Uniform entropy scalings of filtrations

Stéphane Laurent

Uniform entropy scalings of filtrations

Stéphane Laurent

July 3, 2014

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 an example we derive a proper uniform entropy scaling for a next-jump time filtration. As a side note, we use the example of the next-jump time filtrations to write down a case study of Vershik's theory of intrinsic topology on Bratteli graphs.

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

4 Central and noncentral ergodic measures on the next-jump graph **8**
4.1 Statement of the problem . . . 9
4.2 Central and noncentral ergodic measures: bare-hands method . . . 10
4.3 Central and noncentral ergodic central measures: Vershik's method . . . 10

5 The uniformly scaled entropy **13**
5.1 Definition . . . 13
5.2 Properties . . . 14

6 Entropy of next-jump time filtrations **16**
6.1 Entropy of V_n . . . 17
6.2 Entropy of the filtration . . . 17

1 Introduction

This is the first paper written in the probabilistic language about Vershik & Gorbulsky's theory of scaled entropy introduced in [12]. 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 [9] (where I-cosiness was used to derive this standardness criterion). Section 4 is a digression: we use the calculations of section 3 to study the intrinsic topology induced by the Bratteli graph of the next-jump time filtrations. This provides a complete case study of Vershik's recent theory introduced in [14], which is a by-product of standardness. The reader only interested in the entropy could skip this section, whereas the reader only interested in this illustration of Vershik's new theory could only read sections 2 to 4. Section 5 introduces the definition of uniform entropy scalings. In section 6 we pursue the work of section 3 by studying uniform entropy scalings for the next-jump time filtrations.

2 Vershik's standardness criterion

In the probabilistic literature, *standardness* of a filtration $\mathcal{F} = (\mathcal{F}_n)_{n\leqslant 0}$ 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 [2, 6, 7, 8]). 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 [7]. In section 5 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

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

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 $\mathbb{E}\big[\rho(X,x)\big] < \infty$ 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 [10] and [11]) 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 $\mathbb{E}\big[\rho(X,x)\,|\,\mathcal{F}_{-1}\big]$ is finite, therefore $\rho'\big(\mathcal{L}(X\,|\,\mathcal{F}_{-1}),\delta_x\big) < \infty$ 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 $E^{(n-1)} = \big(E^{(n)}\big)'$, denoting as before by E' the space of integrable probability measures on any separable metric space E. Note that $(\pi_n X)_{n\leqslant 0}$ 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 $\left(\rho^{(n)}(\pi_n X, \pi_n Y)\right)_{n\leqslant 0}$ is a submartingale. In particular the expectation $\mathbb{E}\left[\rho^{(n)}(\pi_n X, \pi_n Y)\right]$ is increasing with n.*

Finally, in order to state Vershik's standardness criterion, one introduces the *dispersion* $\operatorname{disp} X$ 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 $\operatorname{disp} \pi_n X \longrightarrow 0$ 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.

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 ambiant 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 5 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 5 about the uniformly scaling entropy generalize the following propositions and theorem which are provided in [7].

Proposition 2.2. *Let $\mathcal{F}$ be a filtration, $n_0 \leqslant 0$ be an integer, and denote by $\mathcal{F}^{n_0]} = (\mathcal{F}_{n_0+n})_{n\leqslant 0}$ 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 $(\mathcal{B}_k)_{k\geqslant 1}$ is an increasing sequence of sub-σ-fields of $\mathcal{F}_0$ then*

$$[\forall k \geqslant 1,\, V(\mathcal{B}_k)] \implies V\Big(\bigvee_{k\geqslant 1} \mathcal{B}_k\Big).$$

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 $(\pi_n X)_{n\leqslant 0}$ satisfies the Vershik property if and only if the random variable X satisfies the Vershik property.*

Proposition 2.2 is a consequence of proposition 5.12. Proposition 2.3 is a consequence of proposition 5.10 and proposition 5.11. Theorem 2.4 is a particular case of theorem 5.6.

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 $\pi_n X$ does 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

$(\pi_n X)_{n\leqslant 0}$. 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 $(X_n)_{n\leqslant 0}$: just take for X_n any random variable generating the σ-field $\mathcal{F}_n$ for every $n \leqslant 0$. Vershik's standardness criterion can be rephrased to a more practical criterion by considering such a Markov process $(X_n)_{n\leqslant 0}$, 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 $n \leqslant 0$. 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

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

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 $\psi_n \colon A_n \to A_0^{(n)}$ such that $\pi_n X_0 = \psi_n(X_n)$ and

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

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 $\rho_n(X'_n, X''_n) \to 0$ 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

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

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 $(X_n)_{n\leqslant 0}$. Denote by A_n the state space of X_n for every $n \leqslant 0$ 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 $\psi_n \colon A_n \to A_0^{(n)}$ such that $\pi_n X_0 = \psi_n(X_n)$ and*

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

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

2) *The Vershik property $V(X_0)$ is equivalent to $\mathbb{E}[\rho_n(X'_n, X''_n)] \to 0$ 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 $(\pi_n X_0)_{n\leqslant 0}$, 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 $\pi_n X_0 = \mathcal{L}(\pi_{n+1} X_0 \mid \mathcal{F}_n)$ for $n < 0$ is a $\sigma(X_n)$-measurable random variable by the Markov property. Therefore, for each $n \leqslant 0$, 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 $\rho^{(n+1)}\big(\psi_{n+1}(x_{n+1}), \psi_{n+1}(x'_{n+1})\big) = \rho_{n+1}(x_{n+1}, x'_{n+1})$, then the Kantorovich distance $\rho_n(x_n, x'_n)$ is given by

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

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

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

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, $\psi_n(x_n) \neq \psi(x'_n)$, thereby showing 3). The last claim of 3), asserting equivalence between $V(X_0)$ and $\mathcal{F}$, stems from theorem 2.4. □

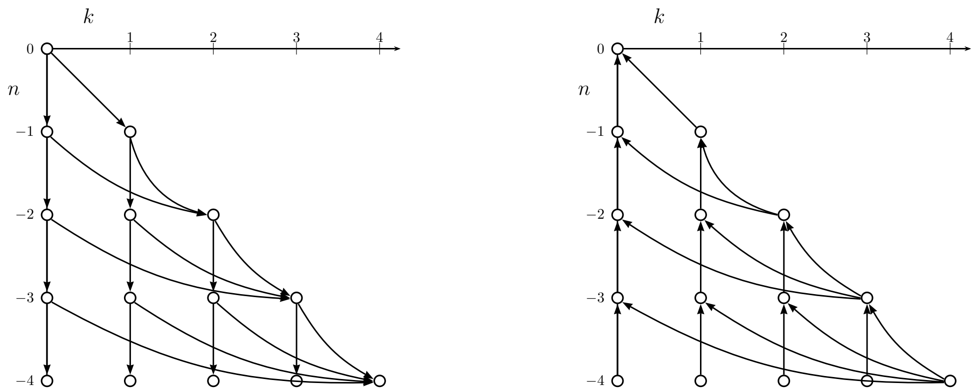
Obviously we can similarly state lemma 2.5 for X_k instead of X_0, for any $k \leqslant 0$. When the identifiability condition $(\star)$ does not hold, then in order to prove standardness of $\mathcal{F}$, it is sufficient to check that $V(X_k)$ holds true for every $k \leqslant 0$. This is a consequence of proposition 6.2 in [7].

3 The next-jump time filtrations

In section 6 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 [9] 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). The distances ρ_n will be the starting point of section 4 where we will apply the theory recently introduced by Vershik in [14] to the next-jump graph (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 [9]. Here we define these filtrations as those generated by a Markov process on the vertices of a Bratteli graph.



(a) Random walk from $n = 0$ to $n = -\infty$

(b) Random walk from $n = -\infty$ to $n = 0$

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 $k \in \{0, \ldots, |n|\}$, 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 $(\gamma_n)_{n \leqslant 0}$ consisting of edges γ_n such that γ_n connects a vertex at level n to a vertex at level $n-1$ for every $n \leqslant 0$. 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 $(V_n)_{n \leqslant 0}$. When this causes no possible confusion we identify a vertex to its label. We study the case when the process $(V_0, V_{-1}, \ldots)$ is the Markov chain whose transition distributions are defined from a given $[0,1]$-valued sequence $(p_n)_{n\leqslant 0}$ satisfying $p_0 = 1$, by

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

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 $\bigotimes_{n\leqslant -1}(1-p_n, p_n)$ by denoting by $(1-p,p)$ the Bernoulli probability measure with probability of success p.

The time-directed process $(V_n)_{n\leqslant 0}$ is Markovian too. The next-jump time process $(Z_n)_{n\leqslant 0}$ defined in [9] is obtained from V_n by putting $Z_0 = 0$ and $Z_n = -V_{n+1}$ for $n \leqslant -1$. Hence the filtration $\mathcal{F}$ generated by the Markov process $(V_n)_{n\leqslant 0} = (Z_{n-1})_{n\leqslant 0}$ shares the same standardness status as the one studied in [9] 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 [9] that

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

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

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

Obviously the identifiability condition $(\star)$ defined in section 2.3 cannot hold for $(V_n)_{n\leqslant 0}$ 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 $n \leqslant -2$.

An important particular case is the one when $p_n = (|n|+1)^{-1}$. In this case, V_n has the uniform distribution on $\{0,\ldots,|n|\}$ for every $n \leqslant 0$ and the filtration $\mathcal{F}$ generated by $(V_n)_{n\leqslant 0}$ is Kolmogorovian and not standard in this case. This results from the standardness criterion provided by theorem 3.7, which was proved in [9] with the help of the I-cosiness criterion, and which is proved in the present paper with the help of Vershik's criterion.

The following proposition about the tail σ- field $\mathcal{F}_{-\infty}$ is a rewriting of proposition 3.1 in [9], to which we refer for a detailed proof.

Proposition 3.1. *The sequence $(V_n)_{n\leqslant 0}$ 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*

$$p_{n_0} = 1 \text{ and } p_n = 0 \text{ for every } n < n_0 \text{ for some } n_0 \leqslant 0 \tag{$*$}$$

   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 $\mathcal{F}_m = \{\varnothing, \Omega\}$ for every $m \leqslant n_0$.

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

Throughout this section, we denote by $(V_n)_{n\leqslant 0}$ the next-jump time process with jumping probabilities $(p_n)_{n\leqslant 0}$ and we denote by $\mathcal{F}$ the filtration it generates. Discarding the elementary case $(*)$, it is shown in [9] with the help of the I-cosiness citerion 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 $k \leqslant 0$, define the random vector $X_k = (V_k, \ldots, V_0)$ and denote by $\mathcal{B}_k = \sigma(V_k, \ldots, V_0)$ 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 $n \leqslant k$, and $V_n = |n|$ almost surely when $p_n = 1$, therefore $\pi_n X_k$ is a degenerate random variable too, and $\mathrm{disp}(\pi_n X_k) = 0$. Consequently, $\mathcal{F}$ satisfies Vershik's standardness criterion by proposition 2.3(a). □

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

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

In this case, $\mathcal{F}$ is generated by the process $(\pi_n V_{-1})_{n\leqslant 0}$, 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 $\left(V_{n_0+n} - |n_0|\right)_{n\leqslant 0}$ is the next-jump time process with jumping probabilities $(p_{n_0+n})_{n\leqslant 0}$.

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

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

has the same distribution than $(V_{n-1})_{n\leqslant 0}$ where $(V_n)_{n\leqslant 0}$ is the next-jump time process with jumping probabilities $(p'_n)_{n\leqslant 0}$ 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. □

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 $\mathbb{V}_n = \{0, \ldots, |n|\}$ 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,

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

for every $n \leqslant -2$, 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 salo 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 $d_n \to 0$.*

Proof. Consequence of lemma 2.5 and lemma 3.3. □

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. Moreover this lemma will be used in section 4 in a situation when only the kernels P_n are given.

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

1) If $n \leqslant -1$ and $x, x' \leqslant |n| - 1$, then $\rho_n(x, x') = \rho_{n+1}(x, x')$.

2) If $n \leqslant -2$ and $x' \leqslant |n| - 1$, then $\rho_n(|n|, x') = P_{n+1}\big(|n|, \rho_{n+1}(\cdot, x')\big)$.

3) If $n \leqslant -3$ and $x' \leqslant |n| - 2$, then $\rho_n(|n|, x') = \rho_{n+1}(|n+1|, x')$.

4) If $n \leqslant -1$, then $\rho_{n-1}(|n-1|, |n|) = (1 - p_n)P_{n+1}\big(|n|, \rho_n(|n|, \cdot)\big)$.

5) If $n \leqslant -2$, then $P_n\big(|n-1|, \rho_{n-1}(|n-1|, \cdot)\big) = (1 - p_n^2)P_{n+1}\big(|n|, \rho_n(|n|, \cdot)\big)$.

6) For every $n \leqslant -1$, $P_n\big(|n-1|, \rho_{n-1}(|n-1|, \cdot)\big) = 2p_{-1}(1 - p_{-1}) \prod_{m=n}^{-2}(1 - p_m^2)$.

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

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

valid for $0 \leqslant k < |n|$ and $n \leqslant -2$. 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. □

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 $n \leqslant -1$.*

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

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. □

4 Central and noncentral ergodic measures on the next-jump graph

This section is a digression in the present paper. It provides a complete case study of the theory introduced in the recent work of Vershik [14]. We are mainly motivated to write it because our previous investigation on the metrics ρ_n is a good opportunity to provide this illustration Vershik's new theory. This theory mainly deals with the identification of *ergodic central probability measures* on a Bratteli graph (or more precisely, on the space of paths of a Brattel graphs). Any probability measure μ on the space of paths a Bratteli graph without multiple edges can be interpreted as the law of the process $(V_n)_{n \leqslant 0}$ where V_n is the vertex at level n of the path taken at random according to μ. Denoting by $\mathcal{F}$ the filtration generated by $(V_n)_{n \leqslant 0}$, then μ is said to be *ergodic* when $\mathcal{F}$ is Kolmogorovian. It is said to be *central* if for every $n < 0$, the conditional distribution $\mathcal{L}\big((V_{n+1}, \ldots, V_0) \,|\, \mathcal{F}_n\big)$ is uniform on the paths connecting the vertex V_n to the root vertex V_0. Therefore, the centrality assumption means the process $(V_n)_{n \leqslant 0}$ is Markovian and is governed by the transition kernels P_n given by

$$P_n(v_{n-1}, v_n) = \Pr(V_n = v_n \,|\, V_{n-1} = v_{n-1}) = \frac{\dim(v_n)}{\dim(v_{n-1})},$$

where $\dim(v)$ denotes the number of paths connecting a vertex v to the root vertex $\varnothing$.

Thus, saying that μ is central means that $(V_n)_{n\leqslant 0}$ is a Markov chain governed by a certain system of transition kernels which is intrinsic to the graph. Although Vershik's theory is mainly focused on central measures, it also works when one considers a given arbitrary system of the transition laws instead of the one corresponding to the centrality assumption. We will illustrate this theory by applying Vershik's theorems to the case of the transition laws of the next-jump time processes (problem 4.1). Vershik's method is based on the metrics ρ_n which are straightforward to derive for our Bratelli graph B with the help of lemma 3.5. This method will be applied in section 4.3 after we investigate the problem by a bare-hands approach in section 4.2.

Centrality of our Bratteli graph B corresponds to the transition laws of the next-jump time process with jumping probabilities $(p_n)_{n\leqslant 0}$ given by $p_n \equiv \frac{1}{2}$ for all $n < 0$. Indeed, for a vertex v_n at level n of our Bratteli graph B, one has $\dim(v_n) = 1$ if $v_n = 0$ and $\dim(v_n) = 2^{v_n - 1}$ for $v_n \in \{1, \ldots, |n|\}$. Therefore, centrality here means that the transition kernels P_n are given by

$$\begin{cases} P_n(v_{n-1}, v_{n-1}) = 1 & \text{if } v_{n-1} \neq |n-1| \\ P_n(|n-1|, v_n) = \begin{cases} \frac{1}{2^{|n|}} & \text{if } v_n = 0 \\ \frac{1}{2^{|n|-v_n+1}} & \text{if } v_n \in \{1, \ldots, |n|\} \end{cases} \end{cases}, \tag{4.1}$$

and we recognize the transition laws of the next-jump time process in case $p_n \equiv \frac{1}{2}$.

Now, under the ergodicity assumption, the law of V_n is the almost sure limit of the conditional law $\mathcal{L}(V_n \mid \mathcal{F}_m)$ when $m \to -\infty$, as a consequence of the convergence theorem for reverse martingales, that is to say

$$\Pr(V_n = v_n) = \lim_{m\to-\infty} \Pr(V_n = v_n \mid V_m) \quad \text{almost surely} \tag{4.2}$$

for every vertex v_n at level n. Though the probability transitions $\Pr(V_n = v_n \mid V_m = v_m)$ are determined under the centrality assumption, in spite of (4.2) the additional ergodicity assumption does uniquely determine the distribution of the Markov process $(V_n)_{n\leqslant 0}$. For example, the next-jump time process in case $p_n \equiv \frac{1}{2}$ for every $n < 0$ is governed by the transition kernels (4.1) as well as the next-jump time process with jumping probabilities $p_n \equiv 0$ for every $n < 0$. Indeed, in this case $V_n = 0$ for every $n \leqslant 0$, and the Markov chain is indeed governed by the transition kernels (4.1) because $P_n(\cdot, 0) = \delta_0$ and $P_n(\cdot, v_{n-1})$ has no importance when v_{n-1} is outside the support of V_{n-1}.

4.1 Statement of the problem

Thus, Vershik's theory introduced in [14] deals with the problem of identifying all possible laws of a Markov chain $(V_n)_{n\leqslant 0}$ that is governed by a given system of transitions kernels and generates a Kolmogorovian filtration. We will illustrate it by investigating the following problem.

Problem 4.1. Let B be the Bratteli graph under study (figure 1) and consider the transition kernels P_n of the next-jump time process defined by a jumping probabilities sequence $(p_n)_{n\leqslant 0}$:

$$\begin{cases} P_n(v_{n-1}, v_{n-1}) = 1 & \text{if } v_{n-1} \neq |n-1| \\ P_n(|n-1|, v_n) = \begin{cases} p_n & \text{if } v_n = |n| \\ (1-p_n)\cdots(1-p_{k-1})p_k & \text{if } v_n = |k| \in \{0, \ldots, |n|-1\} \end{cases} \end{cases} \tag{4.3}$$

What are all the possible Markov chains $(V_n)_{n\leqslant 0}$ governed by these transition laws and generating a Kolmogorovian filtration ?

One can check, and this is shown in [9], that any Markov chain governed by the transition kernels (4.3) satisfies

$$\mathcal{L}(V_n \mid V_m = k) = \mathcal{L}(V_n \mid V_{n-1} = |n-1|) \quad \text{for every } m < n \text{ and } |n-1| \leqslant k \leqslant |m|. \tag{4.4}$$

4.2 Central and noncentral ergodic measures: bare-hands method

Before investigating problem 4.1 under the light of Vershik's theory, we will solve it by a bare-hands approach with the help of the following elementary lemma.

Lemma 4.2. *Let $(V_n)_{n\leqslant 0}$ be a Markov chain on the vertices of a Bratteli graph B. If (4.2) holds and if $\Pr(V_n = x_n \text{ i.o.}) = 1$ for a sequence of vertices $(x_n)_{n\leqslant 0}$ where each x_n is a vertex at level n, and such that $\Pr(V_n = v_n \mid V_m = x_m)$ has a limit as $m \to -\infty$, then $\Pr(V_n = v_n) = \lim_{m\to-\infty} \Pr(V_n = v_n \mid V_m = x_m)$ for every vertex v_n at level n.*

Proof. Under the assumption $\Pr(V_n = x_n \text{ i.o.}) = 1$, the random sequence $\big(\Pr(V_n = v_n \mid V_m)\big)_{m\leqslant n}$ almost surely has a subsequence which is also a subsequence of the convergent deterministic sequence $\big(\Pr(V_n = v_n \mid V_m = x_m)\big)_{m\leqslant n}$. □

The solution of problem 4.1 corresponding to $k = 0$ in theorem below was already mentioned after equation 4.1. It is the degenerate case $V_n \equiv 0$. More generally, for the solution corresponding to the parameter k, one has $V_n = k$ almost surely for every $n \leqslant -k$, and this solution satisfies condition $(*)$ of proposition 3.1.

Theorem 4.3. *The set of distributions solving problem 4.1 is parameterized by $k \in \mathbb{N} \cup \{+\infty\}$. The solution distribution with parameter k is the one for which $\lim_{n\to-\infty} V_n = k$, and this is the distribution of the next-jump time process with jumping probabilities $(p'_n)_{n\leqslant 0}$ given by*

$$p'_n = p_n \quad \text{for } n > -k, \quad p'_{-k} = 1, \quad \text{and} \quad p'_{-k-1} = p'_{-k-2} = \ldots = 0$$

when $k < +\infty$, and $(p'_n)_{n\leqslant 0} = (p_n)_{n\leqslant 0}$ when $k = +\infty$.

Proof. The Markov process $(V_n)_{n\leqslant 0}$ is almost surely decreasing whatever its distribution is. The limit $V_{-\infty} = \lim_{n\to-\infty} V_n$ is a degenerate random variable under the ergodicity assumption, and it is either $V_{-\infty} = k \in \mathbb{N}$ or $V_{-\infty} = +\infty$.

If $V_{-\infty} = k \in \mathbb{N}$, then there is only one possible distribution satisfying (4.3) in view of lemma 4.2, and this solution satisfies $\Pr(V_n = v_n) = \Pr(V_n = v_n \mid V_{n-1} = |n-1|)$ in view of property (4.4). Then we can check that this is the distribution of the next-jump time process defined by the jumping probabilities sequence $(p'_n)_{n\leqslant 0}$ given in the theorem. Note that the transition kernels given by (4.3) are not the same as the ones given by (3.1) for this sequence of jumping probabilities $(p'_n)_{n\leqslant 0}$, but this makes sense because they only differ for values of V_n outside of its support.

If $V_{-\infty} = +\infty$ then the event $\{V_n = |n|\}$ occurs for infinitely many n, because any possible trajectory of $(V_n)_{n\leqslant 0}$ realizing only finitely many events $\{V_n = |n|\}$ is bounded. Therefore $\Pr(V_n = v_n) = \lim_{m\to-\infty} \Pr(V_n = v_n \mid V_m = |m|)$ by lemma 4.2. Hence,$\Pr(V_n = v_n) = \Pr(V_n = v_n \mid V_{n-1} = |n-1|)$ by (4.4), and we recognize the distribution of the next-jump time process defined by the jumping probabilities $(p_n)_{n\leqslant 0}$. □

4.3 Central and noncentral ergodic central measures: Vershik's method

Vershik's theory introduced in [14] distinguishes two mutually exclusive possible situations in a problem such as problem 4.1. Under the first possible situation (precompactness of the intrinsic topology), problem 3.1 is fully solved, and we get more (standardness of the filtrations generated by the solution Markov chains). Under the opposite situation (non-precompactness of the intrinsic topology), the theory partially solves the problem. We will state these results in theorem 4.6. For our problem 4.1 we will meet the two possible situations, depending on the asymptotic behavior of the jumping probabilities sequence $(p_n)_{n\leqslant 0}$.

Denote by $\mathrm{Vert}(B)$ the set of vertices of a Bratteli graph B. Theorem 1 in [14] relates the ergodic central measures to the accumulation points of $(\mathrm{Vert}(B), \rho_B)$, defined as the points of $I(B) := \widehat{\mathrm{Vert}(B)} \setminus \mathrm{Vert}(B)$, where $\widehat{\mathrm{Vert}(B)}$ is the completion of $\mathrm{Vert}(B)$ under the *intrinsic pseudometric* ρ_B introduced in [14], which is a by-product of Vershik's standardness criterion. As we previously said, centrality means that we consider the system of transition kernels (4.1), but one can also define the intrinsic pseudometric in the same way for any other given system of transition kernels.

Perhaps we should reserve the term *intrinsic* to the central case, but in the present paper we will allow ourselves to use it for the pseudometric similarly defined from another given system of transition kernels.

In order to define ρ_B, Vershik firstly defines the *intrinsic pseudometric* ρ_n on the set $\mathbb{V}_n$ of vertices at each level n. It is initiated by a pseudometric ρ_{-1} on $\mathbb{V}_{-1}$ and then ρ_n is nothing but the pseudometric ρ_n studied in section 3.2, which takes its origin in Vershik's standardness criterion (lemma 2.5).

In our example, 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 $(\mathbb{V}_n, \rho_n) \to (\mathbb{V}_{n-1}, \rho_{n-1})$ 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 $n \leqslant -2$:

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

where the expression of d_n is given in lemma 3.6 for every $n \leqslant -1$ 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 $n \leqslant -2$ is explicitely 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 $(\mathbb{V}_n, \rho_n)$ 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 $n < 0$).

In lemma below and in theorem 4.6, for a given system of transition kernels P_n we denote by $P_{n|m}$ the product kernel $\prod_{k=m+1}^{n} P_k$ for $m < n$. Thus $\mathcal{L}(V_n \,|\, V_m) = P_{n|m}(V_m, \cdot)$ for any Markov chain $(V_n)_{n \leqslant 0}$ governed by the P_n.

Definition 4.4. The intrinsic pseudometric ρ_B on the space of vertices $\mathrm{Vert}(B) = \bigcup_{n \leqslant 0} \mathbb{V}_n$ induced by a system of transition kernels P_n is firstly defined by $\rho_B(v_n, v_n') = \rho_n(v_n, v_n')$ for all vertices v_n and v_n' at a same level n, where ρ_n is the iterated Kantorovich pseudometric initiated at time 0 by a given pseudometric ρ_0. Then it is defined by

$$\rho_B(v_m, v_n') = P_{n|m}[v_m, \rho_n(\cdot, v_n')]$$

for all vertices v_m at level m and v_n' at level n with $m < n \leqslant -1$.

Theorem 4.5. *Let B be the Bratteli graph under study and consider the intrinsic pseudometric ρ_B induced by the transition kernels P_n of the next-jump time process with jumping probabilities sequence $(p_n)_{n \leqslant 0}$, given by (4.3).*

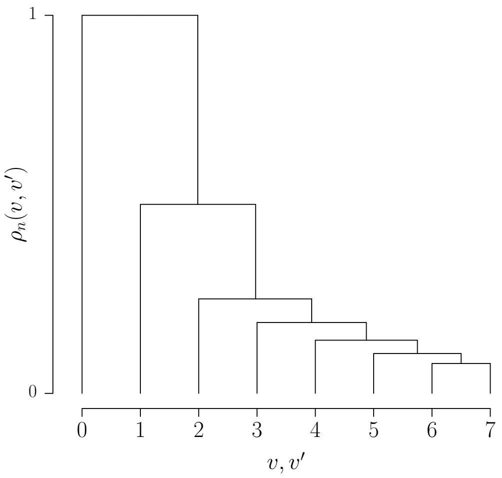


Figure 2: The space $(\mathbb{V}_n, \rho_n)$.

1) *The intrinsic pseudometric is given by* $\rho_B(v_m, v'_n) = \rho_m(v_m, v'_n)$ *for all vertices* v_m *at level* m *and* v'_n *at level* $n > m$.

2) *Under the identifiability condition* (3.2) *(given in lemma 3.3), the intrinsic pseudometric generates a precompact topology if and only if* $\sum p_n^2 = \infty$. *The accumulation set* $I(B)$ *is* $\mathbb{N} \cup \{+\infty\}$ *in the precompact case and* $\mathbb{N}$ *otherwise.*

Proof. We firstly derive the intrinsic pseudometric. Assume $v = v_m$ is a vertex at level m and $v' = v'_n$ is a vertex at level $n > m$. If $v_m \leqslant |m|$ then $\mathcal{L}(V_n \mid V_m = v_m)$ is the Dirac distribution at the vertex at level n whose label is the same as v_m and then $\rho_B(v, v') = \rho_n(v, v') = \rho_m(v, v')$ by the canonical isometries $(\mathbb{V}_n, \rho_n) \to (\mathbb{V}_{n-1}, \rho_{n-1})$. If $v_m \geqslant |m|$ then we know by (4.4) that $\mathcal{L}(V_n \mid V_m = v_m) = \mathcal{L}(V_n \mid V_{n-1} = |n-1|)$. Therefore by equality 2) in lemma 3.5, $\rho_B(v_m, v'_n) = \rho_{n-1}(|n-1|, v'_n)$ and we know that $\rho_{n-1}(|n-1|, v'_n) = \rho_m(v_m, v'_n)$ (see table 1).

Now we check 2). The ρ_n are metrics under the identifiability condition (3.2) and two vertices are at 0 distance under ρ_B if and only if the have the same label. Therefore the quotient space $\mathrm{Vert}(B)/\rho_B$ is isomorphic to $(\mathbb{N}, \rho_{-\infty})$ where $\rho_{-\infty}$ is the direct limit of ρ_n as $n \to -\infty$, defined by $\rho_{-\infty}(m, n) = \rho_{-m}(m, n)$ when $m \leqslant n$. It is precompact if and only if $+\infty$ is an accumulation point, which is equivalent to $\rho_{-\infty}(k, k+1) \to 0$ as $k \to +\infty$. We know that $\rho_{-\infty}(k, k+1) = (1 - p_{-k+1}) d_{-k+2}$ and it is not diffcult to check that it goes to 0 if and only if $\sum p_n^2 = \infty$. ☐

Now we have everything required to apply Vershik's following theorem.

Theorem 4.6 (Vershik [14])**.** *Assume we are looking for a Markov chain* $(V_n)_{n \leqslant 0}$ *on the vertices of a Bratteli graph* B*, which generates a Kolmogorovian filtration and which is governed by some given transition kernels* P_n *satisfying the identifiability condition* $P_n(v, \cdot) \neq P_n(v', \cdot)$ *for* $v \neq v'$ *and every* $n < 0$. *Define the intrinsic pseudometric* ρ_B *induced by these transition laws. Define the accumulation points of* $\mathrm{Vert}(B)$ *as the points of* $I(B) := \widehat{\mathrm{Vert}(B)} \setminus \mathrm{Vert}(B)$*, where* $\widehat{\mathrm{Vert}(B)}$ *is the completion of* $\mathrm{Vert}(B)$.

1) *For every* $x \in I(B)$ *and every sequence* $(x_n)_{n \leqslant 0}$ *converging to* x*, where each* x_n *is a vertex at level* n*, the limit* $P_{n|m}(x_m, \cdot)$ *at* $m \to -\infty$ *exists for every* $n \leqslant 0$ *and only depends on* x*, and one defines a possible distribution of* $(V_n)_{n \leqslant 0}$ *by setting*

$$\Pr(V_n = v_n) = \lim_{m \to -\infty} P_{n|m}(x_m, v_n)$$

for every $n < 0$. *Moreover this Markov chain* $(V_n)_{n \leqslant 0}$ *generates a standard filtration.*

2) *If the intrinsic pseudometric induces a precompact topology on* $\mathrm{Vert}(B)$*, then every possible distribution of* $(V_n)_{n \leqslant 0}$ *is given by an accumulation point* x *as in 1).*

Thus, note that Vershik's theory only provides Markov chains $(V_n)_{n\leqslant 0}$ solving problem 4.1 that generate a standard filtration, a property stronger than the desired Kolmogorov property.

Theorem 4.7. *Let B be the Bratteli graph under study and consider the conditional laws $\mathcal{L}(V_n \mid V_{n-1} = v)$ of the next-jump time process defined by the jumping probabilities sequence $(p_n)_{n\leqslant 0}$ satisfying the identifiability condition* (3.2)*. The accumulation set $I(B)$ has been derived in theorem 4.5.*

1) *For each accumulation point $k \in I(B)$, the distribution of $(V_n)_{n\leqslant 0}$ given by 1) in theorem 4.6 is the distribution of the next-jump time process given by the jumping probabilities sequence $(p'_n)_{n\leqslant 0}$*
$$p'_n = p_n \text{ for } n > -k, \quad p'_{-k} = 1, \quad \text{and } p'_{-k-1} = p'_{-k-2} = \ldots = 0$$
when $k < +\infty$, and $(p'_n)_{n\leqslant 0} = (p_n)_{n\leqslant 0}$ in the case when $k = +\infty$. We know in addition that the filtration of $(V_n)_{n\leqslant 0}$ is standard for these distributions.

2) *In the precompact case $\sum p_n^2 = \infty$, there is no other solution than those given by 1).*

Proof. Assertion 1) is not difficult to check with the help of property (4.4), and assertion 2) is an application theorem 4.6. ☐

We have seen in theorem 4.3 that there is one other solution in the non-precompact case: the one which was naturally parameterized by its limit $k = +\infty$ in theorem 4.3, and for which the associated filtration is not standard by theorem 3.7. This solution is not provided by theorem 4.6 because $k = +\infty$ is not an accumulation point in the non-precompact case (theorem 4.5). In the precompact case, Vershik's theorem provides more than the set of solutions: it also asserts that corresponding filtrations are standard.

5 The uniformly scaled entropy

In this section we introduce scaled entropies of filtrations by following Vershik and Gorbulsky [12], except that we use the probabilistic language and we restrict our attention to scalings which are not ϵ-dependent. Theorem 5.6 is our main result, it provides a more general claim than theorem 2.4. We denote by $H(X)$ the entropy of a discrete random variable X, defined as $H(X) = -\sum \mu_i \log \mu_i$ where μ is the law of X.

5.1 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 5.1. Let X be an integrable random variable taking its values in a Polish metric space (E, ρ). The ϵ-entropy of X is
$$H^{\varepsilon}(X) = \inf\{H(F) \mid \mathbb{E}[\rho(X, F)] < \epsilon\}$$
where the infimum is taken over E-valued but discrete $\sigma(X)$-measurable random variables F.

The *scaling* $c\colon (-\mathbb{N}) \to]0, \infty[$ in definition below is termed as *uniform scaling* because Vershik and Gorbulsky more generally allow ε-dependent scaling $c \mapsto c(\varepsilon, n)$. 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 5.2. 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 $H_n^\varepsilon(X, \mathcal{F}) = H^\varepsilon(\pi_n X)$, shorter denoted by $H_n^\varepsilon(X)$ 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 $c\colon (-\mathbb{N}) \to]0, \infty[$, termed as *uniform scaling*.

2) The limit

$$h_c(X, \mathcal{F}) = \limsup_{\varepsilon \to 0} \limsup_{n \to 0} \frac{H_n^\varepsilon(X)}{c(n)}$$

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 5.3. 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 5.11 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 5.4. As already mentioned in the definition, the ϵ-entropy $H_n^\varepsilon(X)$ 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 $(\pi_n X)_{n \leqslant 0}$ of the Vershik progressive predictions of X. Indeed, it is easy to see that the value of $H_n^\varepsilon(X)$ 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 [2]).

Remark 5.5. We could replace the ϵ-entropy $H^\varepsilon(X)$ by

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

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

5.2 Properties

The goal of this section is to prove theorem 5.6, which is a deep generalization of theorem 2.4. It will be used in section 6 to study the uniform entropy scalings of the next-jump time filtrations.

Theorem 5.6. *Let $\mathcal{F}$ be a filtration, $X \in L^1(\mathcal{F}_0; E)$ where E is a Polish space, and $c\colon (-\mathbb{N}) \to]0, \infty[$ 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 $(\pi_n X)_{n \leqslant 0}$.*

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 $\sigma(\pi_n X; n \leqslant 0)$ when we consider $\mathcal{F}$ as the underlying filtration (see remark 5.4). Theorem 5.6 is particularly useful when the $\pi_n X$ are discrete random variables with finite entropy, because it gives the upper bound $h_c(\mathcal{F}^X) \leqslant 1$ for any scaling $c(n) \sim H(\pi_n X)$.

Theorem 5.6 will be derived from the following series of lemmas and propositions.

Lemma 5.7. *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 [2]. □

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 [7].

Lemma 5.8. *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 $W^n = (\pi_n X, \ldots, \pi_{-1} X, X)$ for some $n \leqslant 0$. Consider the metric $\bar{\rho}_n = \frac{1}{|n|+1} \sum_{k=n}^{k=0} \rho^{(k)}$ on the state space of W^n. Then $\pi_n W^n = \phi(\pi_n X)$ where 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 $n \leqslant 0$ and $k \in \{n, \ldots, 0\}$, one has $\pi_k W_n = g_k^n(\pi_n X, \ldots, \pi_k X)$ for some functions g_k^n related by the fact that $g_{k-1}^n(\mu_n, \ldots, \mu_{k-1})$ is the distribution of $g_k^n(\mu_n, \ldots, \mu_{k-1}, M_k)$ where $M_k \sim \mu_{k-1}$. Therefore

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

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

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

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. □

Lemma 5.9. *Let $\mathcal{F}$ be a filtration, $(X_k)_{k \geqslant 1}$ be a sequence in $L^1(\mathcal{F}_0; E)$ where E is Polish. If $X_k \to X$ 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 $k \geqslant 1$, then for every $\epsilon_0 > 0$ there exists k_0 such that $H_n^{\epsilon_0}(X_{k_0}) \geqslant H_n^{2\epsilon_0}(X)$ for every $n \leqslant 0$.*

Proof. Let $k_0 = k(\epsilon_0)$ such that $\mathbb{E}[\rho(X_{k_0}, X)] \leqslant \epsilon_0$, hence $H_n^{\epsilon_0}(X_{k_0}) \geqslant H_n^{2\epsilon_0}(X)$ for every n by definition of $H_n^{\varepsilon}(\cdot)$ and lemma 2.1. (Actually the lemma is true for every $k \leqslant k_0$ if we take $\mathbb{E}[\rho(X_k, X)] \leqslant \epsilon_0$ for every $k \geqslant k_0$.) □

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

Proposition 5.10. *Let $c\colon (-\mathbb{N}) \to]0, \infty[$ be a scaling. If, under the same hypotheses as lemma 5.9, there exists $\ell \geqslant 0$ such that $h_c(X_k, \mathcal{F}) \leqslant \ell$ for every $k \geqslant 1$ then $h_c(X, \mathcal{F}) \leqslant \ell$.*

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 $\limsup_{n\to-\infty} \frac{H_n^{2\epsilon_0}(X)}{c(n)} > \ell + 1$. Therefore there is k_0 such that $\limsup_{n\to-\infty} \frac{H_n^{\epsilon_0}(X_{k_0})}{c(n)} > \ell + 1$ by lemma 5.9. But $\epsilon \mapsto H_n^{\epsilon}(X_{k_0})$ is decreasing, therefore inequality $\limsup_{n\to-\infty} \frac{H_n^{\epsilon}(X_{k_0})}{c(n)} > \ell+1$ holds for every $\epsilon \leqslant \epsilon_0$, a contradiction of the assumption of the lemma.

Knowing now that $a < \infty$, we check that $\ell \geqslant a$. Given $\delta > 0$, the definition of the superior limit provides ϵ_0 such that $\limsup_{n\to-\infty} \frac{H_n^{2\epsilon_0}(X)}{c(n)} > a - \delta$. By lemma 5.9 and since $\epsilon \mapsto H_n^{\epsilon}(X_{k_0})$ is decreasing, one gets $\limsup_{n\to-\infty} \frac{H_n^{\epsilon}(X_{k_0})}{c(n)} > a - \delta$. for some k_0 and every $\epsilon \leqslant \epsilon_0$. Finally $\ell \geqslant a$. $\square$

Proposition 5.11. *Let $\mathcal{F}$ be a filtration and $X \in L^1(\mathcal{F}_0; E)$ where E is a Polish. Then $h_c(\sigma(X), \mathcal{F}) = h_c(X, \mathcal{F})$ for any scaling $c\colon (-\mathbb{N}) \to]0, \infty[$.*

Proof. If $Y = f(X)$ for some Lipschitz function f then it is easy to check that $h_c(Y, \mathcal{F}) \leqslant h_c(X, \mathcal{F})$ with the help of lemma 5.7. The result follows from proposition 5.10 and from the density of the set of random variables $f(X)$, f Lipschitzian, in $L^1\big(\sigma(X), [0, 1]\big)$ (see lemma 2.15 in [6]). $\square$

Now we can quickly prove theorem 5.6.

Proof of theorem 5.6. Let $\mathcal{B}_n = \sigma(\pi_n X, \ldots, \pi_{-1} X, X)$. By lemma 5.8 and proposition 5.11, $h_c(\mathcal{B}_n, \mathcal{F}) = h_c(X, \mathcal{F})$. Then the theorem follows from proposition 5.10. $\square$

Proposition 5.12 below is another corollary of lemma 5.8, generalizing propostion 2.2.

Proposition 5.12. *Let $\mathcal{F}$ be a filtration, $n_0 \leqslant 0$ be an integer, and denote by $\mathcal{F}^{n_0]} = (\mathcal{F}_{n_0+n})_{n\leqslant 0}$ the filtration $\mathcal{F}$ truncated at n_0. Let $c\colon (-\mathbb{N}) \to]0, \infty[$ be a scaling and denote $c^{n_0]} = (c_{n_0+n})_{n\leqslant 0}$ 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

$$H_n^{\epsilon}(X_{n_0}, \mathcal{F}^{n_0]}) = H_{n_0+n}^{\epsilon}(X_{n_0}, \mathcal{F}) \tag{5.1}$$

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

Conversely, if X_0 is an integrable $\mathcal{F}_0$-measurable random variable, then one has $h_c(X_0, \mathcal{F}) \leqslant h_c(W^{n_0}, \mathcal{F})$ by proposition 5.11, where $W^{n_0} = (\pi_{n_0} X_0, \ldots, X_0)$. But lemma 5.8 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 $n \leqslant 0$. Hence equality (5.1) 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 $h_c(X_0, \mathcal{F}) \leqslant h_{c^{n_0]}}(\pi_{n_0} X_0, \mathcal{F}^{n_0]})$. This provides the inequality $h_c(\mathcal{F}) \leqslant h_{c^{n_0]}}(\mathcal{F}^{n_0]})$. $\square$

6 Entropy of next-jump time filtrations

In this section, we consider, for a given sequence $(p_n)_{n\leqslant 0}$ of jumping probabilities, the next-jump time process $(V_n)_{n\leqslant 0}$ introduced in section 3 and its filtration $\mathcal{F}$. 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 $\sum p_n^2 < \infty$. Preliminarily, we study the entropy of the random variables V_n.

6.1 Entropy of 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:

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

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 $k \leqslant |n|+1$ 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

$$H(V_{n-1}) = h(p_{n-1}) + (1-p_{n-1})H(V_n), \tag{6.1}$$

yielding

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

Now, note that $p_{-V_n} > 0$ almost surely, because for every $k \in \{0, \ldots, |n|\}$, 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 6.1. *When $\mathcal{F}$ is Kolmogorovian but not standard,* $\lim H(V_n) = +\infty$ *and* $\lim \frac{H(V_{n-1})}{H(V_n)} = 1$.

Proof. According to proposition 3.1, $V_n \to +\infty$ almost surely when $\mathcal{F}$ is Kolmogorovian. Therefore $V_n = |n|$ for infinitely many n, because any possible trajectory of $(V_n)_{n\leqslant 0}$ realizing only finitely many events $\{V_n = |n|\}$ is bounded. In addition, by theorem 3.7, $p_n \to 0$, hence $\frac{h(p_{-V_n})}{p_{-V_n}} \to +\infty$ because $h(x)/x \to +\infty$ when $x \to 0^+$. We deduce from the recursive relation (6.1) that $\frac{H(V_{n-1})}{H(V_n)} \to 1$, by noting that $p_n \to 0$ in the Kolmogorovian but not standard case. □

6.2 Entropy of the filtration

We have seen in section 3 that the assumption of lemma 6.2 below is fulfilled for the random variable V_{-1} of the next-jump time process $(V_n)_{n\leqslant 0}$ in case of nonstandardness. The proof of this lemma involves *Fano's inequality* (see [3]), whose statement is:

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

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 6.2. *Let $\mathcal{F}$ be a filtration, (E, ρ) a Polish metric space, $X \in L^1(\mathcal{F}_0; E)$, and $c\colon (-\mathbb{N}) \to]0, \infty[$ 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 $n \leqslant 0$ as long as $x \neq x'$. Then*

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

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 $H_n^\epsilon(X) = H(F_n) \leqslant H(\pi_n X)$ 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, $H(\pi_n X) - H(F_n) = H(\pi_n X \mid F_n)$, and by Fano's inequality,

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

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

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

and consequently

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

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

□

Now consider the next-jump time process $(V_n)_{n\leqslant 0}$ and its filtration $\mathcal{F}$. Recall that we have seen in lemma 6.1 that $\lim H(V_n) = +\infty$ in the case when $\mathcal{F}$ is Kolmogorovian but not standard.

Proposition 6.3. *Assume $\mathcal{F}$ is Kolmogorovian but not standard. Let $c\colon (-\mathbb{N}) \to]0,\infty[$ be a uniform scaling such that $c(n) \sim H(V_n)$. Then one always has $h_c(\mathcal{F}) \leqslant 1$, 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 $n \leqslant -2$). By lemma 3.3(1) and theorem 5.6,

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

with $X = V_{-1}$, for any scaling c. In the uniform case $p_n = (|n|+1)^{-1}$, we know by lemma 6.2 that $h_c(X,\mathcal{F}) = 1$ for $c(n) \sim H(V_n) = \log(|n|+1)$. In other cases, one obviously has

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

and then $h_c(X,\mathcal{F}) \leqslant 1$ 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 6.1. If $p'_{-1} = 0$, we similarly conclude by using lemma 3.3(3) after noting that p'_n cannot be zero for every $n \leqslant -1$ in the Kolmogorovian but non-standard case. □

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 $H(\pi_n X \mid F_n)$ in the proof of lemma 6.2. We have used Fano's inequality which is a general majoration of the conditional entropy. Generalized Fano's inequalities studied in [4] 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.

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] 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)

[3] Gray, R.M. : Entropy and Information Theory. Springer, 2nd ed, 2011.

[4] Han, T.S. Verdú, S: Generalizing the Fano inequality. IEEE Transactions on Information Theory **40**, 1247–1251.

[5] Janvresse, E., Laurent, S., de la Rue, T.: Standardness of monotonic Markov filtrations. Preprint 2014.

[6] Laurent, S.: On standardness and I-cosiness. Séminaire de Probabilités XLIII, Springer Lecture Notes in Mathematics 2006, 127–186 (2010)

[7] 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)

[8] Laurent, S.: Further comments on the representation problem for stationary processes. Statist. Probab. Lett. 80, 592–596 (2010).

[9] Laurent, S: Standardness and nonstandardness of next-jump time filtrations. Electronic Communications in Probability 18 (2013), no. 56, 1–11.

[10] 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)

[11] 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)

[12] Vershik, A.M, Gorbulsky, A.D.: Scaled entropy of filtrations of σ-fields. Theory of Probability and its Applications, **52**, 3, 493–508, 2008.

[13] Vershik, A.M.: Smooth and non-smooth AF-algebras and problem on invariant measures. arXiv:1304.2193 (2013) `http://arxiv.org/abs/1304.2193`

[14] Vershik, A.M.: Intrinsic metric on graded graphs, standardness, and invariant measures. Zapiski Nauchn. Semin. POMI 421, 58–67 (2014).