



# Persistent random walks, variable length Markov chains and piecewise deterministic Markov processes \*

Peggy Cénac, Brigitte Chauvin, Samuel Herrmann, Pierre Vallois

## ► To cite this version:

Peggy Cénac, Brigitte Chauvin, Samuel Herrmann, Pierre Vallois. Persistent random walks, variable length Markov chains and piecewise deterministic Markov processes \*. Markov Processes and Related Fields, Polymath, 2013, 19 (1), pp.1-50. hal-01285853

**HAL Id: hal-01285853**

**<https://hal.archives-ouvertes.fr/hal-01285853>**

Submitted on 9 Mar 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Persistent random walks, variable length Markov chains and piecewise deterministic Markov processes \*

Peggy CÉNAC<sup>1</sup>, Brigitte CHAUVIN<sup>2</sup>, Samuel HERRMANN<sup>1</sup>, and Pierre VALLOIS<sup>3</sup>

<sup>1</sup>Institut de Mathématiques de Bourgogne (IMB) - UMR CNRS 5584,  
Université de Bourgogne, B.P. 47 870  
21078 Dijon Cedex, France

<sup>2</sup>Laboratoire de Mathématiques de Versailles (LMV) - UMR CNRS 8100  
45 avenue des Etats-Unis  
78035 Versailles Cedex, France

<sup>3</sup>Université de Lorraine, Institut de Mathématiques Elie Cartan,  
INRIA-BIGS, CNRS UMR 7502,  
BP 70239, F-54506 Vandoeuvre-lès-Nancy Cedex, France

April 1, 2013

## Abstract

A classical random walk  $(S_t, t \in \mathbb{N})$  is defined by  $S_t := \sum_{n=0}^t X_n$ , where  $(X_n)$  are i.i.d. When the increments  $(X_n)_{n \in \mathbb{N}}$  are a one-order Markov chain, a short memory is introduced in the dynamics of  $(S_t)$ . This so-called “persistent” random walk is no longer Markovian and, under suitable conditions, the rescaled process converges towards the integrated telegraph noise (ITN) as the time-scale and space-scale parameters tend to zero (see [11, 17, 18]). The ITN process is effectively non-Markovian too. The aim is to consider persistent random walks  $(S_t)$  whose increments are Markov chains with variable order which can be infinite. This variable memory is enlightened by a one-to-one correspondence between  $(X_n)$  and a suitable Variable Length Markov Chain (VLMC), since for a VLMC the dependency from the past can be unbounded. The key fact is to consider the non Markovian letter process  $(X_n)$  as the margin of a couple  $(X_n, M_n)_{n \geq 0}$  where  $(M_n)_{n \geq 0}$  stands for the memory of the process  $(X_n)$ . We prove that, under a suitable rescaling,  $(S_n, X_n, M_n)$  converges in distribution towards a time continuous process  $(S^0(t), X(t), M(t))$ . The process  $(S^0(t))$  is a semi-Markov and Piecewise Deterministic Markov Process whose paths are piecewise linear.

*2010 Mathematics Subject Classification.* 60J10, 60J27, 60F05, 60G17, 60G40, 60K15.

*Key words and phrases.* Persistent random walk. Variable length Markov chain. Integrated telegraph noise. Piecewise Deterministic Markov Processes. Semi Markov processes. Variable memory. Simple and double infinite combs.

---

\*Supported in part by the ANR grants : "Malliavin, Stein and Stochastic Equations with Irregular Coefficients" [ANR-10-BLAN-0121], "Mathematical Analysis of Neuronal Dynamics" [ANR-09-BLAN-0008-01] and "Approches Spatio-Temporelles pour la modélisation du Risque" [ANR- 08-BLAN-0314-01].

# 1 Introduction

Classical random walks are defined by

$$S_t := \sum_{n=0}^t X_n, \quad (1.1)$$

for  $t \in \mathbb{N}$  and for i.i.d. increments  $(X_n)_{n \in \mathbb{N}}$ . It is well known that a suitable rescaling of the random walk permits to obtain the standard Brownian motion as the time-scale and space-scale parameters tend to zero. When the increments  $(X_n)_{n \in \mathbb{N}}$  are defined as a one-order Markov chain, a short memory in the dynamics of the stochastic paths is introduced: the process is called in the literature the *persistent* random walk or a correlated random walk or also a Kac walk (see [8, 14, 19, 20]). The random walk is no longer Markovian and, under suitable conditions, the rescaled process converges towards the integrated telegraph noise (ITN), see [11, 17] and [18]. The ITN process is effectively non-Markovian too.

Our aim is to define processes  $(X_n)_{n \in \mathbb{N}}$  with variable memory and thus to generalize this convergence result to random walks whose increments are higher order Markov chains. When  $(X_n)_{n \in \mathbb{N}}$  is a Markov chain of finite order, it is natural to think that the limit process should be very close to the integrated telegraph noise. That is why we are mostly interested in constructing infinite length Markov chain or in dealing with *Variable Length Markov Chains* (VLMC) for which the dependency from the past is non bounded.

A VLMC can be defined as follows (this probabilistic presentation comes from [2], other more statistic points of view can be found in [15, 9]). Let  $\mathcal{L} = \{0, 1\}^{-\mathbb{N}}$  be the set of left-infinite words on the alphabet  $\{0, 1\}$ . Consider a complete (each node has 0 or 2 children) binary tree whose finite leaves  $\mathcal{C}$  are words on the alphabet  $\{0, 1\}$ . To each leaf  $c$  (not necessarily finite) is attached a Bernoulli distribution denoted by  $q_c$ . Each leaf is called a *context* and this probabilized tree is called a *context tree*. See for instance the simple infinite comb in Figure 2: the set of leaves  $\mathcal{C}$  is defined by

$$\mathcal{C} := \{0^n 1, n \geq 0\} \cup \{0^\infty\}$$

where  $0^n 1$  represents the sequence  $00 \dots 01$  composed with  $n$  characters '0' and one character '1'. By convention  $0^0 1 = 1$ . The set of leaves contains one infinite leaf  $0^\infty$  and a countable set of finite leaves  $0^n 1$ . The *prefix function*  $\overleftarrow{\text{pref}} : \mathcal{L} = \{0, 1\}^{-\mathbb{N}} \rightarrow \mathcal{C}$  indicates the length of the last run of '0': for instance,

$$\overleftarrow{\text{pref}}(\dots 1000) = 0001 = 0^3 1.$$

For a general context tree and for any left-infinite word  $U$ , we define  $\overleftarrow{\text{pref}}(U)$  in a similar way as the first suffix of  $U$  reading from right to left appearing as a leaf of the context tree. The associated VLMC is the  $\mathcal{L}$ -valued Markov chain  $(U_n)_{n \geq 0}$  defined by the transitions

$$\mathbb{P}(U_{n+1} = U_n \ell | U_n) = q_{\overleftarrow{\text{pref}}(U_n)}(\ell) \quad (1.2)$$

where  $\ell \in \{0, 1\}$  is any letter. Notice that the VLMC is entirely determined by the data  $q_c, c \in \mathcal{C}$ . Moreover the order of dependence (the memory) depends on the past itself.

For a given VLMC  $(U_n)_{n \geq 0}$ , define  $X_n$  as the last letter of  $U_n$  for any  $n \geq 0$ . When the context tree associated with  $(U_n)$  is infinite, then the letter process  $(X_n)_{n \geq 0}$  is non Markovian, because the transition probabilities (1.2) indicate that  $X_{n+1}$  depends on a variable and unbounded number of previous letters. The corresponding random walk  $(S_t)$  defined by (1.1) is non Markovian anymore, it is somehow *very persistent*, so we investigate the following natural questions: is the random walk of the same nature as in the one-order Markov case?

Is the rescaled process convergent to some limit process? Is the limiting process analog to the ITN?

Recall that  $X_n$  is the last letter of a VLMC  $(U_n)$ . The key point of view is the following: we consider the non Markovian letter process  $(X_n)$  as the margin of a couple  $(X_n, M_n)_{n \geq 0}$  where  $(M_n)_{n \geq 0}$  stands for the memory of the process  $(X_n)$ . It is reasonable to believe that  $M_n = |\overleftarrow{\text{pref}}(U_n)|$  is a good candidate, where the notation  $|w|$  stands for the length of a word  $w$ . More precisely in the particular case of a two-letter alphabet  $\mathcal{A}$ , the Markov chain  $(X_n, M_n)_{n \geq 0}$  valued in the state space  $\mathcal{A} \times \mathbb{N}^*$  is defined by the transition probabilities: let  $\ell, \ell' \in \mathcal{A}, \ell \neq \ell'$ ,

$$\begin{cases} Q((\ell, n), (\ell, n+1)) = 1 - \alpha_{\ell, n}, \\ Q((\ell, n), (\ell', 1)) = \alpha_{\ell, n}. \end{cases}$$

Note that  $\alpha_{\ell, k}$  is the probability of changing letter after a run of length  $k$  of letter  $\ell$ , that is

$$\alpha_{\ell, k} = \mathbb{P}(X_{n+1} \neq \ell | X_n = \ell, M_n = k). \quad (1.3)$$

Introducing the sequence of breaking times:

$$T_0 = 0, \quad T_{k+1} = \inf\{n > T_k, X_n \neq X_{T_k}\}$$

it is easy to see that  $(X_n, T_n)_{n \geq 0}$  is a semi-Markov process (see [4, Chapter 10] and [10, 12]). In Section 2, we consider a Markov chain  $(X_n, M_n)_{n \geq 0}$ , where  $(X_n)_{n \geq 0}$  is a letter process, the letters belong to an alphabet  $\mathcal{A} := \{a_1, a_2, \dots, a_K\}$ , and  $(M_n)_{n \geq 0}$  stands for the memory of the process  $(X_n)$ . The state space associated with this Markov chain is  $\{a_1, a_2, \dots, a_K\} \times \overline{\mathbb{N}}^*$ . We give in Section 2.2 the properties of  $(X_n, M_n)_{n \geq 0}$  and we determine necessary and sufficient conditions for existence and unicity of a stationary probability measure, in Section 2.2.2. We would like to emphasize that  $(X_n)$  is non-Markovian in general.

In Section 3, we consider two particular cases of VLMC, associated with the simple infinite comb and the double infinite comb. In each of these two cases, the stationary measure can be explicitly calculated, in [2] for the simple comb and in the Appendix for the double comb. We make precise the correspondence between the process  $(X_n, M_n)_{n \geq 0}$  defined in Section 2 and the VLMC  $(U_n)$  whose the last letter is  $X_n$ . Namely, we establish the dictionary between the stationary measure for the Markov chain  $(X_n, M_n)_{n \geq 0}$  and the stationary measure for the VLMC (see Theorems 3.1 and 3.3). Thanks to these results, we do not have to worry about the point of view (couple letter/memory or VLMC) when considering the persistent

random walk  $S_t := \sum_{n=0}^t X_n$ , under the stationary regime.

Section 4 is devoted to the study of  $(S_n)$ . In particular, we determine the explicit distribution of the r.v.  $S_n$ , see Proposition 4.1. Although the result is complicated, we are able to determine explicitly the generating function of the r.v.  $S_{\tau+1}$ , where  $\tau$  is a geometric r.v. independent from  $(X_n, M_n)_{n \geq 0}$ . One way to compare the process  $(S_n)$  with a classical random walk is to analyse how both processes fluctuate at infinity. We have the two following limit theorems, see Section 4.4:

$$\lim_{n \rightarrow \infty} \frac{S_n}{n} = \Xi, \quad \text{and} \quad \sqrt{n} \left( \frac{S_n}{n} - \Xi \right) \rightarrow \mathcal{N}(0, \sigma)$$

where  $\mathcal{N}(0, \sigma)$  is a Gaussian distribution with 0 mean and variance  $\sigma^2$ ,  $\Xi$  and  $\sigma$  are constants which can be expressed in terms of the model parameters.

Finally in Section 5 we study the persistent random walk. After a convenient scaling, it converges towards a ‘‘generalized ITN’’ as proved in Theorem 5.1. More precisely, we focus on

the limit in law of Markov chains of the type  $(X_n^\varepsilon, M_n^\varepsilon)$  which depends on a small parameter  $\varepsilon > 0$ . We suppose that  $X_n^\varepsilon$  takes its values in  $\{-1, 1\}$ ,  $X_0^\varepsilon = 1$  and

$$\mathbb{P}\left(X_{n+1}^\varepsilon = 1 \mid X_n^\varepsilon = -1, M_n^\varepsilon = k\right) = f_1(k\varepsilon)\varepsilon + o(\varepsilon) \quad (1.4)$$

$$\mathbb{P}\left(X_{n+1}^\varepsilon = -1 \mid X_n^\varepsilon = 1, M_n^\varepsilon = k\right) = f_2(k\varepsilon)\varepsilon + o(\varepsilon) \quad (1.5)$$

where  $f_1, f_2 : [0, \infty[ \rightarrow \mathbb{R}$  are non negative and right continuous functions. Note that (1.4) and (1.5) mean that  $(X_n^\varepsilon)$  has a *conservative* behaviour: if  $X_n^\varepsilon = 1$  (resp.  $X_n^\varepsilon = -1$ ) the probability that  $X_{n+1}^\varepsilon$  changes, i.e.  $X_{n+1}^\varepsilon = -1$  (resp.  $X_{n+1}^\varepsilon = 1$ ) is small for convenient  $f_1, f_2$  and is measured by the parameter  $\varepsilon$ .

Under additional assumptions, see the beginning of Section 5 for details, it is actually possible to rescale the triplet  $(X_n^\varepsilon, M_n^\varepsilon, S_n^\varepsilon)$  so that it converges as  $\varepsilon \rightarrow 0$ . For simplicity, we only present the scaling procedure concerning  $S_n^\varepsilon$ . The process  $(S^\varepsilon(t), t \geq 0)$  is piecewise linear and satisfies

$$S^\varepsilon(n\varepsilon) = \varepsilon S_n^\varepsilon, \quad \text{for any } n \in \mathbb{N}. \quad (1.6)$$

We prove (see Theorem 5.1 for a more complete result) that  $(S^\varepsilon(t), t \geq 0)$  converges in distribution, as  $\varepsilon \rightarrow 0$  to  $(S^0(t), t \geq 0)$  where

$$S^0(t) = \int_0^t (-1)^{N^0(s)} ds.$$

Here,  $(N^0(t))$  is the counting process with jump times  $(\xi_n)_{n \geq 0}$ :

$$N^0(t) = \sum_{n \geq 0} \mathbb{1}_{\{\xi_n \leq t\}},$$

where  $(\xi_{n+1} - \xi_n, n \geq 0)$  is a sequence of independent r.v. such that  $\xi_0 = 0$  and

$$\begin{aligned} \mathbb{P}(\xi_{2n+1} - \xi_{2n} \geq t) &= \exp\left(-\int_0^t f_2(u) du\right) \\ \mathbb{P}(\xi_{2n+2} - \xi_{2n+1} \geq t) &= \exp\left(-\int_0^t f_1(u) du\right) \end{aligned}$$

for any  $t \geq 0$  and  $n \geq 0$ , where  $f_1, f_2$  satisfies (5.49).

The process  $(S^0(t), t \geq 0)$  is called the *Generalized Integrated Telegraph Noise* (see [11] for the ITN). It is both a semi-Markov process and a Piecewise Deterministic Markov Process [6, 7, 5] and its trajectories look like a zig-zag.

## 2 The Markov chain $(X_n, M_n)_{n \geq 0}$

### 2.1 Definition

Let us consider the finite set  $\mathcal{A} = \{a_1, \dots, a_K\}$  with  $K > 1$  elements. To each  $a_i$  is associated a sequence  $(\alpha_{i,n})_{n \geq 1} \in ]0, 1[^{\mathbb{N}^*}$  where  $\mathbb{N}^* = \{1, 2, 3, \dots\}$ . We can now introduce the Markov chain  $(X_n, M_n)_{n \in \mathbb{N}}$  valued in the state space  $\{a_1, \dots, a_K\} \times \mathbb{N}^*$  with transition probabilities

$$\begin{cases} Q\left((a_i, n), (a_i, n+1)\right) = 1 - \alpha_{i,n}, \\ Q\left((a_i, n), (a_j, 1)\right) = \alpha_{i,n} p_{i,j}, \quad 1 \leq i \neq j \leq K, \quad n \geq 1, \end{cases} \quad (2.1)$$

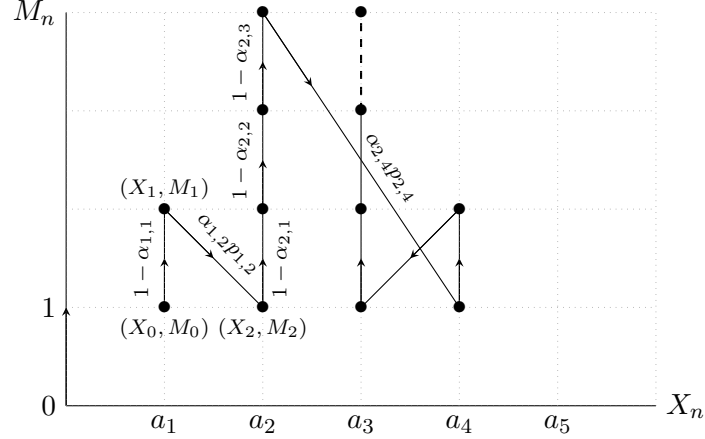


Figure 1: A path description of the process  $(X_n, M_n)_{n \geq 0}$

where  $\mathcal{P} := (p_{i,j})$  is a given  $K \times K$  transition matrix satisfying  $p_{i,i} = 0$  for all  $1 \leq i \leq K$ ,  $p_{i,j} > 0$  for all  $i \neq j$  and  $\sum_{j=1}^K p_{i,j} = 1$  for all  $i$ . In fact,  $p_{i,j}$  is the probability to move from  $a_i$  to  $a_j$  knowing that we leave  $a_i$ .

Moreover, in order to deal with VLMC later on, we extend the definition of the Markov chain to the state space  $\{a_1, \dots, a_K\} \times \bar{\mathbb{N}}^*$  with  $\bar{\mathbb{N}}^* = \mathbb{N}^* \cup \{\infty\}$ . Therefore we introduce  $\alpha_{i,\infty} \in ]0, 1[$  for all  $1 \leq i \leq K$  such that

$$\begin{cases} Q((a_i, \infty), (a_i, \infty)) = 1 - \alpha_{i,\infty} \\ Q((a_i, \infty), (a_j, 1)) = \alpha_{i,\infty} p_{i,j}, \quad i \neq j. \end{cases} \quad (2.2)$$

Note that  $\alpha_{i,k}$  is the probability of changing letter after a run of length  $k$  of  $a_i$ , that is

$$\alpha_{i,k} = \mathbb{P}(X_{n+1} \neq a_i | X_n = a_i, M_n = k). \quad (2.3)$$

There are strong links between  $(X_n)$  and  $(M_n)$ . In particular, if  $M_0 = 1$ ,  $M_n$  can be expressed with  $X_0, \dots, X_n$ . Indeed, if the sequence  $(X_j)_{j=0, \dots, n}$  is constant then  $M_n = n + 1$  and  $M_n = \inf\{1 \leq i \leq n; X_{n-i} \neq X_n\}$  otherwise. In other words, one has

$$\begin{aligned} M_n &= 1 + \sup\{0 \leq i \leq n, X_{n-j} = X_n, \forall j \in \{0, \dots, i\}\} \\ &= \inf\{0 \leq i \leq n, X_{n-i} \neq X_n\}. \end{aligned} \quad (2.4)$$

Let us explain how moves  $(X_n, M_n)$  in the case  $M_0 = 1$  and  $X_0 = a_i$ . The variable  $M_n$  increases by one unit at each time until  $X_n$  switches to  $a_j \neq a_i$ . At that first jump time, the memory is reset to 1 and so on... So that  $M_n$  represents the variable memory of  $(X_t)_{0 \leq t \leq n}$  since it counts the last consecutive stays (at  $X_n$ ) before  $n$ . Moreover the dynamics of the jumps of  $X_n$  is governed by the value of  $M_n$ . In Figure 1, we have drawn the following trajectory of  $(X_n, M_n)$  corresponding to the values  $a_1, a_1, a_2, a_2, a_2, a_2, a_4, a_4, a_3, a_3, a_3, a_3 \dots$  of  $X_n$ .

Let us note that in the particular case:  $K = 2$ ,  $a_1 = 0$ ,  $a_2 = 1$  and  $\alpha_{j,n} = \alpha_j$  for all  $n \geq 1$ , then  $(X_n)_{n \geq 0}$  is a sequence of independent Bernoulli random variables.

## 2.2 Properties of the Markov chain $(X, M)$

First we investigate under which conditions either  $(X_n)$  or  $(M_n)$  is Markov. Secondly we prove existence of invariant probability measure and finally we present a path description for the process  $(X_n)$ .

### 2.2.1 Link between the margins

A natural question arising about a 2-dimensional Markov chain  $(X_n, M_n)_{n \in \mathbb{N}}$  is to know whether the margins are Markovian too. The following proposition says that in general case, neither  $(X_n)_{n \in \mathbb{N}}$  nor  $(M_n)_{n \in \mathbb{N}}$  is a Markov chain.

**Proposition 2.1.** *Assume that  $M_0 = 1$ .*

- (i) *The margin process  $(X_n)_{n \in \mathbb{N}}$  is Markovian if and only if for all  $1 \leq i \leq K$ ,  $n \mapsto \alpha_{i,n}$  is constant. In that case the transition matrix  $Q^X$  of  $X$  is given by:*

$$Q^X(i, j) = \begin{cases} 1 - \alpha_{i,1} & \text{if } j = i \\ \alpha_{i,1} p_{i,j} & \text{if } j \neq i. \end{cases} \quad (2.5)$$

- (ii) *The margin process  $(M_n)_{n \in \mathbb{N}}$  is Markovian for any initial condition  $X_0$  if and only if for all  $n \geq 1$ , the function  $i \mapsto \alpha_{i,n}$  is constant. In that case, the transition matrix  $Q^M$  of  $M$  is*

$$Q^M(n, j) = \begin{cases} 1 - \alpha_{1,n} & \text{if } j = n + 1 \\ \alpha_{1,n} & \text{if } j = 1. \end{cases}$$

PROOF.

- (i) For a given vector  $(x_0, \dots, x_n) \in \{a_1, \dots, a_K\}^{n+1}$ , let us first denote

$$\delta_{i,n} := \mathbb{P}(X_{n+1} = a_i | X_n = x_n, \dots, X_0 = x_0).$$

According to (2.4) let us introduce:

$$m_n = 1 + \sup\{0 \leq i \leq n : x_{n-j} = x_n, \forall j \in \{0, \dots, i\}\}. \quad (2.6)$$

We have to distinguish two cases.

- (a) If  $x_n = a_i$  then  $M_n = m_n$  and therefore  $\delta_{i,n} = 1 - \alpha_{i,m_n}$ . We can choose different values of  $x_2, \dots, x_{n-1}$  such that  $m_n = 1, 2, \dots, n$ . Hence if  $(X_n)$  is Markovian then  $\delta_{i,n}$  is independent of  $n$  and  $(x_0, \dots, x_{n-1})$  and  $(\alpha_{i,k})_{k \geq 1}$  is constant.
- (b) If  $x_n = a_j \neq a_i$  then  $\delta_{i,n} = \alpha_{j,m_n} p_{j,i} = \alpha_{j,1} p_{j,i}$  implying that  $(X_n)$  is effectively Markovian.

- (ii) Let us study the process  $(M_n)_{n \geq 0}$ . Set

$$d_{i,n} := \mathbb{P}(X_0 = i, M_0 = 1, M_1 = 2, \dots, M_n = n + 1, M_{n+1} = 1).$$

We have

$$d_{i,n} = \mathbb{P}(X_0 = i, M_0 = 1, X_1 = i, M_1 = 2, \dots, X_n = i, M_n = n + 1, M_{n+1} = 1).$$

Since  $(X_n, M_n)$  is a Markov chain, using (2.1) and (2.3) we get

$$d_{i,n} = (1 - \alpha_{i,1}) \times \dots \times (1 - \alpha_{i,n}) \alpha_{i,n+1}.$$

Suppose that  $(M_n)$  is a Markov chain, with transition matrix  $Q^M$ . Then

$$\begin{aligned} d_{i,n} &= Q^M(n+1, 1) \mathbb{P}(X_0 = i, M_0 = 1, \dots, M_n = n+1) \\ &= Q^M(n+1, 1) (1 - \alpha_{i,1}) \times \dots \times (1 - \alpha_{i,n}). \end{aligned}$$

Consequently,  $\alpha_{i,n+1} = Q^M(n+1, 1)$  is independent of  $i$  and thus  $\alpha_{i,n+1} = \alpha_{1,n+1}$  for all  $i$  and  $n$ .

As for the converse, since  $i \mapsto \alpha_{i,n}$  is constant, it is clear that (2.1) implies that  $(M_n)$  is Markov. ■

**Remark 2.2.** The one-dimensional memory process  $(M_n)_{n \in \mathbb{N}}$  could be replaced by a  $K$ -dimensional process. For each state  $a_i$ , define

$$\mathcal{M}_n^{(a_i)} = \inf\{0 \leq k \leq n; X_{n-k} \neq a_i\}.$$

There is a one-to-one correspondence between  $(X_n, M_n)$  and  $(\mathcal{M}_n^{(a_1)}, \dots, \mathcal{M}_n^{(a_K)})$ . Consequently the vector memory  $(\mathcal{M}_n^{(a_1)}, \dots, \mathcal{M}_n^{(a_K)})$  is a Markov chain. For instance  $(X_n, M_n) = (a_3, 4)$  corresponds to  $(\mathcal{M}_n^{(a_1)}, \dots, \mathcal{M}_n^{(a_K)}) = (0, 0, 4, 0, \dots, 0)$ . Indeed if the  $k^{\text{th}}$  coordinate of the vector does not vanish then  $X_n = a_k$ . This permits to recover  $X_n$  via  $(\mathcal{M}_n^{(a_1)}, \dots, \mathcal{M}_n^{(a_K)})$ , as for  $M_n$ , we have  $M_n = \mathcal{M}_n^{X_n}$ .

### 2.2.2 Invariant probability measure for $(X_n, M_n)_{n \geq 0}$

Let us now investigate the existence of an invariant probability measure. It is convenient to introduce for all  $1 \leq i \leq K$ ,

$$\Theta_i := \sum_{n \geq 1} \prod_{k=1}^{n-1} (1 - \alpha_{i,k}), \quad (2.7)$$

and for  $m \geq 1$ ,

$$\mathcal{P}_i(m) := \prod_{k=1}^{m-1} (1 - \alpha_{i,k}), \quad (2.8)$$

with the convention  $\prod_1^0 = 1$ .

$\mathcal{P}_i(k+1)$  represents the conditional probability that the process  $(X_n)$  stays at least a time interval of length  $k$  in the same state  $i$

$$\mathcal{P}_i(k+1) = \mathbb{P}(X_1 = \dots = X_k = i | X_0 = i, M_0 = 1).$$

**Proposition 2.3.** Let us denote  $\mathcal{P} = (p_{i,j})$  a given irreducible transition matrix.

(i) Then the Markov chain  $(X_n, M_n)_{n \geq 0}$  with transition probabilities defined by (2.1) and (2.2) admits a invariant probability measure  $\nu$  on the space  $\{a_1, \dots, a_K\} \times \overline{\mathbb{N}}^*$  if and only if  $\Theta_1, \dots, \Theta_K$  defined by (2.7) are all finite. This invariant probability measure is unique.

(ii) Moreover, if we denote by  $v^*$  the unique positive vector associated with the largest eigenvalue of  ${}^t\mathcal{P} = (p_{j,i})$  by Frobenius's theorem, then  $\nu(a_i, \infty) = 0$  for all  $1 \leq i \leq K$  and  $n \geq 1$ ,

$$\nu(a_i, n) = \frac{v_i^*}{\langle \Theta, v^* \rangle} \mathcal{P}_i(n)$$

where  $\Theta = {}^t(\Theta_1, \dots, \Theta_K)$  and  $\langle \Theta, v^* \rangle = \sum_{i=1}^K \Theta_i v_i^*$ .

**Remark 2.4.** The invariant measure  $\nu$  can be decomposed in the following way: for  $1 \leq i \leq K$  and  $n \geq 1$ ,

$$\nu(a_i, n) = \nu^X(a_i) \nu_i(n), \quad (2.9)$$

where

$$\nu^X(a_i) = \frac{\Theta_i v_i^*}{\langle \Theta, v^* \rangle} \quad \text{and} \quad \nu_i(n) = \frac{\mathcal{P}_i(n)}{\Theta_i}.$$

If  $(X_0, M_0) \sim \nu$ , then, for any  $n \geq 1$ ,  $\nu^X$  is the law of  $X_n$ , and  $\nu_i$  is the conditional distribution of  $M_n$ , given  $X_n = i$ .

Let us consider the particular case when for all  $1 \leq i \leq K$  and  $n \geq 1$ ,

$$1 - \alpha_{i,n} = \frac{\rho_i}{n}, \quad \text{with } \rho_i > 0.$$



After straightforward calculations, we obtain  $\Theta_i = e^{\rho_i}$  and

$$\nu^X(a_i) = \frac{v_i^* e^{\rho_i}}{\langle \Theta, v^* \rangle}, \quad \nu_i(n) = \frac{\rho_i^{n-1}}{(n-1)!} e^{-\rho_i}.$$

In other words, if  $(X_0, M_0) \sim \nu$  then the distribution of the couple  $(X_n, M_n)$  can be described as follows:  $X_n$  is chosen first with the probability  $\nu^X$  and afterwards, conditionally on  $X_n = a_i$ ,  $M_n$  is Poisson distributed with parameter  $\rho_i$ .

PROOF OF PROPOSITION 2.3.

For notational simplicity, we shall fix  $a_i = i$  for all  $1 \leq i \leq K$ .

**Step 1 — Invariant measure:** Let  $\nu$  be a non-negative measure. Since  $(X_n, M_n)$  is valued in the state space  $\{1, 2, 3, \dots, K\} \times \bar{\mathbb{N}}^*$ ,  $\nu$  is an invariant measure if and only if

$$\begin{aligned} \nu(i, k) &= \sum_{\ell \geq 1} \left\{ \nu(i, \ell) Q\left((i, \ell), (i, k)\right) + \sum_{j \neq i} \nu(j, \ell) Q\left((j, \ell), (i, k)\right) \right\} \\ &= \nu(i, k-1)(1 - \alpha_{i, k-1}) \mathbb{1}_{\{k > 1\}} + \mathbb{1}_{\{k=1\}} \sum_{j \neq i} p_{j, i} \sum_{\ell \geq 1} \nu(j, \ell) \alpha_{j, \ell}, \end{aligned} \quad (2.10)$$

for any  $1 \leq i \leq K$ , and

$$\nu(i, \infty) = \nu(i, \infty)(1 - \alpha_{i, \infty}). \quad (2.11)$$

Obviously (2.11) implies that  $\nu(i, \infty) = 0$  for all  $1 \leq i \leq K$ . Relation (2.10), with  $k \geq 2$  is equivalent to  $\nu(i, k) = \nu(i, k-1)(1 - \alpha_{i, k-1})$  which implies for  $k \geq 2$

$$\nu(i, k) = \nu(i, 1) \prod_{r=1}^{k-1} (1 - \alpha_{i, r}) = \nu(i, 1) \mathcal{P}_i(k). \quad (2.12)$$

The particular situation  $k = 1$  in (2.10) and (2.12) leads to

$$\nu(i, 1) = \sum_{j \neq i} p_{j, i} \left( \sum_{\ell \geq 1} \alpha_{j, \ell} \mathcal{P}_j(\ell) \right) \nu(j, 1). \quad (2.13)$$

Using (2.12) and (2.7) we get:

$$\sum_{1 \leq i \leq K, n \geq 1} \nu(i, n) = \sum_{i=1}^K \nu(i, 1) \sum_{n \geq 1} \mathcal{P}_i(n) = \sum_{i \geq 1} \nu(i, 1) \Theta_i.$$

Finally  $\nu$  is a probability measure iff  $\nu(i, k)$  is given by (2.12) for any  $k \geq 1$ , the vector  ${}^t(\nu(1, 1), \dots, \nu(K, 1))$  solves (2.13) and

$$\sum_{i=1}^K \nu(i, 1) \Theta_i = 1. \quad (2.14)$$

**Step 2 — Necessary condition:** Assume that

$$\Theta_i < \infty, \quad \forall i \in \{1, \dots, K\}. \quad (2.15)$$

Writing  $\alpha_{j, l} = -(1 - \alpha_{j, l}) + 1$ , we develop the expression (2.13) using (2.15):

$$\begin{aligned} \nu(i, 1) &= \sum_{j \neq i} p_{j, i} \left\{ \sum_{\ell \geq 1} \mathcal{P}_j(\ell) - \mathcal{P}_j(\ell + 1) \right\} \nu(j, 1) \\ &= \sum_{j \neq i} p_{j, i} \nu(j, 1). \end{aligned}$$

The vector  $v := {}^t(\nu(1, 1), \nu(2, 1), \dots, \nu(K, 1))$  satisfies

$$v = {}^t\mathcal{P}v, \quad \text{with } \mathcal{P} = (p_{i,j}). \quad (2.16)$$

Let  $v^*$  be the unique positive vector associated with the largest eigenvalue of  ${}^t\mathcal{P} = (p_{j,i})$  by Frobenius' theorem, then there exists  $\lambda > 0$  such that

$$(\nu(1, 1), \nu(2, 1), \dots, \nu(K, 1)) = \lambda {}^t v^*.$$

Using (2.14) we deduce:

$$\sum_{i=1}^K \nu(i, 1) \Theta_i = \lambda \sum_{i=1}^K \Theta_i v_i^* = \lambda \langle \Theta, v^* \rangle.$$

Hence  $\lambda = 1/\langle \Theta, v^* \rangle$  and by (2.12),  $\nu$  is determined by  $\nu(i, n) = \frac{v_i^*}{\langle \Theta, v^* \rangle} \mathcal{P}_i(n)$ , which gives existence and unicity of  $\nu$ .

**Step 3 — Sufficient condition:** Conversely let us assume the existence of an invariant probability measure  $\nu$ . We shall prove (2.15). Obviously (2.13) implies that if  $\nu(i, 1) = 0$  for some  $i$ , then  $\nu(j, 1) = 0$  for all  $j$ . Therefore  $\nu = 0$  which contradicts the fact that  $\nu$  is a probability measure. Hence  $\nu(i, 1) > 0$  for all  $1 \leq i \leq K$ . It is clear that (2.14) implies (2.15).  $\blacksquare$

Since the Markov chain  $(X_n, M_n)_{n \in \mathbb{N}}$  admits an invariant probability measure, we can extend its definition to  $\mathbb{Z}$  (instead of  $\mathbb{N}$ ) such that it is stationary. This extension will be useful to connect with certain Variable Length Markov Chains (defined later in Section 3).

**Remark 2.5.** *Since  $\nu$  is the invariant probability measure of the Markov chain  $(X_n, M_n)_{n \in \mathbb{Z}}$  then  $(X_{-n}, M_{-n})_{n \in \mathbb{Z}}$  is a Markov chain with invariant probability measure  $\nu$  and transition probabilities  $\widehat{Q}$  where:*

$$\nu(x)Q(x, y) = \nu(y)\widehat{Q}(y, x), \quad \forall x, y \in \{a_1, \dots, a_K\} \times \mathbb{N}^*.$$

From (2.1) and Proposition 2.3 we easily obtain

$$\begin{cases} \widehat{Q}((a_i, n+1), (a_i, n)) = 1, & \forall i \in \{1, \dots, K\}, n \geq 1, \\ \widehat{Q}((a_j, 1), (a_i, n)) = \frac{v_i^*}{v_j^*} p_{i,j} \alpha_{i,n} \mathcal{P}_i(n), & i \neq j, n \geq 1. \end{cases}$$

### 2.2.3 Paths description of $X$

From now on, for notational simplicity, we only consider the case  $K = 2$ . The trajectory  $n \mapsto X_n$  is determined as soon as the transition times between the different states are known. Let us define  $T_0 = 0$  and the sequence of stopping times for  $n \geq 1$ ,

$$T_n = \inf \left\{ i \geq T_{n-1} : X_i \neq X_{T_{n-1}} \right\}. \quad (2.17)$$

**Proposition 2.6.** *(i) Let us assume that  $\Theta_1$  and  $\Theta_2$  defined by (2.7) are finite. Then the random variables  $(T_{n+1} - T_n)_{n \geq 1}$  are almost surely finite and independent.*

(ii) (a) If  $X_0 = a_2$  and  $M_0 = m \geq 1$ . Then for all  $i \geq 1$  and  $n \geq 1$ ,

$$\mathbb{P}(T_1 = i) = \alpha_{2,m+i-1} \prod_{j=m}^{m+i-2} (1 - \alpha_{2,j}), \quad (2.18)$$

and

$$\mathbb{P}(T_{2n+1} - T_{2n} = i) = \alpha_{2,i} \mathcal{P}_2(i), \quad (2.19)$$

$$\mathbb{P}(T_{2n} - T_{2n-1} = i) = \alpha_{1,i} \mathcal{P}_1(i). \quad (2.20)$$

(b) If  $X_0 = a_1$  and  $M_0 = m \geq 1$  then (2.18) and (2.19) (resp. (2.20)) are still valid after replacing  $(\alpha_{2,\bullet})$  by  $(\alpha_{1,\bullet})$  (resp.  $(\alpha_{1,\bullet})$  by  $(\alpha_{2,\bullet})$ ).

**Remark 2.7.** 1. Note that, if  $X_0 = a_2$  and  $M_0 = m$  then for all  $n \geq 1$ ,

$$\mathbb{P}(T_1 \geq i) = \prod_{j=m}^{m+i-2} (1 - \alpha_{2,j}), \quad \mathbb{P}(T_{2n+1} - T_{2n} \geq i) = \mathcal{P}_2(i)$$

and  $\mathbb{P}(T_{2n} - T_{2n-1} \geq i) = \mathcal{P}_1(i)$ .

2. Between two consecutive jump times, the memory increases linearly

$$M_{T_n+t} = 1 + t, \quad 0 \leq t < T_{n+1} - T_n, \quad n \geq 1. \quad (2.21)$$

PROOF OF PROPOSITION 2.6

Let us consider  $X_0 = a_2$  and  $M_0 = m$ . Then

$$\mathbb{P}(T_1 = i) = \mathbb{P}(X_1 = a_2, X_2 = a_2, \dots, X_{i-1} = a_2, X_i = a_1).$$

Using the Markov property, we deduce

$$\mathbb{P}(T_1 = i) = \prod_{j=m}^{m+i-2} Q((a_2, j), (a_2, j+1)) Q((a_2, m+i-1), (a_1, 1)).$$

Equation (2.18) is therefore a direct consequence of (2.1).

Using  $\alpha_{2,m+i-1} = 1 - (1 - \alpha_{2,m+i-1})$ , it is easy to deduce that

$$\sum_{i \geq 1} \alpha_{2,m+i-1} \prod_{j=m}^{m+i-2} (1 - \alpha_{2,j}) = 1.$$

This shows that  $\mathbb{P}(T_1 < \infty) = 1$ .

Moreover, conditioning by  $X_0 = a_2$  and  $M_0 = m$ , for  $j \geq 1$  and  $i \geq 1$ , one has

$$\begin{aligned} \mathbb{P}(T_2 - T_1 = j, T_1 = i) &= \mathbb{P}(X_1 = \dots = X_{i-1} = a_2, X_i = \dots = X_{j+i-1} = a_1, X_{j+i} = a_2) \\ &= \alpha_{2,m+i-1} \prod_{j=m}^{m+i-2} (1 - \alpha_{2,j}) \prod_{\ell=1}^{j-1} (1 - \alpha_{1,\ell}) \alpha_{1,j} \\ &= \mathbb{P}(T_1 = i) \mathbb{P}(T_2 - T_1 = j), \end{aligned}$$

which leads to the independence between  $T_1$  and  $T_2 - T_1$ . The independence of  $T_3 - T_2$  and  $(T_1, T_2 - T_1)$  can be proved similarly. The proof of (ii) (a) of Proposition 2.6 follows by induction. The proof for (ii) (b) is analog.  $\blacksquare$

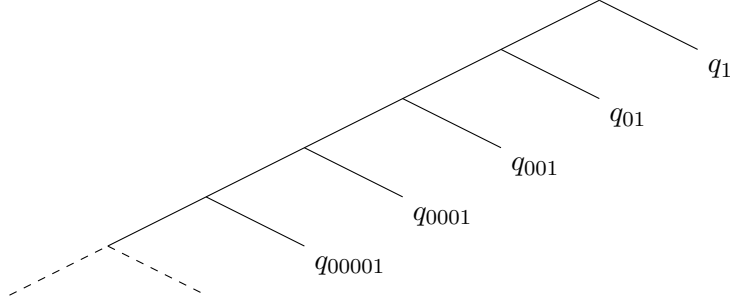


Figure 2: Infinite simple comb probabilized context tree.

**Remark 2.8.** Suppose that  $(X_n)$  takes its values in  $\{a_1, \dots, a_K\}$  and for simplicity  $X_0 = a_1$ ,  $M_0 = 1$ . Similarly to Proposition 2.6, we can prove:

$$\mathbb{P}(T_1 \geq i) = \mathcal{P}_1(i), \quad i \geq 1 \quad (2.22)$$

where the stopping time  $T_1$  and  $\mathcal{P}_1$  are defined by (2.17) and (2.8) respectively.

Recall that it has been supposed that the Markov chain associated with the transition matrix  $\mathcal{P}$  is irreducible and is therefore positive recurrent. It can be easily proved that the Markov chain  $(X_n, M_n)$  is recurrent if and only if

$$\lim_{m \rightarrow \infty} \mathcal{P}_i(m) = 0, \quad \forall i \in \{1, \dots, K\}. \quad (2.23)$$

Moreover since the positive recurrence of  $(X_n, M_n)$  is equivalent to the existence of an invariant measure, Proposition 2.3 yields the following necessary and sufficient condition:  $(X_n, M_n)$  is positive recurrent if and only if  $\Theta_1, \dots, \Theta_K$  are finite. This is coherent with the easy calculation

$$\mathbb{E}(T_1) = \sum_{i \geq 1} \mathbb{P}(T_1 \geq i) = \sum_{i \geq 1} \mathcal{P}_1(i) = \Theta_1.$$

### 3 The variable length Markov Chain $(U_n)_{n \geq 0}$

In this section, the relation between the Markov chain  $(X_n, M_n)_{n \in \mathbb{Z}}$  valued in  $\{0, 1\} \times \overline{\mathbb{N}}^*$  and the VLMC  $(U_n)_{n \geq 0}$  introduced in Section 1 is highlighted by the Theorems 3.1 and 3.3. For two very particular variable length Markov chains, we prove that these two models are equivalent. We consider two cases of VLMC for two specific context trees: the simple infinite comb and the double infinite comb.

From now on and until the end of this paper, for the sake of simplicity, we only consider the case  $K = 2$ .

#### 3.1 The simple infinite comb

Let us consider the alphabet  $\{a_1, a_2\}$  with  $a_1 = 0$  and  $a_2 = 1$ . We associate with a Markov chain of type  $(X_n, M_n)_{n \in \mathbb{Z}}$  defined in Section 2 a unique VLMC and vice versa. Abusing words, this VLMC is called the *infinite comb*. We refer to [2] for a complete definition. It is proved in [2] that in the irreducible case *i.e.* when  $q_{0^\infty}(0) \neq 1$ ,  $(U_n)_{n \geq 0}$  has a unique stationary probability measure  $\pi$  on the set of left-infinite words  $\mathcal{L}$  if and only if  $\Theta_1$  is finite. Similarly, if  $\Theta_1 < \infty$ , Proposition 2.3 implies that  $(X_n, M_n)_{n \in \mathbb{Z}}$  has a unique invariant probability measure. The following theorem enlightens the links between the VLMC  $(U_n)_{n \geq 0}$  and the chain  $(X_n, M_n)_{n \in \mathbb{Z}}$  and their respective stationary probability measure.

**Theorem 3.1 (infinite comb).**

- (i) Let  $(X_n, M_n)_{n \in \mathbb{Z}}$  be a stationary Markov chain valued in  $\{0, 1\} \times \overline{\mathbb{N}}^*$ , with transition probabilities (2.1) and (2.2), with  $p_{1,2} = p_{2,1} = 1$ . We suppose  $\Theta_1 < \infty$  (where  $\Theta_1$  is defined in (2.7)) and  $\forall n \in \overline{\mathbb{N}}^*$ ,

$$\alpha_{2,n} = \alpha_2. \quad (3.1)$$

We define for all  $n \in \mathbb{N}$ ,

$$U_n = \dots X_{n-2} X_{n-1} X_n. \quad (3.2)$$

Then,  $(U_n)_{n \geq 0}$  is a stationary variable length Markov chain associated with the infinite comb with

$$q_1(0) = \alpha_2, \quad q_{0^n 1}(1) = \alpha_{1,n}, \quad q_{0^\infty}(1) = \alpha_{1,\infty}. \quad (3.3)$$

The initial distribution is given by  $U_0 \stackrel{(d)}{=} \dots X_{-2} X_{-1} X_0$ .

- (ii) Conversely consider a stationary VLMC  $(U_n)_{n \geq 0}$  satisfying (3.3). For  $n \geq 0$ , define  $X_n$  as the last letter of  $U_n$  and  $(M_n)_{n \geq 0}$  as in (2.4). Then  $\Theta_1 < \infty$  (where  $\Theta_1$  is defined in (2.7)) and  $(X_n, M_n)_{n \geq 0}$  is a stationary Markov chain with transitions (2.1), (2.2) and (3.3) and initial data  $(X_0, M_0)$ . A stationary Markov chain  $(X_n, M_n)_{n \in \mathbb{Z}}$  can therefore be defined using the classical procedure of extension from  $\mathbb{N}$  to  $\mathbb{Z}$ .

The following tabular resumes the correspondence between these two models and could be considered as a dictionary (in the case:  $a_1 = 0$  and  $a_2 = 1$ ).

$(X_n, M_n)_{n \in \mathbb{Z}}$	$(U_n)_{n \geq 0}$
$\nu$	$\pi$
$Q((a_i, k), (a_i, k+1)) = 1 - \alpha_{i,k}$	$q_{a_i^k a_j}(a_i)$ for $j \neq i$
$Q((a_i, k), (a_j, 1)) = \alpha_{i,k}$	$q_{a_i^k a_j}(a_j)$

PROOF.

- (i) Due to Definition (3.2) of the process  $(U_n)_{n \geq 0}$ , for all  $s \in \{0, 1\}$ , the events  $\{U_{n+1} = U_n s\}$  and  $\{X_{n+1} = s\}$  are equal. Therefore  $(U_n)_{n \geq 0}$  is a Markov chain as soon as

$$\delta_{s,u} := \mathbb{P}(U_{n+1} = U_n s | U_n = u) = \mathbb{P}(X_{n+1} = s | X_n = u_0, \dots, X_{n-k} = u_{-k}, \dots)$$

only depends on  $s \in \{0, 1\}$  and  $u$ , where  $u = \dots u_{-1} u_0 \in \{0, 1\}^{-\mathbb{N}}$ .

Suppose first that  $u_0 = 1$ . Since  $M_n \in \mathbb{N}^*$ ,  $(\alpha_{2,n})_{n \geq 1}$  is constant and  $\overleftarrow{\text{pref}}(u) = 1$ , then (2.1) and (3.3) imply that

$$\delta_{s,u} = (1 - \alpha_2) \mathbb{1}_{\{s=1\}} + \alpha_2 \mathbb{1}_{\{s=0\}} = q_1(s).$$

Let us now consider the case  $u_0 = 0$ . Recall (see Proposition 2.3) that  $M_n \in \mathbb{N}^*$ . Consequently, there exists  $m \in \mathbb{N}^*$  such that  $u = \dots 10^m$ . Then  $M_n = m$ ,  $\overleftarrow{\text{pref}}(U_n) = 0^m 1$  and

$$\mathbb{P}(X_{n+1} = s | X_n = 0, M_n = m, \dots) = (1 - \alpha_{1,m}) \mathbb{1}_{\{s=0\}} + \alpha_{1,m} \mathbb{1}_{\{s=1\}} = q_{0^m 1}(s).$$

Next, we prove that  $(U_n)_{n \geq 0}$  is stationary. Note that (3.2) yields  $U_n = \psi((X_{n-i})_{i \geq 0})$  a.s. where  $\psi((x_{-n})_{n \geq 0}) = \dots x_{-2} x_{-1} x_0$ . Therefore, for any  $\ell \geq 0$ ,

$$\mathbb{E} \left[ f \left( U_{n+1-\ell}, \dots, U_{n+1} \right) \right] = \mathbb{E} \left[ f \left( \psi((X_{n+1-\ell-i})_{i \geq 0}), \dots, \psi((X_{n+1-i})_{i \geq 0}) \right) \right].$$

Since  $(X_n, M_n)_{n \in \mathbb{Z}}$  is stationary, then  $(X_{m+1-i})_{i \geq 0} \stackrel{(d)}{=} (X_{m-i})_{i \geq 0}$  for any  $m \in \mathbb{Z}$ . This implies that  $(U_n)_{n \geq 0}$  is stationary.

- (ii) Let us now assume that  $(U_n)_{n \geq 0}$  is a stationary VLMC. Let  $x, x' \in \{0, 1\}$ ,  $k, k' \geq 1$ ,  $n \in \mathbb{N}$  and

$$\delta' := \mathbb{P}(X_{n+1} = x', M_{n+1} = k' | X_n = x, M_n = k, \dots).$$

Then

$$\begin{aligned} \delta' &= \mathbb{1}_{\left\{\frac{k'=1}{x \neq x'}\right\}} \mathbb{P}(U_{n+1} = U_n x' | U_n = \dots x' x^k) + \mathbb{1}_{\left\{\frac{k'=k+1}{x=x'}\right\}} \mathbb{P}(U_{n+1} = U_n x | U_n = \dots (1-x) x^k) \\ &= \mathbb{1}_{\{k'=1\}} \left[ \mathbb{1}_{\left\{\frac{x=0}{x'=1}\right\}} \mathbb{P}(U_{n+1} = U_n 1 | U_n = \dots 1 0^k) + \mathbb{1}_{\left\{\frac{x=1}{x'=0}\right\}} \mathbb{P}(U_{n+1} = U_n 0 | U_n = \dots 0 1^k) \right] \\ &\quad + \mathbb{1}_{\{k'=k+1\}} \left[ \mathbb{1}_{\{x=x'=1\}} \mathbb{P}(U_{n+1} = U_n 1 | U_n = \dots 0 1^k) + \mathbb{1}_{\{x=x'=0\}} \mathbb{P}(U_{n+1} = U_n 0 | U_n = \dots 1 0^k) \right] \\ &= \mathbb{1}_{\{k'=1\}} \left[ \mathbb{1}_{\left\{\frac{x=0}{x'=1}\right\}} q_{0^k 1}(1) + \mathbb{1}_{\left\{\frac{x=1}{x'=0}\right\}} q_1(0) \right] + \mathbb{1}_{\{k'=k+1\}} \left[ \mathbb{1}_{\{x=x'=1\}} q_1(1) + \mathbb{1}_{\{x=x'=0\}} q_{0^k 1}(0) \right]. \end{aligned}$$

Using (3.3) we get

$$\delta' = \mathbb{1}_{\{k'=1\}} \left[ \mathbb{1}_{\left\{\frac{x=0}{x'=1}\right\}} \alpha_{1,k} + \mathbb{1}_{\left\{\frac{x=1}{x'=0}\right\}} \alpha_2 \right] + \mathbb{1}_{\{k'=k+1\}} \left[ \mathbb{1}_{\{x=x'=1\}} (1 - \alpha_2) + \mathbb{1}_{\{x=x'=0\}} (1 - \alpha_{1,k}) \right].$$

Then (2.1) follows directly with  $\alpha_{2,n} = \alpha_2$ .  $\blacksquare$

The following result is a corollary of Proposition 2.3. It enables us to compare the expression of the invariant measure  $\nu$  from the model  $(X_n, M_n)_{n \in \mathbb{Z}}$  with the invariant measure  $\pi$  for the VLMC *infinite comb* (see Section B in the Appendix for notations about VLMC).

**Corollary 3.2.** *Under the condition  $\Theta_1 < \infty$ , there exists a unique invariant probability measure  $\nu$  for the Markov chain  $(X_n, M_n)$  given by  $\nu(a_1, \infty) = \nu(a_2, \infty) = 0$  and for all  $m \geq 1$ ,*

$$\nu(a_1, m) = \frac{1}{\Theta_1 + \Theta_2} \mathcal{P}_1(m) \quad \text{and} \quad \nu(a_2, m) = \frac{\alpha_2 (1 - \alpha_2)^{m-1}}{1 + \alpha_2 \Theta_1}, \quad (3.4)$$

where  $\Theta_1 = 1/\alpha_2$ . In particular one gets

$$\nu(a_2, \mathbb{N}^*) = \pi(a_2) = \frac{1}{1 + \alpha_2 \Theta_1}.$$

PROOF. Proposition 2.3 with

$$\mathcal{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad v^* = \frac{1}{2} (1, 1)$$

lead to (3.4) and

$$\nu(a_2, \mathbb{N}^*) = \sum_{m \geq 1} \nu(a_2, m) = \frac{1}{1 + \alpha_2 \Theta_1}.$$

Consequently one has

$$\nu(a_2, \mathbb{N}^*) = \frac{1}{1 + q_1(0) \sum_{n \geq 1} \prod_{k=1}^{n-1} (1 - q_{0^k 1}(1))} = \frac{1}{\sum_{n \geq 0} \prod_{k=0}^{n-1} q_{0^k 1}(0)} = \pi(a_2),$$

which is fortunately (!) the invariant measure obtained in [2].  $\blacksquare$

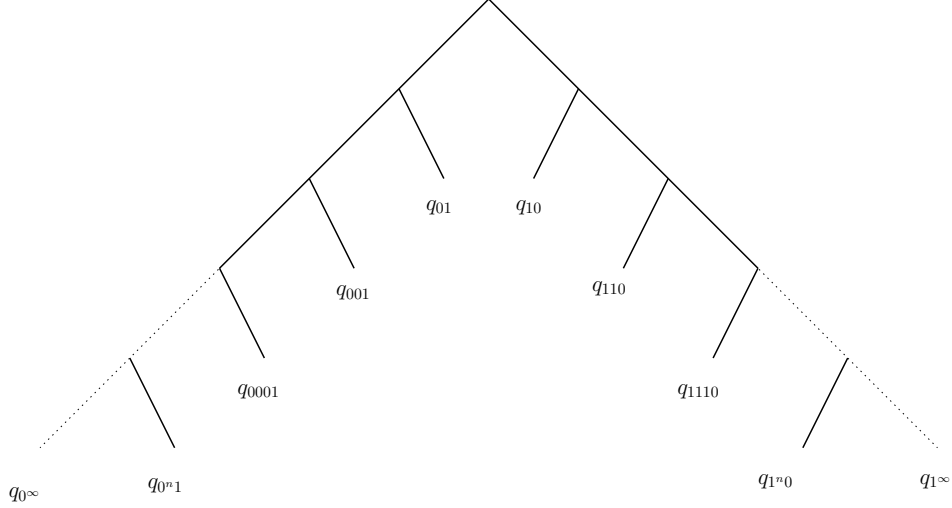


Figure 3: infinite double comb probabilized context tree.

### 3.2 The double infinite comb

Let us now present the double infinite comb. Consider the probabilized context tree given on 3 (hereafter called double infinite comb). In this case, there are two infinite leaves  $0^\infty$  and  $1^\infty$  and countably many finite leaves  $0^n 1$  and  $1^n 0$ ,  $n \in \mathbb{N}$ , so that

$$\mathcal{C} = \{0^n 1, n \geq 1\} \cup \{1^n 0, n \geq 1\} \cup \{0^\infty\} \cup \{1^\infty\}.$$

The data of a corresponding VLMC consists thus in Bernoulli probability measures on  $\{0, 1\}$ :

$$q_{0^\infty}, q_{1^\infty}, \text{ and } q_{0^n 1}, q_{1^n 0}, n \in \mathbb{N}^*.$$

We refer to Appendix B to see that the finiteness of  $\Theta_1$  and  $\Theta_2$  implies the existence of a unique invariant measure for this VLMC.

#### Theorem 3.3 (double infinite comb).

- (i) Let  $(X_n, M_n)_{n \in \mathbb{Z}}$  be a stationary Markov chain with transition probabilities (2.1) and (2.2). We suppose  $\Theta_1 < \infty$  and  $\Theta_2 < \infty$  (where  $\Theta_i$  is defined in (2.7)). Then the process  $(U_n)_{n \geq 0}$  defined by (3.2) is a stationary variable length Markov chain associated with the double infinite comb with

$$q_{1^n 0}(0) = \alpha_{2,n}, \quad q_{1^\infty}(0) = \alpha_{2,\infty}, \quad q_{0^n 1}(1) = \alpha_{1,n}, \quad q_{0^\infty}(1) = \alpha_{1,\infty}. \quad (3.5)$$

The initial data is given by  $U_0 = \dots X_{-2} X_{-1} X_0$ .

- (ii) Conversely let  $(U_n)_{n \geq 0}$  be a stationary VLMC satisfying (3.5). For  $n \geq 0$ , define  $X_n$  by the last letter of  $U_n$  and  $(M_n)_{n \geq 0}$  as in (2.4). Then  $(X_n, M_n)_{n \geq 0}$  is a stationary Markov chain with transitions (2.1), (2.2) and (3.5) and with initial data  $(X_0, M_0)$ . This stationary Markov chain can be extended on the time space  $\mathbb{Z}$  as usual.

The arguments for the proof are similar to those presented in Theorem 3.1.

As for the simple infinite comb, the invariant measure of the first margin of the Markov chain  $(X_n, M_n)_{n \in \mathbb{Z}}$  corresponding to the double infinite comb can be compared with the invariant measure  $\pi$  for the VLMC *double infinite comb* calculated in Appendix B.

**Corollary 3.4.** *Under the condition  $\Theta_1 < \infty$  and  $\Theta_2 < \infty$ , there exists a unique invariant probability measure  $\nu$  for the Markov chain  $(X_n, M_n)$  given by  $\nu(a_1, \infty) = \nu(a_2, \infty) = 0$  and for all  $m \geq 1$  and  $i = 1, 2$ ,*

$$\nu(a_i, m) = \frac{1}{\Theta_1 + \Theta_2} \mathcal{P}_i(m). \quad (3.6)$$

Consequently one gets

$$\nu(a_2, \mathbb{N}^*) = \frac{\Theta_2}{\Theta_1 + \Theta_2} = \pi(a_2).$$

PROOF. Again (3.6) is a direct consequence of Proposition 2.3. Suming up it comes

$$\nu(a_2, \mathbb{N}^*) = \frac{\Theta_2}{\Theta_1 + \Theta_2},$$

with

$$\Theta_1 = \sum_{n \geq 1} \prod_{k=1}^{n-1} (1 - q_{0^{k_1}}(1)) = \sum_{n \geq 0} \prod_{k=1}^n q_{0^{k_1}}(0)$$

and

$$\Theta_2 = \sum_{n \geq 1} \prod_{k=1}^{n-1} (1 - q_{1^{k_0}}(0)) = \sum_{n \geq 0} \prod_{k=1}^n q_{1^{k_0}}(1),$$

which is exactly the calculation of  $\pi(a_2)$  in Appendix B.

**Remark 3.5.** *The results developed in Theorem 3.1 and Theorem 3.3 can be generalized to context trees which are based on a finite alphabet  $\{a_1, \dots, a_K\}$  and composed with a finite number of combs. The corresponding Markov chain  $(X_n, M_n)_{n \in \mathbb{Z}}$  is then valued in the state space  $\{a_1, \dots, a_K\} \times \overline{\mathbb{N}}^*$ .*

Of particular interest are variable length Markov chains  $(U_n)$  associated with the infinite comb or the double infinite comb. While the sequence  $(X_n)$  formed by the last letters of the process  $U_n = \dots X_{n-1} X_n$  is not a Markov process, except for very particular  $q_c$ , the previous theorems show that it suffices to add a memory process  $(M_n)$  to get a Markov chain  $(X_n, M_n)$ . Note that  $(U_n)$  takes its value in the non-countable space  $\mathcal{L}$  and Theorems 3.1 and 3.3 allow to associate by a one to one correspondence a Markov chain  $(X_n, M_n)$  which is valued in the countable set  $\{0, 1\} \times \overline{\mathbb{N}}^*$ . This reduction of the size of the state space (which becomes here minimal) is made possible by the particular shape of the context tree: for instance, the VLMC associated with the bamboo blossom defined in [2] is not equivalent to a Markov Chain  $(X_n, M_n)$  with a real memory process. Nevertheless for suitable VLMC we suggest to introduce the following application

$$(U_n)_n \mapsto (\overleftarrow{\text{pref}}(U_n))_n,$$

which should permit to generalize the reduction of the state space. The image process is not Markovian in the general case, even under the stationary distribution for  $U_n$ . A conjecture: the process  $(\overleftarrow{\text{pref}}(U_n))_n$  is Markovian (and thus defines an automaton) if and only if the associated context tree has a completeness property, studied in a companion paper [3].

## 4 Distribution of the persistent random walk

By definition, a random walk  $(S_n)_{n \geq 0}$  is a process whose increments are independent. It is often pertinent, for instance in modeling, to begin with the increments and second to study the associated random walk. Let us give an example coming from finance. Suppose that  $S_n$



is the price at time  $n$  of an asset. In the Cox, Ross and Rubinstein model, the non-arbitrage condition implies that the relative increments  $\left(\frac{S_n - S_{n-1}}{S_{n-1}}; n \geq 1\right)$  are independent.

We study here a class of *additive* processes  $(S_n)$  of the type

$$S_n = \sum_{k=0}^n X_k, \quad n \geq 0, \quad (4.1)$$

where the increments  $(X_n)$  are not independent. A tentative of considering increments with short dependency has been already developed in [17] and [18]. In these studies, the authors have supposed that  $(X_n)$  is a Markov chain. We would like to go further here introducing *variable length memory* between the increments.

We consider in this section, a Markov chain  $(X_n, M_n)_{n \geq 0}$  with transition probability (2.1) and (2.2) and we assume

$$K = 2, a_1 = -1, a_2 = 1.$$

The process  $(S_n)$  defined by (4.1) is called a *persistent random walk*. This terminology comes from [8].

A path description of  $(S_n)$  is given in Section 4.1, putting ahead the breaking times  $(T_n)_{n \geq 1}$ . We give in Section 4.2 the explicit distribution of  $S_n$ . Although the law of  $S_n$  is complicated, we can determine explicitly the distribution and the generating function of the position of the persistent random walk at an exponential independent random time. The double generating function will play an important role in Section 5.

We end this section studying how  $S_n$  fluctuates as  $n \rightarrow \infty$ . Indeed it is not so far from the persistent walk with one-order Markovian increments. We prove a law of large number and a central limit theorem. We recover the classical setting where  $(X_n)$  is a Markov chain. We have introduced variable memory to  $(X_n)$ , but it seems that it is not sufficient to obtain new asymptotic behavior: it would be therefore very interesting to investigate the behaviour of the random walk when *mixing assumptions* are relaxed, *i.e.* when the length of the memory increases significantly to give a real *persistent memory effect to the random walks*.

## 4.1 Paths description

Since  $X_n$  is  $\{-1, 1\}$ -valued, it is clear that the trajectory of  $(S_n)_{n \geq 0}$  is a sequence of straight lines with slopes  $\pm 1$ , and the instants of breaks are  $(T_n)_{n \geq 1}$  which were introduced in (2.17).

Let us assume that  $S_0 = X_0 = 1$ , then the trajectory increases step 1 by step 1 till  $T_1 - 1$  where it reaches a first local maximum. After that time, it decreases and reaches a local minimum at time  $T_2 - 1$  and so on. The trajectory of  $(S_n)_{n \in \mathbb{N}}$  corresponds to the linear interpolation between the sequence of points  $(W_n, Z_n)_{n \geq 0}$  where  $W_0 = 0, Z_0 = 1$  and for  $n \geq 1$ ,

$$(W_n, Z_n) = (T_n - 1, S_{T_n} - (-1)^n) = \left( T_n - 1, \sum_{k=1}^n (-1)^{k-1} (T_k - T_{k-1}) \right).$$

If  $S_0 = X_0 = -1$ , then the behaviour of the process is similar and reduces on a succession of increasing and decreasing parts. The trajectory  $(S_t)_{t \geq 0}$  is a linear interpolation between  $(W_n, Z'_n)_{n \geq 0}$  where  $W_0 = 0, Z'_0 = -1$  and for  $n \geq 1$ ,

$$(W_n, Z'_n) = (T_n - 1, S_{T_n} + (-1)^n) = \left( T_n - 1, \sum_{k=1}^n (-1)^k (T_k - T_{k-1}) \right).$$

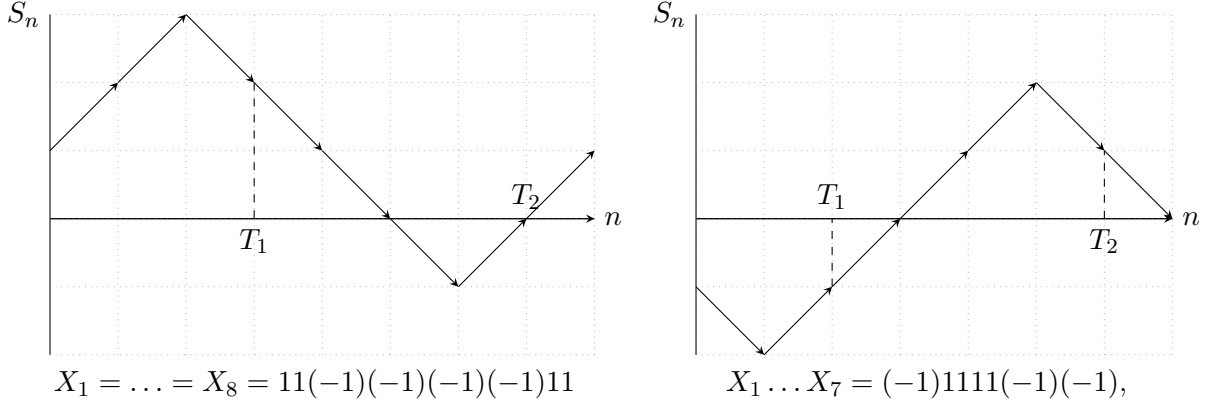


Figure 4: Trajectories of  $(S_t)_{t \geq 0}$  when either  $S_0 = 1$  or  $S_0 = -1$ .

Note that  $Z'_n = -Z_n$ .

Let us introduce the counting process  $(N_t)_{t \in \mathbb{N}}$  whose jump times are  $T_n$ :

$$N_t = \sup\{n \geq 1 : T_n \leq t\} = \sum_{n \geq 1} \mathbb{1}_{\{T_n \leq t\}}, \quad t \in \mathbb{N}. \quad (4.2)$$

From now on, we suppose that  $(X_0, M_0) = (1, 1)$ . Note that the case  $(X_0, M_0) = (-1, 1)$  can be deduced from the former case changing  $X$  in  $-X$ .

The counting process  $(N_t)_{t \geq 0}$  will play an important role in the study of  $(S_t)_{t \geq 0}$  (see Section 5) and  $(S_t)_{t \geq 0}$  can be expressed via  $(N_t)_{t \geq 0}$  as:

$$S_t = \sum_{n=0}^t (-1)^{N_n}. \quad (4.3)$$

There is a one-to-one correspondence between  $(M_t)_{t \geq 0}$  and  $(T_n)_{n \geq 0}$ :

$$\{k; M_k = 1\} = \{T_n; n \geq 0\}. \quad (4.4)$$

$\{N_s; s \leq t\}$  can be expressed via  $\{M_s; s \leq t\}$  and vice and versa. Indeed, (4.2) (4.4) and (2.4) imply

$$N_t = \sum_{k=1}^t \mathbb{1}_{\{M_k=1\}} \quad \text{and} \quad M_t = 1 + \sup\{n \geq 0 : N_{t-n} = N_t\}, \quad t \in \mathbb{N}. \quad (4.5)$$

## 4.2 Distribution of the persistent random walk at a fixed time

In this section we give the explicit distribution of the persistent random walk at any fixed time.

We recall that  $(X_n, M_n)_{n \geq 0}$  is a  $\{-1, 1\} \times \overline{\mathbb{N}}^*$ -valued Markov chain with transitions matrix  $Q$  defined by (2.1) and starting values  $(X_0, M_0) = (1, 1)$ . Therefore the law of  $(X_n, M_n)$  is given by  $Q^n$ . However the calculation of  $Q^n$  is untractable. This leads to restrict ourselves to the law of  $X_n$ .

Let us define

$$\mathcal{N}(m, b) := \left\{ u \in (\mathbb{N}^*)^m : u_1 + \dots + u_m = b \right\}, \quad m \geq 1, \quad b \geq 1.$$

and

$$A_i(m, b) = \sum_{u \in \mathcal{N}(m, b)} \mathcal{P}_i(u_1) \dots \mathcal{P}_i(u_m) \alpha_{i, u_1} \times \dots \times \alpha_{i, u_m}, \quad i = 1, 2 \quad (4.6)$$

with  $A_i(m, b) = 0$  for  $0 \leq b < m$  and  $A_i(0, b) = \mathbb{1}_{\{b=0\}}$ .

The distribution of the random walk  $(S_n)_{n \geq 1}$  can be directly linked to the occupation measure  $L_n(1)$  of the increments  $(X_n)_{n \geq 1}$  in the following way:

**Proposition 4.1. (distribution of  $S_n$ )** *Suppose that  $(X_0, M_0) = (1, 1)$ .*

(i) *Let us introduce the local time*

$$L_n(1) := \sum_{k=1}^n \mathbb{1}_{\{X_k=1\}}, \quad (4.7)$$

*then the random walk satisfies for  $n \geq 1$ ,*

$$S_n = 1 + 2L_n(1) - n. \quad (4.8)$$

*Consequently, for any  $0 \leq k \leq n$ :*

$$\eta_n(k) := \mathbb{P}(S_n = 1 + 2k - n) = \mathbb{P}(L_n(1) = k). \quad (4.9)$$

(ii) *Moreover, for  $0 \leq k \leq n$ , we have*

$$\eta_n(k) = \eta_n^{(1)}(k) + \eta_n^{(2)}(k) \quad (4.10)$$

*with*

$$\eta_n^{(1)}(k) = \sum_{1 \leq m \leq (k+1) \wedge (n-k)} A_2(m, k+1) \sum_{\ell=1}^{n-k-m+1} A_1(m-1, n-k-\ell) \mathcal{P}_1(\ell) \quad (4.11)$$

$$\eta_n^{(2)}(k) = \sum_{0 \leq m \leq k \wedge (n-k)} A_1(m, n-k) \sum_{\ell=1}^{k-m+1} A_2(m, k+1-\ell) \mathcal{P}_2(\ell). \quad (4.12)$$

PROOF.

(i) Using the definition of  $L_n(1)$ , it comes

$$\begin{aligned} S_n &= 1 + \sum_{i=1}^n \mathbb{1}_{\{X_i=1\}} - \sum_{i=1}^n \mathbb{1}_{\{X_i=-1\}} \\ &= 1 + \sum_{i=1}^n \mathbb{1}_{\{X_i=1\}} - \left( n - \sum_{i=1}^n \mathbb{1}_{\{X_i=1\}} \right) \\ &= 1 + 2L_n(1) - n. \end{aligned}$$

(ii) In order to compute  $\eta_n(k)$ , it is convenient to use the family of stopping times  $(T_n)$  introduced in (2.17). The probability of the event  $\{L_n(1) = k\}$  can be decomposed into two parts, according to the fact that time  $n$  arrives on the way up or on the way down:

$$\eta_n(k) = \sum_{m \geq 0} \eta_n^{(1)}(k, m) + \sum_{m \geq 1} \eta_n^{(2)}(k, m) \quad (4.13)$$

where

$$\eta_n^{(1)}(k, m) := \mathbb{P}\left(L_n(1) = k, T_{2m} \leq n < T_{2m+1}\right), \quad m \geq 0$$

and

$$\eta_n^{(2)}(k, m) := \mathbb{P}\left(L_n(1) = k, T_{2m-1} \leq n < T_{2m}\right), \quad m \geq 1.$$

**First step — Computation of  $\eta_n^{(1)}(k, m)$  for  $n \geq k$ .** Suppose first that  $m \geq 1$ . On the set  $\{L_n(1) = k, T_{2m} \leq n < T_{2m+1}\}$ , we define for  $0 \leq i < m$ , the length of the  $i^{\text{th}}$  ascent  $W_i := T_{2i+1} - T_{2i}$ ,  $W_m := n + 1 - T_{2m}$  and the length of the  $i^{\text{th}}$  descent  $V_i := T_{2i} - T_{2i-1}$  for  $1 \leq i \leq m$ . Then

$$W_0 + W_1 + \cdots + W_m + V_1 + \cdots + V_m = n + 1, \quad W_0 + W_1 + \cdots + W_m = k + 1. \quad (4.14)$$

Therefore for  $\underline{W} = (W_0, \dots, W_m)$  and  $\underline{V} = (V_1, \dots, V_m)$  we get

$$\eta_n^{(1)}(k, m) = \sum_{w \in \mathcal{N}(m+1, k+1)} \sum_{v \in \mathcal{N}(m, n-k)} \mathbb{P}(\underline{W} = w, \underline{V} = v). \quad (4.15)$$

Using the distributions of  $T_{2i+1} - T_{2i}$  and  $T_{2i+2} - T_{2i+1}$  given in Proposition 2.6, we obtain

$$\begin{aligned} \mathbb{P}(\underline{W} = w, \underline{V} = v) &= \mathcal{P}_2(w_1) \alpha_{2, w_1} \mathcal{P}_1(v_1) \alpha_{1, v_1} \times \cdots \\ &\quad \times \mathcal{P}_2(w_m) \alpha_{2, w_m} \mathcal{P}_1(v_m) \alpha_{1, v_m} \mathcal{P}_2(u_{m+1}). \end{aligned} \quad (4.16)$$

It is clear that (4.15) and (4.16) imply

$$\eta_n^{(1)}(k, m) = \widehat{A}_2(m+1, k+1) A_1(m, n-k), \quad (4.17)$$

where  $A_i$  is defined by (4.6) and for  $m \geq 2$  and  $i \in \{1, 2\}$ ,

$$\widehat{A}_i(m, b) := \sum_{w \in \mathcal{N}(m, b)} \mathcal{P}_i(w_1) \times \cdots \times \mathcal{P}_i(w_m) \alpha_{i, w_1} \times \cdots \times \alpha_{i, w_{m-1}}, \quad (4.18)$$

and  $\widehat{A}_i(1, b) = \mathcal{P}_i(b)$ .

If  $m = 0$ , then  $n = k$ ,  $\eta_n^{(1)}(k, m) = \mathcal{P}_2(n+1)$ . Therefore (4.17) holds with  $m = 0$ .

**Step 2 — Computation of  $\eta_n^{(2)}(k, m)$ .** Similarly, define on  $\{L_n(1) = k, T_{2m-1} \leq n < T_{2m}\}$ ,  $W_i := T_{2i+1} - T_{2i}$  for  $0 \leq i < m$ ,  $V_i := T_{2i} - T_{2i-1}$  for  $1 \leq i < m$  and  $V_m := n + 1 - T_{2m-1}$  then:

$$W_0 + W_1 + \cdots + W_{m-1} + V_1 + \cdots + V_m = n + 1, \quad W_0 + W_1 + \cdots + W_{m-1} = k + 1.$$

For  $\underline{W} = (W_0, \dots, W_{m-1})$  and  $\underline{V} = (V_1, \dots, V_m)$  we get

$$\begin{aligned} \eta_n^{(2)}(k, m) &= \sum_{w \in \mathcal{N}(m, k+1)} \sum_{v \in \mathcal{N}(m, n-k)} \mathbb{P}(\underline{W} = w, \underline{V} = v) \\ &= A_2(m, k+1) \widehat{A}_1(m, n-k). \end{aligned} \quad (4.19)$$

Combining (4.13), (4.17) and (4.19) leads to

$$\eta_n(k) = \sum_{m=0}^{k \wedge (n-k)} \widehat{A}_2(m+1, k+1) A_1(m, n-k) + \sum_{m=1}^{(k+1) \wedge (n-k)} A_2(m, k+1) \widehat{A}_1(m, n-k) \quad (4.20)$$

In order to prove (4.10), it suffices to express  $\widehat{A}_i$  in terms of  $A_i$ . For  $b \geq m > 1$  we observe that

$$\mathcal{N}(m, b) = \left\{ (w, w_m) : w \in \mathcal{N}(m-1, j), w_m = b-j, m-1 \leq j \leq b-1 \right\}.$$

Hence, for  $b \geq m > 1$ ,

$$\widehat{A}_i(m, b) = \sum_{j=m-1}^{b-1} A_i(m-1, j) \mathcal{P}_i(b-j) = \sum_{\ell=1}^{b-m+1} A_i(m-1, b-\ell) \mathcal{P}_i(\ell). \quad (4.21)$$

Observe that (4.21) is still valid if  $m = 1$ , since  $A_i(0, b) = \mathbb{1}_{\{b=0\}}$  and  $\widehat{A}_i(1, b) = \mathcal{P}_i(b)$ .

The decomposition (4.21) permits to transform (4.20) into (4.10).  $\blacksquare$

**Remark 4.2.** *In the particular situation  $\alpha_{2,k} = \alpha_2$  for any  $k \geq 1$ , which is associated with the simple infinite comb (Section 3), then the distribution of  $S_n$  given by (4.9) and (4.10) can be simplified since*

$$A_2(m, b) = \binom{b-1}{m-1} (1-\alpha_2)^{b-m} \alpha_2^m = \alpha_2 \widehat{A}_2(m, b), \quad b \geq m \geq 1.$$

Of course by symmetry we get also a similar expression of  $A_1$  if  $\alpha_{1,k} = \alpha_1$  for any  $k \geq 1$ . Combining both identities, Proposition 4.1 gives the distribution of  $S_n$  when  $X_n$  is a Markov chain. Let us just note that the associated VLMC is very particular and the generating function of  $S_n$  was already presented in [17].

**Corollary 4.3.** *Suppose that  $\alpha_{1,k} = \alpha_1$  and  $\alpha_{2,k} = \alpha_2$  for any  $k \geq 1$ . This means that  $(X_n)$  is a  $\{-1, 1\}$ -valued Markov chain with transition matrix*

$$\begin{pmatrix} 1-\alpha_1 & \alpha_1 \\ \alpha_2 & 1-\alpha_2 \end{pmatrix}.$$

Then one has

$$\begin{aligned} \mathbb{P}(L_n(1) = k) &= \sum_{m=1}^{(k+1) \wedge (n-k)} \binom{k}{m-1} \binom{n-k-1}{m-1} \alpha_1^{m-1} (1-\alpha_1)^{n-k-m} \alpha_2^m (1-\alpha_2)^{k+1-m} \\ &+ \sum_{m=1}^{k \wedge (n-k)} \binom{k}{m} \binom{n-k-1}{m-1} \alpha_1^m (1-\alpha_1)^{n-k-m} \alpha_2^m (1-\alpha_2)^{k-m}. \end{aligned}$$

**Remark 4.4.** (i) *Note that we have actually proved a more complete result than (4.11) and (4.12):*

$$\mathbb{P}(L_n(1) = k, T_{2m} \leq n < T_{2m+1}) = \widehat{A}_2(m+1, k+1) A_1(m, n-k), \quad (4.22)$$

for  $0 \leq m \leq k \wedge (n-k)$  and

$$\mathbb{P}(L_n(1) = k, T_{2m-1} \leq n < T_{2m}) = A_2(m, k+1) \widehat{A}_1(m, n-k), \quad (4.23)$$

for  $1 \leq m \leq (k+1) \wedge (n-k)$ , where  $\widehat{A}_1$  and  $\widehat{A}_2$  are defined by (4.18).

(ii) *We deduce from (4.22) that*

$$\mathbb{P}(T_{2m} \leq n < T_{2m+1}) = \sum_{k=m}^{n-m} \widehat{A}_2(m+1, k+1) A_2(m, n-k).$$

Since the left hand side equals  $\mathbb{P}(T_{2m} \leq n, T_{2m+1} - T_{2m} > n - T_{2m})$ , Proposition 2.6 and Remark 2.7 imply

$$\mathbb{E}[\mathbb{1}_{\{T_{2m} \leq n\}} \mathcal{P}_2(n - T_{2m})] = \sum_{k=m}^{n-m} \widehat{A}_2(m+1, k+1) A_2(m, n-k).$$

Recall that  $T_{2m} \geq 2m$ . Then taking successively  $n = 2m$ ,  $n = 2m+1$  and so on, we are theoretically able to determine the law of  $T_{2m}$ .

As it is said in Remark 4.4, Proposition 4.1 contains in an hidden way the distribution of  $T_{2m}$  and  $T_{2m+1}$ . However it is actually possible to determine differently the distribution of these two random variables. It is convenient to introduce the notations:

$$\begin{aligned}\Delta g(n) &= g(n) - g(n+1), \quad n \geq 0, \\ \varphi * \psi(n) &= \sum_{k=0}^n \varphi(k)\psi(n-k) \quad n \geq 0, \\ \theta : \mathbb{N} &\rightarrow \mathbb{N} \quad \theta(n) = n+1,\end{aligned}$$

where  $g, \varphi, \psi : \mathbb{N} \rightarrow \mathbb{N}$ .

**Proposition 4.5.** *Let  $\xi_1, \dots, \xi_k$  be  $k$  independent  $\mathbb{N}$ -valued random variables. Denote for any  $n \geq 0$ ,*

$$f_i(n) := \mathbb{P}(\xi_i \geq n).$$

*We introduce  $\mathbb{A}_r^k$  the set of all subsets of  $\{1, \dots, k\}$  containing  $r$  elements. Then*

$$\mathbb{P}(\xi_1 + \dots + \xi_k \geq n) = h_k(n)$$

where

$$h_k = \sum_{r=1}^k \sum_{A \in \mathbb{A}_r^k} \Delta^{r-1}(f^{*A} \circ \theta^{k-r})$$

and  $f^{*A} = f_{i_1} * \dots * f_{i_r}$  when  $A = \{i_1, \dots, i_r\}$ .

We do not prove Proposition 4.5 since it does not play a main role in our study.

**Remark 4.6.** 1. *If  $\xi$  is geometrically distributed with parameter  $1 - \rho$  (i.e.  $\mathbb{P}(\xi = n) = (1 - \rho)\rho^n$ ,  $n \geq 1$ ,  $\rho \in ]0, 1[$ ) then the function  $f$  associated with  $\xi$  is  $f(n) = \rho^n$ ,  $n \geq 0$ .*

2. *Suppose that  $\xi_1 = T_{2m} - T_{2m-1}$  (resp.  $\xi_2 = T_{2m-1} - T_{2m-2}$ ) where  $m \geq 1$ , then Remark 2.7 implies that*

$$\mathbb{P}(\xi_1 - 1 \geq n) = \mathcal{P}_i(n+1), \quad i = 1, 2, \quad n \geq 0,$$

where  $\mathcal{P}_i$  has been defined by (2.8)

**Definition 4.7.** *Let  $\rho \in ]0, 1[$ . A  $\mathbb{N}$ -valued random variable  $\xi_\rho$  is said to be pseudo-Poisson distributed with parameter  $\rho > 0$  when for all  $n \geq 0$ :*

$$f_\rho(n) = \mathbb{P}(\xi_\rho \geq n) = \frac{\rho^n}{n!}.$$

It is clear that if  $\alpha_{i,k} = 1 - \frac{\rho_i}{k}$  where  $\rho_i \in ]0, 1[$ , then  $\mathcal{P}_i(n) = \frac{\rho_i^n}{n!}$ . Therefore  $\xi_i - 1$  (cf item 2. of Remark 4.6) is pseudo-Poisson with parameter  $\rho_i$ .

It is immediate to prove that:

$$f_\rho * f_{\rho'} = f_{\rho+\rho'}.$$

Reasoning by induction on  $k$  and using Proposition 4.5, we get the following result.

**Proposition 4.8.** *Suppose that  $\xi_1, \dots, \xi_k$  are independent, and  $\xi_i$  is pseudo-Poisson with parameter  $\rho_i$ . Then:*

$$\mathbb{P}(\xi_1 + \dots + \xi_k \geq n) = h_k(n), \quad n \geq 0,$$

where

$$h_k(n) = \sum_{r=1}^k \sum_{A \in \mathbb{A}_r^k} \sum_{\ell=0}^{r-1} \binom{r-1}{\ell} \frac{(-1)^\ell}{(n+k+\ell-r)!} \left( \sum_{i \in A} \rho_i \right)^{n+k+\ell-r}. \quad (4.24)$$

In the particular case  $\rho_1 = \dots = \rho_k = \rho$ ,

$$h_k(n) = k \sum_{t=0}^{k-1} \frac{\rho^{n+t}}{(n+t)!} \binom{k-1}{t} \left( \sum_{\ell=0}^t \binom{t}{\ell} (-1)^\ell (\ell + k - t)^{n-1+t} \right)$$

**Remark 4.9.** Suppose that  $\alpha_{1,k} = 1 - \frac{\rho}{k}$ ,  $k \geq 1$ . Then

$$\mathbb{P} \left( \sum_{i=1}^k (T_{2i} - T_{2i-1} - 1) \geq n \right) = h_k(n), \quad n \geq 0,$$

where  $h_k$  is given by (4.24).

### 4.3 Distribution of the persistent random walk at an independent time

As shows Proposition 4.1, the law of  $S_n$  is rather complicated. In the study of a Markov chain, it can be interesting to stop it at a random time. For instance, a Markov chain stopped at a geometric time independent from the Markov chain remains a Markov chain.

Let us consider a geometric random variable  $\tau + 1$  with parameter  $\rho \in ]0, 1[$  and independent from  $(X_n, M_n)$ :

$$\mathbb{P}(\tau = k) = \rho^k (1 - \rho), \quad k \geq 0. \quad (4.25)$$

In this section we first determine in Theorem 4.10 below the generating function  $\Phi(\lambda, \rho)$  of  $S_\tau$ :

$$\Phi(\lambda, \rho) := \mathbb{E}[\lambda^{S_\tau}] = (1 - \rho) \sum_{k \geq 0} \rho^k \mathbb{E}[\lambda^{S_k}], \quad 0 \leq \rho \leq 1. \quad (4.26)$$

This would allow to deduce the generating function of  $S_k$  for any  $k$  since:

$$\mathbb{E}[\lambda^{S_k}] = \frac{1}{k!} \frac{\partial^k}{\partial \rho^k} \left( \frac{\Phi(\lambda, \rho)}{1 - \rho} \right) \Big|_{\rho=0}. \quad (4.27)$$

Since we have already calculated the law of  $S_k$  we do not go further in this direction.

In Section 5, we will prove that under certain conditions, the persistent random walk  $(S_n)$  converges to a Markov process  $(S(t))_{t \in \mathbb{R}_+}$ . The following Theorem 4.10 will be used to calculate the Laplace transform of  $S(\xi)$ , where  $\xi$  is an exponential random variable independent of  $(S(t))_{t \in \mathbb{R}_+}$ . Theoretically, the following theorem permits to deduce the law of  $S_\tau$  but it is in practice impossible to determine it explicitly. However, using the law of  $L_n(1)$  for any  $n$ , given in Proposition 4.1, we present in Proposition 4.13 below the distribution of  $L_\tau(1)$ . Recall that from (4.8),  $S_\tau = 1 + 2L_\tau(1) - \tau$ . Since  $\tau$  is a random time, we cannot deduce from this identity the distribution of  $S_\tau$ .

**Theorem 4.10.** Let  $0 < \rho < \lambda < 1$ . Then the generating function of  $S_\tau$ , where  $(S_n)_{n \geq 0}$  and  $\tau$  are independent, is equal to

$$\mathbb{E}[\lambda^{S_\tau}] = \frac{(\rho - 1) \left\{ \lambda \rho \left( \widehat{\mathcal{P}}_1 \left( \frac{\rho}{\lambda} \right) + \widehat{\mathcal{P}}_2(\lambda \rho) \right) + (\lambda \rho - 1) \widehat{\mathcal{P}}_1 \left( \frac{\rho}{\lambda} \right) \widehat{\mathcal{P}}_2(\lambda \rho) \right\}}{\rho(\lambda \rho - 1) \widehat{\mathcal{P}}_2(\lambda \rho) + \lambda \rho(\rho - \lambda) \widehat{\mathcal{P}}_1 \left( \frac{\rho}{\lambda} \right) + (\lambda \rho - 1)(\rho - \lambda) \widehat{\mathcal{P}}_1 \left( \frac{\rho}{\lambda} \right) \widehat{\mathcal{P}}_2(\lambda \rho)} \quad (4.28)$$

where  $\widehat{\mathcal{P}}_i$  is defined for  $0 < x < 1$  by

$$\widehat{\mathcal{P}}_i(x) = \sum_{k \geq 1} \mathcal{P}_i(k) x^k, \quad i = 1, 2. \quad (4.29)$$

**Remark 4.11.** If  $\alpha_{2,k} = \alpha_2$  for any  $k \geq 1$  (recall that in that case  $S_n$  is the persistent random walk associated with the simple infinite comb), the function  $\widehat{\mathcal{P}}_2$  satisfies

$$\widehat{\mathcal{P}}_2(x) = \sum_{k \geq 1} (1 - \alpha_2)^{k-1} x^k = \frac{x}{1 - (1 - \alpha_2)x}.$$

Therefore (4.28) becomes

$$\mathbb{E}[\lambda^{S_\tau}] = \frac{\lambda(\rho - 1) \left(1 - \alpha_2 \widehat{\mathcal{P}}_1\left(\frac{\rho}{\lambda}\right)\right)}{\lambda\rho - 1 + \alpha_2\lambda(\rho - \lambda)\widehat{\mathcal{P}}_1\left(\frac{\rho}{\lambda}\right)}.$$

Moreover, if  $\alpha_{1,k} = 1 - \alpha_1/k$ , then

$$\widehat{\mathcal{P}}_1(x) = \sum_{k \geq 1} \frac{\alpha_1^{k-1} x^k}{(k-1)!} = x e^{\alpha_1 x},$$

and

$$\mathbb{E}[\lambda^{S_\tau}] = \frac{(\rho - 1) (\lambda - \alpha_2 \rho e^{\alpha_1 \rho / \lambda})}{\lambda\rho - 1 + \alpha_2 \rho (\rho - \lambda) e^{\alpha_1 \rho / \lambda}}.$$

We begin with a preliminary result (Lemma 4.12). The proof of Theorem 4.10 will be given later on. For  $i \in \{1, 2\}$  and  $0 < x < 1$ , let us define the generating function

$$\mathcal{G}^{(i)}(x) := \sum_{k \geq 1} \mathcal{P}_i(k) \alpha_{i,k} x^k.$$

**Lemma 4.12.** (i) For  $i = 1, 2$  and  $0 < x < 1$ , the generating function  $L^{(i)}$  satisfies

$$\mathcal{G}^{(i)}(x) = 1 + \left(\frac{x-1}{x}\right) \widehat{\mathcal{P}}_i(x), \quad (4.30)$$

where  $\widehat{\mathcal{P}}_i(x)$  has been defined by (4.29).

(ii) Moreover for  $m \geq 1$ ,

$$\sum_{b \geq m} A_i(m, b) x^b = \left(\mathcal{G}^{(i)}(x)\right)^m, \quad \sum_{b \geq m} \widehat{A}_i(m, b) x^b = \left(\mathcal{G}^{(i)}(x)\right)^{m-1} \widehat{\mathcal{P}}_i(x), \quad (4.31)$$

where  $A_i$  (resp.  $\widehat{A}_i$ ) is defined by (4.6) (resp. (4.18)).

PROOF OF LEMMA 4.12.

(i) Let  $0 < x < 1$ . We have

$$\begin{aligned} \mathcal{G}^{(i)}(x) &= - \sum_{k \geq 1} \mathcal{P}_i(k) (1 - \alpha_{i,k}) x^k + \sum_{k \geq 1} \mathcal{P}_i(k) x^k \\ &= - \frac{1}{x} \sum_{k \geq 1} \mathcal{P}_i(k+1) x^{k+1} + \widehat{\mathcal{P}}_i(x) = - \frac{1}{x} (\widehat{\mathcal{P}}_i(x) - x) + \widehat{\mathcal{P}}_i(x). \end{aligned}$$

(ii) For  $m \geq 1$ ,

$$\begin{aligned} \sum_{b \geq m} A_i(m, b) x^b &= \sum_{b \geq m, u \in \mathcal{N}(m, b)} \mathcal{P}_i(u_1) \dots \mathcal{P}_i(u_m) \alpha_{i,u_1} \times \dots \alpha_{i,u_m} x^{u_1 + \dots + u_m} \\ &= \sum_{u \in (\mathbb{N}^*)^m} \left(\mathcal{P}_i(u_1) \alpha_{i,u_1} x^{u_1}\right) \dots \left(\mathcal{P}_i(u_m) \alpha_{i,u_m} x^{u_m}\right) \\ &= \left(\mathcal{G}^{(i)}(x)\right)^m. \end{aligned}$$

The proof of the second equality in (4.31) is similar to the first one. ■



PROOF OF THEOREM 4.10. Let  $0 < \rho < \lambda < 1$ . Using (4.8) together with the independence between  $S_n$  and  $\tau$  yield

$$\begin{aligned}\mathbb{E}[\lambda^{S_\tau}] &= (1 - \rho) \sum_{n \geq 0} \mathbb{E}[\lambda^{S_n}] \rho^n = \lambda(1 - \rho) \sum_{n \geq 0} \mathbb{E}[\lambda^{2L_n(1)}] \left(\frac{\rho}{\lambda}\right)^n \\ &= \lambda(1 - \rho) \sum_{k \geq 0} \lambda^{2k} \sum_{n \geq k} \eta_n(k) \left(\frac{\rho}{\lambda}\right)^n.\end{aligned}\quad (4.32)$$

See Proposition 4.1 for the definition of  $\eta_k(n)$ . Using the decomposition (4.13) and equality (4.17) lead to the following decomposition

$$\mathbb{E}[\lambda^{S_\tau}] = \lambda(1 - \rho)(\mathcal{E}_1 + \mathcal{E}_2), \quad (4.33)$$

where  $\mathcal{E}_i$  corresponds to the part related to  $\eta_n^{(i)}$  (cf (4.11) and (4.12)) i.e. :

$$\begin{aligned}\mathcal{E}_1 &= \sum_{k \geq 0} \lambda^{2k} \sum_{n \geq k} \sum_{m=0}^{k \wedge (n-k)} \eta_n^{(1)}(k, m) \left(\frac{\rho}{\lambda}\right)^n \\ &= \sum_{k \geq 0} (\lambda\rho)^k \sum_{m=0}^k \widehat{A}_2(m+1, k+1) \sum_{n \geq m+k} A_1(m, n-k) \left(\frac{\rho}{\lambda}\right)^{n-k}.\end{aligned}$$

By (4.31), we get

$$\begin{aligned}\mathcal{E}_1 &= \sum_{k \geq 0} (\lambda\rho)^k \sum_{m=0}^k \widehat{A}_2(m+1, k+1) \left(\mathcal{G}^{(1)}\left(\frac{\rho}{\lambda}\right)\right)^m \\ &= \frac{1}{\lambda\rho} \sum_{m \geq 0} \left(\mathcal{G}^{(1)}\left(\frac{\rho}{\lambda}\right)\right)^m \sum_{k \geq m} \widehat{A}_2(m+1, k+1) (\lambda\rho)^{k+1} \\ &= \frac{1}{\lambda\rho} \sum_{m \geq 0} \left(\mathcal{G}^{(1)}\left(\frac{\rho}{\lambda}\right) \mathcal{G}^{(2)}(\lambda\rho)\right)^m \widehat{\mathcal{P}}_2(\lambda\rho) = \frac{\widehat{\mathcal{P}}_2(\lambda\rho)}{\lambda\rho \left(1 - \mathcal{G}^{(1)}\left(\frac{\rho}{\lambda}\right) \mathcal{G}^{(2)}(\lambda\rho)\right)}.\end{aligned}$$

In a similar way, we compute  $\mathcal{E}_2$ :

$$\begin{aligned}\mathcal{E}_2 &= \sum_{k \geq 0} \lambda^{2k} \sum_{n \geq k} \sum_{m=1}^{(k+1) \wedge (n-k)} \eta_n^{(2)}(k, m) \left(\frac{\rho}{\lambda}\right)^n \\ &= \sum_{k \geq 0} (\lambda\rho)^k \sum_{m=1}^{k+1} A_2(m, k+1) \sum_{n \geq m+k} \widehat{A}_1(m, n-k) \left(\frac{\rho}{\lambda}\right)^{n-k} \\ &= \frac{1}{\lambda\rho} \widehat{\mathcal{P}}_1\left(\frac{\rho}{\lambda}\right) \sum_{m \geq 1} \left(\mathcal{G}^{(1)}\left(\frac{\rho}{\lambda}\right)\right)^{m-1} \sum_{k \geq m-1} A_2(m, k+1) (\lambda\rho)^{k+1} \\ &= \frac{\widehat{\mathcal{P}}_1\left(\frac{\rho}{\lambda}\right) \mathcal{G}^{(2)}(\lambda\rho)}{\lambda\rho \left(1 - \mathcal{G}^{(1)}\left(\frac{\rho}{\lambda}\right) \mathcal{G}^{(2)}(\lambda\rho)\right)}.\end{aligned}$$

Now (4.33) yields

$$\mathbb{E}[\lambda^{S_\tau}] = \frac{\lambda(1 - \rho) \widehat{\mathcal{P}}_2(\lambda\rho) + \widehat{\mathcal{P}}_1\left(\frac{\rho}{\lambda}\right) \mathcal{G}^{(2)}(\lambda\rho)}{\lambda\rho \left(1 - \mathcal{G}^{(1)}\left(\frac{\rho}{\lambda}\right) \mathcal{G}^{(2)}(\lambda\rho)\right)}$$

which, combined with (4.30), implies (4.28). ■

**Proposition 4.13.** *Let  $k \geq 0$ . The random variable  $L_\tau(1)$  satisfies*

$$\begin{aligned} \mathbb{P}(L_\tau(1) = k) &= (1 - \rho)\rho^k \left\{ g_1(\rho) \sum_{m=1}^{k+1} A_2(m, k+1) f_1(\rho)^{m-1} \right. \\ &\quad \left. + \sum_{m=0}^k \sum_{\ell=1}^{k-m+1} A_2(m, k+1-\ell) \mathcal{P}_2(\ell) f_1(\rho)^m \right\} \end{aligned} \quad (4.34)$$

with

$$f_i(\rho) = \sum_{k \geq 1} \mathcal{P}_i(k) \alpha_{i,k} \rho^k \quad \text{and} \quad g_i(\rho) = \sum_{k \geq 1} \mathcal{P}_i(k) \rho^k, \quad i = 1, 2. \quad (4.35)$$

Moreover

$$f_i(\rho) = \left(1 - \frac{1}{\rho}\right) g_i(\rho) + 1. \quad (4.36)$$

PROOF. Let us first recall (cf Proposition 4.1) that

$$\eta_n(k) := \mathbb{P}(L_n(1) = k) = \eta_n^{(1)}(k) + \eta_n^{(2)}(k),$$

where  $\eta_n^{(1)}$  resp.  $\eta_n^{(2)}$  is defined by (4.11) resp. (4.12). In a similar way, we decompose the following probability

$$\eta(k) := \mathbb{P}(L_\tau(1) = k) = \eta^{(1)}(k) + \eta^{(2)}(k) \quad (4.37)$$

where

$$\eta^{(i)}(k) = (1 - \rho) \sum_{n \geq 0} \rho^n \eta_n^{(i)}(k), \quad i = 1, 2.$$

We shall only present the details of calculation for  $\eta^{(1)}(k)$  ( $\eta^{(2)}(k)$  can be determined similarly). By definition

$$\eta^{(1)}(k) = (1 - \rho) \sum A_2(m, k+1) A_1(m-1, n-k-\ell) \mathcal{P}_1(\ell) \rho^n, \quad (4.38)$$

the sum is taken over all combinations of indexes  $n$ ,  $m$ , and  $\ell$  satisfying

$$n \geq 0, \quad m \geq 1, \quad m \leq k+1, \quad m \leq n-k, \quad \ell \geq 1, \quad \ell \leq n-k-m+1.$$

Let us first fix the indexes  $m$  and  $\ell$  with

$$1 \leq m, \quad m \leq k+1, \quad \ell \geq 1. \quad (4.39)$$

Then we compute the sum with respect to  $n$ . We therefore introduce

$$\psi_{m,\ell}(n) := \sum_{n \geq \ell+k+m-1} A_1(m-1, n-k-\ell) \rho^n.$$

By the change of variable  $i = n - k - \ell - m + 1$ , we get

$$\begin{aligned} \psi_{m,\ell}(n) &= \sum_{i \geq 0} A_1(m-1, m-1+i) \rho^{k+\ell} \rho^{m-1+i} \\ &= \rho^{k+\ell} \sum_{i \geq 0} \sum_{u_1, \dots, u_{m-1}} \mathcal{P}_1(u_1) \times \dots \times \mathcal{P}_1(u_{m-1}) \alpha_{1,u_1} \times \dots \times \alpha_{1,u_{m-1}} \\ &\quad \times \rho^{u_1 + \dots + u_{m-1}} \mathbb{1}_{\{u_1 + \dots + u_{m-1} = m-1+i\}} \\ &= \rho^{k+\ell} \sum_{u_1, \dots, u_{m-1}} \mathcal{P}_1(u_1) \times \dots \times \mathcal{P}_1(u_{m-1}) \alpha_{1,u_1} \times \dots \times \alpha_{1,u_{m-1}} \rho^{u_1 + \dots + u_{m-1}} \\ &= \rho^{k+\ell} (f_1(\rho))^{m-1}, \end{aligned}$$

where  $f_1$  is defined by (4.35). Let us just note that, in the particular case  $m = 1$ , we get  $A_1(0, n - k - \ell) = \mathbb{1}_{\{n-k-\ell=0\}}$  and  $\psi_{1,\ell}(k) = \rho^{k+\ell}$ . Using (4.38) we obtain the following sum over all indexes  $m$  and  $\ell$  satisfying (4.39):

$$\eta^{(1)}(k) = (1 - \rho) \sum A_2(m, k + 1) \mathcal{P}_1(\ell) \rho^{k+\ell} (f_1(\rho))^{m-1},$$

when  $m, \ell$  verify (4.39). Then

$$\eta^{(1)}(k) = (1 - \rho) \rho^k g_1(\rho) \left( \sum_{m=1}^{k+1} A_2(m, k + 1) f_1(\rho)^{m-1} \right) \quad (4.40)$$

where  $g_1$  is defined by (4.35).

It can be proved

$$\eta^{(2)}(k) = (1 - \rho) \rho^k \sum_{m=0}^k \sum_{\ell=1}^{k-m+1} A_2(m, k + 1 - \ell) \mathcal{P}_2(\ell) f_1(\rho)^m. \quad (4.41)$$

Obviously (4.37), (4.40) and (4.41) imply (4.34). Let us finally prove (4.36):

$$\begin{aligned} f_i(\rho) &= \sum_{k \geq 1} \mathcal{P}_i(k) \alpha_{i,k} \rho^k = \sum_{k \geq 1} \mathcal{P}_i(k) (1 - (1 - \alpha_{i,k})) \rho^k = g_i(\rho) - \sum_{k \geq 1} \mathcal{P}_i(k + 1) \rho^k \\ &= g_i(\rho) - \frac{1}{\rho} \sum_{k \geq 2} \mathcal{P}_i(k) \rho^k = g_i(\rho) - \frac{1}{\rho} (g_i(\rho) - \rho) = \left(1 - \frac{1}{\rho}\right) g_i(\rho) + 1. \quad \blacksquare \end{aligned}$$

#### 4.4 Large time behavior

The law of  $S_n$  has been given explicitly in Proposition 4.1 but it is very complicated. This leads us to investigate the asymptotic behaviour of  $S_n$  as  $n \rightarrow \infty$ .

**Proposition 4.14.** *Assume that  $\Theta_i < \infty$ ,  $i = 1, 2$ , where  $\Theta_i$  is defined by (2.7).*

(i) *The ratio  $\frac{S_n}{n}$  converges a.s. and in  $L^1$  to  $\frac{\Theta_2 - \Theta_1}{\Theta_1 + \Theta_2}$  as  $n \rightarrow \infty$ .*

(ii) *Moreover, if  $\sum_{k \geq 1} k \mathcal{P}_i(k) < \infty$  for  $i = 1, 2$ , then the Central Limit Theorem holds:*

$$\frac{1}{\sqrt{n} \Upsilon} \left( S_n - n \frac{\Theta_2 - \Theta_1}{\Theta_1 + \Theta_2} \right) \quad (4.42)$$

*converges in distribution to a standard Gaussian random variable as  $n \rightarrow \infty$  and the constant  $\Upsilon$  is defined by*

$$\Upsilon = \frac{4}{\Theta_1 + \Theta_2} \mathbb{E} \left[ \left( T_1 - \frac{\Theta_2 T_2}{\Theta_1 + \Theta_2} \right)^2 \right] \quad (4.43)$$

*where the stopping times  $T_1$  and  $T_2$  are defined by (2.17) and  $X_0 = M_0 = 1$ .*

**Remark 4.15.** 1. *Let us first note that, under the condition presented in (ii) we can also prove the existence of a constant  $C \in \mathbb{R}$  such that*

$$\lim_{n \rightarrow \infty} \left\{ \mathbb{E}(S_n) - n \frac{\Theta_2 - \Theta_1}{\Theta_1 + \Theta_2} \right\} = C. \quad (4.44)$$

2. In the particular case  $\Theta_1 = \Theta_2 < \infty$ , Proposition 4.14 implies that  $\lim_{n \rightarrow \infty} \frac{\mathbb{E}(S_n)}{n} = 0$ . If moreover  $\sum_{k \geq 0} k\mathcal{P}_i(k) < \infty$ , we have a more precise result which says that  $\frac{1}{\sqrt{n}}S_n$  converges in distribution to a Gaussian random variable.
3. Under the conditions  $\Theta_i < \infty$  and  $\sum_{k \geq 1} k\mathcal{P}_i(k) < \infty$ , we observe therefore that the rates of convergence for the first and the second order limit theorems are similar to the rates in the setting of the classical Bernoulli random walk. The persistency does not change the long time behaviour.
4. The assumption  $\sum_{k \geq 1} k\mathcal{P}_i(k) < \infty$  is quite strong and force a relatively strong mixing in the sequence  $(X_n)$ . Open and interesting questions occur when this assumption is not satisfied. In terms of VLMC, it corresponds to the case when the expectation of the length of  $\overleftarrow{\text{pref}}(U_n)$  is infinite.

#### PROOF OF PROPOSITION 4.14

- (i) Proposition 2.3 ensures that, under the condition  $\Theta_i < \infty$ , for  $i \in \{1, 2\}$ , the process  $(X_n, M_n)_{n \geq 0}$  is an ergodic Markov chain with invariant probability  $\nu$ . The ergodic theorem, Corollary 3.2 and (2.7) imply the following almost sure convergence result:

$$\lim_{n \rightarrow \infty} \frac{L_n(1)}{n} = \nu(1, \mathbb{N}) = \frac{\Theta_2}{\Theta_1 + \Theta_2} \quad a.s., \quad (4.45)$$

where  $L_n(1)$  is defined by (4.7). Since  $L_n(1)/n$  is a bounded random variable, the almost sure convergence implies the moment convergence. Therefore, by (4.8) and (4.45), we obtain

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E}(S_n)}{n} = \lim_{n \rightarrow \infty} \frac{1 + 2\mathbb{E}(L_n(1))}{n} - 1 = \frac{2\Theta_2}{\Theta_1 + \Theta_2} - 1 = \frac{\Theta_2 - \Theta_1}{\Theta_1 + \Theta_2}.$$

- (ii) Let us consider the Markov chain  $(X_n, M_n)_{n \geq 0}$  starting at  $(1, 1)$  and denote  $Q$  the associated transition probability and  $\nu$  the invariant measure. We define

$$\sigma = \inf \left\{ n \geq 1 : (X_n, M_n) = (1, 1) \right\}. \quad (4.46)$$

Since the Markov chain is recurrent irreducible and positive, the stopping time  $\sigma$  is almost surely finite. Moreover if  $\mathbb{E}[\sigma^2] < \infty$ , Theorem 17.2.2 in [13] implies that (4.42) holds with the constant

$$\Upsilon := \nu(1, 1) \mathbb{E} \left[ \left( \sum_{k=1}^{\sigma} \left( X_k - \frac{\Theta_2 - \Theta_1}{\Theta_1 + \Theta_2} \right) \right)^2 \right]. \quad (4.47)$$

According to Definition (2.17) of the stopping times  $(T_n)$ , one has  $\sigma = T_2$  and consequently

$$\sum_{k=1}^{\sigma} X_k = \sum_{k=1}^{T_2-1} X_k + \sum_{k=T_1}^{T_2-1} X_k + X_{T_2} = T_1 - 1 - (T_2 - T_1) + 1 = 2T_1 - T_2.$$

From (3.6) and (4.47), we deduce (4.43). It remains to prove that  $\sigma$  is square integrable. Since  $\sigma = T_1 + (T_2 - T_1)$  and  $T_2 - T_1 \geq 0$ ,  $\mathbb{E}(\sigma^2) < \infty$  if and only if  $\mathbb{E}[T_1^2] < \infty$  and

$\mathbb{E}[(T_2 - T_1)^2] < \infty$ . Using Proposition 2.6 we have:

$$\begin{aligned}
\mathbb{E}[T_1^2] &= \sum_{n \geq 1} n^2 \mathcal{P}_2(n) \alpha_{2,n} = - \lim_{N \rightarrow \infty} \sum_{n=1}^N n^2 \mathcal{P}_2(n) ((1 - \alpha_{2,n}) - 1) \\
&= - \lim_{N \rightarrow \infty} \left( \sum_{n=1}^N n^2 \mathcal{P}_2(n+1) - \sum_{n=1}^N n^2 \mathcal{P}_2(n) \right) \\
&\leq \mathcal{P}_2(1) + \lim_{N \rightarrow \infty} \sum_{n=2}^N (n^2 - (n-1)^2) \mathcal{P}_2(n) \\
&\leq 1 + \lim_{N \rightarrow \infty} \sum_{n=2}^N (2n-1) \mathcal{P}_2(n) \leq 1 + 2 \sum_{n \geq 1} n \mathcal{P}_2(n) < \infty.
\end{aligned}$$

Using (2.19) and similar arguments, we obtain that  $\mathbb{E}[(T_2 - T_1)^2] < \infty$ . ■

## 5 From persistent random walk to generalized integrated telegraph noise (GITN).

Let  $(X_n, M_n)_{n \geq 0}$  be a  $\{-1, 1\} \times \overline{\mathbb{N}}^*$ -valued Markov chain satisfying (2.1) and (2.2) and let  $(S_n)_{n \geq 0}$  be the associated persistent random walk, see (4.1). We assume in this section that the transition probabilities  $(\alpha_{i,n})$  depend on a small parameter  $\varepsilon > 0$  and  $\varepsilon$  appears also both in a time scale and a space scale of the persistent random walk. We prove that there exists a normalization expressed in terms of  $\varepsilon$  so that  $(X_n, M_n, S_n)$  converges in distribution as  $\varepsilon \rightarrow 0$ . This limit is a time continuous process. Such a procedure has been already performed in [11] when the increments are a Markov chain.

More precisely we suppose that the transition probabilities satisfy

$$\alpha_{i,n} = f_i(n\varepsilon)\varepsilon + \tilde{\alpha}_{i,n,\varepsilon}\varepsilon, \quad n \geq 1, \quad i = 1, 2 \quad (5.48)$$

where  $f_1$  and  $f_2$  are positive right-continuous functions with left limits satisfying

$$\int_0^\infty f_i(u) du = \infty, \quad i = 1, 2 \quad (5.49)$$

and  $\tilde{\alpha}_{i,n,\varepsilon} \in \mathbb{R}$  with  $\lim_{\varepsilon \rightarrow 0} \sup_{i,n} |\tilde{\alpha}_{i,n,\varepsilon}| = 0$ . It is clear that for any  $i, n$  fixed,  $\lim_{\varepsilon \rightarrow 0} \alpha_{i,n} = 0$ . Therefore  $X_k$  changes from  $-1$  to  $1$  (for instance) with a small probability. The trend of  $(X_k)$  is to stay at the same level.

Let us now introduce the scaling procedure. For any  $\varepsilon > 0$  and for any  $t \in \varepsilon\mathbb{N}$ , we define the processes

$$S^\varepsilon(t) = \varepsilon S_{\frac{t}{\varepsilon}}, \quad M^\varepsilon(t) = \varepsilon M_{\frac{t}{\varepsilon}} \quad \text{and} \quad X^\varepsilon(t) = X_{\frac{t}{\varepsilon}}. \quad (5.50)$$

Note that  $(S_n)$  depends on  $\varepsilon$ , since the two families of coefficients  $(\alpha_{1,n})$  and  $(\alpha_{2,n})$  depend on  $\varepsilon$ . For the sake of simplicity, we do not mention the dependency with respect to  $\varepsilon$ . We extend the definition of the process  $(S^\varepsilon(t), t \in \varepsilon\mathbb{N})$  to  $t \in \mathbb{R}_+$  by linear interpolation and we extend the definition of the processes  $(X^\varepsilon(t), t \in \varepsilon\mathbb{N})$  and  $(M^\varepsilon(t), t \in \varepsilon\mathbb{N})$  into piecewise constant right continuous with left limits functions. In order to describe the asymptotic behavior of  $(S^\varepsilon(t), t \geq 0)$  as  $\varepsilon \rightarrow 0$ , it suffices to study the asymptotic properties of the times of trend changes. Indeed  $t \rightarrow S^\varepsilon(t)$  admits a 1 slope till the stopping time  $\varepsilon T_1$ , with  $T_1$  defined by (2.17). After that instant, the paths admits a  $-1$  slope till  $\varepsilon T_2$  and so on... The increments change periodically from  $-1$  to  $1$  and vice versa.

As  $\varepsilon \rightarrow 0$ , we shall prove that the limit process  $(S^0(t), t \geq 0)$  is still piecewise linear. More precisely it starts at  $t = 0$  with a slope equal to 1. At a random time  $e_1$  the slope changes and becomes equal to  $-1$ , at random time  $e_1 + e_2$  we observe a new change of slope and so on... We are therefore particularly interested in the description of the distribution of  $(e_n)_{n \geq 1}$ .

**Theorem 5.1.** *1. Let us consider a sequence  $(e_n)_{n \geq 1}$  of independent random variables such that for  $n \geq 1$ ,*

$$\mathbb{P}(e_{2n-1} > t) = \exp\left(-\int_0^t f_2(u) du\right), \quad \mathbb{P}(e_{2n} > t) = \exp\left(-\int_0^t f_1(u) du\right), \quad (5.51)$$

where  $f_1$  and  $f_2$  have been introduced in (5.48). Let

$$N^0(t) := \sum_{n \geq 1} \mathbb{1}_{\{e_1 + \dots + e_n \leq t\}}, \quad \text{for any } t \geq 0$$

be the counting process,

$$m(t) := t - \sup\{e_1 + \dots + e_k : e_1 + \dots + e_k \leq t\} = t - T_{N^0(t)}$$

the associate age process (spent life) and finally

$$S^0(t) = \int_0^t (-1)^{N^0(s)} ds, \quad t \geq 0. \quad (5.52)$$

the so-called Generalized Integrated Telegraph Noise (GITN).

2. Let  $(X_n, M_n)_{n \geq 0}$  be a  $\{-1, 1\} \times \mathbb{N}^*$ -valued Markov chain whose probability transition satisfies (2.1) and is  $\varepsilon$ -dependent in the sense of (5.48). We assume  $X_0 = M_0 = 1$ .

(i) For all  $n \geq 1$ , the sequence of times between two consecutive slope changes  $(\varepsilon T_1, \varepsilon(T_2 - T_1), \dots, \varepsilon(T_n - T_{n-1}))$  converges in distribution towards  $(e_1, \dots, e_n)$  as  $\varepsilon \rightarrow 0$ , where the sequence  $(T_k)_{k \geq 0}$  is defined by (2.17).

(ii) The following convergence in distribution in Skorohod's topology holds

$$(S^\varepsilon(t), X^\varepsilon(t), M^\varepsilon(t), t \geq 0) \xrightarrow{\varepsilon \rightarrow 0} (S^0(t), (-1)^{N^0(t)}, m(t), t \geq 0), \quad (5.53)$$

where  $S^\varepsilon(t)$ ,  $M^\varepsilon(t)$  and  $X^\varepsilon(t)$  are defined by (5.50).

Moreover  $(S^0(t), (-1)^{N^0(t)}, m(t), t \geq 0)$  and  $(-1)^{N^0(t)}, m(t), t \geq 0)$  are Markov processes.

**Remark 5.2.** (i) In the case  $X_0 = -1$ , the family of processes  $(S^\varepsilon(t))_{t \geq 0}$  converges in distribution to  $(S^0(t))_{t \geq 0}$  as  $\varepsilon$  goes to zero, where for any  $t \geq 0$ ,

$$S^0(t) = -\int_0^t (-1)^{\tilde{N}^0(s)} ds, \quad \text{and} \quad \tilde{N}^0(t) = \sum_{n \geq 1} \mathbb{1}_{\{e_1 + \dots + e_{n+1} \leq t\}}.$$

In the particular case where the functions  $f_1$  and  $f_2$  are constant, it has been proved in [11] that a particular solution of the telegraph equation can be represented in terms of  $S^0(t)$ . That explains that  $(S^0(t))$  defined by (5.52) is called the Generalized Integrated Telegraph Noise (GITN).

(ii) In the classical integrated telegraph noise [11], the random variables  $(e_n, n \geq 0)$  are exponentially distributed, therefore  $(S^0(t), N^0(t))$  is Markovian. For the generalized situation, this property is not true anymore, we need to consider some additional information. This information is given by  $D_-$  the left derivate of the GITN which is directly related to the age process

$$m(t) = t - \sup\{s \geq 0 : D_- S^0(s) \neq D_- S^0(t)\}.$$

(iii-a) Davis wrote in [6] that "almost all the continuous-time stochastic process models of applied probability consist of some combination of the following: diffusion, deterministic motion and random jumps". According to Theorem 5.1, between two consecutive random jumps the GITN moves in a deterministic way and therefore belongs to the family of the so-called Piecewise Deterministic Markov Processes, see for instance [6, 7, 5].

(iii-b) The possible values of  $X^0(t)$  are  $\{-1, 1\}$ . It is possible to deal with the case where  $X^0(t) \in \{a_1, \dots, a_K\}$ . In that case  $X^0(t)$  is a Markov chain indexed by  $\mathbb{R}_+$  and  $\{a_1, \dots, a_K\}$ -valued. This situation has been already treated in [11], when the functions  $(f_i)_{1 \leq i \leq K}$  are constant.

(iii-c)  $(S^0(t); t \geq 0)$  is a semi-Markov process, see [4, 12]. In [12] (Theorem 3.3 in Chapter 4) it has been proved that  $(X^\varepsilon(t); t \geq 0)$  converges to the semi-Markov process  $(X^0(t), t \geq 0)$ . This result is weaker than ours since we have considered the convergence of  $(S^\varepsilon(t), M^\varepsilon(t), X^\varepsilon(t))_{t \geq 0}$ .

PROOF.

**Step 1 — Convergence of the jump times.** Let us define  $\mathcal{R}_n^\varepsilon := (\varepsilon T_1, \varepsilon T_2, \dots, \varepsilon T_n)$  for  $n \geq 1$ . According to Proposition 2.6,  $(T_n - T_{n-1})_{n \geq 1}$  is a sequence of independent random variables. In order to prove the convergence in distribution of  $\mathcal{R}_n^\varepsilon$  as  $\varepsilon$  tends to 0, it suffices to analyze the behaviour of  $\varepsilon(T_n - T_{n-1})$  where  $n \geq 0$  is given. Recall that  $T_0 = 0$ . Remark 2.7 and (2.19) yield:

$$\begin{aligned} \mathbb{P}(\varepsilon(T_{2n+1} - T_{2n}) > t) &= \mathbb{P}\left(T_{2n+1} - T_{2n} > \frac{t}{\varepsilon}\right) = \mathbb{P}\left(T_{2n+1} - T_{2n} > \left\lfloor \frac{t}{\varepsilon} \right\rfloor\right) \\ &= (1 - \alpha_{2,1}) \times \dots \times (1 - \alpha_{2, \lfloor \frac{t}{\varepsilon} \rfloor}), \end{aligned}$$

where  $\lfloor a \rfloor$  stands for the integer part of  $a$ . Defining

$$\delta_\varepsilon(t) := \log \left\{ \mathbb{P}\left(\varepsilon(T_{2n+1} - T_{2n}) > t\right) \right\} = \sum_{j=1}^{\lfloor t/\varepsilon \rfloor} \log(1 - \alpha_{2,j}),$$

and using (5.48), we get

$$\delta_\varepsilon(t) = \sum_{j=1}^{\lfloor t/\varepsilon \rfloor} \log\left(1 - \varepsilon f_2(j\varepsilon) - \tilde{\alpha}_{2,j,\varepsilon}\right).$$

Due to the continuity of the function  $f_2$  and to the uniform limit of  $\tilde{\alpha}$  to zero,

$$\lim_{\varepsilon \rightarrow 0} \delta_\varepsilon(t) = - \lim_{\varepsilon \rightarrow 0} \varepsilon \sum_{j=1}^{\lfloor t/\varepsilon \rfloor} f_2(j\varepsilon) = - \int_0^t f_2(u) du. \quad (5.54)$$

Hence for any  $t \geq 0$ ,

$$\lim_{\varepsilon \rightarrow 0} \mathbb{P}(\varepsilon(T_{2n+1} - T_{2n}) > t) = \exp\left(-\int_0^t f_2(u) du\right).$$

The same arguments lead to

$$\lim_{\varepsilon \rightarrow 0} \mathbb{P}(\varepsilon(T_{2n+2} - T_{2n+1}) > t) = \exp\left(-\int_0^t f_1(u) du\right).$$

We conclude that  $\mathcal{R}_n^\varepsilon$  converges in distribution towards  $(e_1, e_1 + e_2, \dots, e_1 + e_2 + \dots + e_n)$ , for any  $n \geq 1$ .

**Step 2— Duality and convergence of the counting process.** Let us define the following right-continuous counting process:

$$N^\varepsilon(t) = \sup\{n \geq 0 : \varepsilon T_n \leq t\} = \sum_{n \geq 1} \mathbb{1}_{\{\varepsilon T_n \leq t\}}. \quad (5.55)$$

In order to prove (5.53) we first point out the convergence of the counting process  $N^\varepsilon$  towards  $N^0$ . The one-to-one correspondence between  $(N^\varepsilon(t))_{t \geq 0}$  and  $(T_n)_{n \geq 1}$  implies that for any  $0 < t_1 < \dots < t_k$ , the convergence in distribution of  $(N^\varepsilon(t_1), \dots, N^\varepsilon(t_k))$  as  $\varepsilon$  tends to zero is a consequence of the convergence of  $\mathcal{R}_n^\varepsilon$ . Indeed

$$\mathbb{P}(N^\varepsilon(t_1) = j_1, \dots, N^\varepsilon(t_k) = j_k) = \mathbb{P}(\varepsilon T_{j_1} \leq t_1 < \varepsilon T_{j_1+1}, \dots, \varepsilon T_{j_k} \leq t_k < \varepsilon T_{j_k+1})$$

and consequently

$$\lim_{\varepsilon \rightarrow 0} \mathbb{P}(N^\varepsilon(t_1) = j_1, \dots, N^\varepsilon(t_k) = j_k) = \mathbb{P}(E_{j_1} \leq t_1 < E_{j_1+1}, \dots, E_{j_k} \leq t_k < E_{j_k+1}),$$

where  $E_n = \sum_{k=1}^n e_k$ . In order to obtain the convergence of the counting processes, it suffices to use a tightness criterium (see, for instance, [1, Theorem 15.2 p. 125]). Let  $s < t$  and let us denote  $\tau_{st} := \lfloor t/\varepsilon \rfloor - \lfloor s/\varepsilon \rfloor$  then

$$\begin{aligned} d_{s,t}^\varepsilon &:= \mathbb{P}(N^\varepsilon(t) > N^\varepsilon(s)) = 1 - \mathbb{P}(N^\varepsilon(t) = N^\varepsilon(s)) \\ &= 1 - \mathbb{P}(N^\varepsilon(t) = N^\varepsilon(s), N^\varepsilon(s) \in 2\mathbb{N}) - \mathbb{P}(N^\varepsilon(t) = N^\varepsilon(s), N^\varepsilon(s) \in 2\mathbb{N} + 1). \end{aligned} \quad (5.56)$$

Since  $X_0 = 1$ , if  $N^\varepsilon(s) \in 2\mathbb{N}$  we have on one hand  $X_{\lfloor s/\varepsilon \rfloor} = 1$  and on the other hand  $M_{\lfloor s/\varepsilon \rfloor} \leq \lfloor s/\varepsilon \rfloor + 1$ . Assuming  $M_{\lfloor s/\varepsilon \rfloor} = \ell + 1$  with  $0 \leq \ell \leq \lfloor s/\varepsilon \rfloor$  then

$$\begin{aligned} P_{st}(\ell) &:= \mathbb{P}\left(N^\varepsilon(t) = N^\varepsilon(s) \mid M_{\lfloor s/\varepsilon \rfloor} = \ell + 1, X_{\lfloor s/\varepsilon \rfloor} = 1\right) \\ &= \mathbb{P}\left(X_{\lfloor s/\varepsilon \rfloor + 1} = 1, \dots, X_{\lfloor s/\varepsilon \rfloor + \tau_{st}} = 1 \mid M_{\lfloor s/\varepsilon \rfloor} = \ell + 1, X_{\lfloor s/\varepsilon \rfloor} = 1\right) \\ &= (1 - \alpha_{2,\ell+1})(1 - \alpha_{2,\ell+2}) \dots (1 - \alpha_{2,\ell+\tau_{st}}). \end{aligned} \quad (5.57)$$

Then it comes,

$$\begin{aligned} \mathbb{P}(N^\varepsilon(t) = N^\varepsilon(s), N^\varepsilon(s) \in 2\mathbb{N}) &= \sum_{\ell=0}^{\lfloor s/\varepsilon \rfloor} \mathbb{P}\left(N^\varepsilon(t) = N^\varepsilon(s), N^\varepsilon(s) \in 2\mathbb{N}, M_{\lfloor s/\varepsilon \rfloor} = \ell + 1\right) \\ &= \sum_{\ell=0}^{\lfloor s/\varepsilon \rfloor} \prod_{k=1}^{\tau_{st}} (1 - \alpha_{2,k+\ell}) \mathbb{P}\left(N^\varepsilon(s) \in 2\mathbb{N}, M_{\lfloor s/\varepsilon \rfloor} = \ell + 1\right) \\ &\geq \inf_{0 \leq \ell \leq \lfloor s/\varepsilon \rfloor} \prod_{k=1}^{\tau_{st}} (1 - \alpha_{2,k+\ell}) \mathbb{P}(N^\varepsilon(s) \in 2\mathbb{N}). \end{aligned} \quad (5.58)$$



Similar arguments are used in the odd case  $N^\varepsilon(s) \in 2\mathbb{N}+1$ . In this situation  $X_{\lfloor s/\varepsilon \rfloor} = -1$  and the sequence  $(\alpha_{2,\bullet})$  in (5.57) is therefore replaced by  $(\alpha_{1,\bullet})$ . Combining (5.58) with (5.56), we obtain

$$d_{s,t}^\varepsilon \leq 1 - \inf_{0 \leq \ell \leq \lfloor s/\varepsilon \rfloor} \prod_{k=1}^{\tau_{st}} (1 - \alpha_{2,k+\ell}) \mathbb{P}(N^\varepsilon(s) \in 2\mathbb{N}) - \inf_{0 \leq \ell \leq \lfloor s/\varepsilon \rfloor} \prod_{k=1}^{\tau_{st}} (1 - \alpha_{1,k+\ell}) \mathbb{P}(N^\varepsilon(s) \in 2\mathbb{N}+1).$$

By (5.48), we get

$$\begin{aligned} d_{s,t}^\varepsilon &\leq 1 - \inf_{i=1,2} \left\{ \inf_{0 \leq \ell \leq \lfloor s/\varepsilon \rfloor} \prod_{k=1}^{\tau_{st}} \left( 1 - \varepsilon f_i(\varepsilon(k+\ell)) \right) \right\} + o(\varepsilon) \\ &\leq 1 - \inf_{i=1,2} \left\{ \left( 1 - \varepsilon \sup_{0 \leq u \leq t+\varepsilon} f_i(u) \right)^{\tau_{st}} \right\} + o(\varepsilon) \\ &\leq 1 - \left( 1 - \varepsilon \sup_{0 \leq u \leq t+\varepsilon} f_1(u) \vee f_2(u) \right)^{\tau_{st}} + o(\varepsilon). \end{aligned}$$

Since  $\varepsilon \tau_{st} \leq t - s + \varepsilon$ , for any  $\delta > 0$ ,  $N > 0$ , we can find  $\varepsilon_0 > 0$  such that  $d_{s,t}^\varepsilon \leq \delta$  for all  $\varepsilon \leq \varepsilon_0$  and  $t, s \leq N$ . We deduce that the set of all the distributions of  $N^\varepsilon$ ,  $\varepsilon \in ]0, 1]$ , is weakly relatively compact and obtain finally the convergence in law of  $N^\varepsilon$  towards  $N^0$ .

**Step 3— Convergence of  $(S^\varepsilon, X^\varepsilon, M^\varepsilon)$ .** We have just proved that  $(N^\varepsilon(t))_{t \geq 0}$  converges in distribution towards  $(N^0(t))_{t \geq 0}$ . The paths of these processes belong to the Skorohod space  $\mathbb{D}$ . The two main ingredients of the proof are the following. First we note that  $S^\varepsilon(t)$ ,  $X^\varepsilon(t)$  and  $M^\varepsilon(t)$  can be expressed continuously in terms of the process  $(N^\varepsilon(s), s \leq t)$  and secondly we use the convergence of  $N^\varepsilon$ . For the process  $S^\varepsilon(t)$ , we introduce the mapping  $F_1 : \mathbb{D}(0, 1) \rightarrow \mathcal{C}(0, 1)$  defined for  $t \in [0, 1]$  by

$$F_1(f)(t) = \int_0^t \cos(\pi f(s)) ds.$$

Since  $N^\varepsilon$  is  $\mathbb{N}$ -valued, we get

$$F_1(N^\varepsilon)(t) = \int_0^t \cos(\pi N^\varepsilon(s)) ds = \int_0^t (-1)^{N^\varepsilon(s)} ds.$$

Note that (5.50) combined with (4.3) imply that  $S^\varepsilon(t) = \varepsilon + \int_0^t (-1)^{N^\varepsilon(s+\varepsilon)} ds$ . Finally the definition of  $S^\varepsilon(t)$  leads to

$$|S^\varepsilon(t) - F_1(N^\varepsilon)(t)| = \left| \varepsilon + \int_0^t (-1)^{N^\varepsilon(s+\varepsilon)} ds - \int_0^t (-1)^{N^\varepsilon(s)} ds \right| \leq 3\varepsilon. \quad (5.59)$$

For the process  $X^\varepsilon$ , we observe that  $X^\varepsilon(t) = F_2(N^\varepsilon(t)) := \cos(\pi N^\varepsilon(t))$  and the memory process is linked to the age process of  $N^\varepsilon$ :

$$\left| M^\varepsilon(t) - \left( t - \inf\{s \geq 0 : N^\varepsilon(s) = N^\varepsilon(t)\} \right) \right| \leq \varepsilon.$$

Let us just note that for  $\Phi(x) = \cos(\frac{\pi}{2}x) \mathbb{1}_{\{-1,1\}}(x)$  which is a continuous function, we get

$$t - \inf\{s \geq 0 : N^\varepsilon(s) = N^\varepsilon(t)\} = \int_0^t \Phi(N^\varepsilon(t) - N^\varepsilon(s)) ds = F_3(N^\varepsilon)(t)$$

where

$$F_3 : f \rightarrow \left( \int_0^t \Phi(f(t) - f(s)) ds, t \geq 0 \right).$$

In order to prove (5.53), it suffices to use the convergence in distribution of  $N^\varepsilon$  towards  $N^0$  developed in Step 2 and the continuity in the Skorohod topology of the three functions  $F_1$ ,  $F_2$  and  $F_3$  (see Lemma A.1, A.2 and A.3). Finally we note that  $(-1)^{N^0(t)} = F_1(N^0(t))$ ,  $S^0(t) = F_2(N^0(t))$  and  $m(t) = F_3(N^0(t))$ .  $\blacksquare$

**Examples.** For some particular  $f_1$ , the related random variable  $e_{2n}$  has a distribution which belongs to well-known families of laws.

- If  $f_1$  is a constant function then the sequence  $(e_{2n})$  is exponentially distributed.
- If  $f_1(x) = \alpha\lambda x^{\alpha-1}$  with  $\alpha > 0$  and  $\lambda > 0$  then the law of  $e_{2n}$  corresponds to the Weibull distribution with parameters  $(\alpha, \lambda)$ .
- If  $f_1(x) = \frac{\lambda}{x} \mathbb{1}_{\{x \geq x_0\}}$  with  $x_0 > 0$ , then we deal with the Pareto distribution for  $e_{2n}$ .

It has been shown in [11] that the density part of the distribution of  $S(t)$  can be expressed via Bessel functions. Here, we have a weakened result which says that we are only able to determine the Laplace transform of  $S(\tau)$  (see, Proposition 5.3 below). Being unable to invert this transformation, the distribution of  $S(t)$  is unknown. Although the path description of  $(S(t))_{t \geq 0}$  is very easy, only few properties related to the GITN are known.

**Proposition 5.3.** *Let  $(S^0(t))_{t \in \mathbb{R}_+}$  be the GITN defined by (5.52) then the double Laplace transform defined by*

$$\mathcal{L}(r, \gamma) := \int_0^\infty e^{-rt} \mathbb{E} \left[ e^{-\gamma S^0(t)} \right] dt, \quad r > 0, \quad \gamma > 0, \quad (5.60)$$

is equal to

$$\frac{-(r + \gamma)\mathcal{R}(r - \gamma, f_1)\mathcal{R}(r + \gamma, f_2) + \mathcal{R}(r - \gamma, f_1) + \mathcal{R}(r + \gamma, f_2)}{(r - \gamma)\mathcal{R}(r - \gamma, f_1) + (r + \gamma)\mathcal{R}(r + \gamma, f_2) - (r^2 - \gamma^2)\mathcal{R}(r - \gamma, f_1)\mathcal{R}(r + \gamma, f_2)},$$

where

$$\mathcal{R}(z, f_i) = \int_0^\infty e^{-zt - \int_0^t f_i(u) du} dt, \quad z \in \mathbb{R}, \quad i = 1, 2. \quad (5.61)$$

**Remark 5.4.** (i) *In the particular constant case, that is  $f_1(t) = f_1$  and  $f_2(t) = f_2$  for all  $t \geq 0$ , the stochastic process corresponds to the so-called integrated telegraph noise introduced in [11]. For this process, we get  $\mathcal{R}(z, f_i) = (z + f_i)^{-1}$  for  $i = 1, 2$ . The double Laplace transform  $\mathcal{L}$  becomes*

$$\mathcal{L}(r, \gamma) = \frac{f_0 + g_0 + r - \gamma}{r^2 - \gamma^2 + (r - \gamma)g_0 + (r + \gamma)f_0}.$$

*This identity was already obtained by Weiss in [20] and presented in [11] (see Remark 3.10).*

(ii) *Let  $\xi$  be an exponential r.v. with parameter  $r$  independent from  $(S^0(t), t \geq 0)$ . Then  $\mathcal{L}(r, \gamma)$  is the Laplace transform of  $S^0(\xi)$ :*

$$\mathcal{L}(r, \gamma) = \mathbb{E}[e^{-\gamma S^0(\xi)}].$$

**PROOF OF PROPOSITION 5.3.** Recall that  $S^\varepsilon(t)$  is the piecewise continuous process defined by (5.50). By Theorem 5.1 and the Lebesgue convergence theorem, we just need to study

the convergence of  $\mathcal{L}^\varepsilon(r, \gamma)$  the double Laplace transform of  $S^\varepsilon(t)$ . As  $\varepsilon \rightarrow 0$ , we get

$$\begin{aligned}\mathcal{L}^\varepsilon(r, \gamma) &= \int_0^\infty e^{-rt} \mathbb{E}\left[e^{-\gamma S^\varepsilon(t)}\right] dt = \sum_{k \geq 0} \int_{k\varepsilon}^{(k+1)\varepsilon} e^{-rt} \left(\mathbb{E}[e^{-\gamma \varepsilon S_k}] + o(\varepsilon)\right) dt \\ &= \frac{1 - e^{r\varepsilon}}{r} \sum_{k \geq 0} \left(e^{-r\varepsilon}\right)^k \mathbb{E}\left[e^{-\gamma \varepsilon S_k}\right] + o(\varepsilon) = \frac{1}{r} \mathbb{E}[(e^{-\gamma \varepsilon})^{S_\tau}] + o(\varepsilon),\end{aligned}\quad (5.62)$$

where  $\tau + 1$  is a geometrically distributed random variable, independent of the process  $(S_n)$ :

$$\mathbb{P}(\tau = n) = (e^{-r\varepsilon})^n (1 - e^{-r\varepsilon}).$$

Obviously (5.62) shows that  $r\mathcal{L}^\varepsilon(r, \gamma)$  and  $\mathbb{E}[e^{-\gamma \varepsilon S_\tau}]$  have the same limit as  $\varepsilon \rightarrow 0$ . Note that choosing  $\lambda = e^{-\gamma \varepsilon}$  and  $\rho = e^{-r\varepsilon}$  in Theorem 4.10 gives the value of  $\mathbb{E}[e^{-\gamma \varepsilon S_\tau}]$ . Due to the specific form of (4.28) we are lead to prove the following intermediate result:

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \widehat{\mathcal{P}}_i(e^{-\varepsilon z}) = \mathcal{R}(z, f_i). \quad (5.63)$$

where  $\widehat{\mathcal{P}}_i$  (resp.  $\mathcal{R}(z, f_i)$ ) is defined by (4.29) (resp. (5.61)).

Indeed, according to the definition of  $\widehat{\mathcal{P}}_i$  we easily get

$$\varepsilon \widehat{\mathcal{P}}_i(e^{-\varepsilon z}) = \frac{z\varepsilon}{1 - e^{-\varepsilon z}} e^{-\varepsilon z} \int_0^\infty \mathcal{P}_i\left(\left\lfloor \frac{t}{\varepsilon} \right\rfloor + 1\right) e^{-zt} dt.$$

Using (5.54) (where the index 2 is replaced by  $i$ ) yields

$$\lim_{\varepsilon \rightarrow 0} \mathcal{P}_i\left(\left\lfloor \frac{t}{\varepsilon} \right\rfloor + 1\right) = \lim_{\varepsilon \rightarrow 0} e^{\delta_\varepsilon(t)} = e^{-\int_0^t f_i(u) du}.$$

Then, the dominated convergence theorem implies (5.63). Since

- $S^\varepsilon(t)$  converges in distribution to  $S^0(t)$  as  $\varepsilon \rightarrow 0$
- $\rho - 1 \sim -r\varepsilon$  and  $\lambda\rho - 1 \sim -(r + \gamma)\varepsilon$  as  $\varepsilon \rightarrow 0$

then (5.62) and Theorem 4.10 imply

$$\begin{aligned}\mathcal{L}(r, \gamma) &= \lim_{\varepsilon \rightarrow 0} \mathcal{L}^\varepsilon(r, \gamma) = \frac{1}{r} \lim_{\varepsilon \rightarrow 0} \mathbb{E}[(e^{-\gamma \varepsilon})^{S_\tau}] \\ &= \lim_{\varepsilon \rightarrow 0} \frac{-(r + \gamma)R_1^\varepsilon(r - \gamma)R_2^\varepsilon(r + \gamma) + R_1^\varepsilon(r - \gamma) + R_2^\varepsilon(r + \gamma)}{(r - \gamma)R_1^\varepsilon(r - \gamma) + (r + \gamma)R_2^\varepsilon(r + \gamma) - (r^2 - \gamma^2)R_1^\varepsilon(r - \gamma)R_2^\varepsilon(r + \gamma)}\end{aligned}$$

where  $R_1^\varepsilon(z) = \varepsilon \widehat{\mathcal{P}}_1(e^{-z\varepsilon})$ .

It is clear that Proposition 5.3 is a straightforward consequence of (5.63) and the above identity.  $\blacksquare$

## A Continuity in the Skorohod space

Let us denote  $\mathbb{D}([0, 1])$  the Skorohod space *i.e.* the space of functions which are right-continuous and have left-hand limits.  $\mathbb{D}$  is a complete metric space for the following distance (see [1, Theorem 14.2])

$$d(f, g) = \inf_{\lambda \in \Lambda} \max\left\{\|\lambda\|, \|f - g \circ \lambda\|_\infty\right\}, \quad (A.1)$$

where

$$\|\lambda\| = \sup_{s \neq t} \left| \log \frac{\lambda(t) - \lambda(s)}{t - s} \right|,$$

$\|\cdot\|_\infty$  is the uniform norm and  $\Lambda$  is the space of strictly increasing, continuous mappings of  $[0, 1]$  into itself.

**Lemma A.1.** *Let  $\Phi : \mathbb{R} \rightarrow \mathbb{R}$  be a continuous function, then  $f \in \mathbb{D}([0, 1]) \rightarrow \Phi \circ f$  is continuous in the Skorohod topology.*

PROOF. Let  $f \in \mathbb{D}([0, 1])$ . Then there exists  $M > 0$  such that  $|f(t)| \leq M$  for all  $t \in [0, 1]$ . For  $\varepsilon > 0$ , due to the uniform continuity of  $\Phi$ , there exists  $\delta > 0$  such that: for any  $(x, y) \in [-2M, 2M]^2$  satisfying  $|x - y| < \delta$  we have  $|\Phi(x) - \Phi(y)| < \varepsilon$ . Let us consider now a function  $g \in \mathbb{D}([0, 1])$  such that  $d(f, g) < \delta \wedge M$ . Therefore, there exists  $\lambda \in \Lambda$  such that  $\|\lambda\|_\infty < \delta$  and  $\|f - g \circ \lambda\|_\infty < \delta$ . Consequently

$$\|\Phi(f) - \Phi(g \circ \lambda)\|_\infty < \varepsilon.$$

Continuity of  $\Phi$  at  $f$  follows from the definition of Skorohod's distance. ■

**Lemma A.2.** *The mapping  $f \in \mathbb{D}([0, 1]) \rightarrow \left( \int_0^t f(u) du, t \geq 0 \right)$  is continuous in the Skorohod topology.*

PROOF. First let us recall that any function belonging to the Skorohod space is integrable. We denote  $I_f(t) = \int_0^t f(u) du$ . Let  $f, g \in \mathbb{D}([0, 1])$  such that  $d(f, g) < \delta$  and choose  $\lambda \in \Lambda$  with  $\|\lambda\| < \delta$  and  $\|f - g \circ \lambda\|_\infty < \delta$ , we get

$$\begin{aligned} |I_f(t) - I_g \circ \lambda(t)| &= \lim_{n \rightarrow \infty} \left| \sum_{k=1}^n \frac{t}{n} f\left(\frac{kt}{n}\right) - g \circ \lambda\left(\frac{kt}{n}\right) \left\{ \lambda\left(\frac{kt}{n}\right) - \lambda\left(\frac{(k-1)t}{n}\right) \right\} \right| \\ &\leq \lim_{n \rightarrow \infty} \left| \frac{t}{n} \sum_{k=1}^n (f - g \circ \lambda)\left(\frac{kt}{n}\right) \right| \\ &+ \lim_{n \rightarrow \infty} \left| \sum_{k=1}^n g \circ \lambda\left(\frac{kt}{n}\right) \left\{ \lambda\left(\frac{kt}{n}\right) - \lambda\left(\frac{(k-1)t}{n}\right) - \frac{t}{n} \right\} \right|. \end{aligned} \quad (\text{A.2})$$

By definition of the norm on the Skorohod space, we have

$$e^{-\|\lambda\|} < \frac{\lambda(t) - \lambda(s)}{t - s} < e^{\|\lambda\|}, \quad \text{for } 0 \leq s < t \leq 1.$$

Consequently for any  $0 \leq s < t \leq 1$ , we have

$$|\lambda(t) - \lambda(s) - (t - s)| \leq (t - s) \max\left(e^{\|\lambda\|} - 1, 1 - e^{-\|\lambda\|}\right) \leq (t - s)(e^{\|\lambda\|} - 1). \quad (\text{A.3})$$

Combining (A.2) and (A.3) yields to

$$|I_f(t) - I_g \circ \lambda(t)| \leq \|f - g \circ \lambda\|_\infty + \|g\|_\infty (e^{\|\lambda\|} - 1), \quad \forall t \in [0, 1].$$

We deduce that  $d(f, g) < \delta$  implies

$$d(I_f, I_g) \leq \max\{\delta, \delta + \delta e^\delta \|g\|_\infty\} = \delta(1 + e^\delta \|g\|_\infty).$$

As a result  $f \rightarrow \int_0^1 f(u) du$  is a continuous mapping. ■

Using similar arguments as those presented in the proofs of Lemma A.1 and Lemma A.2, we obtain the following continuity result.

**Lemma A.3.** *Let  $\Phi$  be a continuous function, then the mapping*

$$f \in \mathbb{D}([0, 1]) \longrightarrow \left( \int_0^t \Phi(f(s) - f(s)) ds, t \geq 0 \right)$$

*is continuous in Skorohod's topology.*

## B Invariant measure for the double infinite comb

Consider the probabilized context tree given on Figure 3. In this case, there are two infinite leaves  $0^\infty$  and  $1^\infty$  and a countable number of leaves  $0^n 1$  and  $1^n 0$ ,  $n \in \mathbb{N}$ . Suppose that  $\pi$  is a stationary measure on  $\mathcal{L}$ . Denote by  $\mathcal{W}$  the set of finite words on the alphabet  $\{0, 1\}$ . For any finite word  $w \in \mathcal{W}$ , we denote by  $\pi(w) := \pi(\mathcal{L}w)$  the measure of the cylinder  $\mathcal{L}w$  denoting the set of left infinite words ending with  $w$ . We first compute  $\pi(w)$  as a function of  $\pi(1)$  when the reversed word of  $w$  is any context or any internal node. Applying equation (1.2) to  $U_n = \dots 10^n$ , it comes for any  $n \geq 1$ ,

$$\pi(10^n) = \pi(10^{n-1})q_{0^{n-1}1}(0).$$

An immediate induction yields, for any  $n \geq 1$ ,

$$\pi(10^n) = \pi(10) \prod_{k=1}^{n-1} q_{0^{k-1}1}(0) = \pi(10) \prod_{k=1}^{n-1} (1 - \alpha_{1,k}) = \pi(10)\mathcal{P}_1(n). \quad (\text{B.1})$$

In the same way,

$$\pi(01^n) = \pi(01)\mathcal{P}_2(n), \quad (\text{B.2})$$

The stationary probability of a reversed context is thus necessarily given by Formulae (B.1) and (B.2). Now, if  $0^n$  is any internal node of the context tree but 0, we need going down along the branch in the context tree to reach the contexts; using then the disjoint union  $\pi(0^{n+1}) = \pi(0^n) - \pi(10^n)$ , by induction, it comes for any  $n \geq 2$ ,

$$\pi(0^n) = \pi(0) - \pi(10) \sum_{k=1}^{n-1} \mathcal{P}_1(k). \quad (\text{B.3})$$

The same holds for any internal node  $1^n$  but 1,

$$\pi(1^n) = \pi(1) - \pi(10) \sum_{k=1}^{n-1} \mathcal{P}_2(k). \quad (\text{B.4})$$

where we have used  $\pi(01) = \pi(10)$  (coming from the invariance of  $\pi$ ). The stationary probability of a reversed internal node of the context tree is thus necessarily given by Formulae (B.3) and (B.4).

It remains to compute  $\pi(10)$  and then  $\pi(0)$  (and consequently  $\pi(1)$ ). The denumerable partition of the whole probability space given by all cylinders based on leaves in the context tree implies  $1 - \pi(0^\infty) - \pi(1^\infty) = \pi(10) + \pi(100) + \dots + \pi(01) + \pi(011) + \dots$ , *i.e.*

$$1 - \pi(0^\infty) - \pi(1^\infty) = \pi(10) \sum_{n \geq 1} (\mathcal{P}_1(n) + \mathcal{P}_2(n)). \quad (\text{B.5})$$

This leads to the following statement that covers all cases of existence, unicity and non-triviality for a stationary probability measure for the double infinite comb. In the generic case (named *irreducible* case hereunder), we give a necessary and sufficient condition on the data for the existence of a stationary probability measure; moreover, when a stationary probability exists, it is unique. The *reducible* case is much more singular and gives rise to nonunicity.

**Proposition B.1. (Stationary probability measures for a double infinite comb)**

Let  $(U_n)_{n \geq 0}$  be a VLMC defined by a probabilized double infinite comb.

(i) Irreducible case: Assume that  $q_{0^\infty}(0) \neq 1$  and  $q_{1^\infty}(1) \neq 1$ .

- (a) Existence: The Markov process  $(U_n)_{n \geq 0}$  admits a stationary probability measure on  $\mathcal{L}$  if and only if the numerical series  $\Theta_1$  and  $\Theta_2$  converge.
- (b) Unicity: Assume that the series  $\Theta_1$  and  $\Theta_2$  converge. Then, the stationary probability measure  $\pi$  on  $\mathcal{L}$  is unique; it is characterized by

$$\pi(0) = \frac{\Theta_1}{\Theta_1 + \Theta_2}, \quad \pi(10) = \frac{1}{\Theta_1 + \Theta_2} \quad (\text{B.6})$$

and Formulae (B.1), (B.2) (B.3), (B.4).

(ii) Reducible cases: Assume that  $q_{0^\infty}(0) = 1$  and  $q_{1^\infty}(1) \neq 1$ .

- (a) If at least one of the series  $\Theta_1$  and  $\Theta_2$  diverges, then the trivial probability measure  $\pi$  on  $\mathcal{L}$  defined by  $\pi(0^\infty) = 1$  is the unique stationary probability measure.
- (b) If the series  $\Theta_1$  and  $\Theta_2$  converge, then there is a one parameter family of stationary probability measures on  $\mathcal{L}$ . More precisely, for any  $a \in [0, 1]$ , there exists a unique stationary probability measure  $\pi_a$  on  $\mathcal{L}$  such that  $\pi_a(0^\infty) = a$ . The probability  $\pi_a$  is characterized by

$$\pi_a(0) = \frac{a\Theta_2 + \Theta_1}{\Theta_1 + \Theta_2}, \quad \pi_a(10) = \frac{1 - a}{\Theta_1 + \Theta_2}$$

and Formulae (B.1), (B.2) (B.3), (B.4).

Assume that  $q_{0^\infty}(0) \neq 1$  and  $q_{1^\infty}(1) = 1$ . Then the same results as in (ii.a) and (ii.b) hold, exchanging the role of 0 and 1.

Assume that  $q_{0^\infty}(0) = 1$  and  $q_{1^\infty}(1) = 1$ .

- (c) If at least one of the series  $\Theta_1$  and  $\Theta_2$  diverges, then there is a one parameter family of stationary probability measures on  $\mathcal{L}$ . More precisely, for any  $a \in [0, 1]$ , there exists a unique stationary probability measure  $\pi_a$  on  $\mathcal{L}$  such that  $\pi_a(0^\infty) = a$ . The probability  $\pi_a$  is characterized by  $\pi_a(0^n) = a$  and  $\pi_a(1^n) = 1 - a$  for every  $n \geq 1$  and  $\pi_a(w) = 0$  as soon as  $w$  contains one 0 and one 1.
- (d) If the series  $\Theta_1$  and  $\Theta_2$  converge, then there is a two parameters family of stationary probability measures on  $\mathcal{L}$ . More precisely, for any  $a \in [0, 1]$  and  $b \in [0, 1]$ , there exists a unique stationary probability measure  $\pi_{a,b}$  on  $\mathcal{L}$  such that  $\pi_{a,b}(0^\infty) = a$  and  $\pi_{a,b}(1^\infty) = b$ . The probability  $\pi_{a,b}$  is characterized by

$$\pi_{a,b}(0) = \frac{a\Theta_2 + (1 - b)\Theta_1}{\Theta_1 + \Theta_1}, \quad \pi_{a,b}(10) = \frac{1 - a - b}{\Theta_1 + \Theta_2}$$

and Formulae (B.1), (B.2) (B.3), (B.4).

PROOF.

- (i) Assume that  $q_{0^\infty}(0) \neq 1$ ,  $q_{1^\infty}(1) \neq 1$  and that  $\pi$  is a stationary probability measure. By definition of probability transitions,  $\pi(0^\infty) = \pi(0^\infty)q_{0^\infty}(0)$  and  $\pi(1^\infty) = \pi(1^\infty)q_{1^\infty}(1)$  so that  $\pi(0^\infty)$  and  $\pi(1^\infty)$  necessarily vanish. Thus, thanks to (B.5),  $\pi(10) \neq 0$ , the series  $\Theta_1 + \Theta_2$  converges and so do  $\Theta_1$  and  $\Theta_2$ . This also implies

$$1 = \pi(10)(\Theta_1 + \Theta_2).$$

Passing to the limit in (B.3) implies  $\pi(0) = \pi(10)\Theta_1$ . Thus Formula (B.6) is valid. Moreover, when  $\bar{w}$  is any context or any internal node of the context tree,  $\pi(w)$  is necessarily given by Formulae (B.6), (B.1), (B.2), (B.3) and (B.4). Since the cylinders  $\mathcal{L}w$ ,  $w \in \mathcal{W}$  span the  $\sigma$ -algebra on  $\mathcal{L}$ , there is at most one stationary probability measure. This proves the *only if* part of (i.a), the unicity and the characterization claimed in (i.b).

Reciprocally, when the series converge, Formulae (B.6), (B.1), (B.2) (B.3), (B.4) define a probability measure on the semiring spanned by cylinders, which extends to a stationary probability measure on the whole  $\sigma$ -algebra on  $\mathcal{L}$ . This proves the *if* part of (i.a).

To deal with the reducible cases, recall the three following equations (which hold when the series converge) :

$$\begin{cases} 1 - \pi(0^\infty) - \pi(1^\infty) &= \pi(10)(\Theta_1 + \Theta_2) \\ \pi(0^\infty) &= \pi(0) - \pi(10)\Theta_1 \\ \pi(1^\infty) &= \pi(1) - \pi(10)\Theta_2 \end{cases}$$

- (ii) Assume that  $q_{0^\infty}(0) = 1$  and  $q_{1^\infty}(1) \neq 1$ . First, as above,  $q_{1^\infty}(1) \neq 1$  implies  $\pi(1^\infty) = 0$ . Next, Formula (B.5) is always valid so that the divergence of at least one of the series forces  $\pi(10)$  to vanish. This gives  $\pi(0^\infty) = 1$ . With the assumption  $q_{0^\infty}(0) = 1$ , one immediately sees that this trivial probability is stationary, proving (ii.a).

To prove (ii.b), assume furthermore that the series  $\Theta_1$  and  $\Theta_2$  converge and let  $a \in [0, 1]$ . As before, any stationary probability measure  $\pi$  is completely determined by  $\pi(0)$  and  $\pi(10)$ . As above,  $\pi(1^\infty) = 0$  and if we fix  $\pi(0^\infty) = a$ , the system (i) reduces to

$$\begin{cases} 1 - a &= \pi(10)(\Theta_1 + \Theta_2) \\ a &= \pi(0) - \pi(10)\Theta_1 \end{cases}$$

This gives the characterisation of (ii.b). Formulae (B.1), (B.2) (B.3), (B.4) standardly extend  $\pi_a$  to the whole  $\sigma$ -algebra on  $\mathcal{L}$  and  $\pi_a$  is clearly stationary.

(ii.c) Assume that  $q_{0^\infty}(0) = 1$  and  $q_{1^\infty}(1) = 1$ . As previously, Formula (B.5) is valid so that the divergence of at least one of the series forces  $\pi(10)$  to vanish. Let  $a \in [0, 1]$  and fix  $\pi(0^\infty) = a$ , the system (i) reduces to  $\pi(0^\infty) = \pi(0) = a$  and  $\pi(1^\infty) = \pi(1) = 1 - a$ . The invariance of this measure may be easily checked.

To prove (ii.d), assume furthermore that the series  $\Theta_1$  and  $\Theta_2$  converge and let  $a \in [0, 1]$  and  $b \in [0, 1]$ . If we fix  $\pi(0^\infty) = a$  and  $\pi(1^\infty) = b$ , the system (i) is equivalent to

$$\begin{cases} \pi(0) - \pi(10)\Theta_1 &= a \\ \pi(0) + \pi(10)\Theta_2 &= 1 - b \end{cases}$$

As  $\Theta_1 \geq 1$  and  $\Theta_2 \geq 1$ , this system has a unique solution given by

$$\pi_{a,b}(0) = \frac{a\Theta_2 + (1-b)\Theta_1}{\Theta_1 + \Theta_2} \quad \text{and} \quad \pi_{a,b}(10) = \frac{1-a-b}{\Theta_1 + \Theta_2}.$$

■

## Acknowledgements

We are very grateful to F. Paccaut and N. Pouyanne who made the present of the calculation of the invariant measure for the double comb to us.

## References

- [1] P. Billingsley. Convergence of probability measures. *Wiley Series in Probability and Statistics: Probability and Statistics*, John Wiley & Sons Inc., Second edition, 1999.
- [2] P. Cénac, B. Chauvin, F. Paccaut, N. Pouyanne. Variable length Markov chains and dynamical sources. *Séminaire de Probabilités*. Springer. Vol. 44, pp. 1–39, 2012.
- [3] P. Cénac, B. Chauvin, F. Paccaut, N. Pouyanne. Automata and VLMC. Size of the M-complete tree. *preprint*, 2012.
- [4] E.Çinlar. Introduction to stochastic processes. *Prentice-Hall Inc.*, 1975.
- [5] C.Cocozza-Thivent. Processus de renouvellement markovien, Processus de Markov déterministes par morceaux. Online book available on the webpage:  
<http://perso-math.univ-mlv.fr/users/cocozza.christiane/recherche-page-perso/PresentationRMetPDMP.html>
- [6] H.A. Davis. Piecewise-deterministic Markov processes: a general class of nondiffusion stochastic models. With discussion *J J. Roy. Statist. Soc. Ser. B*, 46, n.3 pp. 353–388, 1984.
- [7] H.A.Davis. Markov Models and Optimization. *Chapman & Hall*, 1993.
- [8] E. C.Eckstein, J. A.Goldstein, and M.Leggas. The mathematics of suspensions: Kac walks and asymptotic analyticity. In *Proceedings of the Fourth Mississippi State Conference on Difference Equations and Computational Simulations (1999)*, volume 3 of *Electron. J. Differ. Equ. Conf.*, pp. 39–50 (electronic), San Marcos, TX, 2000. Southwest Texas State Univ.
- [9] A. Galves and E. Löcherbach. Stochastic chains with memory of variable length. *TICSP Series*, 38, pp. 117–133, 2008.
- [10] B.Harlamov. Continuous semi-Markov processes *Applied Stochastic Methods Series*, ISTE London, John Wiley & Sons, 2008.
- [11] S. Herrmann and P. Vallois. From persistent random walk to the telegraph noise. *Stochastics and Dynamics* 10, no. 2, pp. 161–196, 2010.
- [12] J.Jansson and R.Manca. Applied Semi-Markov Processes. *Springer Science+Business Media*, 2006.
- [13] S.Meyn and R.L.Tweedie. Markov chains and stochastic stability. *Cambridge University Press*, Second edition, 2009.
- [14] Eric Renshaw and Robin Henderson. The correlated random walk. *J. Appl. Probab.*, 18(2):403–414, 1981.
- [15] J. Rissanen. A universal data compression system. *IEEE Trans. Inform. Theory* 29, no. 5, pp. 656–664, 1983.
- [16] E. Seneta. Non-negative matrices and Markov chains. *Springer Series in Statistics*, Revised reprint of the second (1981) edition, Springer-Verlag, New York, 2006.
- [17] C.S.Tapiero and P.Vallois. Memory-based persistence in a counting random walk process. *Phys. A.*, 386, Nr.1, pp 303–307, 2007.
- [18] C.S.Tapiero and P.Vallois. A claims persistence process and insurance. *J. Insurance Math. Econom.*, 44, Nr.3, pp 367–373, 2009.
- [19] George H. Weiss. *Aspects and applications of the random walk*. Random Materials and Processes. North-Holland Publishing Co., Amsterdam, 1994.
- [20] George H. Weiss. Some applications of persistent random walks and the telegrapher’s equation. *Phys. A*, 311(3-4):381–410, 2002.