing property of H . \square

The following proposition is a generalization of proposition 6.1 in [10].

Proposition 4.17. *Let \mathcal{F} be a filtration and $(V_n)_{n \leq 0}$ be a superinnovation of \mathcal{F} , that is to say, each V_n is a random variable independent of \mathcal{F}_{n-1} and satisfies $\mathcal{F}_n \subset \mathcal{F}_{n-1} \vee \sigma(V_n)$. 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 4.16), $h_c(\mathcal{G}) = \lim h_c(\mathcal{B}_m)$ 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 $h_c(X) = h_c(\pi_m X)$ for every \mathcal{F}_0 -measurable integrable random variable X . \square

Proposition 4.18. *Let $(X_n)_{n \leq 0}$ be a Markov process and let \mathcal{F} be the filtration it generates. Then $h_c(\mathcal{F}) = \lim h_c(X_n)$ for any scaling c .*

Proof. By lemma 3.41 in [9], there exists a superinnovation $(V_n)_{n \leq 0}$ of \mathcal{F} satisfying the additional property $\sigma(X_{n-1}, V_n) \supset \sigma(X_n)$. By Proposition 4.17, we know that $h_c(\mathcal{F}) = h_c(\mathcal{G})$. By Proposition 4.16 (left-continuity of the scaled entropy), we know that $h_c(\mathcal{G}) = \lim h_c(\mathcal{B}_m)$ where $\mathcal{B}_m = \sigma(X_m, V_{m+1}, \dots, V_0)$. By noting that $\pi_m X$ is $\sigma(X_m)$ -measurable for every \mathcal{B}_m -measurable random variable X , one gets $h_c(\mathcal{B}_m) = h_c(X_m)$. More precisely, we should write $h_c(\mathcal{B}_m; \mathcal{G}) = h_c(X_m; \mathcal{G})$. But $h_c(X_m; \mathcal{G}) = h_c(X_m; \mathcal{F})$ because \mathcal{F} is immersed in \mathcal{G} . \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 and its filtration \mathcal{F} . Using the Shannon entropy as the measure of entropy, we study the entropy of \mathcal{F} in the Kolmogorovian nonstandard case, that is, in view of Proposition 3.1 and Theorem 3.7, the case when $\sum p_n = \infty$ and $\sum p_n^2 < \infty$. Two non-standard next-jump time filtrations defined by two distinct jumping probabilities sequences are not locally isomorphism, therefore there is no real interest to compare them with the scaling entropy. But this case-study provides a nice illustration because of the simplicity of the pseudometrics ρ_n .

5.1 Entropy of V_n

Preliminarily, we study the entropy of the random variables V_n . The entropy of V_n can be recursively obtained from the conditional entropy formula

$$H(V_n, V_{n-1}) = H(V_n) + H(V_{n-1} | V_n) = H(V_{n-1}) + H(V_n | V_{n-1}),$$

by deriving the two conditional entropies:

$$H(V_{n-1} | V_n) = h(p_{n-1}) \quad \text{and} \quad H(V_n | V_{n-1}) = p_{n-1} H(V_n)$$

where $h(\theta) = -\theta \log \theta - (1 - \theta) \log(1 - \theta)$ is the entropy of a Bernoulli variate with parameter θ . The first formula obviously comes from $H(V_{n-1} | V_n = k) = h(p_{n-1})$ for every k . The second formula comes from the obvious equality $H(V_n | V_{n-1} = k) = 0$ for $k \leq |n| + 1$ and from the equality $H(V_n | V_{n-1} = |n| + 1) = H(V_n)$ which holds because the conditional distribution $\mathcal{L}(V_n | V_{n-1} = |n| + 1)$ equals the unconditional distribution $\mathcal{L}(V_n)$. Thus we finally get the recursive relation

$$H(V_{n-1}) = h(p_{n-1}) + (1 - p_{n-1})H(V_n), \tag{5.1}$$

yielding

$$H(V_n) = h(p_n) + (1 - p_n)h(p_{n+1}) + (1 - p_n)(1 - p_{n+1})h(p_{n+2}) + \dots + (1 - p_n) \dots (1 - p_{-2})h(p_{-1}).$$

Now, note that $p_{-V_n} > 0$ almost surely, because for every $k \in \{0, \dots, |n|\}$, the event $\{V_n = k\}$ is included in the event $\{V_{-k} = k\}$ and the latter event has probability p_k . Moreover $h(0) = 0$, and finally $H(V_n)$ is also given by

$$H(V_n) = \mathbb{E} \left[\frac{h(p_{-V_n})}{p_{-V_n}} \right].$$

Lemma 5.1. *When \mathcal{F} is Kolmogorovian and $p_n \rightarrow 0$, $\lim H(V_n) = +\infty$ and $\lim \frac{H(V_{n-1})}{H(V_n)} = 1$.*

Proof. According to Proposition 3.1, $V_n \rightarrow +\infty$ almost surely when \mathcal{F} is Kolmogorovian. Therefore $V_n = |n|$ for infinitely many n , because any possible trajectory of $(V_n)_{n \leq 0}$ realizing only finitely many events $\{V_n = |n|\}$ is bounded. In addition, assuming $p_n \rightarrow 0$, one has $\frac{h(p-V_n)}{p-V_n} \rightarrow +\infty$ because $h(x)/x \rightarrow +\infty$ when $x \rightarrow 0^+$. We deduce from the recursive relation (5.1) that $\frac{H(V_{n-1})}{H(V_n)} \rightarrow 1$. \square

5.2 Entropy of the filtration

We have seen in Section 3 that the assumption of Lemma 5.2 below is fulfilled for the random variable V_{-1} of the next-jump time process $(V_n)_{n \leq 0}$ in case of nonstandardness. The proof of this lemma involves *Fano's inequality* (see [5]), whose statement is:

$$H(X | Y) \leq h(\Pr(X \neq Y)) + \Pr(X \neq Y) \log N$$

for any pair of discrete random variables X and Y taking no more than N values, where $H(X | Y)$ denotes the conditional entropy and $h(p) = -p \log p - (1-p) \log(1-p)$ denotes the entropy of a Bernoulli variate with parameter p .

Lemma 5.2. *Let \mathcal{F} be a filtration, (E, ρ) a Polish metric space, $X \in L^1(\mathcal{F}_0; E)$, and $c: (-\mathbb{N}) \rightarrow]0, \infty[$ a uniform scaling. Assume that every $\pi_n X$ takes its values in a finite subset $B^{(n)}$ of $E^{(n)}$ and there exists $\epsilon_0 > 0$ such that $\rho^{(n)}(x, x') > \epsilon_0$ for every $n \leq 0$ as long as $x \neq x'$. Then*

$$h_c(X; \mathcal{F}) = \limsup_{n \rightarrow -\infty} \frac{H(\pi_n X)}{\log \#B^{(n)}}$$

for any scaling $c(n) \sim \log \#B^{(n)}$.

Proof. Let $\delta < \frac{1}{2}$ and set $\epsilon = \delta \epsilon_0 < \epsilon_0$. For every n one has $H_n^\epsilon(X) = H(F_n) \leq H(\pi_n X)$ where F_n is a $\sigma(\pi_n X)$ -measurable random variable satisfying $\Pr(\pi_n X \neq F_n) < \delta$. By the conditional entropy formula, $H(\pi_n X) - H(F_n) = H(\pi_n X | F_n)$, and by Fano's inequality,

$$H(\pi_n X | F_n) \leq h(p_n) + p_n \log \#B^{(n)}$$

where $p_n = \Pr(\pi_n X \neq F_n)$. Therefore

$$H(\pi_n X | F_n) \leq h(\delta) + \delta \log \#B^{(n)},$$

and consequently

$$\frac{H_n^\epsilon(X)}{\log \#B^{(n)}} \leq \frac{H(\pi_n X)}{\log \#B^{(n)}} \leq \frac{H_n^\epsilon(X) + h\left(\frac{\epsilon}{\epsilon_0}\right)}{\log \#B^{(n)}} + \frac{\epsilon}{\epsilon_0}$$

for every n and every $\epsilon < \epsilon_0/2$, yielding the assertion of the lemma. \square

Now consider the next-jump time process $(V_n)_{n \leq 0}$ and its filtration \mathcal{F} . Recall that we have seen in Lemma 5.1 that $\lim H(V_n) = +\infty$ in the case when \mathcal{F} is Kolmogorovian but not standard.

Proposition 5.3. *Assume \mathcal{F} is Kolmogorovian but not standard. Let $c: (-\mathbb{N}) \rightarrow]0, \infty[$ be a uniform scaling such that $c(n) \sim H(V_n)$. Then one always has $h_c(\mathcal{F}) \leq 1$, and one has $h_c(\mathcal{F}) = 1$ in the uniform case $p_n = (|n| + 1)^{-1}$ (this is the case when V_n has the uniform distribution).*

Proof. In the nonstandard case, there are, in view of Theorem 3.7, finitely many values of n for which $p_n = 1$.

First assume the identifiability assumption (3.2) ($p_{-1} \in]0, 1[$ and $p_n < 1$ for every $n \leq -2$). By Lemma 3.3(1) and Theorem 4.9,

$$h_c(\mathcal{F}) = h_c(X; \mathcal{F})$$

with $X = V_{-1}$, for any scaling c . In the uniform case $p_n = (|n| + 1)^{-1}$, we know by Lemma 5.2 that $h_c(X; \mathcal{F}) = 1$ for $c(n) \sim H(V_n) = \log(|n| + 1)$. In other cases, one obviously has

$$\frac{H_n^c(X)}{c(n)} \leq \frac{H(\pi_n X)}{c(n)}$$

and then $h_c(X; \mathcal{F}) \leq 1$ for $c(n) \sim H(V_n)$.

In the case when there are several n for which $p_n = 1$, take the smallest one and call it n_0 . Set $p'_n = p_{n_0+n}$. If $p'_{-1} > 0$, then by Lemma 3.3(2) and by the previous case, the proposition holds for $c(n) = H(V_{n_0+n})$ but it holds for $c(n) \sim H(V_n)$ too because of Lemma 5.1. If $p'_{-1} = 0$, we similarly conclude by using Lemma 3.3(3) after noting that p'_n cannot be zero for every $n \leq -1$ in the Kolmogorovian but non-standard case. \square

We have derived a proper uniform scaling for the uniform case only. To derive one for the general case, we should improve the asymptotic estimate of $H(\pi_n X | \mathcal{F}_n)$ in the proof of Lemma 5.2. We have used Fano's inequality which is a general majoration of the conditional entropy. Generalized Fano's inequalities studied in [6] do not seem to be helpful for the general case. It would be interesting to know whether there is a case for which there is a proper uniform scaling $c(n) = o(H(V_n))$, and a case for which there is no proper uniform scaling.

Note that the problem of deriving $h_c(\mathcal{F})$ can be elementary rephrased as a problem about the measure μ on \mathbb{N} defining the next-jump time process in the case when the identifiability condition is fulfilled (see above Proposition 3.1).

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 η_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 η_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) -adic.

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. The converse is easy to prove with the help of Proposition 2.3.

Throughout this section, when an (r_n) -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 $\ell_0 = 1$).

For such filtrations, Vershik defined the exponential entropy, originally in [16], as follows.

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

1. The (r_n) -adic entropy, or the exponential entropy, of X with respect to \mathcal{F} is 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 specified, is the Shannon entropy.

2. The (r_n) -adic entropy, or 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.

With the help of Lemma 6.3, we will see that this definition makes sense (when H is the Shannon entropy). 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 $(\eta_n)_{n \leq 0}$. By a result similar to Proposition 4.16, which is given in [18] but not in the present paper, a standard (r_n) -adic filtration has zero entropy (recalling that standardness means the existence of a generating process of innovations).

Obviously, $h_c(\mathcal{F}) \leq h(\mathcal{F})$ when c is the scaling $c(n) = \ell_n$. The main results of this section are :

1. Theorem 6.4, whose credit is given to Gorbulsky ([4]), is about the coincidence between the exponential entropy and the scaled entropy with scaling $c(n) = \ell_n$ under the slowness condition (Δ) about the poly-adicity sequence $(r_n)_{n \leq 0}$.
2. Theorem 6.10, due to Vershik ([16, 18]), states that poly-adic filtrations have a zero exponential entropy under the fastness condition (∇) about the poly-adicity sequence $(r_n)_{n \leq 0}$.
3. Theorem 6.12, due to Vershik ([16, 18]), gives the value of the exponential entropy for the split-word filtrations under the slowness condition (Δ) .

Gorbulsky showed Theorem 6.4 in the dyadic case $r_n \equiv 2$ only. Our proof of the generalization to condition (Δ) essentially uses the same mathematics.

In addition to the points listed above, in this section we investigate the entropy of some filtrations of unordered pairs.

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 [13]. 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 [13] for details about the group G_n of tree automorphisms and its action on the set of ℓ_n -word A^{ℓ_n} . and

Lemma 6.2. *Let \mathcal{F} be an (r_n) -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 orbit of a random word X_n on A having length ℓ_n under the action of the group of tree automorphisms G_n . Using this identification and starting with the 0–1 metric ρ 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^{ℓ_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 Γ and Γ' , where $\delta_n(w, w')$ is the Hamming distance between the ℓ_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 σ -field, η_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 σ -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 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 (when H is the Shannon entropy): 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.

6.2 Gorbulsky's theorem

In [4], Gorbulsky proved Theorem 6.4 below in the case when $r_n \equiv 2$. We show that this result more generally holds for sequences (r_n) satisfying 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 (Δ) , and it is also fulfilled in the case when $r_n = |n| + 1$.

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

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

for every \mathcal{F}_0 -measurable random variable X taking only finitely many values.

Consequently the scaled entropy $h_c(\mathcal{F})$ is the same as the 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.

Only Lemma 6.8 below will be involved in the proof of Theorem 6.4. It is a consequence of Lemma 6.6 and Lemma 6.7. 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 such that $v_n \searrow 0$. If $\sum u_n v_n < \infty$ 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

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

□

When $\Gamma \in A^{\ell_n}/G_n$ is an orbit, we denote by $|\Gamma|$ the number of words it contains.

Lemma 6.6. *Assume that condition (Δ) is fulfilled by the poly-adicity sequence (r_n) . For any pair of orbits Γ and Γ' ,*

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

with $\lim_{\epsilon \rightarrow 0} d(\epsilon) = 0$.

Proof. Consider a word w of length ℓ_n with its tree structure as shown by Figure 3. We denote by $\Gamma(w)$ its G_n -orbit.

The word w at level n is the concatenation of the r_{n+1} words w_i of length ℓ_{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

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

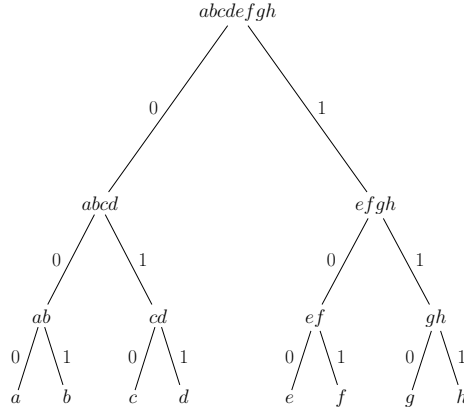


Figure 3: The tree structure of $abcdefgh$.

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

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

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,

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

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

Under the (Δ) 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 orbit Γ , the log-number of words in an ϵ -neighbourhood of Γ does not exceed*

$$\log |\Gamma| + \gamma(\epsilon) \ell_n + \log(\#A - 1) \epsilon \ell_n$$

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

Proof. The ϵ -ball around a word w of length ℓ_n is enumerated by taking $k = 0, \dots, \epsilon \ell_n$ positions in w and changing all letters at these positions. Then the number of words it contains is

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

and

$$\log \sum_{k=0}^{\epsilon \ell_n} \binom{\ell_n}{k} \leq \gamma(\epsilon) \ell_n$$

by the classical large deviations inequality for independent symmetric Bernoulli variables (Corollary 2.20 in [7]). \square

Lemma 6.8. *Under condition (Δ) , for any orbit Γ , the log-number of orbits in the ϵ -neighborhood of Γ does not exceed a value L_n^ϵ satisfying*

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

Proof. The number of orbits in the ϵ -neighbourhood $V_\epsilon(\Gamma)$ of Γ is less than the number of words in $V_\epsilon(\Gamma)$ divided by the minimal length of an orbit Γ' in $V_\epsilon(\Gamma)$. Applying the two previous lemmas gives the desired result:

$$L_n^\epsilon \leq (\log |\Gamma| + \gamma(\epsilon)\ell_n + \log(\#A - 1)\epsilon\ell_n) - (\log |\Gamma| - d(\epsilon)\ell_n)$$

□

Proof of Theorem 6.4.

In the proof we use notation $H(E) = -\sum_{x \in E} \mu(x) \log \mu(x)$ where μ 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 ϵ -entropy as defined in Lemma 4.7. Let $\{A_i, B\}$ be a partition achieving $H_n^\epsilon(X)$, with $A_i \subset \mathcal{V}_\epsilon(\Gamma_i)$ and $\mu(B) \leq \epsilon$. One has

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

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.

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

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

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 4.5). □

6.3 ∇ -adic filtrations

The (∇) condition is stronger than the negation of the (Δ) condition, because this condition is the divergence of a certain sequence whereas the (Δ) 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 [18].

Theorem 6.10. *Let \mathcal{F} be a (∇) -adic filtrations. Then it has zero exponential entropy when we use the min-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 χ_n^p of the action of G_n on A^{ℓ_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 (∇) 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 β_n is the quantity going to ∞ under the (∇) 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 (∇) 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 min-entropy. Then the result for the min-entropy follows from this lemma, and the result for the Shannon entropy H follows because of $H \leq H_0$. \square

6.4 Δ -adic split-word filtrations

The poly- Δ -adic filtrations of the split-word processes with independent letters were studied in [14], [3], [2], [9]. In the more general case of stationary letters, standardness of these filtrations is closely connected, as shown in [13], to the notion of scale of an automorphism introduced by Vershik in [17]. Theorem 6.12 below, which is a rephrasing of theorem 4.1 in [18], provides the exponential entropy of these filtrations under condition (Δ) .

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

- X_n is a random word on A of length ℓ_n ;
- η_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 η_n -th letter of X_{n-1} treated as an (r_n) -word on A^{ℓ_n} .

Obviously the filtration \mathcal{F} generated by $(X_n, \eta_n)_{n \leq 0}$ is a (r_n) -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 defining the law of X_n the projection of this measure on ℓ_n consecutive coordinates. In this case and when A is finite, 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, we can see that θ_b exists for any split-word process with the help of Lemma 6.3.

With the terminology of [8], \mathcal{F} is an Δ -adic filtration on the Bratteli graph shown on Figure 4, called the graph of the ordered pairs by Vershik (in contrast with the graph of unordered pairs that we will see in Section 6.5).

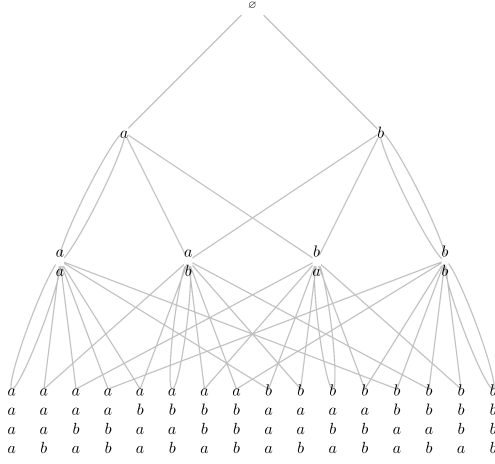


Figure 4: The Bratteli graph of the ordered pairs.

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 6.12. *For the scaling $c(n) = \ell_n$ and under condition (Δ) , the scaled entropy is $h_c(\mathcal{F}) = \theta_b$ when we use the Shannon entropy in logarithmic base b as the underlying measure of entropy.*

Proof. By Proposition 4.17 and Proposition 4.18, 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 tree automorphisms. Therefore $H(\pi_n X_0) = H(X_n) - H(X_n | \pi_n X_0)$, and $H(X_n | \pi_n X_0)$ 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 (Δ) condition is $\lim_{k \rightarrow -\infty} \sum_{k=-\infty}^m \frac{\log r_{k+1}!}{\ell_k} = 0$, and the proof is over. \square

6.5 Dyadic filtrations of unordered pairs

Let (X_n, ϵ_n) 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 6.12. 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 4.9, the scaled entropy $h_c(\mathcal{F}^{X_0})$ of the filtration 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 5 below, and which can be found in [22].

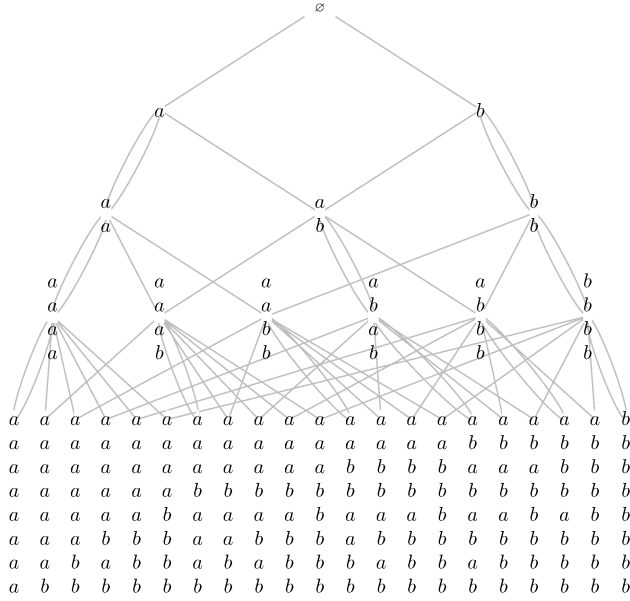


Figure 5: The Brattelli graph of the unordered pairs ([22]).

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 5 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 Proposition 4.17) of \mathcal{F}^{X_0} and which is a process of innovations of that 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 2.4. Moreover we know that $h_c(\mathcal{F}^{X_0}) = h_c(\mathcal{G}^{X_0})$ by Proposition 4.17. In fact it can be shown by using Proposition 4.17 and lemma 5.3 in [8] 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 [8]) to a Brattelli graph endowed with a central probability measure.

The bounds for $h_c(X_0)$ we give in Proposition 6.15 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 [18], 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 [18], it is stated in lemma 3.7 but the given value of γ is not correct.

Lemma 6.13. *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 6.14. *For an alphabet with two letters, the base 2 logarithm of the number χ_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 χ_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 γ . □

Proposition 6.15. *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 γ is given in Lemma 6.14 and θ_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 6.12, with the equality $H(\pi_n X_0) = H(X_n) - H(X_n | \pi_n X_0)$. But this time we bound from above the conditional entropy $H(X_n | \pi_n X_0)$ 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 6.13 and Lemma 6.14, 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. □

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 $\theta_2 < \frac{1}{4}$.

References

- [1] Bolley, F.: Separability and completeness for the Wasserstein distance. Séminaire de Probabilités XLI, Springer Lecture Notes in Mathematics 1934, 371–377 (2008)
- [2] Ceillier, G.: The filtration of the split-words process. Probability Theory and Related Fields 153, Issue 1–2, 269–292 (2012).
- [3] 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, 265–305 (2001)
- [4] 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.
- [5] Gray, R.M. : Entropy and Information Theory. Springer, 2nd ed, 2011.
- [6] Han, T.S., Verdú, S: Generalizing the Fano inequality. IEEE Transactions on Information Theory 40, 1247–1251.
- [7] van der Hofstad, R.: Random Graphs and Complex Networks. To appear.
- [8] É. Janvresse, S. Laurent, T. de la Rue: Standardness of the filtration of a monotonic Markov process. arXiv:1501.02166 (2015). To appear in: Markov Processes and Related Fields.

- [9] Laurent, S.: On standardness and I-cosiness. Séminaire de Probabilités XLIII, Springer Lecture Notes in Mathematics 2006, 127–186 (2010)
- [10] Laurent, S.: On Vershikian and I-cosy random variables and filtrations. *Teoriya Veroyatnostei i ee Primeneniya* 55, 104–132 (2010). Also published in: *Theory Probab. Appl.* 55, 54–76 (2011)
- [11] Laurent, S.: Further comments on the representation problem for stationary processes. *Statist. Probab. Lett.* 80, 592–596 (2010).
- [12] Laurent, S.: Standardness and nonstandardness of next-jump time filtrations. *Electronic Communications in Probability* 18 (2013), no. 56, 1–11.
- [13] 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)
- [14] Smorodinsky, M.: Processes with no standard extension. *Israel Journal of Mathematics* 107, 327–331 (1998)
- [15] 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)
- [16] Vershik, A.M.: Continuum of pairwise nonisomorphic diadic sequences. *Funktsional’nyi Analiz i Ego Prilozheniya* 5:3, 16–18 (1971). English translation: *Functional Analysis and Its Applications* 5:3, 182–184 (1971)
- [17] Vershik, A.M.: Four definitions of the scale of an automorphism. *Funktsional’nyi Analiz i Ego Prilozheniya*, 7:3, 1–17 (1973). English translation: *Functional Analysis and Its Applications*, 7:3, 169–181 (1973)
- [18] 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)
- [19] Vershik, A.M, Gorbulsky, A.D.: Scaled entropy of filtrations of σ -fields. *Theory of Probability and its Applications*, **52**, 3, 493–508, 2008.
- [20] Vershik, A.M.: Smooth and non-smooth AF-algebras and problem on invariant measures. arXiv:1304.2193 (2013) <http://arxiv.org/abs/1304.2193>
- [21] Vershik, A.M.: Intrinsic metric on graded graphs, standardness, and invariant measures. *Zapiski Nauchn. Semin. POMI* 421, 58–67 (2014).
- [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).