ENTROPY OF KILLED-RESURRECTED STATIONARY MARKOV CHAINS

SERVET MARTÍNEZ,* *Universidad de Chile*

Abstract

We consider a strictly substochastic matrix or a stochastic matrix with absorbing states. By using quasi-stationary distributions we show that there is an associated canonical Markov chain that is built from the resurrected chain, the absorbing states, and the hitting times, together with a random walk on the absorbing states, which is necessary for achieving time stationarity. Based upon the 2-stringing representation of the resurrected chain, we supply a stationary representation of the killed and the absorbed chains. The entropies of these representations have a clear meaning when one identifies the probability measure of natural factors. The balance between the entropies of these representations and the entropy of the canonical chain serves to check the correctness of the whole construction.

Keywords: Markov chains; entropy; quasi-stationary distributions

2010 Mathematics Subject Classification: Primary 60J10
                                          Secondary 28D20

1. Introduction

Our starting point is an irreducible strictly substochastic matrix P_I on a countable set I. It defines a killed Markov chain when adding a cemetery that is an absorbing state. One of our purposes is to explore how we can study the entropy of this chain.

The problem can be posed for a Markov chain that is absorbed in a class of states, which is not necessarily a singleton. It is in this enlarged setting that we study the entropy. In this study we use the following concepts:

- the quasi-stationary distribution (QSD) of the matrix P_I. It exists and it is unique when I is finite, and when I is infinite we assume there exists some QSD;

- the Markov chain defined by resurrecting the absorbed chain with the QSD;

- the 2-stringing of the resurrected Markov chain.

In Proposition 1 we show that every QSD defines a canonical stationary distribution associated with an absorbed chain. A construction of this associated stationary chain is given in Proposition 2, showing that it can be recovered from the resurrected chain with the QSD and some additional random elements: the killing on the orbits, the transition to the absorbing

---

states, and a walk on the set of absorbing states. In this chain the absorbing states share the same transition probabilities.

The 2-stringing of the resurrected Markov chain is used to supply stationary Markov representations of the killed and the absorbed Markov chains in an appropriate way, to compute their entropies and provide a clear interpretation. This is done in Sections 5.1 and 5.2 and in Propositions 3 and 4. The entropies are interpreted by identifying the probability measure on the fibers of some natural factors. The entropy of the killed chain is the entropy of the resurrected chain plus the entropy of being alive or killed, and in the absorbed case we must add the entropy of the states where they are absorbed. These additional terms are given by the Abramov–Rokhlin formula on some factors. We note that since the killed and the absorbed trajectory are finite, then almost all the orbits of the stationary representations of the killed and the absorbed Markov chains contain all the killed or absorbed trajectories.

Finally, in Proposition 5 the entropy of the associated stationary chain is decomposed into the entropies of the absorbed chain and of the walk on the set of absorbing states. This last element serves to complete the understanding of the stationary representation of the absorbed chain: it gives the return time to I and the weights of the absorbing states that are necessary for stationarity. The main parameter of the whole construction is the reciprocal of the mean length of the walk, which is also the weight of the set of states I and the Perron–Frobenius eigenvalue of P_I; see the relations (9), (4) and (1).

Before proceeding further, we shall illustrate the associated Markov chain by considering a simplified model. Towards this end, the states in I will be called observable while the absorbing states will be called reservoirs; a transition from a reservoir to an observable state is called an outbreak. The dynamics within the reservoirs cannot be observed, it is only seen at the observable states. In a first approach this occurs in disease epidemics that emerge and disappear in time. A disease may emerge at some location due to human contact with a biological reservoir of some microorganism, diffuse in some geographical area, and when finished retreat unseen to some biological reservoir again. In this case the observable states represent the sizes of the healthy and infected populations in geographic areas, while the reservoirs correspond to the biological reservoirs of the microorganism. From our result the associated Markov chain requires a walk over the reservoirs in order to be stationary in time.

As usual, we use the capital letter H for the entropy of a discrete random variable and h for the entropy of a stationary chain.

Even if it is not usual, we use 'trajectory' to refer to a visit of a finite sequence of states, and 'orbit' for a bilateral sequence of states, that is, for a point in a bilateral product space.

2. Killed and absorbed chains

Let $P_I = (P(i, j) : i, j \in I)$ be an irreducible strictly substochastic matrix on a countable set I. As usual, we add a state $\partial \notin I$ called a cemetery, and the extension of P_I to $I \cup \{\partial\}$ is denoted by P, which satisfies $P(i, \partial) = 1 - \sum_{j \in I} P(i, j)$ for $i \in I$, and the absorption condition $P(\partial, \partial) = 1$. Strictly, substochasticity is equivalent to $\sum_{i \in I} P(i, \partial) > 0$. By irreducibility the states in I are transient. The process defined by P_I is identified with the chain absorbed at a unique cemetery ∂.

The existence of a unique cemetery models the killing when this phenomenon can be interpreted similarly for to all states, for instance in extinction where a unique ∂ has a clear meaning. But there can be several ways of being killed or hitting a boundary, and this is expressed by the existence of a set of absorbing states which is not necessary a singleton.

So, we consider a more general situation. Let $P = (P(a, b) : a, b \in I \cup \mathcal{E})$ be a stochastic matrix on the countable set $I \cup \mathcal{E}$ whose restriction to I is $P_I = (P(i, j) : i, j \in I)$, and such that all the states in $\mathcal{E}$ are absorbing and are attained from I. This last means that $\sum_{i \in I} P(i, \varepsilon) > 0$ for every $\varepsilon \in \mathcal{E}$. We retrieve the one-point absorption when $\mathcal{E} = \{\partial\}$.

Let $\mathcal{X} = (\mathcal{X}_n)$ be a Markov chain with transition matrix P; it will be called absorbed chain. By $\mathbb{P}_a$ we mean the law of this chain when starting from $a \in I \cup \mathcal{E}$, and $\mathbb{E}_a$ denotes the associated mean expected value. Let $\tau_{\mathcal{E}} = \inf\{n \geq 1 : \mathcal{X}_n \in \mathcal{E}\}$ be the first return time to $\mathcal{E}$. If I is finite, the hypotheses made on the chain imply $\mathbb{P}_i(\tau_{\mathcal{E}} < \infty) = 1$ for all $i \in I$. In the countable case we assume that $\mathbb{P}_i(\tau_{\mathcal{E}} < \infty) = 1$ for all $i \in I$.

Let $\mathcal{X}^{(K)} = (\mathcal{X}_n : 0 \leq n < \tau_{\mathcal{E}})$ and $\mathcal{X}^{(A)} = (\mathcal{X}_n : 0 \leq n \leq \tau_{\mathcal{E}})$ be, respectively, the killed and the absorbed trajectory, both starting from $\mathcal{X}_0$. The first one finishes when it is killed, and the second one is stopped at the state where it is absorbed.

2.1. Quasi-stationary distributions

A QSD $\mu = (\mu(i) : i \in I)$ associated with P_I is a probability measure μ on I such that for all $i \in I$, $\mathbb{P}_\mu(\mathcal{X}_n = i \mid \tau_{\mathcal{E}} > n) = \mu(i)$. By writing this equality for $n = 1$, we check that the row vector $\mu^\top$ is a strictly positive left eigenvector of P_I properly normalized (the sum of its components is 1), with eigenvalue $\gamma = \mathbb{P}_\mu(\tau_{\mathcal{E}} > 1) \in (0, 1)$, that is,

$$\mu^\top P_I = \gamma \mu^\top, \text{ with } \gamma = \sum_{i,j \in I} \mu(i) P(i, j) = \mathbb{P}_\mu(\tau_{\mathcal{E}} > 1). \tag{1}$$

It follows that $\mathbb{P}_\mu(\tau_{\mathcal{E}} > k) = \gamma^k$ for all $k \geq 0$. So, if μ is a QSD then the survival time is Geometric$(1 - \gamma)$ distributed; see Lemma 2.2 in [8]. In the finite case there is a unique QSD (see [4]); it corresponds to the normalized left Perron–Frobenius eigenvector, and γ is the associated eigenvalue. The properties of QSD depend on the killed trajectory $\mathcal{X}^{(K)} = (\mathcal{X}_n : 0 \leq n < \tau_{\mathcal{E}})$. In the infinite case QSDs can exist or not (because the positive left eigenvectors can be of infinite mass), and when they exist there could be more than one (even a continuum of them). From now on we fix some QSD μ which, as just discussed, exists and is unique in the finite case and we assume its existence in the infinite case.

Let us give some independence properties between the time of killing and the absorption state. In Theorem 2.6 in [3] the independence relation $\mathbb{P}_\mu(\mathcal{X}_n = i, \tau_{\mathcal{E}} > n) = \mu(i) \gamma^n$ for all $i \in I$ and $n \geq 0$ was stated. Let us prove that when starting from μ, the pair $(\mathcal{X}_{\tau_{\mathcal{E}}-1}, \mathcal{X}_{\tau_{\mathcal{E}}})$, consisting of the last visited state before absorption and the absorption state, is independent of the random time $\tau_{\mathcal{E}}$. For $n \geq 1$, $i \in I$, and $\varepsilon \in \mathcal{E}$, we have

$$\begin{aligned}
\mathbb{P}_\mu(\mathcal{X}_{\tau_{\mathcal{E}}-1} = i, \mathcal{X}_{\tau_{\mathcal{E}}} = \varepsilon, \tau_{\mathcal{E}} = n) &= \mathbb{P}_\mu(\mathcal{X}_{n-1} = i, \mathcal{X}_n = \varepsilon, \tau_{\mathcal{E}} = n) \\
&= \mathbb{P}_\mu(\mathcal{X}_n = \varepsilon \mid \mathcal{X}_{n-1} = i)\mathbb{P}_\mu(\mathcal{X}_{n-1} = i, \tau_{\mathcal{E}} > n - 1) \\
&= P(i, \varepsilon)\,\mu(i)\,\mathbb{P}_\mu(\tau_{\mathcal{E}} > n - 1) = P(i, \varepsilon)\,\mu(i)\,\gamma^{n-1}.
\end{aligned}$$

Then, the independence relation follows. We can be more precise: we have

$$\mathbb{P}_\mu(\mathcal{X}_{\tau_{\mathcal{E}}-1} = i, \mathcal{X}_{\tau_{\mathcal{E}}} = \varepsilon) = P(i, \varepsilon)\,\mu(i) \left(\sum_{l \geq 1} \gamma^{l-1} \right) = P(i, \varepsilon)\,\mu(i)\,(1 - \gamma)^{-1}.$$

Since $\mathbb{P}_\mu(\tau_{\mathcal{E}} = n) = (1 - \gamma)\gamma^{n-1}$, the desired relation holds:

$$\mathbb{P}_\mu(\mathcal{X}_{\tau_{\mathcal{E}}-1} = i, \mathcal{X}_{\tau_{\mathcal{E}}} = \varepsilon, \tau_{\mathcal{E}} = n) = \mathbb{P}_\mu(\mathcal{X}_{\tau_{\mathcal{E}}-1} = i, \mathcal{X}_{\tau_{\mathcal{E}}} = \varepsilon)\mathbb{P}_\mu(\tau_{\mathcal{E}} = n).$$

The above computations also show that the exit law of I when starting from μ satisfies

$$
\mathbb{P}_\mu(\mathcal{X}_{\tau_\mathcal{E}} = \varepsilon) = (1 - \gamma)^{-1} \left(\sum_{i \in I} \mu(i) P(i, \varepsilon) \right),
$$

$$
\mathbb{P}_\mu(\mathcal{X}_{\tau_\mathcal{E}} = \varepsilon, \tau_\mathcal{E} = n) = \left(\sum_{i \in I} \mu(i) P(i, \varepsilon) \right) \gamma^{n-1}.
$$
(2)

These properties depend on the absorbed trajectory $\mathcal{X}^{(A)} = (\mathcal{X}_n : 0 \le n \le \tau_\mathcal{E})$.

3. Associated stationary chain

Let $\rho = (\rho(a) : a \in I \cup \mathcal{E})$ be a probability vector. We define the stochastic matrix $P^\rho = (P^\rho(a, b) : a, b \in I \cup \mathcal{E})$ by

$$
P^\rho(i, b) = P(i, b) \text{ if } i \in I, b \in I \cup \mathcal{E}, \qquad P^\rho(\varepsilon, b) = \rho(b) \text{ if } \varepsilon \in \mathcal{E}, b \in I \cup \mathcal{E}.
$$

So, in P^ρ the ε-row is $P^\rho(\varepsilon, \bullet) = \rho^\top$ for all $\varepsilon \in \mathcal{E}$.

Proposition 1. *Every QSD μ of P_I determines a probability distribution $\pi = (\pi(a) : a \in I \cup \mathcal{E})$ given by*

$$
\text{for all } \varepsilon \in \mathcal{E}, \ \pi(\varepsilon) = \sum_{i \in I} \mu(i) P(i, \varepsilon), \text{ and for all } i \in I, \ \pi(i) = \gamma \mu(i),
$$
(3)

which is a stationary distribution of the matrix $P^\pi = (P^\pi(a, b) : a, b \in I \cup \mathcal{E})$. In a reciprocal way, every distribution $\widetilde{\pi}$ that satisfies $\widetilde{\pi}^\top = \widetilde{\pi}^\top P^{\widetilde{\pi}}$ is defined by a QSD μ as in (3). So, if P_I has a unique QSD (as in the finite case) then there is a unique distribution π that satisfies $\pi^\top = \pi^\top P^\pi$.

Proof. The QSD μ satisfies $\mu^\top P_I = \gamma \mu^\top$ with $\gamma \in (0, 1)$ and $\sum_{i \in I} \mu(i) = 1$. The vector π is a probability distribution because, from (3) and (1), $\pi(I) = \sum_{i \in I} \pi(i)$ and $\pi(\mathcal{E}) = \sum_{\varepsilon \in \mathcal{E}} \pi(\varepsilon)$ satisfy

$$
\pi(I) = \sum_{i \in I} \gamma \mu(i) = \gamma \text{ and } \pi(\mathcal{E}) = \sum_{i \in I} \mu(i) P(i, \mathcal{E}) = 1 - \gamma.
$$
(4)

We check that π is stationary for P^π. For $\varepsilon \in \mathcal{E}$ and $j \in I$ we have

$$
\left(\pi^\top P^\pi \right)(\varepsilon) = \pi(\varepsilon) \sum_{\delta \in \mathcal{E}} \pi(\delta) + \sum_{i \in I} \pi(i) P(i, \varepsilon) = \pi(\varepsilon)(1 - \gamma) + \gamma \pi(\varepsilon) = \pi(\varepsilon),
$$

$$
\left(\pi^\top P^\pi \right)(j) = \pi(j) \sum_{\varepsilon \in \mathcal{E}} \pi(\varepsilon) + \sum_{i \in I} \pi(i) P(i, j) = \pi(j)(1 - \gamma) + \gamma \pi(j) = \pi(j).
$$

Then $\pi^\top = \pi^\top P^\pi$ holds.

Now we check that a probability distribution $\widetilde{\pi}$ that satisfies $\widetilde{\pi}^\top = \widetilde{\pi}^\top P^{\widetilde{\pi}}$ is necessarily defined by a QSD μ as in (3). For $j \in I$ we have

$$
\widetilde{\pi}(j) = \left(\widetilde{\pi}^\top P^{\widetilde{\pi}} \right)(j) = \widetilde{\pi}(j) \sum_{\delta \in \mathcal{E}} \widetilde{\pi}(\delta) + \sum_{i \in I} \widetilde{\pi}(i) P(i, j),
$$

and so $\widetilde{\pi}(j) \left(1 - \sum_{\delta \in \mathcal{E}} \widetilde{\pi}(\delta) \right) = \sum_{i \in I} \widetilde{\pi}(i) P(i, j)$. Then, the restriction $\widetilde{\pi}_I = (\widetilde{\pi}(i) : i \in I)$ satisfies $\gamma \widetilde{\pi}_I^\top = \widetilde{\pi}_I^\top P_I$ for some γ. So, $\widetilde{\pi}_I$ is a strictly positive left eigenvector with finite mass, so

$\mu = \gamma^{-1}\widetilde{\pi}_I$. Then, $\widetilde{\pi}_I = \gamma\mu = (\pi(i): i \in I)$ is given by the second term in (3). On the other hand, we have

$$\widetilde{\pi}(\varepsilon) = \big(\widetilde{\pi}^\top P^{\widetilde{\pi}}\big)(\varepsilon) = \widetilde{\pi}(\varepsilon)\sum_{\delta\in\mathcal{E}}\widetilde{\pi}(\delta) + \sum_{i\in I}\widetilde{\pi}(i)P(i,\varepsilon).$$

Then, $\widetilde{\pi}(\varepsilon)\big(1 - \sum_{\delta\in\mathcal{E}}\widetilde{\pi}(\delta)\big) = \sum_{i\in I}\widetilde{\pi}(i)P(i,\varepsilon)$, so $\widetilde{\pi}(\varepsilon)\gamma = \gamma\sum_{i\in I}\mu(i)P(i,\varepsilon)$. This gives the equality $\widetilde{\pi}(\varepsilon) = \sum_{i\in I}\mu(i)P(i,\varepsilon)$, so $\widetilde{\pi} = \pi$, which finishes the proof. $\qquad\square$

From the equality $\pi(I) = \gamma$ in (4), we shall use $\pi(I)$ in what follows to refer to the Perron–Frobenius eigenvalue of P_I.

Observe that (2) can be written as

$$\mathbb{P}_\mu(\mathcal{X}_{\tau_\mathcal{E}} = \varepsilon) = \pi(\varepsilon\mid\mathcal{E}) \quad \text{and} \quad \mathbb{P}_\mu(\mathcal{X}_{\tau_\mathcal{E}} = \varepsilon, \tau_\mathcal{E} = n) = \pi(\varepsilon\mid\mathcal{E})(1 - \pi(I))\pi(I)^{n-1}.$$

We denote by $\mathbb{X} = (X_n)$ the Markov chain evolving with the transition kernel P^π and call it the associated stationary chain. By an abuse of notation we shall denote P^π by P, and so, from now on, $P(\varepsilon, b) = \pi(b)$ for all $\varepsilon \in \mathcal{E}$, $b \in I \cup \mathcal{E}$. All the concepts developed in the absorbed case depended only on the trajectory $\mathcal{X}^{(A)} = (\mathcal{X}_n : n \le \tau_\mathcal{E})$, which is equally distributed as $(X_n : n \le \tau_\mathcal{E})$ when starting from $X_0 = \mathcal{X}_0 \in I$. Hence, there is no confusion if we continue denoting by $\mathbb{P}_a$ the law of the chain $\mathbb{X}$ starting from $a \in I \cup \mathcal{E}$ and by $\mathbb{E}_a$ its associated mean expected value.

Since $\mathbb{X} = (X_n)$ has transition probability kernel P and stationary distribution π, its entropy is given by (see Proposition 12.3 in [5], p. 69)

$$h(\mathbb{X}) = -\sum_{a\in I\cup\mathcal{E}}\pi(a)\sum_{b\in I\cup\mathcal{E}}P(a,b)\log P(a,b).$$

Then,

$$\begin{aligned}
h(\mathbb{X}) = &-\sum_{\delta\in\mathcal{E}}\pi(\delta)\sum_{a\in I\cup\mathcal{E}}\pi(a)\log\pi(a) - \sum_{i\in I}\pi(i)\sum_{b\in I\cup\mathcal{E}}P(i,b)\log P(i,b)\\
= &-\pi(\mathcal{E})\sum_{i\in I}\pi(i)\log\pi(i) - \pi(\mathcal{E})\sum_{\delta\in\mathcal{E}}\pi(\delta)\log\pi(\delta)\\
&-\sum_{i,j\in I}\pi(i)P(i,j)\log P(i,j) - \sum_{i\in I,\delta\in\mathcal{E}}\pi(i)P(i,\delta)\log P(i,\delta).
\end{aligned} \tag{5}$$

Further, we will compare this entropy to the entropies of some random sequences appearing in the chain.

4. Elements of the associated stationary chain

The object of this section is to show how one can retrieve the chain $\mathbb{X}$ from the absorbed trajectories and some walks on the set of absorbing states. To this end, the behavior of the chain $\mathbb{X}$ is first decomposed along its visits to I and to $\mathcal{E}$ in a separated way.

4.1. Decoupling the stationary chain

Let $\tau_I = \inf\{n \ge 1 : X_n \in I\}$ be the first return time of $\mathbb{X}$ to I. Now, consider the stochastic matrix $Q = (Q(i,j) : i,j \in I)$ given by $Q(i,j) = \mathbb{P}_i(X_{\tau_I} = j)$. By using that $\mathbb{P}_\varepsilon(X_{\tau_I} = j) = \pi(j\mid I) = \mu(j)$ for all $\varepsilon \in \mathcal{E}$, $j \in I$, we get

$$\begin{aligned}
Q(i,j) &= \mathbb{E}_i\Big(\mathbf{1}_{\{X_{\tau_I}=j,\tau_I=1\}}\Big) + \mathbb{E}_i\Big(\mathbf{1}_{\{X_{\tau_I}=j,\tau_I>1\}}\Big)\\
&= P(i,j) + \mathbb{P}_i(\tau_I > 1)\mathbb{P}_i\big(X_{\tau_I} = j\mid X_{\tau_I-1}, \tau_I > 1\big)\\
&= P(i,j) + P(i,\mathcal{E})\mu(j).
\end{aligned} \tag{6}$$

Let $\mathbb{Y} = (Y_n : n \in \mathbb{Z})$ be a Markov chain with transition matrix Q. It is straightforwardly checked that μ is a stationary measure for $\mathbb{Y}$.

Remark 1. For a substochastic matrix P_I the matrix $Q = (Q(i,j) = P(i,j) + P(i,\mathcal{E})\mu(j) : i,j \in I)$ was defined in [7] and called the resurrected matrix from P_I with distribution μ. It was a key concept used in [7] to prove the existence of QSD for geometrically absorbed Markov chains taking values in an infinite countable set.

The chain $\mathbb{Y}$ can be constructed as follows. Let $\Xi = \{\xi_l : l \in \mathbb{Z}\}$ be the ordered sequence given by

$$\{\xi_l : l \in \mathbb{Z}\} = \{n \in \mathbb{Z} : X_n \in I\} \text{ with } \xi_{l-1} < \xi_l, \ \xi_{-1} < 0 \leq \xi_0.$$

Then, $(\xi_l - \xi_{l-1} : l \in \mathbb{Z})$ is a renewal stationary sequence with interarrival times distributed as $\mathbb{P}(\xi_l - \xi_{l-1} = \bullet) = \mathbb{P}_\mu(\tau_I = \bullet)$, $l \neq 0$. By definition, $(X_{\xi_l} : l \in \mathbb{Z})$ is a stationary sequence distributed as $\mathbb{Y} = (Y_n : n \in \mathbb{Z})$, so $(X_{\xi_l} : l \in \mathbb{Z})$ is a copy of $\mathbb{Y}$.

The random sequence $\mathbf{b} = (b_l : l \in \mathbb{Z}, \ l \neq 0)$ defined by $b_l = 1$ if $\xi_l - \xi_{l-1} = 1$ and $b_l = 0$ if $\xi_l - \xi_{l-1} > 1$ is a collection of independent and identically distributed (i.i.d.) Bernoulli random variables, with

$$\mathbb{P}(b_l = 1) = \pi(I) \text{ and } \mathbb{P}(b_l = 0) = \sum_{i \in I} \mu(i)P(i,\mathcal{E}) = \pi(\mathcal{E}).$$

(Recall that $\pi(I) + \pi(\mathcal{E}) = 1$.) When $X_0 \in I$, we find $\tau_{\mathcal{E}} = \inf\{l \geq 1 : b_l = 0\}$.

Remark 2. Every irreducible matrix stochastic matrix Q with stationary distribution μ can be written as in (6). In fact, let $\chi = (\chi(i) : i \in I)$ be a non-null vector, $\chi \neq 0$, that satisfies

$$\text{for all } i \in I: \quad 0 \leq \chi(i) < 1 \text{ and } \chi(i) \leq \min\{Q(i,j)\mu(j)^{-1} : j \in I\}.$$

This can be achieved because μ is strictly positive. Define $P_I = Q - \chi\mu^\top$, so

$$\text{for all } i,j \in I: \quad P(i,j) = Q(i,j) - \chi(i)\mu(j). \tag{7}$$

To avoid the trivial situation we can assume that the vector χ also satisfies that for every $i \in I$ and for some (or for all) $j \in I$ for which $Q(i,j) > 0$, we have $P(i,j) > 0$. This allows us to take χ ensuring that P_I is irreducible. From the construction, $P(i,j) \in [0,1)$ and, since $\chi \neq 0$, we get

$$\sum_{j \in I} P(i,j) = 1 - \chi(i) \in (0,1], \qquad \sum_{i,j \in I} (1 - P(i,j)) > 0.$$

Hence, P is strictly substochastic, it is not trivial, and when adding the cemetery ∂ we have $\chi(i) = P(i,\partial)$. So,

$$\mu^\top P = \mu^\top (Q - \chi\mu^\top) = \left(1 - \sum_{i \in I} \mu(i)P(i,\partial)\right)\mu^\top;$$

that is, $\mu^\top$ is the Perron–Frobenius left eigenvector of P with eigenvalue $\pi(I) = \sum_{i,j \in I} \mu(i)P(i,j)$ (see (1) and (4)). From (7) it follows that $Q(i,j) = P(i,j) + P(i,\partial)\mu(j)$, so (6) is satisfied.

The restriction of $P = P^\pi$ to the absorption states $\mathcal{E}$ satisfies

$$\text{for all } \varepsilon, \delta \in \mathcal{E}: \quad P(\varepsilon,\delta) = \pi(\delta), \text{ so } \mathbb{P}(X_1 = \delta \mid X_0 = \varepsilon, X_1 \in \mathcal{E}) = \pi(\delta \mid \mathcal{E}).$$

The transition law to an absorbing point after being in $X_{t-1} = i \in I$ is given by

$$\text{for all } \delta \in \mathcal{E}: \quad \mathbb{P}(X_t = \delta \mid X_t \in \mathcal{E}, X_{t-1} = i) = P(i, \delta)/P(i, \mathcal{E}).$$

So, if $X_{-1} \in I$ and $X_0 \in \mathcal{E}$, the total sojourn time at $\mathcal{E}$ is τ_I, and it is distributed as a Geometric($\pi(I)$). Then, immediately after the entrance to $\mathcal{E}$ the chain $\mathbb{X}$ makes a walk on $\mathcal{E}$ of length $\tau_I - 1$ (a quantity that could vanish). To describe it, take $\mathbb{G} = (G_n : n \in \mathbb{Z})$ a Bernoulli chain with probability vector $\pi(\bullet \mid \mathcal{E})$. Let us consider the finite random sequence $V = (G_l : 1 \le l < \tau_I)$ (with V empty if $\tau_I = 1$), which is distributed as

$$\mathbb{P}(V = \emptyset) = \mathbb{P}(\tau_I = 1) = \pi(I), \tag{8}$$
$$\mathbb{P}(V = (\delta_1, \ldots, \delta_{k-1})) = \mathbb{P}(G_1 = \delta_1, \ldots, G_{k-1} = \delta_{k-1}, \tau_I = k)$$
$$= \left(\prod_{l=1}^{k-1} \pi(\delta_l \mid \mathcal{E})\right) \pi(\mathcal{E})^{k-1} \pi(I) = \left(\prod_{l=1}^{k-1} \pi(\delta_l)\right) \pi(I)$$
$$\text{for } k \ge 2, \ (\delta_1, \ldots, \delta_{k-1}) \in \mathcal{E}^{k-1}.$$

Notice that the last equality also holds when $\tau_I = k = 1$ because an empty product satisfies $\prod_{l=1}^{k-1} = 1$. We have $(X_t, 1 \le t < \tau_I \mid X_{-1} \in I, X_0 \in \mathcal{E}) \sim V$, and V is called a walk on $\mathcal{E}$. Note that $\tau_I - 1 \mid \tau_I > 1$ is equally distributed as τ_I. The exit law from $\mathcal{E}$ is $\mathbb{P}(X_{\tau_I} \in \bullet) \sim \mu$. In fact, for all $\delta \in \mathcal{E}$,

$$\mathbb{P}_\delta(X_{\tau_I} = i) = \sum_{\varepsilon \in \mathcal{E}} \mathbb{P}_\delta(X_{\tau_I} = i, X_{\tau_I - 1} = \varepsilon)$$
$$= \sum_{\varepsilon \in \mathcal{E}} \mathbb{P}_\delta(X_{\tau_I - 1} = \varepsilon) \mathbb{P}_\varepsilon(X_1 = i \mid X_1 \in I)$$
$$= \pi(i \mid I) = \mu(i).$$

Notice that

$$\text{for all } \delta \in \mathcal{E}, \ \mathbb{E}_\delta(\tau_I) = \pi(I)^{-1}. \tag{9}$$

We consider a sequence of i.i.d. random variables $\mathcal{T} = (\mathcal{T}_n : n \in \mathbb{Z})$ which are Geometric($\pi(I)$) $- 1$ distributed, that is, $\mathbb{P}(\mathcal{T}_n = l) = \pi(I)(1 - \pi(I))^l$ for $l \ge 0$. The construction of i.i.d. walks on $\mathcal{E}$ is made as follows. One takes an increasing sequence of times $(t_n : n \in \mathbb{Z})$ with $t_{n+1} - t_n = \mathcal{T}_n$ and such that $t_n \to \infty$ if $n \to \infty$ and $t_n \to -\infty$ if $n \to -\infty$. We define $V^n = (G_{t_n}, \ldots, G_{t_{n+1}-1}) = (V_1^n, \ldots, V_{\mathcal{T}_n}^n)$. So, $\mathbb{V} = (V^n : n \ge \mathbb{Z})$ is an i.i.d. sequence of walks on $\mathcal{E}$. The walk V^n is empty when $\mathcal{T}_n = 0$.

When $\mathcal{E} = \{\partial\}$ is a singleton, we have $\pi(\partial \mid \mathcal{E}) = 1$, $\mathbb{G} = (G_n : n \in \mathbb{Z})$ is the orbit with the unique symbol $G_n = \partial$ for all n, and the random sequence $V = (G_l : 1 \le l < \tau_I)$ has the symbol ∂ repeated $|\tau_I| - 1$ times.

4.2. Retrieving the stationary chain

Let $\mathbb{Y} = (Y_n : n \in \mathbb{Z})$ be a stationary Markov chain with transition matrix Q. Our purpose is to construct a copy of $\mathbb{X}$ from $\mathbb{Y}$ by adding a series of random operations.

Let $\mathbb{P}$ be a probability measure governing the law of $\mathbb{Y}$ when it starts from the stationary distribution μ, the sequences $\mathbb{G}, \mathcal{T}$ and so $\mathbb{V}$, and also the random element $\mathbb{B}^{I,I}$ and $\mathbb{D}^I$ defined below.

Let $\mathbb{B}^{I,I} = \big((B_l^{i,j} : l \in \mathbb{Z}); i, j \in I\big)$ be an independent array of Bernoulli random variables such that $B_l^{i,j} \sim B^{i,j}$ for $l \in \mathbb{Z}$, where

$$\mathbb{P}(B^{i,j} = 1) = \theta_{i,j}, \qquad \mathbb{P}(B^{i,j} = 0) = \bar{\theta}_{i,j} = 1 - \theta_{i,j},$$

$$\theta_{i,j} = \frac{P(i,j)}{P(i,j) + P(i,\mathcal{E})\mu(j)} = \frac{P(i,j)}{Q(i,j)}. \tag{10}$$

Let $\tau_\partial = \inf\{l \geq 1 : B_l^{Y_{l-1}, Y_l} = 0\}$. For $k \geq 1$, $i_0, \ldots, i_{k-1} \in I$ we have

$$\mathbb{P}(Y_0 = i_0, Y_1 = i_1, \ldots, Y_{k-1} = i_{k-1}, \tau_\partial = k)$$
$$= \mu(i_0)\left(\prod_{l=1}^{k-1} P(i_{l-1}, i_l)\right)\left(\sum_{j \in I} P(i_{k-1}, \mathcal{E})\mu_j\right) = \mu(i_0)\left(\prod_{l=1}^{k-1} P(i_{l-1}, i_l)\right)P(i_{k-1}, \mathcal{E}).$$

Hence, the sequence $Y^{(K)} = (Y_l : 0 \leq l < \tau_\partial)$ is distributed as a killed chain $\mathcal{X}^{(K)}$ starting from μ.

Now take an independent array $\mathbb{D}^I = \big((D_l^i : l \in \mathbb{Z}); i \in I\big)$ of random variables taking values in $\mathcal{E}$ and with law

$$\text{for all } \delta \in \mathcal{E}: \quad \mathbb{P}(D_l^i = \delta) = P(i, \delta)/P(i, \mathcal{E}). \tag{11}$$

For $k \geq 1$, $i_0, \ldots, i_{k-1} \in I$, $\delta \in \mathcal{E}$ we set

$$\mathbb{P}(Y_0 = i_0, Y_1 = i_1, \ldots, Y_{k-1} = i_{k-1}, D_k^{i_{k-1}} = \delta, \tau_\partial = k)$$
$$= \mu(i_0)\left(\prod_{l=1}^{k-1} P(i_{l-1}, i_l)\right)P(i_{k-1}, \mathcal{E})\big(P(i_{k-1}, \delta)/P(i_{k-1}, \mathcal{E})\big)$$
$$= \mu(i_0)\left(\prod_{l=1}^{k-1} P(i_{l-1}, i_l)\right)P(i_{k-1}, \delta).$$

Then, the sequence $Y^{(A)} = (Y_0, \ldots, Y_{\tau_\partial - 1}, D_{\tau_\partial}^{Y_{\tau_\partial - 1}})$ is distributed as an absorbed chain $\mathcal{X}^{(A)}$ starting from μ.

Let us construct a chain $\mathbb{S}^s = (\mathbb{S}_t^s : t \in \mathbb{Z})$ from $\mathbb{Y}$, $\mathbb{B}^{I,I}$, $\mathbb{D}^Y$, $\mathbb{G}$, and $\mathcal{T}$ (and so also $\mathbb{V}$), having the same distribution as $\mathbb{X}$. First, define a random sequence $\mathbb{S} = (\mathbb{S}_t : t \in \mathbb{Z})$ as follows. We set $T_0 = 0$, $\mathbb{S}_0 = Y_0$ (so $\mathbb{S}_0 = Y_0 \in I$ is distributed as μ), and:

I In a sequential way on $n \geq 0$ we make the following construction. Assume at step n that T_n has been defined; then, put $\mathbb{S}_{T_n} = Y_n$ and go to step $n + 1$.

Ia If $B_{n+1}^{Y_n, Y_{n+1}} = 1$ put $T_{n+1} = T_n + 1$, $\mathbb{S}_{T_{n+1}} = Y_{n+1}$, and go to step $n + 2$.

Ib If $B_{n+1}^{Y_n, Y_{n+1}} = 0$ put $T_{n+1} = T_n + \mathcal{T}_n + 2$, define $\mathbb{S}_{T_n + 1} = D_{n+1}^{Y_n}$, $\mathbb{S}_{T_n + 1 + l} = V_l^n$ for $1 \leq l < \mathcal{T}_n$ (it is empty when $\mathcal{T}_n = 0$), and $\mathbb{S}_{T_{n+1}} = Y_{n+1}$. Then continue with step $n + 2$.

II Similarly, in a sequential way on $n < 0$ we make the following construction for step n:

IIa If $B_{n+1}^{Y_n, Y_{n+1}} = 1$ put $T_n = T_{n+1} - 1$, $\mathbb{S}_{T_n} = Y_n$, and continue with step $n - 1$.

IIb If $B_{n+1}^{Y_n, Y_{n+1}} = 0$ put $T_n = T_{n+1} - (\mathcal{T}_n + 2)$, $\mathbb{S}_{T_n + 1} = D_{n+1}^{Y_n}$, $\mathbb{S}_{T_n + 1 + l} = V_l^n$ for $1 \leq l < \mathcal{T}_n$, and $\mathbb{S}_{T_n} = Y_n$. Then, continue with step $n - 1$.

Let $\mathbb{S} = (\mathbb{S}_t : t \in \mathbb{Z})$ be the random sequence resulting from this construction, and let $\mathbb{T} = (T_n : n \in \mathbb{Z})$, recalling that $T_0 = 0$. By an abuse of notation we also denote by $\mathbb{T} = \{T_n : n \in \mathbb{Z}\}$ the set of these values. By definition, $\mathbb{T} = \{t \in \mathbb{Z} : \mathbb{S}_t \in I\}$ is the set of random points where $\mathbb{S}$ is in I. In Proposition 2 we will prove that $(\mathbb{S}, \mathbb{T})$ is a regenerative process (see [2], pp. 169–170), that is, for all $l \geq 0$ the process $(\mathbb{S}_{\bullet + T_l} : \bullet \geq 0; T_{n+1} - T_n, \ n \geq l)$ has the same distribution as $(\mathbb{S}_\bullet : \bullet \geq 0; T_{n+1} - T_n, \ n \geq 0)$ and it is independent of $(T_n : n \leq l)$.

The cycles of this regenerative process, $(\mathbb{S}_{T_n}, \ldots, \mathbb{S}_{T_{n+1}-1})$, $n \in \mathbb{Z}$, are i.i.d., and so all of them have the same distribution as $(\mathbb{S}_0, \ldots, \mathbb{S}_{T_1-1})$. By shifting the process $\mathbb{S}^s = (\mathbb{S}_t^s : t \in \mathbb{Z})$ by a random time chosen uniformly in $\{0, \ldots, T_1 - 1\}$ and conditionally independent of the rest of the process, we get a stationary process $\mathbb{S}^s = (\mathbb{S}_t^s : t \in \mathbb{Z})$ (see Theorem 6.4 in [2]). So, $\mathbb{S}_0^s$ takes values in $I \cup \mathcal{E}$ and, from the next proposition, it is distributed as π (different than $\mathbb{S}_0$, which takes values in I and is distributed as μ).

Proposition 2. *The process* $(\mathbb{S}, \mathbb{T})$ *is regenerative and the associated stationary process* $\mathbb{S}^s$ *is equally distributed as* $\mathbb{X}$.

The proof can be found in the Appendix.

5. Stationary representation of killed and absorbed chains

The stationary Markov chain $\mathbb{Y} = (Y_n : n \in \mathbb{Z})$ with transition matrix Q and stationary distribution μ has entropy

$$h(\mathbb{Y}) = -\sum_{i \in I} \mu(i) \sum_{j \in I} Q(i, j) \log Q(i, j). \tag{12}$$

To get stationary representations of the killed and the absorbed chains we will use the 2-stringing form of $\mathbb{Y}$. Let us recall this notion. Consider the stochastic matrix $Q^{[2]}$, with set of indexes I^2, given by $Q^{[2]}((i, j), (l, k)) = Q(l, k)\mathbf{1}(l = j)$. Its stationary distribution satisfies $\nu((i, j)) = \mu(i)Q(i, j)$ for $(i, j) \in I^2$. In fact, by using $\sum_{i \in I} \mu(i)Q(i, j) = \mu(j)$ we get

$$\sum_{(i,j) \in I^2} \nu((i, j))Q^{[2]}((i, j), (l, k)) = \sum_{i \in I} \mu(i)Q(i, l)Q(l, k) = \mu(l)Q(l, k) = \nu((l, k)).$$

The stationary chain $\mathbb{Y}^{[2]} = ((Y_n^1, Y_n^2) : Y_{n-1}^2 = Y_n^1, \ n \in \mathbb{Z})$ evolving with $Q^{[2]}$ is the 2-stringing of $\mathbb{Y}$. We write it by $\mathbb{Y}^{[2]} = ((Y_{n-1}, Y_n) : n \in \mathbb{Z})$. It is well known that it is conjugated to $\mathbb{Y}$ by the (1-coordinate) mapping $\Upsilon(((Y_{n-1}, Y_n) : n \in \mathbb{Z})) = (Y_n : n \in \mathbb{Z})$. (This property was stated in a general form in Lemma 1 in [9].) Being conjugated by a mapping means that the mapping is one-to-one, measure preserving, and commutes with the shift on $\mathbb{Z}$. Since Υ is clearly one-to-one and shift commuting, we only check that it is measure preserving. Taking $(i_l : l = 0, \ldots, k) \in I^{k+1}$, we have

$$\mathbb{P}((Y_{n-1}, Y_n) \in \mathbb{Y}^{[2]} : (Y_{l-1}, Y_l) = (i_{l-1}, i_l) : l = 1, \ldots, k)$$
$$= \mu(i_0)Q(i_0, i_1) \prod_{l=1}^{k-1} Q(i_l, i_{l+1}) = \mathbb{P}(Y_l = i_l : l = 0, \ldots, k).$$

The orbits $((Y_{n-1}, Y_n) : n \in \mathbb{Z}) \in \mathbb{Y}^{[2]}$ can be identified with the orbits $(Y_n : n \in \mathbb{Z}) \in \mathbb{Y}$.

5.1. The killed chain

The stationary representation of the killed chain will be a stationary Markov chain on the set of states $I^2 \times \{0, 1\} = \{(i, j, a) : (i, j) \in I^2, \ a \in \{0, 1\}\}$. Prior to defining the transition matrix we introduce the function

$$\varphi(i, j, a) = (\theta_{i,j}\mathbf{1}(a = 1) + \overline{\theta}_{i,j}\mathbf{1}(a = 0)), \qquad (i, j, a) \in I^2 \times \{0, 1\}.$$

The transition matrix $\mathcal{K} = \big(\mathcal{K}((i, j, a), (l, k, b)) : (i, j, a), (l, k, b) \in I^2 \times \{0, 1\}\big)$ is defined by

$$\mathcal{K}((i, j, a), (l, k, b)) = \begin{cases} 0 \text{ if } l \neq j, \\ Q(l, k)\varphi(l, k, b) = P(l, k)\mathbf{1}(b = 1) + P(l, \mathcal{E})\mu(k)\mathbf{1}(b = 0) \text{ if } l = j. \end{cases}$$

It can be straightforwardly checked that this is a stochastic matrix: we claim its stationary distribution $\zeta = (\zeta(i, j, a) : (i, j, a) \in I^2 \times \{0, 1\})$ is given by

$$\begin{aligned} \zeta(i, j, a) &= \mu(i)Q(i, j)\varphi(i, j, a) \\ &= \mu(i)P(i, j)\mathbf{1}(a = 1) + \mu(i)P(i, \mathcal{E})\mu(j)\mathbf{1}(a = 0). \end{aligned} \tag{13}$$

By using $\sum_{i \in I} \mu(i)Q(i, l) = \mu(l)$ we get the desired property,

$$\begin{aligned} \sum_{(i,j,a) \in I^2 \times \{0,1\}} \zeta(i, j, a)\mathcal{K}((i, j, a), (l, k, b)) &= \left(\sum_{i \in I} \mu(i)Q(i, l)\right) Q(l, k)\varphi(l, k, b) \\ &= \mu(l)Q(l, k)\varphi(l, k, b) = \zeta(l, k, b), \end{aligned}$$

so the claim follows.

The killed Markov chain presented in its stationary form is denoted

$$\mathbb{Y}^{(\mathcal{K})} = ((Y_{n-1}, Y_n, B_n) : n \in \mathbb{Z});$$

it takes values in $I^2 \times \{0, 1\}$ and has transition matrix $\mathcal{K}$. The component B_n is called the label at n. By hypothesis, P_I is irreducible so also $\mathcal{K}$ is an irreducible matrix. Then, the Markov shift $\mathbb{Y}^{(\mathcal{K})}$ is ergodic (see Proposition 8.12 in [5]).

It is straightforward to check that the mapping

$$\Upsilon^{(\mathcal{K})} : \mathbb{Y}^{(\mathcal{K})} \to \mathbb{Y}, \ \ ((Y_{n-1}, Y_n, B_n) : n \in \mathbb{Z}) \to (Y_n : n \in \mathbb{Z}) \tag{14}$$

is a factor, which means that it is measure preserving and commutes with the shift on $\mathbb{Z}$.

Remark 3. We show that $\mathbb{Y}^{(\mathcal{K})}$ models the killed Markov chain. Let $\mathcal{N} = \{n \in \mathbb{Z} : B_n = 0\}$ and write it as $\mathcal{N} = \{n_l : l \in \mathbb{Z}\}$, where n_l is increasing with l, and $n_{-1} < 0 \leq n_0$. Note that $\mathbb{P}(0 \in \mathcal{N}) = \pi(\mathcal{E})$.

The orbit $((Y_{n-1}, Y_n, B_n) : n \in \mathbb{Z})$ in $\mathbb{Y}^{(\mathcal{K})}$ is denoted in the simpler form $((Y_{n-1}, B_n) : n \in \mathbb{Z})$ and we can divide it into the disjoint connected pieces $(Y, B)_l^{(K)} = ((Y_{n_l}, 1), \ldots, (Y_{n_{l+1}-2}, 1), (Y_{n_{l+1}-1}, 0)), l \in \mathbb{Z}$. The component Y_{n_l} is distributed with law μ for all l, and one can identify $(Y, B)_l^{(K)}$ with $Y_l^{(K)} = (Y_{n_l}, \ldots, Y_{n_{l+1}-1})$, a piece of the orbit $Y = (Y_n : n \in \mathbb{Z})$

starting from μ at n_l and killed at $n_{l+1} - 1$. We get that $Y_l^{(K)} \sim \mathcal{X}^{(K)}$ for all $l \neq -1$ when $\mathcal{X}^{(K)}$ starts from μ. In fact, for $s \geq 0$, $i_0, \ldots, i_s \in I$, we have

$$
\mathbb{P}\big(\mathcal{X}^{(K)} = (i_0, \ldots, i_s)\big) = \mu(i_0) \prod_{r=0}^{s-1} P(i_r, i_{r+1}) P(i_s, \partial)
$$

$$
= \mu(i_0) \left(\prod_{r=0}^{s-1} Q(i_r, i_{r+1}) \theta_{i_r, i_{r+1}} \right) \left(\sum_{m \in I} Q(i_s, m) \overline{\theta}_{i_s, m} \right) = \mathbb{P}\big(Y_l^{(K)} = (i_0, \ldots, i_s)\big).
$$

Let $s \geq 0$, $(i_0, \ldots, i_s) \in I^{s+1}$. For almost all the orbits $Y \in \mathbb{Y}^{(K)}$ we have

$$
\mathbb{P}(n_0 = 0, Y_0^{(K)} = (i_0, \ldots, i_s)) = \pi(\mathcal{E}) \mu(i_0) \prod_{r=0}^{s-1} P(i_r, i_{r+1}) P(i_s, \partial) > 0.
$$

Since the killed trajectories are finite, the class of killed trajectories is countable. From the ergodic theorem, and since $\mathbb{Y}^{(K)}$ is ergodic, it follows that $\mathbb{P}$-a.e. (almost everywhere) the orbits of $\mathbb{Y}^{(K)}$ contain all the killed trajectories of the chain.

The entropy of the killed chain satisfies

$$
h\big(\mathbb{Y}^{(K)}\big) = -\sum_{(i,j) \in I^2} \mu(i) Q(i, j) \sum_{k \in I} Q(j, k) \big(\theta_{j,k} \log (Q(j, k) \theta_{j,k}) + \overline{\theta}_{j,k} \log (Q(j, k) \overline{\theta}_{j,k})\big)
$$

$$
= -\sum_{(j,k) \in I^2} \mu(j) Q(j, k) \log Q(j, k) + \sum_{(j,k) \in I^2} \mu(j) Q(j, k) H(B^{j,k}),
$$

where $H(B^{j,k}) = -\big(\theta_{j,k} \log \theta_{j,k} + \overline{\theta}_{j,k} \log \overline{\theta}_{j,k}\big)$ is the entropy of the Bernoulli random variable $B^{j,k}$. Hence,

$$
h\big(\mathbb{Y}^{(K)}\big) = h(\mathbb{Y}) + \Delta(B), \quad \text{with } \Delta(B) = \sum_{(j,k) \in I^2} \mu(j) Q(j, k) H(B^{j,k}). \tag{15}
$$

The quantity $\Delta(B) = h(\mathbb{Y}^{(K)}) - h(\mathbb{Y})$ is the conditional entropy of $\mathbb{Y}^{(K)}$ given the factor $\mathbb{Y}$ (see Lemma 2 and Definition 3 in [6]). To be more precise, given an orbit $Y = (Y_n : n \in \mathbb{Z})$ of $\mathbb{Y}$, the fiber given by (14) satisfies $(\Upsilon^{(K)})^{-1}\{Y\} = \{(B_n^{Y_{n-1}, Y_n} : n \in \mathbb{Z}) \in \{0, 1\}^{\mathbb{Z}}\}$, and it is distributed as a sequence of independent Bernoulli variables given by (10); we denote it by $(\mathbf{P})_{\mathbf{Y}}$. We have

$$
H_{(\mathbf{P})_{\mathbf{Y}}}\big(B_1^{Y_0, Y_1}\big) = -\big(\theta_{Y_0, Y_1} \log \theta_{Y_0, Y_1} + \overline{\theta}_{Y_0, Y_1} \log \overline{\theta}_{Y_0, Y_1}\big). \tag{16}
$$

Let us summarize the results on the entropy of $\mathbb{Y}^{(K)}$.

Proposition 3. *The entropy of the stationary representation $\mathbb{Y}^{(K)}$ of the killed chain satisfies*

$$
h\big(\mathbb{Y}^{(K)}\big) = h(\mathbb{Y}) + \Delta(B), \tag{17}
$$

$$
\Delta(B) = \int H_{(\mathbf{P})_{\mathbf{Y}}}\big(B_1^{Y_0, Y_1}\big) \, \mathrm{d}\mathbb{P}(Y) = \sum_{(i,j) \in I^2} \mu(i) Q(i, j) H(B^{i,j}),
$$

and

$$h\big(\mathbb{Y}^{(\mathcal{K})}\big) = - \sum_{i \in I} \mu(i) P(i, \mathcal{E}) \log P(i, \mathcal{E}) - (1 - \pi(I)) \sum_{j \in I} \mu(j) \log \mu(j) \qquad (18)$$
$$- \sum_{i,j \in I} \mu(i) P(i, j) \log P(i, j).$$

Proof. From (15) and (16), and by using the Markov property, we retrieve the Abramov–Rokhlin formula (see [1,6]),

$$\Delta(B) = h\big(\mathbb{Y}^{(\mathcal{K})}\big) - h(\mathbb{Y}) = \int H_{(\mathbf{P})_{\bar{Y}}}\big(B_1^{Y_0, Y_1}\big) \, d\mathbb{P}(Y) = \sum_{i,j \in I} \mu(i) Q(i, j) H(B^{i,j}).$$

This gives (17). The only thing left to prove is (18). By using

$$\sum_{i \in \mathcal{E}} \mu(i) P(i, \mathcal{E}) = 1 - \pi(I), \quad \sum_{j \in I} P(i, j) = 1 - P(i, \mathcal{E}), \quad \sum_{i \in I} \mu(i) P(i, j) = \pi(I) \mu(j)$$

and (12), we get

$$\Delta(B) = - \sum_{i,j \in I} \mu(i) \left(P(i, \mathcal{E}) \mu(j) \log \left(P(i, \mathcal{E}) \mu(j)\right) + P(i, j) \log P(i, j)\right)$$
$$+ \sum_{i,j \in I} \mu(i) Q(i, j) \log Q(i, j)$$
$$= - \sum_{i \in I} \mu(i) P(i, \mathcal{E}) \log P(i, \mathcal{E}) - (1 - \pi(I)) \sum_{j \in I} \mu(j) \log \mu(j)$$
$$+ \sum_{i,j \in I} \mu(i) P(i, j) \log P(i, j) - h(\mathbb{Y}).$$

This shows (18).   □

Remark 4. From (13) we get that there are, in mean,

$$\sum_{i,j \in I} \zeta(i, j, 1) = \sum_{i,j \in I} \mu(i) P(i, j) = \pi(I)$$

sites in $\mathbb{Z}$ where $\mathbb{Y}^{(\mathcal{K})}$ has made a transition with label 1, and a mean

$$\sum_{i,j \in I} \zeta(i, j, 0) = \sum_{i \in I} \mu(i) P(i, \mathcal{E}) = \pi(\mathcal{E})$$

of sites in $\mathbb{Z}$ where $\mathbb{Y}^{(\mathcal{K})}$ has made a transition with label 0, and so resurrects with distribution μ.

5.2. The absorbed chain

Let us construct a stationary representation of the absorbed chain in a similar way as we did for the killed chain. Define $\mathcal{E}^* = \mathcal{E} \cup \{o\}$ with $o \notin \mathcal{E} \cup I$. The absorbed chain

will take values on the set of states $I^2 \times \mathcal{E}^* = \{(i, j, \delta) : i \in I, \ j \in I, \ \delta \in \mathcal{E}^*\}$. The matrix $\mathcal{A} = (\mathcal{A}((i, j, \delta), (l, k, \varepsilon)) : (i, j, \delta), (l, k, \varepsilon) \in I^2 \times \mathcal{E})$ defined by

$$\mathcal{A}((i, j, \delta), (l, k, \varepsilon)) = \begin{cases} 0 \text{ if } l \neq j, \\ Q(l, k)\theta_{l,k} = P(l, k) \text{ if } l = j, \ \varepsilon = o, \\ Q(l, k)\overline{\theta}_{l,k}P(l, \varepsilon)/P(l, \mathcal{E}) = P(l, \varepsilon)\mu(k) \text{ if } l = j, \ \varepsilon \in \mathcal{E} \end{cases}$$

is a stochastic matrix whose stationary distribution $\eta = (\eta(i, j, \delta) : (i, j, r) \in I^2 \times \mathcal{E}^*)$ is given by

$$\eta(i, j, \delta) = \mu(i)Q(i, j)\left(\theta_{i,j}\mathbf{1}(\delta = o) + \overline{\theta}_{i,j}P(i, \delta)/P(i, \mathcal{E})\mathbf{1}(\delta \in \mathcal{E})\right)$$
$$= \mu(i)P(i, j)\mathbf{1}(\delta = o) + \mu(i)P(i, \delta)\mu(j)\mathbf{1}(\delta \in \mathcal{E}).$$

In fact, since $\sum_{\delta \in \mathcal{E}^*}\left(\theta_{i,l}\mathbf{1}(\delta = o) + \overline{\theta}_{i,l}P(i, \delta)/P(i, \mathcal{E})\mathbf{1}(\delta \in \mathcal{E})\right) = 1$, we get the stationarity property

$$\sum_{(i,j,\delta) \in I^2 \times \mathcal{E}^*} \eta(i, j, \delta)\mathcal{A}((i, j, \delta), (l, k, \varepsilon))$$

$$= \left(\sum_{i \in I}\mu(i)Q(i, l)\right)Q(l, k)\left(\theta_{l,k}\mathbf{1}(\varepsilon = o) + \overline{\theta}_{l,k}P(l, \varepsilon)/P(l, \mathcal{E})\mathbf{1}(\varepsilon \in \mathcal{E})\right)$$

$$= \mu(l)Q(l, k)\left(\theta_{l,k}\mathbf{1}(\varepsilon = o) + \overline{\theta}_{l,k}P(l, \varepsilon)/P(l, \mathcal{E})\mathbf{1}(\varepsilon \in \mathcal{E})\right) = \eta(l, k, \varepsilon).$$

We denote by $\mathbb{Y}^{(\mathcal{A})} = ((Y_{n-1}, Y_n, D_n^*) : n \in \mathbb{Z})$ the absorbed Markov chain presented in its stationary form, and taking values in $I^2 \times \mathcal{E}^*$ with transition matrix $\mathcal{A}$. Since $\mathcal{A}$ is irreducible, the Markov shift $\mathbb{Y}^{(\mathcal{A})}$ is ergodic.

It is straightforward to check that the mapping

$$\Upsilon^{(\mathcal{A})} : \mathbb{Y}^{(\mathcal{A})} \to \mathbb{Y}^{(\mathcal{K})}, ((Y_{n-1}, Y_n, D_n^*) : n \in \mathbb{Z}) \to (Y_{n-1}, Y_n, B_n) : n \in \mathbb{Z})$$
$$\text{with } B_n = \mathbf{1}(D_n^* = o) \tag{19}$$

is a factor between $\mathbb{Y}^{(\mathcal{A})}$ and $\mathbb{Y}^{(\mathcal{K})}$.

Remark 5. Let us see that the stationary chain $\mathbb{Y}^{(\mathcal{A})}$ models the absorbed Markov chain. First, denote $\mathcal{N}^* = \{n \in \mathbb{Z} : D_n^* \in \mathcal{E}\}$ and write it by $\mathcal{N}^* = \{n_l : l \in \mathbb{Z}\}$ with n_l increasing in l and $n_{-1} < 0 \leq n_0$. We have $\mathbb{P}(0 \in \mathcal{N}^*) = \pi(\mathcal{E})$. Similarly to Remark 3, an orbit $((Y_{n-1}, Y_n, D_n^*) : n \in \mathbb{Z})$ of $\mathbb{Y}^{(\mathcal{A})}$ is denoted in the form $((Y_{n-1}, D_n^*) : n \in \mathbb{Z})$ and is partitioned into the disjoint connected pieces

$$(Y, D^*)_l^{(\mathcal{A})} = ((Y_{n_l}, o), \dots, (Y_{n_{l+1}-2}, o), (Y_{n_{l+1}-1}, D_{n_{l+1}}^*)) \text{ with } l \in \mathbb{Z}.$$

The component Y_{n_l} is distributed with law μ for all l, and we can identify $(Y, D^*)_l^{(\mathcal{A})}$ with $Y_l^{(\mathcal{A})} = (Y_{n_l}, \dots, Y_{n_{l+1}-1}, D_{n_{l+1}}^*)$ starting from μ. Since the events $\{n \in \mathbb{Z} : D_n^* \in \mathcal{E}\}$ have the same distribution as $\{n \in \mathbb{Z} : B_n = 0\}$ in $\mathbb{Y}^{(\mathcal{K})}$, it can be checked that, for all $l \neq -1$, $Y_l^{(\mathcal{A})} \sim \mathcal{X}^{(\mathcal{A})}$, where $\mathcal{X}^{(\mathcal{A})}$ starts form μ. In fact, for $s \geq 0, i_0, \dots, i_s \in I, \varepsilon \in \mathcal{E}$, we have

$$\mathbb{P}\left(\mathcal{X}^{(\mathcal{A})} = (i_0, \dots, i_s, \varepsilon)\right) = \mu(i_0)\left(\prod_{r=0}^{s-1}P(i_r, i_{r+1})\right)P(i_s, \varepsilon) = \mathbb{P}\left(Y_l^{(\mathcal{A})} = (i_0, \dots, i_s, \varepsilon)\right).$$

Let $s \geq 0$, $(i_0, \ldots, i_s) \in I^{s+1}$, $\varepsilon \in \mathcal{E}$. For almost all the orbits $Y \in \mathbb{Y}^{(\mathcal{A})}$ we have

$$\mathbb{P}\big(n_0 = 0, Y_0^{(\mathcal{A})} = (i_0, \ldots, i_s, \varepsilon)\big) = \pi(\mathcal{E})\mu(i_0) \prod_{r=0}^{s-1} P(i_r, i_{r+1}) P(i_s, \varepsilon) > 0.$$

Since the absorbed trajectories are finite, the class of absorbed trajectories is countable. Then, since $\mathbb{Y}^{(\mathcal{A})}$ is ergodic, we get from the ergodic theorem that, $\mathbb{P}$-a.e., the orbits of $\mathbb{Y}^{(\mathcal{A})}$ contain all the absorbed trajectories of the chain.

The entropy of the absorbed chain satisfies

$$
\begin{aligned}
h\big(\mathbb{Y}^{(\mathcal{A})}\big) = &-\sum_{(i,j)\in I^2}\mu(i)Q(i,j)\sum_{k\in I}Q(j,k)\theta_{j,k}\log\left(Q(j,k)\theta_{j,k}\right) \\
&-\sum_{(i,j)\in I^2}\mu(i)Q(i,j)\sum_{k\in I,\varepsilon\in\mathcal{E}}Q(j,k)\overline{\theta}_{j,k}P(j,\varepsilon)/P(j,\mathcal{E})\log\left(Q(j,k)\overline{\theta}_{j,k}P(j,\varepsilon)/P(j,\mathcal{E})\right) \\
= &-\sum_{(j,k)\in I^2}\mu(j)Q(j,k)\log Q(j,k) - \sum_{j\in I}\mu(j)Q(j,k)(\theta_{j,k}\log\theta_{j,k}+\overline{\theta}_{j,k}\log\overline{\theta}_{j,k}) \\
&-\sum_{(j,k)\in I^2}\mu(j)Q(j,k)\overline{\theta}_{j,k}\sum_{\varepsilon\in\mathcal{E}}P(j,\varepsilon)/P(j,\mathcal{E})\log\left(P(j,\varepsilon)/P(j,\mathcal{E})\right).
\end{aligned}
$$

Then,

$$h\big(\mathbb{Y}^{(\mathcal{A})}\big) = h\big(\mathbb{Y}^{(\mathcal{K})}\big) + \sum_{i\in I}\mu(i)P(i,\mathcal{E})H(D^i), \quad \text{where}$$

$$H(D^i) = -\sum_{\delta\in\mathcal{E}}P(i,\delta)/P(i,\mathcal{E})\log\left(P(i,\delta)/P(i,\mathcal{E})\right) \tag{20}$$

is the entropy of a random variable in $\mathcal{E}$ distributed as the transition probability from $i \in I$ to a state conditioned to be in $\mathcal{E}$. Note that the above expression can also be written as

$$h\big(\mathbb{Y}^{(\mathcal{A})}\big) = h\big(\mathbb{Y}^{(\mathcal{K})}\big) + \sum_{i\in I}\mu(i)(P(i,\mathcal{E})H(D^i) + P(i,I)H(o)),$$

$H(o) = 0$ being the entropy of a constant.

Define

$$\Delta(D) = h\big(\mathbb{Y}^{(\mathcal{A})}\big) - h\big(\mathbb{Y}^{(\mathcal{K})}\big) = \sum_{i\in I}\mu(i)P(i,\mathcal{E})H(D^i).$$

This is the conditional entropy of $\mathbb{Y}^{(\mathcal{A})}$ given the factor $\mathbb{Y}^{(\mathcal{K})}$. To see this, take an orbit $Y^{(\mathcal{K})} = ((Y_{n-1}, Y_n, B_n) : n \in \mathbb{Z})$ of $\mathbb{Y}^{(\mathcal{K})}$. The fiber given by (19) satisfies $(\Upsilon^{(\mathcal{A})})^{-1}\{Y^{(\mathcal{K})}\} = \{(D_n^{Y_n, B_n} : n \in \mathbb{Z}) \in (\mathcal{E}^*)^{\mathbb{Z}}\}$ with $D_n^{Y_n, B_n} \in \mathcal{E}$ when $B_n = 0$ and $D_n^{Y_n, B_n} = o$ when $B_n = 1$. These variables are independently distributed as a Bernoulli $D_n^{Y_n}$ given in (11) if $B_n = 0$ and the constant variable o if $B_n = 1$. This probability measure is denoted by $(\mathbf{P})_{\mathbf{Y}^{(\mathcal{K})}}$. Thus,

$$H_{(\mathbf{P})_{\mathbf{Y}^{(\mathcal{K})}}}\big(D_0^{Y_0, B_0}\big) = \begin{cases} -\sum_{\delta\in\mathcal{E}}P(Y_0,\delta)/P(Y_0,\mathcal{E})\log\left(P(Y_0,\delta)/P(Y_0,\mathcal{E})\right) & \text{if } B_0 = 0, \\ 0 & \text{if } B_0 = 1. \end{cases} \tag{21}$$

Proposition 4. *The entropy of the stationary representation $\mathbb{Y}^{(\mathcal{A})}$ of the absorbed chain satisfies*

$$h\big(\mathbb{Y}^{(\mathcal{A})}\big) = h\big(\mathbb{Y}^{(\mathcal{K})}\big) + \Delta(D) \quad \text{with}$$

$$\Delta(D) = \int H_{(\mathbf{P})_{\mathbf{Y}}}\big(D_0^{Y_0, B_0}\big)\, \mathrm{d}\mathbb{P}(Y^{(\mathcal{K})}) = -\sum_{i \in I} \mu(i) P(i, \mathcal{E}) H(D^i)$$

and

$$
\begin{aligned}
h\big(\mathbb{Y}^{(\mathcal{A})}\big) = {}&-(1 - \pi(I)) \sum_{j \in I} \mu(j) \log \mu(j) - \sum_{i,j \in I} \mu(i) P(i, j) \log P(i, j) \\
&- \sum_{i \in I, \delta \in \mathcal{E}} \mu(i) P(i, \delta) \log P(i, \delta).
\end{aligned}
\tag{22}
$$

Proof. From (20) and (21), and by using the Markov property,

$$
\begin{aligned}
\Delta(D) &= \int H_{(\mathbf{P})_{\mathbf{Y}^{(\mathcal{A})}}}\big(D_0^{Y_0, B_0}\big)\, \mathrm{d}\mathbb{P}(Y^{(\mathcal{K})}) \\
&= -\sum_{i \in I} \mu(i) P(i, \mathcal{E}) \sum_{\delta \in \mathcal{E}} P(i, \delta)/P(i, \mathcal{E}) \log\left(P(i, \delta)/P(i, \mathcal{E})\right) \\
&= -\sum_{i \in I, \delta \in \mathcal{E}} \mu(i) P(i, \delta) \log P(i, \delta) + \sum_{i \in I} \mu(i) P(i, \mathcal{E}) \log P(i, \mathcal{E}).
\end{aligned}
$$

We then use (18) to get the expression in (22). □

Remark 6. Let $\mathcal{X}^{(A)}$ be a trajectory of an absorbed chain, with initial distribution μ in I and finishing after it hits $\mathcal{E}$. It has length $\tau_{\mathcal{E}}$ and it corresponds to an absorbed trajectory of length $\tau_{\mathcal{E}} - 1$ in the process $\mathbb{Y}^{(\mathcal{A})}$ with alphabet $I^2 \times \mathcal{E}^*$. In fact, if $(\mathcal{X}_1, \ldots, \mathcal{X}_l, \varepsilon)$ with $\mathcal{X}_1, \ldots, \mathcal{X}_l \in I$, $\varepsilon \in \mathcal{E}$, is an absorbed trajectory of length $l + 1$, then the associated trajectory in $\mathbb{Y}^{(\mathcal{A})}$ is given by $((\mathcal{X}_r, \mathcal{X}_{r+1}, o), r = 1, \ldots, l-1; (\mathcal{X}_l, j^*, \varepsilon))$ of length l. Here, $j^* \in I$ is an element chosen with distribution μ and it is the starting state of the next absorbed trajectory.

5.3. Entropy balance

The associated stationary chain $\mathbb{X}$ with transition kernel $P = P^\pi$ is retrieved from the stationary chain $\mathbb{Y}$ with transition kernel Q, a collection of Bernoulli variables $\mathbb{B}^{I,I}$ that assign 0 or 1 between the connections of $\mathbb{Y}$, a set of Bernoulli variables $\mathbb{D}^I$ giving the transition from I to $\mathcal{E}$, and a family of walks $\mathbb{V}$ whose components are Bernoulli variables (G_n) distributed as $\pi(\bullet \mid \mathcal{E})$. The length of these walks is Geometric$(\pi(I)) - 1$ distributed, and so they could be empty.

It is straightforward to prove the following equality, relating $h(\mathbb{X})$ given by (5) to the entropies of the elements forming the chain $\mathbb{X}$.

Proposition 5.

$$h(\mathbb{X}) = \pi(I) h\big(\mathbb{Y}^{(\mathcal{A})}\big) + \pi(\mathcal{E})^2 h(\mathbb{G}) + \pi(I)\pi(\mathcal{E}) \log \pi(I) + \pi(\mathcal{E})^2 \log \pi(\mathcal{E}).$$

Below we discuss the way this equality appears. We have reduced the elements forming the chain $\mathbb{X}$ to only two, the absorbed chains $\mathbb{Y}^{(\mathcal{A})}$ and the walks $\mathbb{V}$ with Bernoulli variables G_n. From (22), we have

$$h\big(\mathbb{Y}^{(\mathcal{A})}\big) = -\pi(\mathcal{E}) \sum_{j \in I} \mu(j) \log \mu(j) - \sum_{i,j \in I} \mu(i) P(i, j) \log P(i, j)$$
$$- \sum_{i \in I, \delta \in \mathcal{E}} \mu(i) P(i, \delta) \log P(i, \delta),$$

and the Bernoulli sequence $\mathbb{G} = (G_n)$ has entropy

$$h(\mathbb{G}) = -\sum_{\delta \in \mathcal{E}} \pi(\delta \mid \mathcal{E}) \log \pi(\delta \mid \mathcal{E}) = -\pi(\mathcal{E})^{-1} \sum_{\delta \in \mathcal{E}} \pi(\delta) \log \pi(\delta) + \log \pi(\mathcal{E}).$$

Taking some N, we divide the sequence $(X_1, \ldots, X_N)$ into the set of absorbed chains $\mathcal{X}^{(A)}$ and the set of nonempty walks V in $\mathcal{E}$. The proportion of elements in I approaches $\pi(I)$ as $N \to \infty$. Therefore, from (9) we obtain that for every time $t \in \mathbb{T}$ there are in mean $\sum_{i \in I} \mu(i) P(i, \mathcal{E})(\pi(I)^{-1} - 1)$ points belonging to a walk in $\mathcal{E}$. Since the set of points in $\mathbb{T}$ has a weight $\pi(I)$, we obtain that the proportion of points in $\mathbb{Z}$ belonging to a walk in $\mathcal{E}$ is $\pi(I) \cdot \sum_{i \in I} \mu(i) P(i, \mathcal{E})(\pi(I)^{-1} - 1) = \pi(\mathcal{E})^2$. Hence, the proportion of sites in $(X_1, \ldots, X_N)$ with symbols in $\mathbb{G}$ arising from a walk V in $\mathcal{E}$ approaches $\pi(\mathcal{E})^2$ as $N \to \infty$. We have

$$\pi(\mathcal{E})^2 h(\mathbb{G}) = -\pi(\mathcal{E}) \sum_{\delta \in \mathcal{E}} \pi(\delta) \log \pi(\delta) + \pi(\mathcal{E})^2 \log \pi(\mathcal{E}). \tag{23}$$

Let us compute $\pi(I) h\big(\mathbb{Y}^{(\mathcal{A})}\big)$. Since $\mu(i) = \pi(i)/\pi(I)$ for $i \in I$, we have

$$-\pi(I) \sum_{j \in I} \mu(j) \log \mu(j) = -\sum_{j \in I} \pi(j) \log \pi(j) + \pi(I) \log \pi(I),$$

and so, using (22), we get

$$\pi(I) h\big(\mathbb{Y}^{(\mathcal{A})}\big) = -\pi(\mathcal{E}) \sum_{j \in I} \pi(j) \log \pi(j) + \pi(\mathcal{E}) \pi(I) \log \pi(I)$$
$$- \sum_{i,j \in I} \pi(i) P(i, j) \log P(i, j) - \sum_{i \in I, \delta \in \mathcal{E}} \pi(i) P(i, \delta) \log P(i, \delta).$$

Then, the equality given in Proposition 5 has been proved:

$$h(\mathbb{X}) - (\pi(\mathcal{E})^2 h(\mathbb{G}) + \pi(\mathcal{E})^2 \log \pi(\mathcal{E})) = \pi(I) h\big(\mathbb{Y}^{(\mathcal{A})}\big) + \pi(\mathcal{E}) \pi(I) \log \pi(I).$$

The term $\pi(I)\pi(\mathcal{E}) \log \pi(I)$ has an origin similar to the last term in (23). In fact, from Remark 4, the resurrection weights $\mu(j)$, $j \in I$, appear with frequency $\pi(\mathcal{E})$ in the sequence $\mathbb{Y}$ because this occurs at the sites where there is a jump to $\mathcal{E}$. Since the sequence $\mathbb{Y}$ appears with frequency $\pi(I)$, then the term $-\sum_{i \in I} \mu(i) \log \mu(i)$ appears with frequency $\pi(I)\pi(\mathcal{E})$. Hence, as in (23), we have

$$-\pi(I)\pi(\mathcal{E}) \sum_{i \in I} \mu(i) \log \mu(i) = -\pi(\mathcal{E}) \sum_{i \in I} \pi(i) \log \pi(i) + \pi(I)\pi(\mathcal{E}) \log \pi(I),$$

and since $-\pi(\mathcal{E}) \sum_{i \in I} \pi(i) \log \pi(i)$ is the term present in (5), the extra term given by $\pi(I)\pi(\mathcal{E}) \log \pi(I)$ appears.

Remark 7. From Remark 6 the length of an absorbed trajectory in $\mathbb{Y}^{(\mathcal{A})}$ with alphabet $I^2 \times \mathcal{E}^*$ is the same as the number of elements in I of an absorbed trajectory $\mathcal{X}^{(A)}$ starting from μ and absorbed when hitting $\mathcal{E}$ (this is of length $|\mathcal{X}^{(A)}| - 1$, which counts the visited sites in I, but not the one containing the absorbing state). Since the entropy of a system is the gain of entropy per unit of time, the proportion of symbols given the entropy $h(\mathbb{Y}^{(\mathcal{A})})$ is $\pi(I)$. This explains why the term $\pi(I)h(\mathbb{Y}^{(\mathcal{A})})$ appears.

Appendix A. Proof of Proposition 2

Let us first show that the process $(\mathbb{S}, \mathbb{T})$ is regenerative. Consider a pair of sequence $(a(u) : u \geq 0)$ and $(b(1), \ldots, b(m))$ taking values in $I \cup \mathcal{E}$ and such that $a(0) \in I$. From the construction of $\mathbb{S}$, we have

$$\mathbb{P}(\mathbb{S}_{T_n+l} = b(l), l = 1, \ldots, m \mid \mathbb{S}_{T_n-u} = a(u), u \geq 0) \qquad (24)$$
$$= \mathbb{P}(\mathbb{S}_{T_n+l} = b(l), l = 1, \ldots, m \mid \mathbb{S}_{T_n} = a(0)).$$

Also, we have that $(\mathbb{S}_{T_l} : l \in \mathbb{Z})$ is equally distributed as $(Y_n : n \in \mathbb{Z})$. Then,

$$\mathbb{S}_{T_n} \sim \mu \text{ for all } n \in \mathbb{Z}. \qquad (25)$$

This proves that $(\mathbb{S}, \mathbb{T})$ is regenerative. Therefore, by making a shift on a random number of sites U uniformly chosen in $\{0, \ldots, T_1 - 1\}$ we define a stationary process $\mathbb{S}^s$ given by $\mathbb{S}_t^s = \mathbb{S}_{t+U}$ for $t \in \mathbb{Z}$.

Since the random number U only depends on the length T_1, there is regeneration at the random times in $\mathbb{T} = \{t \in \mathbb{Z} : \mathbb{S}_t^s \in I\}$ (see relations (24) and (25)). Hence, for all $t \in \mathbb{Z}, b \in I \cup \mathcal{E}$, and $(a(u) : n \leq t)$ taking values in $I \cup \mathcal{E}$, we have

$$\mathbb{P}(\mathbb{S}_{t+1}^s = b \mid b\mathbb{S}_u^s = a(u), u \leq t) = \mathbb{P}(\mathbb{S}_{t+1}^s = b \mid \mathbb{S}_u^s = a(u), u = t, \ldots, t - r), \qquad (26)$$

where $r \geq 0$ is the first nonnegative element such that $a(t - r) \in I$.

Again using that U only depends on T_1, we get that, for all $t \in \mathbb{Z}$ and all $a, b \in I \cup \mathcal{E}$,

$$\mathbb{P}(\mathbb{S}_{t+1}^s = b \mid \mathbb{S}_t^s = a) = \mathbb{P}(\mathbb{S}_{u+1} = b \mid \mathbb{S}_u = a) \quad \text{for } u \neq -1, 0.$$

We avoid taking $u = 1$ or $u = 0$ because $\mathbb{S}_0$ only takes values in I.

Let us compute $\mathbb{P}(t \in \mathbb{T})$. It suffices to calculate $\mathbb{P}(t \notin T)/\mathbb{P}(t \in \mathbb{T})$. From (9) we get that, for every time $t \in \mathbb{T}$, there are in mean

$$\sum_{i \in I} \mu(i) \sum_{j \in I} Q(i, j)\overline{\theta}_{i,j}\pi(I)^{-1} = \sum_{i \in I} \mu(i)P(i, \mathcal{E})\pi(I)^{-1}$$

points in $\mathbb{Z} \setminus \mathbb{T}$. Then, from (4) we find

$$\mathbb{P}(t \notin \mathbb{T})/\mathbb{P}(t \in \mathbb{T}) = \sum_{i \in I} \mu(i)P(i, \mathcal{E})\pi(I)^{-1} = \pi(\mathcal{E})/\pi(I).$$

We conclude that

$$\mathbb{P}(t \in \mathbb{T}) = \pi(I) \text{ and } \mathbb{P}(t \notin \mathbb{T}) = \pi(\mathcal{E}). \qquad (27)$$

Hence, $\mathbb{P}(\mathbb{S}_t^s \in I) = \pi(I)$ and $\mathbb{P}(\mathbb{S}_t^s \in \mathcal{E}) = \pi(\mathcal{E})$.

Let $i, j \in I$. We have $\mathbb{P}(\mathbb{S}_t^s = i \mid \mathbb{S}_t^s \in I) = \mathbb{P}(\mathbb{S}_0 = i \mid \mathbb{S}_0 \in I) = \mu(i)$, and so, using (27), we get

$$\mathbb{P}(\mathbb{S}_t^s = i) = \mu(i)\mathbb{P}(\mathbb{S}_t^s \in I) = \mu(i)\pi(I) = \pi(i). \tag{28}$$

Let $i, j \in I$. From the definition of $\theta_{i,j}$ in (10) we get

$$\mathbb{P}(\mathbb{S}_{t+1}^s = j \mid \mathbb{S}_t^s = i) = \theta_{i,j}Q(i, j) = P(i, j). \tag{29}$$

We have $\sum_{j \in I} \mathbb{P}(\mathbb{S}_{t+1}^s = j \mid \mathbb{S}_t^s = i) = 1 - P(i, \mathcal{E})$, so $\mathbb{P}(\mathbb{S}_{t+1}^s \in \mathcal{E} \mid \mathbb{S}_t^s = i) = P(i, \mathcal{E})$. Then, $\mathbb{S}_t^s = i$ jumps to $\mathcal{E}$ with probability $P(i, \mathcal{E})$, and the jump to some particular state $\delta \in \mathcal{E}$ is done with probability

$$\mathbb{P}(\mathbb{S}_{t+1}^s = \delta \mid \mathbb{S}_t^s = i) = P(i, \mathcal{E})P(i, \delta)/P(i, \mathcal{E}) = P(i, \delta). \tag{30}$$

Let $\delta, \varepsilon \in \mathcal{E}$. We have $P(\delta, \mathcal{E}) = \mathbb{P}(V \neq \emptyset) = \pi(\mathcal{E})$, so

$$\mathbb{P}(\mathbb{S}_{t+1}^s = \varepsilon \mid \mathbb{S}_t^s = \delta) = \pi(\varepsilon \mid \mathcal{E})P(\delta, \mathcal{E}) = \pi(\varepsilon \mid \mathcal{E})\pi(\mathcal{E}) = \pi(\varepsilon). \tag{31}$$

Then, by using previous relations and (3), we get

$$\begin{aligned}
\mathbb{P}(\mathbb{S}_t^s = \delta) &= \sum_{i \in I} \mathbb{P}(\mathbb{S}_{t-1}^s = i, \mathbb{S}_t^s = \delta) + \sum_{\varepsilon \in \mathcal{E}} \mathbb{P}(\mathbb{S}_{t-1}^s = \varepsilon, \mathbb{S}_t^s = \delta) \\
&= \sum_{i \in I} \mathbb{P}(\mathbb{S}_t^s = \delta \mid \mathbb{S}_{t-1}^s = i)\pi(i) + \sum_{\varepsilon \in \mathcal{E}} \mathbb{P}(\mathbb{S}_t^s = \delta \mid \mathbb{S}_{t-1}^s = \varepsilon)\mathbb{P}(\mathbb{S}_{t-1}^s = \varepsilon) \\
&= \sum_{i \in I} P(i, \delta)\pi(i) + \pi(\delta) \sum_{\varepsilon \in \mathcal{E}} \mathbb{P}(\mathbb{S}_{t-1}^s = \varepsilon) \\
&= \sum_{i \in I} P(i, \delta)\mu(i)\pi(I) + \pi(\delta)\pi(\mathcal{E}) = \pi(\delta)(\pi(I) + \pi(\mathcal{E})) = \pi(\delta).
\end{aligned}$$

Again from the construction of the process $\mathbb{S}$, it follows that

$$\mathbb{P}(\mathbb{S}_{t+1}^s = \varepsilon \mid \mathbb{S}_t^s = \delta, \mathbb{S}_u^s = a(u), u < t) = \mathbb{P}(\mathbb{S}_{t+1}^s = \varepsilon \mid \mathbb{S}_t^s = \delta) = \pi(\varepsilon). \tag{32}$$

Now, let us compute $\mathbb{P}(\mathbb{S}_t^s = \varepsilon, \mathbb{S}_{t+1}^s = j)$ for $\varepsilon \in \mathcal{E}$, $j \in I$. The pair $(\mathbb{S}_t^s = \varepsilon, \mathbb{S}_{t+1}^s = j)$ has its origin in some pair $(Y_s = i, Y_{s+1} = j)$ satisfying $B_{s+1}^{i,j} = 0$, for some $i \in I$. Then, by summing over all states $i \in I$ and all pieces of trajectories in $\mathcal{E}$ that are built between i and j, and by using (30), (8), and (3), we get

$$\begin{aligned}
\mathbb{P}(\mathbb{S}_t^s = \varepsilon, \mathbb{S}_{t+1}^s = j) &= \sum_{i \in I} \sum_{l \geq 1} \mathbb{P}(\mathbb{S}_{t-l}^s = i; \mathbb{S}_{t-u}^s \in \mathcal{E}, 1 \leq u < l; \mathbb{S}_t^s = \varepsilon, \mathbb{S}_{t+1}^s = j) \\
&= \sum_{i \in I} \mu(i)\mu(j)P(i, \varepsilon)\pi(I) + \sum_{i \in I} \mu(i)\mu(j) \sum_{l \geq 1; \delta_1, \dots, \delta_l \in \mathcal{E}} \left(P(i, \delta_1) \prod_{k=2}^{l-1} \pi(\delta_k) \right) \pi(\varepsilon)\pi(I) \\
&= \pi(\varepsilon)\pi(I)\mu(j) = \pi(\varepsilon)\pi(j).
\end{aligned} \tag{33}$$

From relations (28)–(33) we get that the bivariate marginals of the stationary chains $\mathbb{X} = (X_n)$ and $\mathbb{S}^s = (\mathbb{S}_n^s)$ are the same. So, we will finish by proving that $\mathbb{S}^s$ satisfies the Markov property. In view of the regeneration property (26), this will be proven once we show that

$$\mathbb{P}(\mathbb{S}_{t+1}^s = b \mid \mathbb{S}_u^s = a(u), u = t, \dots, t - r) = \mathbb{P}(X_{t+1} = b \mid X_t = a(t)), \tag{34}$$

where $r \geq 0$ and satisfies $a(t - r) \in I$, $a(t - u) \in \mathcal{E}$ for $u = 1, \ldots, r - 1$. This was shown for the case $r = 0$ in (29) and (30). On the other hand, (32) proves (34) in the case $b \in \mathcal{E}$ and $r > 0$. So, the unique case left to show is $b \in I$ and $r > 0$.

Let $i, j \in I$, $r > 0$, $\delta_u \in \mathcal{E}$ for $u = 0, \ldots, r - 1$. Since $\mathbb{P}(X_{t+1} = j \mid X_{t-u} = \delta_0) = \pi(j)$, to achieve the proof of (34), the unique relation that we are left to show is

$$\mathbb{P}(\mathbb{S}^s_{t+1} = j \mid \mathbb{S}^s_{t-u} = \delta_u, u = 0, \ldots, r - 1, \mathbb{S}^s_{t-r} = i) = \pi(j).$$

We have

$$\mathbb{P}(\mathbb{S}^s_{t+1} = j, \mathbb{S}^s_{t-u} = \delta_u, u = 0, .., r - 1, \mathbb{S}^s_{t-r} = i)$$

$$= \mu(i)(P(i, j) + P(i, \mathcal{E})\mu(j))\theta_{i,j} \frac{P(i, \delta_0)}{P(i, \mathcal{E})} \left(\prod_{u=1}^{r-1} \pi(\delta_l) \right) \pi(I)$$

$$= \mu(i)P(i, \mathcal{E})\mu(j) \frac{P(i, \delta_0)}{P(i, \mathcal{E})} \left(\prod_{u=1}^{r-1} \pi(\delta_l) \right) \pi(I)$$

$$= \mu(i)P(i, \delta_0)\mu(j) \left(\prod_{u=1}^{r-1} \pi(\delta_l) \right) \pi(I),$$

and

$$\mathbb{P}(\mathbb{S}^s_t = i, \mathbb{S}^s_{t-u} = \delta_u : u = 0, \ldots, r - 1) = \left(\sum_{j \in I} \mu(i)P(i, \delta_0)\mu(j) \right) \left(\prod_{u=1}^{r-1} \pi(\delta_l) \right)$$

$$= \mu(i)P(i, \delta_0) \left(\prod_{u=1}^{r-1} \pi(\delta_l) \right).$$

Therefore,

$$\mathbb{P}(\mathbb{S}^s_{t+1} = j \mid \mathbb{S}^s_{t-u} = \delta_u, u = 0, \ldots, r - 1, \mathbb{S}^s_{t-r} = i) = \mu(j)\pi(I) = \pi(j).$$

Then (34) follows. We have proven that the laws of the stationary chains $\mathbb{X} = (X_t)$ and $\mathbb{S}^s = (\mathbb{S}^s_t)$ are the same.

Acknowledgements

References

[1] ABRAMOV, L. AND ROKHLIN, V. (1962). The entropy of a skew product of measure-preserving transformations. *Vestnik Leningrad Univ.* **17**, 5–13 (in Russian).

[2] ASMUSSEN, S. (2003). *Applied Probability and Queues*, 2nd edn. New York, Springer.

[3] COLLET, P., MARTÍNEZ, S. AND SAN MARTÍN, J. (2013). *Quasi-Stationary Distributions: Markov Chains, Diffusions and Dynamical Systems*. Heidelberg, Springer.

[4] DARROCH, J. AND SENETA, E. (1965). On quasi-stationary distributions in absorbing discrete-time finite Markov chains. *J. Appl. Prob.* **2**, 88–100.

[5] DENKER, M., GRILLENBERGER, C. AND SIGMUND, K. (1976). *Ergodic Theory on Compact Spaces*. Berlin, Springer.

[6] DOWNAROWICZ, T. AND SERAFIN, J. (2002). Fiber entropy and conditional variational principles in compact non-metrizable spaces. *Fund. Math.* **172**, 217–247.

[7] FERRARI, P. A., KESTEN, H., MARTÍNEZ, S. AND PICCO, P. (1995). Existence of quasi-stationary distributions. A renewal dynamical approach. *Ann. Prob.* **23**, 501–521.

[8] FERRARI, P. A., MARTÍNEZ, S. AND PICCO, P. (1992). Existence of non-trivial quasi-stationary distributions in the birth-death chain. *Adv. Appl. Prob.* **24**, 795–813.

[9] KEANE, M. AND SMORODINSKY, M. (1979). Finitary isomorphisms of irreducible Markov shifts. *Israel J. Math.* **34**, 281–286.