PART 1

Markov and Semi-Markov Processes

Statistical Topics and Stochastic Models for Dependent Data with Applications,
First Edition. Vlad Stefan Barbu and Nicolas Vergne.

1

Variable Length Markov Chains, Persistent Random Walks: A Close Encounter

We consider a walker on the line that at each step keeps the same direction with a probability that depends on the time already spent in the direction the walker is currently moving. These walks with memories of variable length can be seen as generalizations of directionally reinforced random walks (DRRWs) introduced in Mauldin *et al*. (1996). We give a complete and usable characterization of the recurrence or transience in terms of the probabilities to switch the direction. These conditions are related to some characterizations of existence and uniqueness of a stationary probability measure for a particular Markov chain: in this chapter, we define the general model for words produced by a variable length Markov chain (VLMC) and we introduce a key combinatorial structure on words. For a subclass of these VLMC, this provides necessary and sufficient conditions for existence of a stationary probability measure.

1.1. Introduction

This is the story of the encounter between two worlds: the world of random walks and the world of VLMCs. The meeting point turns around the semi-Markov property of underlying processes.

In a VLMC, unlike fixed-order Markov chains, the probability to predict the next symbol depends on a possibly unbounded part of the past, the length of which depends on the past itself. These relevant parts of pasts are called *contexts*. They are stored in a *context tree*. With each context, a probability distribution is associated, prescribing the conditional probability of the next symbol, given this context.

Chapter written by Peggy CÉNAC, Brigitte CHAUVIN, Frédéric PACCAUT and Nicolas POUYANNE.

VLMCs are now widely used as random models for character strings. They were introduced in Rissanen (1983) to perform data compression. When they have a finite memory, they provide a parsimonious alternative to fixed-order Markov chain models, in which the number of parameters to estimate grows exponentially fast with the order; they are also able to capture finer properties of character sequences. When they have infinite memory – this will be our case of study in this chapter – they are a tractable way to build non-Markov models and they may be considered as a subclass of "*chaînes à liaisons complètes*" (Doeblin and Fortet 1937) or "chains with infinite order" (Harris 1955).

VLMCs are used in bioinformatics, linguistics and coding theory to model how words grow or to classify words. In bioinformatics, both for protein families and DNA sequences, identifying patterns that have a biological meaning is a crucial issue. Using VLMC as a model enables one to quantify the influence of a meaning pattern by giving a transition probability on the following letter of the sequence. In this way, these patterns appear as contexts of a context tree. Note that their length may be unbounded (Bejerano and Yona 2001).

In addition, if the context tree is recognized to be a signature of a family (say, of proteins), this gives an efficient statistical method to test whether or not two samples belong to the same family (Busch *et al*. 2009).

Therefore, estimating a context tree is an issue of interest and many authors (statisticians or not, applied or not) stress the fact that the height of the context tree should not be supposed to be bounded. This is the case in Galves and Leonardi (2008) where the algorithm `CONTEXT` is used to estimate an unbounded context tree, or in Garivier and Leonardi (2011). Furthermore, as explained in Csiszár and Talata (2006), the height of the estimated context tree grows with the sample size, so that estimating a context tree by assuming *a priori* that its height is bounded is not realistic.

There is extensive literature on the construction of efficient estimators of context trees, as well for finite or infinite context trees. This chapter is not a review of statistics issues, which would already be relevant for finite memory VLMC. This is a study of the probabilistic properties of infinite memory VLMC as random processes, and more specifically of the main property of interest for such processes: existence and uniqueness of a stationary measure.

As has already been said, VLMC are a natural generalization to infinite memory of Markov chains. It is usual to index a sequence of random variables forming a Markov chain with positive integers and to make the process grow to the right. The main drawback of this habit for an infinite memory process is that the sequence of the process is read from left to right, whereas the (possibly infinite) sequence giving the past needed to predict the next symbol is read in the context tree from right to left, thus giving rise to confusion and lack of readability. For this reason, in this chapter, the VLMC grows to the left. In this way, both the process sequence and the memory in the context tree are read from left to right.

Classical random walks have *independent* and identically distributed increments. In the literature, *persistent* random walks (PRMs), also called *Goldstein-Kac random walks* or *correlated random walks*, refer to random walks having a Markov chain of finite order as an increment process. For such walks, the dynamics of trajectories has a short memory of given length and the random walk itself is not Markovian anymore. What happens whenever the increments depend on a *non-bounded* past memory?

Consider a walker on $\mathbb{Z}$, allowed to increment its trajectory by -1 or 1 at each step of time. Assume that the probability to keep the current direction ± 1 depends on the time already spent in the said direction – the distribution of increments thus acts as a reinforcement of the dependency from the past. More precisely, the process of increments of such a one-dimensional random walk is a Markov chain on the set of (right-)infinite words, with variable – and unbounded – length memory: a VLMC. The concerned VLMC is defined in section 1.3.1. It is based on a context tree called a *double comb*. Later, section 1.3.2 deals with a two-dimensional persistent random walk defined in an analogous manner on $\mathbb{Z}^2$ by a VLMC based on a context tree called a *quadruple comb*.

These random walks that have an unbounded past memory can be seen as a generalization of "directionally reinforced random walks (DRRW)" introduced by Mauldin *et al*. (1996), in the sense that the persistence times are anisotropic ones. For a one-dimensional random walk associated with a double comb, a complete characterization of recurrence and transience, in terms of changing (or not) direction probabilities, is given in section 1.3.1. More precisely, when one of the random times spent in a given direction (the so-called *persistence times*) is an integrable random variable, the recurrence property is equivalent to a classical drift-vanishing. In all other cases, the walk is transient unless the weight of the tail distributions of both persistent times are equal. In two-dimensional random walk, sufficient conditions of transience of recurrence are given in section 1.3.2.

Actually, because of the very specific form of the underlying driving VLMC, these PRWs turn out to be in one-to-one correspondence with so-called *Markov additive processes*. Section 1.5 examines the close links between PRWs, Markov additive processes, semi-Markov chains and VLMC.

In section 1.2, the definition of a general VLMC and a couple of examples are given. In section 1.3, the PRWs are defined and known results on their recurrence properties are collected. In view of section 1.5 where we show how PRW and VLMC meet through the world of semi-Markov chains, section 1.4 is devoted to results – together with a heuristic approach – on the existence and unicity of stationary measures for a VLMC.

1.2. VLMCs: definition of the model

Let $\mathcal{A}$ be a finite set, called the *alphabet*. Here $\mathcal{A}$ will most often be the standard alphabet $\mathcal{A} = \{0, 1\}$, but also $\mathcal{A} = \{d, u\}$ (for *down* and *up*) or $\mathcal{A} = \{\mathtt{n}, \mathtt{e}, \mathtt{w}, \mathtt{s}\}$ (for the cardinal directions). Let

$$\mathcal{R} = \{\alpha\beta\gamma\cdots : \alpha, \beta, \gamma, \cdots \in \mathcal{A}\}$$

be the set of *right-infinite* words over $\mathcal{A}$, written by simple concatenation. A VLMC on $\mathcal{A}$, defined below and most often denoted by $(U_n)_{n\in\mathbb{N}}$, is a particular type of $\mathcal{R}$-valued discrete time Markov chain where:

– the process evolves between time n and time $n+1$ by adding one letter *on the left* of U_n;

– the transition probabilities between time n and time $n+1$ depend on a finite – but not bounded – prefix[1] of the current word U_n.

Giving a formal frame of such a process leads to the following definitions. For a complete presentation of VLMC, one can also refer to Cénac *et al.* (2012).

As usual, a *tree on* $\mathcal{A}$ is a set $\mathcal{T}$ of finite words – namely a subset of $\cup_{n\in\mathbb{N}}\mathcal{A}^n$ – which contains the empty word $\emptyset$ (the *root* of $\mathcal{T}$) and which is prefix-stable: for all finite words u, v, $uv \in \mathcal{T} \Longrightarrow u \in \mathcal{T}$. A tree is made of *internal nodes* ($u \in \mathcal{T}$ is internal when $\exists\alpha \in \mathcal{A}$, $u\alpha \in \mathcal{T}$) and of *leaves* ($u \in \mathcal{T}$ is a leaf when it has no child: $\forall\alpha \in \mathcal{A}$, $u\alpha \notin \mathcal{T}$).

DEFINITION 1.1 (Context tree).– *A* context tree *on $\mathcal{A}$ is a saturated tree on $\mathcal{A}$ having an at most countable set of infinite branches.*

The tree $\mathcal{T}$ is *saturated* whenever any internal node has $\#(\mathcal{A})$ children: for any finite word u and for any $\alpha \in \mathcal{A}$, $u\alpha \in \mathcal{T} \Longrightarrow (\forall\beta \in \mathcal{A},\ u\beta \in \mathcal{T})$. A right-infinite word on $\mathcal{A}$ is an *infinite branch* of $\mathcal{T}$ when all its finite prefixes belong to $\mathcal{T}$.

Following the vocabulary introduced by Rissanen, a *context* of the tree is a leaf or an infinite branch. A finite or right-infinite word on $\mathcal{A}$ is an *external node* when it is neither internal nor a context. See Figure 1.1 which illustrates these definitions, as well as the pref function defined hereunder.

DEFINITION 1.2 (pref function).– *Let $\mathcal{T}$ be a context tree. If w is any external node or any context, the symbol* $\mathrm{pref}\, w$ *denotes the longest (finite or infinite) prefix of w that belongs to $\mathcal{T}$.*

1 In fact, an infinite prefix might be needed in a denumerable number of cases.

In other words, $\mathrm{pref}\,w$ is the only context c for which $w = c\cdots$ For a more visual presentation, hang w by its head (its left-most letter) and insert it into the tree; the only context through which the word goes out of the tree is its pref.

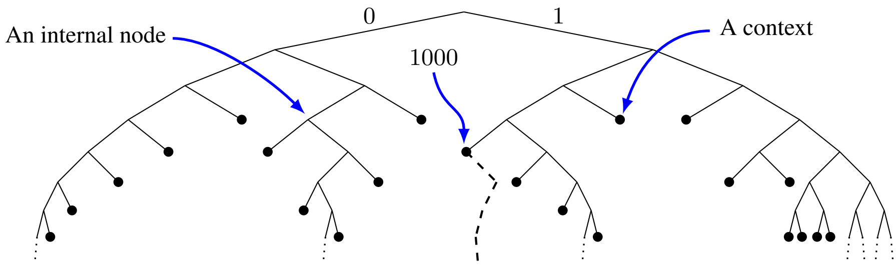


Figure 1.1. *A context tree on the alphabet $\mathcal{A} = \{0, 1\}$. The dotted lines are possibly the beginning of infinite branches. Any word that writes $1000\cdots$, like the one represented by the dashed line, admits 1000 as a pref. For a color version of this figure, see www.iste.co.uk/barbu/data.zip*

With these definitions, it is now possible to define a VLMC.

DEFINITION 1.3 (VLMC).– *Let $\mathcal{T}$ be a context tree. For every context c of $\mathcal{T}$, let q_c be a probability measure on $\mathcal{A}$. The* VLMC *defined by $\mathcal{T}$ and by the $(q_c)_c$ is the $\mathcal{R}$-valued discrete-time Markov chain $(U_n)_{n\in\mathbb{N}}$ defined by the following transition probabilities: $\forall n \in \mathbb{N}$, $\forall \alpha \in \mathcal{A}$,*

$$\mathbb{P}\left(U_{n+1} = \alpha U_n | U_n\right) = q_{\mathrm{pref}(U_n)}\left(\alpha\right). \qquad [1.1]$$

To get a realization of a VLMC as a process on $\mathcal{R}$, take a (random) right infinite word

$$U_0 = X_0 X_{-1} X_{-2} X_{-3} \cdots$$

At each step of time $n \geq 0$, one gets U_{n+1} by adding a random letter X_{n+1} on the left of U_n:

$$\begin{aligned} U_{n+1} &= X_{n+1} U_n \\ &= X_{n+1} X_n \cdots X_1 X_0 X_{-1} X_{-2} \cdots \end{aligned}$$

under the conditional distribution [1.1].

REMARK 1.1.– *Probabilizing* a context tree consists, as in definition 1.3, of endowing it with a family of probability measures on the alphabet, indexed by the set of contexts. This vocabulary is used below.

REMARK 1.2.– Assume that the context tree is finite and denote its height by h; in this condition, the VLMC is just a Markov chain of order h on $\mathcal{A}$. On the contrary, when the context tree is infinite, and this is mainly our case of interest, the VLMC is generally *not* a Markov process on $\mathcal{A}$.

EXAMPLE 1.1.– Take $\mathcal{A} = \{\mathtt{n}, \mathtt{e}, \mathtt{w}, \mathtt{s}\}$ as an (ordered) alphabet, so that the daughters of an internal node are represented, as shown on the left side of Figure 1.2. Making the transition probabilities $\mathbb{P}(U_{n+1} = \alpha U_n | U_n)$ depend only on the length of the largest prefix of the form $\mathtt{n}^k$ ($k \geq 0$) of U_n amounts to taking a comb as a context tree, as shown on the right side of Figure 1.2. Its finite contexts are the $\mathtt{n}^k\alpha$ where $k \geq 0$ and $\alpha \in \mathcal{A} \setminus \{\mathtt{n}\}$.

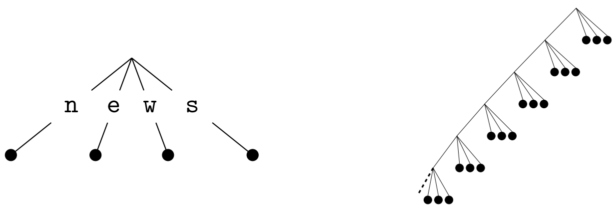


Figure 1.2. *On the left: how one can represent trees on $\mathcal{A} = \{\mathtt{n}, \mathtt{e}, \mathtt{w}, \mathtt{s}\}$. On the right, the so-called* left comb *on $\mathcal{A} = \{\mathtt{n}, \mathtt{e}, \mathtt{w}, \mathtt{s}\}$*

EXAMPLE 1.2.– Take again $\mathcal{A} = \{\mathtt{n}, \mathtt{e}, \mathtt{w}, \mathtt{s}\}$ as an alphabet. Making the transition probabilities $\mathbb{P}(U_{n+1} = \alpha U_n | U_n)$ depend only on the length of the largest prefix of the form α^k ($k \geq 1$) of U_n, where α is *any* letter, amounts to taking a *quadruple comb* as a context tree, as shown on the right side of Figure 1.3. In the same vein, if one takes $\mathcal{A} = \{u, d\}$, the *double comb* is the context tree, as shown on the left side of Figure 1.3. In the corresponding VLMC, the transitions depend only on the length of the last current run u^k or d^k, $k \geq 1$. The double comb and the quadruple comb are used below to define PRWs.

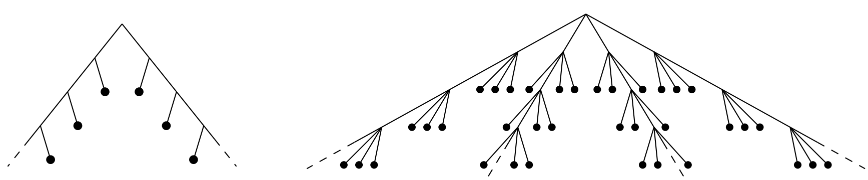


Figure 1.3. *The double comb and the quadruple comb*

EXAMPLE 1.3.– Take $\mathcal{A} = \{0, 1\}$ (naturally ordered for the drawings). The left comb of right combs, shown on the left side of Figure 1.4, is the context tree of a VLMC that makes its transition probabilities depend on the largest prefix of U_n of the form 0^p1^q.

If one has to take into consideration the largest prefix of the form 0^p1^q or 1^p0^q, one has to use the double comb of opposite combs, as shown on the right side of Figure 1.4.

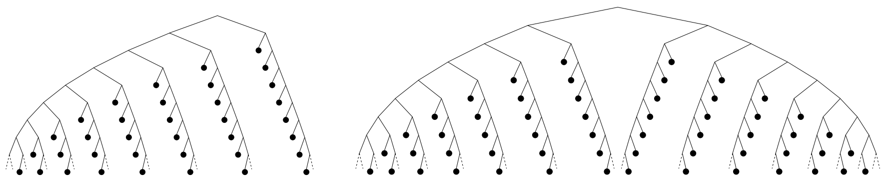


Figure 1.4. *Context trees on $\mathcal{A} = \{0, 1\}$: the left comb of right combs (on the left) and a double comb of opposite combs (on the right)*

DEFINITION 1.4 (Non-nullness).– *A VLMC is called* non-null *when no transition probability vanishes, i.e. when $q_c(\alpha) > 0$ for every context c and for every $\alpha \in \mathcal{A}$.*

Non-nullness appears below as an irreducibility-like assumption made on the driving VLMC of PRWs and for existence and unicity of an invariant probability measure for a general VLMC as well.

1.3. Definition and behavior of PRWs

In this section, the so-called *PRWs* are defined. A PRW is a random walk driven by some VLMC. In dimensions one and two, results on transience and the recurrence of PRW are given. These results are detailed and proven in Cénac *et al*. (2018b, 2013) in dimension one and in Cénac *et al*. (2020) in dimension two.

1.3.1. *PRWs in dimension one*

In this section, we deal with one-dimensional PRWs. Note that, contrary to the classical random walk, a PRW is generally not Markovian. Let $\mathcal{A} := \{d, u\} = \{-1, 1\}$ (d for down and u for up) and consider the *double comb* on this alphabet as a context tree, probabilize it and denote by $(U_n)_n$ a realization of the associated VLMC. The n^{th} increment X_n of the PRW is given as the first letter of U_n: define the persistent random walk $S = (S_n)_{n\geq 0}$ by $S_0 = 0$ and, for $n \geq 1$,

$$S_n := \sum_{\ell=1}^{n} X_\ell, \qquad [1.2]$$

so that for any $n \geq 1$, $m \geq 0$,

$$\begin{aligned} \mathbb{P}\left(S_{m+1} = S_m + 1 | U_m = d^n u \ldots\right) &= q_{d^n u}(u) \\ \mathbb{P}\left(S_{m+1} = S_m - 1 | U_m = u^n d \ldots\right) &= q_{u^n d}(d). \end{aligned}$$

Furthermore, for the sake of simplicity and without loss of generality, we condition the walk to start almost surely (a.s.) from $\{X_{-1} = u, X_0 = d\}$ – this amounts to changing the origin of time. In this model, a walker on a line keeps the same direction with a probability that depends on the discrete time already spent in the direction the walker is currently moving (see Figure 1.5). This model can be seen as a generalization of DRRWs introduced in Mauldin *et al.* (1996).

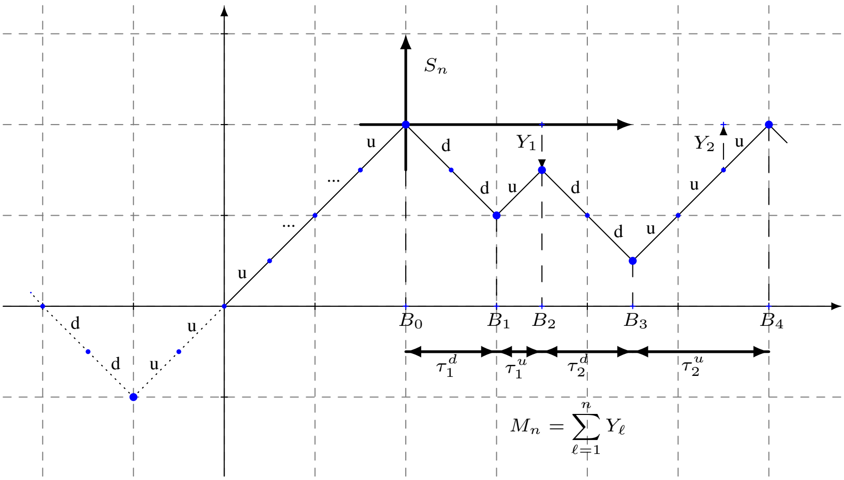


Figure 1.5. *A one-dimensional PRW. For a color version of this figure, see www.iste.co.uk/barbu/data.zip*

Taking different probabilized context trees would lead to different probabilistic impacts on the asymptotic behavior of resulting PRWs. Moreover, the characterization of the recurrent versus transient behavior is difficult in general. We state here exhaustive recurrence criteria for PRWs defined from a double comb.

In order to avoid trivial cases, we assume that S cannot be frozen in one of the two directions with a positive probability. Therefore, we make the following assumption.

ASSUMPTION 1.1 (Finiteness of the length of runs).– For any $\alpha, \beta \in \{u, d\}$, $\alpha \neq \beta$,

$$\lim_{n\to+\infty} \left(\prod_{k=1}^{n} q_{\alpha^k \beta}(\alpha) \right) = 0. \qquad [1.3]$$

Let τ_n^u and τ_n^d be, respectively, the length of the n^{th} rise and of the n^{th} descent.

Then, by a renewal-type property (see Cénac *et al.* 2013, proposition 2.3), $(\tau_n^d)_{n\geq 1}$ and $(\tau_n^u)_{n\geq 1}$ are independent sequences of i.d.d. random variables. Their distribution tails are straightforwardly given by: for any $\alpha, \beta \in \{u, d\}$, $\alpha \neq \beta$ and $n \geq 1$,

$$\mathbb{P}(\tau_1^\alpha \geq n) = \prod_{k=1}^{n-1} q_{\alpha^k \beta}(\alpha). \qquad [1.4]$$

Note that assumption 1.1 amounts to supposing that the *persistence times* τ_n^d and τ_n^u are a.s. finite. The *jump times* (or breaking times) are: $B_0 = 0$ and, for $n \geq 1$,

$$B_{2n} := \sum_{k=1}^{n} \left(\tau_k^d + \tau_k^u\right) \text{ and } B_{2n+1} := B_{2n} + \tau_{n+1}^d. \qquad [1.5]$$

In order to deal with a more tractable random walk built with the possibly unbounded but i.d.d. increments $Y_n := \tau_n^u - \tau_n^d$, we introduce the underlying *skeleton* random walk $(M_n)_{n\geq 1}$, which is the original walk observed at the random times of up-to-down turns:

$$M_n := \sum_{k=1}^{n} Y_k = S_{B_{2n}}. \qquad [1.6]$$

Two main quantities play a key role in the asymptotic behavior, namely the expectations of the lengths of runs: with formula [1.4], let

$$\Theta_d := \mathbf{E}[\tau_1^d] = \sum_{n\geq 1} \prod_{k=1}^{n-1} q_{d^k u}(d) \text{ and } \Theta_u := \mathbf{E}[\tau_1^u] = \sum_{n\geq 1} \prod_{k=1}^{n-1} q_{u^k d}(u). \qquad [1.7]$$

Actually, Θ_d and Θ_u are already discussed in Cénac *et al.* (2013, proposition B1), where it is shown that the driving VLMC of a one-dimensional PRW admits a unique invariant probability measure if and only if $\Theta_d < \infty$ and $\Theta_u < \infty$.

Note that the expectation of Y_1 is well defined in $[-\infty, +\infty]$ whenever at least one of the persistence times τ_1^u or τ_1^d is integrable. Thus, as soon as $\Theta_d < \infty$ or $\Theta_u < \infty$, let

$$\mathbf{d}_M := \mathbf{E}[Y_1] = \underbrace{\Theta_u - \Theta_d}_{\in[-\infty,+\infty]} \qquad [1.8]$$

and

$$\mathbf{d}_S := \frac{\mathbf{E}[\tau_1^u] - \mathbf{E}[\tau_1^d]}{\mathbf{E}[\tau_1^u] + \mathbf{E}[\tau_1^d]} = \frac{\Theta_u - \Theta_d}{\Theta_u + \Theta_d} \in [-1, 1]. \qquad [1.9]$$

An elementary computation shows that $\mathbb{E}(M_n) = n\mathbf{d}_M$ and $\mathbb{E}(S_n) \sim n\mathbf{d}_S$ when n tends to infinity. Thus, $\mathbf{d}_M$ and $\mathbf{d}_S$ appear as asymptotic drifts when the walks $(M_n)_n$ and $(S_n)_n$ respectively, turn out to be transient (see Table 1.1). The behavior of the walk also depends on quantities $J_{\alpha|\beta}$, defined for α and $\beta \in \mathcal{A}, \alpha \neq \beta$ by:

$$J_{\alpha|\beta} := \sum_{n=1}^{\infty} \frac{n\mathbb{P}(\tau_1^\alpha = n)}{\sum_{k=1}^{n} \mathbb{P}(\tau_1^\beta \geq k)}.$$

A complete and usable characterization of the recurrence and the transience of the PRW in terms of the probabilities to persist in the same direction or to switch is given in proposition 1.1. Its proof relies on a criterion of Erickson (1973), applied to the skeleton walk $(M_n)_n$, which is simpler to deal with because its increments are independent.

PROPOSITION 1.1.– Under the non-nullness assumption and assumption 1.1, the random walk $(S_n)_n$ is recurrent or transient as described in Table 1.1.

<table>
<tr><td></td><td colspan="2">$\Theta_u < \infty$</td><td colspan="2">$\Theta_u = \infty$</td></tr>
<tr><td rowspan="2">$\Theta_d < \infty$</td><td rowspan="2">Recurrent
$\mathbf{d}_S = 0$</td><td>Drifting $+\infty$
$\mathbf{d}_S > 0$</td><td colspan="2" rowspan="2">Drifting $+\infty$</td></tr>
<tr><td>Drifting $-\infty$
$\mathbf{d}_S < 0$</td></tr>
<tr><td rowspan="2">$\Theta_d = \infty$</td><td colspan="2" rowspan="2">Drifting $-\infty$</td><td rowspan="2">Recurrent
$J_{u|d} = J_{d|u} = \infty$</td><td>Drifting $+\infty$
$\infty = J_{u|d} > J_{d|u}$</td></tr>
<tr><td>Drifting $-\infty$
$\infty = J_{d|u} > J_{u|d}$</td></tr>
</table>

Table 1.1. *Recurrence versus transience (drifting) for $(S_n)_n$ in dimension one*

The most fruitful situation emerges when both running times τ_1^u and τ_1^d have infinite means. In that case, the recurrence properties of $(S_n)_n$ are related to the behavior of the skeleton random walk $(M_n)_n$ defined in [1.6], the drift of which, $\mathbf{d}_M$, is not defined. Thus, the behavior of $(S_n)_n$ depends on the comparison between the distribution tails of τ_1^u and τ_1^d defined in [1.4], expressed by the quantities $J_{\alpha|\beta}$. Note that the case when both $J_{u|d}$ and $J_{d|u}$ are finite does not appear in the table since it would imply that $\Theta_u < \infty$ and $\Theta_d < \infty$ (see Erickson 1973).

In all three other cases, the drift $\mathbf{d}_S$ is well defined and the PRW is recurrent if and only if $\mathbf{d}_S = 0$. In that case, $\lim_{n\to\infty} \frac{S_n}{n} = \mathbf{d}_S = 0$. Note that modifying one transition q_c transforms a recurrent PRW into a transient one, since $\mathbf{d}_S$ becomes non-zero.

1.3.2. *PRWs in dimension two*

Take the alphabet $\mathcal{A} := \{\mathtt{n}, \mathtt{e}, \mathtt{w}, \mathtt{s}\}$. Here, $(\mathtt{e}, \mathtt{n})$ stands for the canonical basis of $\mathbb{Z}^2$, $\mathtt{w} = -\mathtt{e}$ and $\mathtt{s} = -\mathtt{n}$. Hence, the letters $\mathtt{e}$, $\mathtt{n}$, $\mathtt{w}$ and $\mathtt{s}$ stand for moves to the east, north, west and south, respectively. Having in mind a random walk with increments in $\mathcal{A}$, any word of the form $\alpha\beta$, $\alpha, \beta \in \mathcal{A}, \alpha \neq \beta$ is called a *bend*. For the sake of simplicity, we condition the walk to start a.s. with an $\mathtt{ne}$ bend: $\{X_{-1} = \mathtt{n}, X_0 = \mathtt{e}\}$.

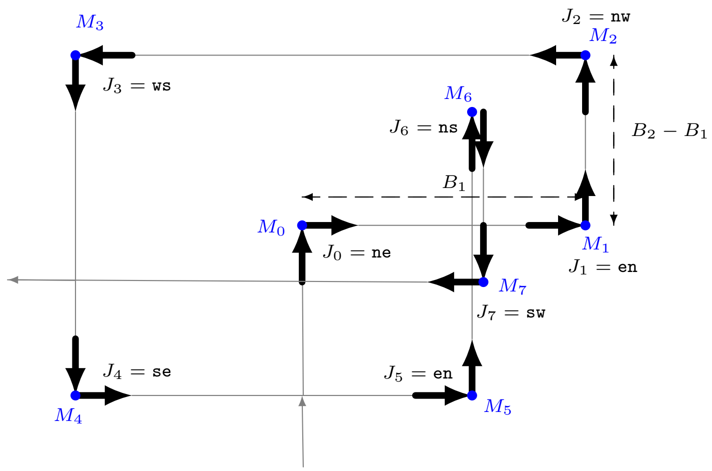


Figure 1.6. *A walk in dimension two. For a color version of this figure, see www.iste.co.uk/barbu/data.zip*

Take a non-null VLMC associated with a quadruple comb on $\mathcal{A}$, as shown in Figure 1.3: the contexts are $\alpha^n\beta$ for $\alpha, \beta \in \mathcal{A}, \alpha \neq \beta$, $n \geq 1$ and the attached probability distributions are denoted by $q_{\alpha^n\beta}$. The two-dimensional PRW $(S_n)_n$ is defined, using this VLMC, as in formula [1.2].

Contrary to the one-dimensional PRWs, as detailed below, the probability to change direction depends on the time spent in the current direction but also on the previous direction. As in dimension one, we intend to avoid that S remains frozen in one of the four directions with a positive probability. Therefore, we make the following assumption, analogous to assumption 1.1 in dimension two.

ASSUMPTION 1.2 (Finiteness of the length of runs).– For any $\alpha, \beta \in \{\mathtt{n}, \mathtt{e}, \mathtt{w}, \mathtt{s}\}$, $\alpha \neq \beta$,

$$\lim_{n \to +\infty} \left(\prod_{k=1}^{n} q_{\alpha^k\beta}(\alpha) \right) = 0. \qquad [1.10]$$

Let $(B_n)_{n\geq 0}$ be the *breaking times* defined inductively by

$$B_0 = 0 \quad \text{and} \quad B_{n+1} = \inf\{k > B_n : X_k \neq X_{k-1}\}. \qquad [1.11]$$

As in dimension one, assumption 1.2 implies that the breaking times B_n are a.s. finite.

Define the so-called *internal chain* $(J_n)_{n\geq 0}$ by $J_0 = \mathtt{ne}$ and, for all $n \geq 1$,

$$J_n := X_{B_{n-1}} X_{B_n}. \qquad [1.12]$$

Let us illustrate these random variables with a small example, in which: $B_1 = 4$, $B_2 = 7$, $J_0 = X_{-1}X_0$, $J_1 = X_{B_0}X_{B_1} = X_0X_4$, $J_2 = X_{B_1}X_{B_2} = X_4X_7$.

n:	-1	0	1	2	3	4	5	6	7
X_n:	n	e	e	e	e	n	n	n	w
		$B_0 = 0$				$B_1 = 4$			$B_2 = 7$
		$J_0 = \mathtt{ne}$				$J_1 = \mathtt{en}$			$J_2 = \mathtt{nw}$

The process $(J_n)_{n\geq 0}$ is an irreducible Markov chain on the set of bends $\mathcal{S} := \{\alpha\beta | \alpha \in \mathcal{A}, \beta \in \mathcal{A}, \alpha \neq \beta\}$. Its Markov kernel is defined by: for every $\beta, \alpha, \gamma \in \mathcal{A}$ with $\beta \neq \alpha$ and $\alpha \neq \gamma$,

$$P(\beta\alpha; \alpha\gamma) := \sum_{n=1}^{\infty} \left(\prod_{k=1}^{n-1} q_{\alpha^k\beta}(\alpha) \right) q_{\alpha^n\beta}(\gamma), \qquad [1.13]$$

the numbers $P(\alpha\beta, \gamma\delta)$ being 0 for every couple of bends not of the previous form. Remark that the non-nullness assumption (see definition 1.4) implies the irreducibility of $(J_n)_n$ and its aperiodicity. The state space $\mathcal{S}$ is finite so that $(J_n)_n$ is positive recurrent: it admits a unique invariant probability measure π_J.

Denote $T_0 = 0$ and $T_{n+1} := B_{n+1} - B_n$ for every $n \geq 0$. These waiting times (also called *persistence times*) are not independent, contrary to the one-dimensional case. The *skeleton random walk* $(M_n)_{n\geq 0}$ on $\mathbb{Z}^2$ – which is the PRW observed at the breaking times – is then defined as

$$M_n := S_{B_n} = \sum_{i=1}^{n} \left(\sum_{k=B_{i-1}+1}^{B_i} X_k \right) = \sum_{i=1}^{n} (B_i - B_{i-1}) X_{B_i}. \qquad [1.14]$$

Note that $(M_n)_n$ is generally not a classical RW with i.d.d. increments. Nevertheless, taking into account the additional information given by the internal

Markov chain $(J_n)_n$, then $(J_n, M_n)_n$ is a Markov additive process (see Çinlar 1972) as it will appear in section 1.5.

Here, $(J_n)_n$ is positive recurrent but this does not imply the recurrence of $(S_n)_n$ or $(M_n)_n$. Moreover, $(S_n)_n$ and $(M_n)_n$ may have different behaviors. Explicit, necessary and sufficient conditions for the recurrence of $(M_n)_n$ in terms of characteristic functions and convergence of suitable series are given in Cénac *et al.* (2020, theorem 2.1). The following proposition states a dichotomy between some recurrence versus transience phenomenon.

THEOREM 1.1.– Under non-nullness assumption, the following dichotomy holds:

i) the series $\sum_n \mathbb{P}(M_n = 0)$ diverges if and only if the process $(M_n)_n$ is recurrent in the following sense:

$$\exists r > 0, \;\; \mathbb{P}\left(\liminf_{n\to\infty} \|M_n\| < r\right) = 1.$$

ii) the series $\sum_n \mathbb{P}(M_n = 0)$ converges if and only if the process $(M_n)_n$ is transient in the following sense:

$$\mathbb{P}\left(\lim_{n\to\infty} \|M_n\| = \infty\right) = 1.$$

Does the recurrence (respectively, the transience) of $(M_n)_n$ and $(S_n)_n$ occur at the same time? The answer to this 20-year-old question is no:

THEOREM 1.2 (Definitive invalidation of the conjecture in Mauldin *et al.* 1996).– There exist recurrent PRWs $(S_n)_n$ having an associated transient skeleton $(M_n)_n$.

Supposing that the persistence time distributions are horizontally and vertically symmetric is a natural necessary condition for the random walk $(S_n)_n$ to be recurrent. One example is given by the DRRW, originally introduced in Mauldin *et al.* (1996) (see Figure 1.7). Some particular values of the transition probabilities $q_{\alpha^n\beta}$ provide counterexamples. It is shown in Cénac *et al.* (2020) that the corresponding distributions of the persistence times must be non-integrable. In section 1.5, this non-integrability will be related to the non-existence of any invariant probability measure for the driving VLMC.

1.4. VLMC: existence of stationary probability measures

Consider a VLMC denoted by $U = (U_n)_{n\geq 0}$, defined by a pair $(\mathcal{T}, q)$ where $\mathcal{T}$ is a context tree on an alphabet $\mathcal{A}$ and $q = (q_c)_{c\in\mathcal{C}}$ a family of probability measures on $\mathcal{A}$, indexed by the contexts of $\mathcal{T}$. A probability measure π on $\mathcal{R}$ is *stationary* or *invariant* (with regard to U) whenever π is the distribution of every U_n as soon as it is the distribution of U_0. The question of interest consists here of finding conditions on

$(\mathcal{T}, q)$ for the process to admit at least one – or a unique one – stationary probability measure. The heuristic presentation aims to show how combinatoric objects, namely the α-*LIS* of contexts, and conditional probabilities, the *cascades*, naturally emerge.

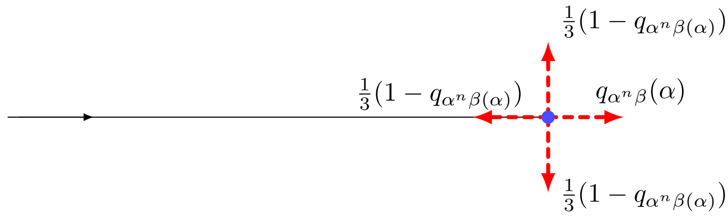


Figure 1.7. *The original directionally reinforced random walk (DRRW). For a color version of this figure, see www.iste.co.uk/barbu/data.zip*

Assume that π is a stationary probability measure on $\mathcal{R}$:

– First step: finite words. Since $\mathcal{R}$ is endowed with the cylinder σ-algebra, π is determined by its values $\pi(w\mathcal{R})$ on the cylinders $w\mathcal{R}$, where w runs over all finite words on $\mathcal{A}$.

– Second step: longest internal suffixes of words. Assume that e is a finite non-internal word and take $a \in \mathcal{A}$. Then, its pref is well defined and, because of formula [1.1], since π is stationary,

$$\pi(\alpha e\mathcal{R}) = q_{\mathrm{pref}(e)}(\alpha) \times \pi(e\mathcal{R}). \qquad [1.15]$$

Iterating this formula as far as possible leads to the following definitions. Consider any non-empty finite word w. It is uniquely decomposed as $w = p\alpha s = \beta_1\beta_2\beta_3 \cdots \beta_\ell \alpha s$, where the α and the β are letters and s is the *longest internal suffix* of w. The integer ℓ is non-negative and $p = \beta_1\beta_2 \cdots \beta_\ell$ is a prefix of w that may be empty – in which case $\ell = 0$.

DEFINITION 1.5 (Lis and α-LIS).– *With these notations, the longest internal suffix s is shortened as the* lis *of w. The word αs is called the* α-LIS *of w.*

DEFINITION 1.6 (Cascade).– *With the notation above, the* cascade *of w is the product*

$$\mathrm{casc}(w) = q_{\mathrm{pref}(\beta_2\cdots\beta_\ell\alpha s)}(\beta_1) q_{\mathrm{pref}(\beta_3\cdots\beta_\ell\alpha s)}(\beta_2) \cdots q_{\mathrm{pref}(\alpha s)}(\beta_\ell). \qquad [1.16]$$

Note that this definition makes sense because all the $\beta_k \cdots \beta_\ell \alpha s$ are non-internal words, $k \geq 2$. Moreover, if $w = \alpha s$ where s is internal, then $\ell = 0$ and $\mathrm{casc}(w) = 1$.

With these definitions, iterating formula [1.15] leads to the following equality, known as the *cascade formula*: for every non-empty finite word w having αs as an α-LIS,

$$\pi\left(w\mathcal{R}\right) = \text{casc}(w) \times \pi\left(\alpha s\mathcal{R}\right). \qquad [1.17]$$

This shows that π is determined by its values on words of the form αs where s is internal and $\alpha \in \mathcal{A}$.

– Third step: finite contexts. Assume that s is an internal word and that $\alpha \in \mathcal{A}$. It is shown in Cénac *et al.* (2018a) that a stationary probability measure never charges infinite words so that, by disjoint union,

$$\pi\left(\alpha s\mathcal{R}\right) = \sum_{\substack{c:\ \text{finite context} \\ c = s\cdots}} \pi\left(\alpha c\mathcal{R}\right) = \sum_{\substack{c:\ \text{finite context} \\ c = s\cdots}} q_c(\alpha)\pi\left(c\mathcal{R}\right). \qquad [1.18]$$

Note that the set of indices may be infinite but the family is summable because π is a finite measure. This shows that π is entirely determined by its values $\pi\left(c\mathcal{R}\right)$ on the finite contexts.

– Fourth step: α-LIS of finite contexts. Cascade formula [1.16] applied to any finite context c (contexts are non-empty words) is written as $\pi\left(c\mathcal{R}\right) = \text{casc}(c)\pi\left(\alpha_c s_c\mathcal{R}\right)$, where $\alpha_c s_c$ is the α-LIS of c. Denote by $\mathcal{S} = \mathcal{S}\left(\mathcal{T}\right)$ the set of finite context α-LIS:

$$\mathcal{S} = \left\{\alpha_c s_c :\ c \text{ finite context}\right\}.$$

If s is an internal word and if $\alpha \in \mathcal{A}$, then formula [1.18] leads to

$$\pi\left(\alpha s\mathcal{R}\right) = \sum_{\substack{c:\ \text{finite context} \\ c = s\cdots}} \text{casc}\left(\alpha c\right)\pi\left(\alpha_c s_c\mathcal{R}\right), \qquad [1.19]$$

showing that π is determined by its values $\pi\left(\alpha_c s_c\mathcal{R}\right)$ on $\mathcal{S}$.

– Fourth step: a (generally infinite) linear system. When w and v are finite words and when $\alpha s \in \mathcal{S}$, the notation

$$w = v\cdots = \cdots[\alpha s]$$

stands for: w has v as a prefix and αs as an α-LIS. Writing formula [1.19] for every $\alpha s \in \mathcal{S}$ and grouping in each of them the terms that arise from contexts having the same α-LIS leads to the following square system (at most countably many unknowns $\pi\left(\alpha s\mathcal{R}\right)$ and as many equations):

$$\forall \alpha s \in \mathcal{S},\ \pi\left(\alpha s\mathcal{R}\right) = \sum_{\beta t \in \mathcal{S}} \pi\left(\beta t\mathcal{R}\right) \left(\sum_{\substack{c:\ \text{finite context} \\ c = s\cdots = \cdots[\beta t]}} \text{casc}\left(\alpha c\right) \right). \qquad [1.20]$$

DEFINITION 1.7 (Matrix Q).– *When $\mathcal{T}$ is a context tree having $\mathcal{S}$ as a context α-LIS set, $Q = Q(\mathcal{T})$ is the $\mathcal{S}$-indexed square matrix defined by:*

$$\forall \alpha s, \beta t \in \mathcal{S},\ Q_{\beta t, \alpha s} = \sum_{\substack{c:\ \text{finite context} \\ c = s \cdots = \cdots [\beta t]}} \operatorname{casc}(\alpha c) \in [0, +\infty]. \qquad [1.21]$$

Thus, system [1.20] tells us that, when π is a stationary measure, the row-vector $(\pi(\alpha s \mathcal{R}))_{\alpha s \in \mathcal{S}}$ appears as a left-fixed vector of the matrix Q.

DEFINITION 1.8 (Cascade series).– *For every $\alpha s \in \mathcal{S}$, denote*

$$\kappa_{\alpha s} = \sum_{\substack{c:\ \text{finite context} \\ c = \cdots [\alpha s]}} \operatorname{casc}(c) \in [0, +\infty].$$

When this series is summable, one says that the cascade series of αs converges. *Whenever the cascades series of all $\alpha s \in \mathcal{S}$ converge, one says that* the cascade series (of the VLMC) converge*s*.

Note that the convergence of (all) the cascade series is sufficient to guarantee the finiteness of Q's entries. Actually, for a general VLMC, as it is made precise in Cénac *et al.* (2018a), the convergence of the cascade series appears as a pivot condition when dealing with existence and unicity of a stationary probability measure. In this chapter, we state a necessary and sufficient condition for a special kind of VLMC: the *stable* ones that have a finite $\mathcal{S}$. The following proposition is proven in Cénac *et al.* (2018a).

PROPOSITION 1.2.– Let $\mathcal{T}$ be a context tree. The following conditions are equivalent:

i) $\forall \alpha \in \mathcal{A}, \forall w \in \mathcal{W}, \alpha w \in \mathcal{T} \Longrightarrow w \in \mathcal{T}$;

ii) if c is a finite context and $\alpha \in \mathcal{A}$, then αc is non-internal;

iii) $\mathcal{T} \subseteq \mathcal{A}\mathcal{T} = \{\alpha w,\ \alpha \in \mathcal{A},\ w \in \mathcal{T}\}$;

iv) for any VLMC $(U_n)_n$ associated with $\mathcal{T}$, the process $(\operatorname{pref}(U_n))_{n \in \mathbb{N}}$ is a Markov chain that has the set of contexts as a state space.

The context tree is referred to as *stable* whenever one of these conditions is fulfilled.

It turns out that the stability of $\mathcal{T}$ together with the non-nullness of the VLMC imply both stochasticity and irreducibility of the matrix Q. Consequently, in the simple case where Q is a finite-dimensional matrix, there exists (because of stochasticity) a unique (because of irreducibility) left-fixed vector for Q. As a result of a much more general result proven in Cénac *et al.* (2018a), this implies the existence and unicity of a stationary probability measure for the VLMC, as stated below.

THEOREM 1.3.– Let $(\mathcal{T}, q)$ be a non-null stable probabilized context tree. If $\#\mathcal{S} < \infty$, then the following are equivalent:

1) the VLMC associated with $(\mathcal{T}, q)$ has a unique stationary probability measure;

2) the cascade series converge (see definition 1.8).

Note that in the non-stable case, the matrix Q is generally not stochastic nor is it even substochastic. Note also that, even in the stable case, when $\#\mathcal{S} = \infty$, the matrix Q may be stochastic, irreducible and positive recurrent, while the VLMC does not admit any stationary probability measure. One can find such an example in remark 3.16, page 20 in Cénac *et al.* (2020), built with a "left comb of left comb".

1.5. Where VLMC and PRW meet

On the one hand, a VLMC is defined by its context tree and its transition probability distributions q_c – in particular the double and the quadruple combs that are stable trees with finitely many context α-LIS.

Necessary and sufficient conditions of existence and uniqueness of stationary probability measures are given in terms of cascade series. On the other hand, for PRW (defined from VLMC), recurrence properties are written in terms of persistence times. Our aim is to build a bridge between these two families of objects and properties. The meeting point turns out to be the semi-Markov processes of α-LIS and bends.

1.5.1. *Semi-Markov chains and Markov additive processes*

Semi-Markov chains are defined following Barbu and Limnios (2008) because of so-called Markov renewal chains (MRCs).

DEFINITION 1.9 (Markov renewal chain).– *A Markov chain $(J_n, T_n)_{n\geq 0}$ with state space $\mathcal{E} \times \mathbb{N}$ is called a (homogeneous) MRC whenever the transition probabilities satisfy:* $\forall n \in \mathbb{N}$, $\forall a, b \in \mathcal{E}$, $\forall j, k \in \mathbb{N}$,

$$\begin{aligned}\mathbb{P}\left(J_{n+1} = b, T_{n+1} = k \middle| J_n = a, T_n = j\right) &= \mathbb{P}\left(J_{n+1} = b, T_{n+1} = k \middle| J_n = a\right)\\ &=: p_{a,b}(k)\end{aligned}$$

and $\forall a, b \in \mathcal{E}$, $p_{a,b}(0) = 0$. *For such a chain, the family* $p = \left(p_{a,b}(k)\right)_{a,b\in\mathcal{A},k\geq 1}$ *is referred to as its* semi-Markov kernel.

DEFINITION 1.10 (Semi-Markov chain).– *Let* $(J_n, T_n)_{n\geq 0}$ *be an MRC with state space* $\mathcal{E} \times \mathbb{N}$. *Assume that* $T_0 = 0$. *For any* $n \in \mathbb{N}$, *let* B_n *be defined by*

$$B_n = \sum_{i=0}^{n} T_i.$$

The semi-Markov chain *associated with* $(J_n, T_n)_{n\geq 0}$ *is the* $\mathcal{E}$*-valued process* $(Z_j)_{j\geq 0}$ *defined by*

$$\forall j \quad \textit{such that} \quad B_n \leq j < B_{n+1}, \qquad Z_j = J_n.$$

Note that the sequence $(B_n)_{n\geq 0}$ is a.s. increasing because of the assumption $p_{a,b}(0) = 0$ (instantaneous transitions are not allowed) that guarantees that $T_n \geq 1$ a.s. for any $n \geq 1$.

The B_n are *jump times*, the T_n are *sojourn times* in a given state and Z_j stagnates at a same state between two successive jump times. The process $(J_n)_n$ is called the *internal (underlying) chain* of the semi-Markov chain $(Z_n)_n$.

The previous definitions make transitions to the same state between time n and time $n + 1$ possible. Nevertheless, one can boil down to the case where $p_{a,a}(k) = 0$ for all $a \in \mathcal{E}, k \in \mathbb{N}$ (see the details in Cénac *et al.* 2018a).

A close notion, *Markov additive processes*, can be found in Çinlar (1972).

1.5.2. *PRWs induce semi-Markov chains*

Let us start with one-dimensional PRW, as defined in section 1.3.1. In this case, at each time $j, j \geq 0$, the increment X_j of the walk S takes d or u as a value (see Figure 1.5). Let us see that $(X_j)_{j\geq 0}$ is a semi-Markov chain, starting from $X_0 = d$. Remember that B_n denotes the nth jump times – see equation [1.5]. Define then $(J_n)_n$ by

$$J_n := X_{B_n}. \qquad [1.22]$$

Moreover, let T_n be the nth waiting time, namely $T_0 = 0$ and, for $n \geq 1$,

$$T_n = B_n - B_{n-1}.$$

These waiting times are related to the persistence times τ by the following formulas: for all $k \geq 1$,

$$T_{2k} := \tau_k^u \text{ and } T_{2k-1} := \tau_k^d. \qquad [1.23]$$

With these notations, $(J_n, T_n)_{n\geq 0}$ is an MRC and its semi-Markov kernel is written as: $\forall \alpha, \beta \in \{u, d\}, \alpha \neq \beta, \forall k \geq 1$,

$$p_{\alpha,\beta}(k) = \left(\prod_{j=1}^{k-1} q_{\alpha^j \beta}(\alpha) \right) q_{\alpha^k \beta}(\beta), \qquad [1.24]$$

as can be straightforwardly checked. Moreover, assumption 1.1 guarantees that the T_n are a.s. finite. Besides, formulas [1.7] are written as

$$\mathbb{E}\left(T_{2k}\right) = \Theta_u \text{ and } \mathbb{E}\left(T_{2k+1}\right) = \Theta_d.$$

The situation in dimension one is summarized by the following proposition.

PROPOSITION 1.3.– For a PRW in dimension one, defined by a VLMC associated with a double comb, the sequence $(X_j)_j$ of the increments is an $\mathcal{A}$-valued semi-Markov chain with MRC $(J_n, T_n)_n$ as defined in [1.22] and [1.23] and its semi-Markov kernel is given by equation [1.24].

Let us now deal with the two-dimensional PRW defined in section 1.3.2. At each time $j, j \geq 0$, the increment X_j of the walk S takes $\mathtt{n}, \mathtt{e}, \mathtt{w}$ or $\mathtt{s}$ as a value. But, as already noted, changing direction depends on the time spent in the current direction but also, contrary to the one-dimensional PRWs, on the previous direction. In other words, the bends play the main role. This gives rise to the process $\left(Z_j\right)_j$, valued in the set of bends $\{\alpha\beta : \alpha, \beta \in \mathcal{A}, \alpha \neq \beta\}$, defined in the following manner: $Z_0 = X_{-1}X_0 = \mathtt{ne}$ and, for $j \geq 1$, $Z_j = \alpha\beta$ if and only if $X_j = \beta$ and the first letter distinct from β in the sequence $X_{j-1}, X_{j-2}, X_{j-3}, \cdots$ is α. Let us see that $(Z_j)_{j\geq 0}$ is a semi-Markov chain. Use here notations $(J_n)_n$, $(B_n)_n$ and $(T_n)_n$ of section 1.3.2.

Note that, contrary to the one-dimensional case, the waiting times T_n are not independent. Nevertheless, $(J_n, T_n)_{n\geq 0}$ is an MRC with semi-Markov kernel

$$p_{\beta\alpha,\alpha\gamma}(k) := \left(\prod_{j=1}^{k-1} q_{\alpha^j \beta}(\alpha) \right) q_{\alpha^k \beta}(\gamma), \qquad [1.25]$$

as can be straightforwardly checked. Summarizing, the following proposition holds.

PROPOSITION 1.4.– For a PRW in dimension two, defined by a VLMC associated with a quadruple comb, the sequence $(Z_j)_j$ of the bends is a semi-Markov chain with MRC $(J_n, T_n)_n$ as defined in section 1.3.2. Its semi-Markov kernel is given by equation [1.25]. In addition, $(J_n, B_n)_n$ is a Markov additive process.

1.5.3. *Semi-Markov chain of the α-LIS in a stable VLMC*

In this section, let us consider a more general case than a double comb or a quadruple comb, namely a stable VLMC. In this case, there is always a semi-Markov chain induced by the process $(U_n)_n$, as described in the following.

Let $(U_n)_{n\geq 0}$ be a stable non-null VLMC such that the series of cascades converge (see definition 1.8). Recall that $\mathcal{S}$ denotes the set of context α-LIS of the VLMC. Let $(C_n)_{n\geq 0}$ be the sequence of contexts, and for $n \geq 0$, let Z_n be the α-LIS of C_n:

$$C_n = \mathrm{pref}(U_n) \quad \text{ and } \quad Z_n = \alpha_{C_n} s_{C_n}.$$

PROPOSITION 1.5.– Let $(B_n)_{n\geq 0}$ be the increasing sequence of times defined by $B_0 = 0$ and for any $n \geq 1$,

$$B_n = \inf\{k > B_{n-1}, |C_k| \leq |C_{k-1}|\} = \inf\{k > B_{n-1}, C_k \in \mathcal{S}\}$$

and let $T_n = B_n - B_{n-1}$ for $n \geq 1$ and $T_0 = 0$. For any $n \geq 0$, let $J_n = Z_{B_n}$. Then

i) B_n and T_n are a.s. finite and for $\alpha s \in \mathcal{S}$, $\mathbb{E}\left(T_n \middle| J_n = \alpha s\right) = \kappa_{\alpha s}$;

ii) $(Z_n)_{n\geq 0}$ is an $\mathcal{S}$-valued semi-Markov chain associated with the MRC $(J_n, T_n)_{n\geq 0}$;

iii) the associated semi-Markov kernel writes: $\forall \alpha s, \beta t \in \mathcal{S}$, $\forall k \geq 1$,

$$p_{\alpha s,\beta t}(k) = \sum_{\substack{c\in\mathcal{C},\ c=t\cdots \\ c=\cdots[\alpha s] \\ |c|=|\alpha s|+k-1}} \mathrm{casc}\,(\beta c)\,.$$

The proof is detailed in Cénac *et al.* (2018a). It relies on the way the VLMC grows between two jump times: at the beginning, letters are added to the current context C_n, the α-LIS does not change and the length of the current context increases one by one. At a certain time (a.s. finite), adding a letter to the current context does not provide a context any more but an external node. At this moment, it happens (it is not trivial and only holds for a stable context tree) that

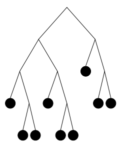
i) the α-LIS of the current context is renewed;

ii) the length of the current context does not grow;

iii) the current context begins by a lis.

These mechanisms explain the expressions of B_n and the formula giving the semi-Markov kernel.

REMARK 1.3.– In the very particular case of the double or quadruple comb, the semi-Markov chain $(Z_n)_n$ contains as much information as the chain $(U_n)_n$. But in general, the semi-Markov chain $(Z_n)_n$ contains less information than the chain $(U_n)_n$. To illustrate this, here is an example with a finite context tree.



α-LIS αs	contexts having αs as an α-LIS
10	10,010,110,0010,0110
000	000
111	111,0111
0011	0011

In this example, 0010 and 0110 are two contexts of the same length, with the same α-LIS 10 and beginning by the same lis 0. Hence, if we know that $J_n = 10$, $B_{n+1} - B_n = 3$ and $J_{n+1} = 10$, then Z_j is uniquely determined between the two successive jump times, whereas there are two possibilities to reconstruct the VLMC $(U_n)_n$. With the notations of proposition 1.5, there are two cascade terms in $p_{10,10}(3)$:

$$\begin{aligned} p_{10,10}(3) &= \mathbb{P}\left(C_{B_n+1} = 010, C_{B_n+2} = 0010, C_{B_n+3} = 10010 | C_{B_n} = 10\right) \\ &\quad + \mathbb{P}\left(C_{B_n+1} = 110, C_{B_n+2} = 0110, C_{B_n+3} = 10110 | C_{B_n} = 10\right) \\ &= q_{10}(0)q_{010}(0)q_{0010}(1) + q_{10}(1)q_{110}(0)q_{0110}(1) \\ &= \text{casc}(10010) + \text{casc}(10110). \end{aligned}$$

1.5.4. *The meeting point*

Summing up, the announced close encounter can be done with the following (commutative) diagram, together with the following explanations.

$$\begin{tikzcd} \text{VLMC } (U_n)_n \arrow[r, "D"] \arrow[d, "L"] & \text{PRW } (S_n)_n \arrow[d, "B"] \\ \text{MRC } \left(J_n^V, T_n\right)_n \arrow[d, "S_V"] & \text{MAP } \left(J_n^W, M_n\right)_n \arrow[l, "N"'] \arrow[d, "S_W"] \\ \text{Semi-Markov } \left(Z_n^V\right)_n \arrow[r, leftrightarrow, "R"] & \text{Semi-Markov } \left(Z_n^W\right)_n \end{tikzcd} \qquad [1.26]$$

The mapping D consists of defining the PRW from the VLMC: the random increments of the PRW are the initial letters of a VLMC. With the notations above, $S_n = \sum_{0 \leq k \leq n} X_k$, where X_k is the initial letter of U_k.

The mapping L associates with a VLMC the process of its successive different α-LIS that turns out to be an MRC when considered together with its jump times T_n (see section 1.5.3). Here, J_n^V is the nth distinct α-LIS of the successive right-infinite words $U_0, U_1, U_2, \cdots$ and T_n is the length of the nth run of identical letters in the sequence $X_0, X_1 X_2, \cdots$ The power V refers to the VLMC.

The mapping B associates with a PRW $(S_n)_n$ the process of its successive different bends (changes of directions). With our notations, J_n^W is the nth distinct bend and M_n is the value of S at the precise moment when the nth bend J_n^W occurs (see section 1.5.2). The power W refers to the PRW.

The mapping S_V only consists of defining a semi-Markov process from an MRC, as stated in section 1.5.1. The mapping S_W is defined in the same manner: it maps a MAP $\left(J_n^W, M_n\right)_n$ to the semi-Markov chain of the MRC $\left(J_n^W, M_n - M_{n-1}\right)_n$.

The mapping N acts on the first coordinate by reversing words: $J_n^V = \overline{J_n^W}$. The notation $\overline{w}$ stands for the reversed word of w: $\overline{ab} = ba$. For the second coordinate, remark first that $M_n - M_{n-1}$ is always of the form $k\alpha$, where k is a positive integer and α an increment vector. The integer T_n is this k.

Finally, the mapping R is simply the reversing of words: $Z_n^V = \overline{Z_n^W}$.

In fact, in these particular situations (double and quadruple combs), the composition $S_V \circ L$ is a bijection (see remark 1.3). Therefore, all these mappings are also one-to-one, showing that all these processes are essentially equivalent.

Now when our different processes are related, let us translate the parameters, properties and assumptions that come from the VLMC world in terms of PRW and vice versa.

Dimension one

The PRW in dimension one is driven by a VLMC based on the so-called double comb, as it was defined in example 1.2. The contexts of this tree are the $u^k d$, which have ud as an α-LIS, and the $d^k u$, which have du as an α-LIS ($k \geq 1$ for both families of contexts). The cascades of the contexts write

$$\operatorname{casc}\left(u^k d\right) = \prod_{j=1}^{k-1} q_{u^j d}(u) \quad \text{and} \quad \operatorname{casc}\left(d^k u\right) = \prod_{j=1}^{k-1} q_{d^j u}(d)$$

and there are two cascade series

$$\kappa_{ud} = \sum_{k \geq 1} \operatorname{casc}\left(u^k d\right) \quad \text{and} \quad \kappa_{du} = \sum_{k \geq 1} \operatorname{casc}\left(d^k u\right).$$

Theorem 1.3 guarantees that, under non-nullness assumption, this VLMC admits an invariant probability measure if and only if $\kappa_{ud} < \infty$ and $\kappa_{du} < \infty$. Since the double comb is a very simple context tree, one can also make a direct computation that leads to the following result: *a non-null double comb VLMC admits a σ-finite stationary measure if and only if* $\text{casc}\left(u^k d\right) \to 0$ *and* $\text{casc}\left(d^k u\right) \to 0$ *when k tends to infinity*.

It turns out that, on the side of the one-dimensional PRW, assumption 1.1 and the expectations of the persistence times τ_1^u and τ_1^d are functions of these cascades so that one can relate the above properties of the VLMC to the results of section 1.3.1 on one-dimensional PRW. The expectations of the waiting times are exactly the sums of cascades: $\kappa_{ud} = \Theta_u$ and $\kappa_{du} = \Theta_d$.

Finally, one can assert:

$$\begin{array}{ccc}
\begin{pmatrix} \text{casc}\left(u^k d\right) \underset{k\to\infty}{\longrightarrow} 0 \\ \text{and} \\ \text{casc}\left(d^k u\right) \underset{k\to\infty}{\longrightarrow} 0 \end{pmatrix} & \Longleftrightarrow & \text{assumption 1.1} \\
\Updownarrow & & \Updownarrow \\
\begin{pmatrix} \text{The VLMC admits} \\ \text{a } \sigma\text{−finite} \\ \text{invariant measure} \end{pmatrix} & \Longleftrightarrow & \begin{pmatrix} \tau_1^u \text{ and } \tau_1^d \\ \text{are a.s. finite} \end{pmatrix}
\end{array}$$

and

$$\begin{array}{ccc}
\begin{pmatrix} \displaystyle\sum_{k\geq 1} \text{casc}\left(u^k d\right) < \infty \\ \text{and} \\ \displaystyle\sum_{k\geq 1} \text{casc}\left(d^k u\right) < \infty \end{pmatrix} & \Longleftrightarrow & \begin{pmatrix} \tau_1^u \text{ and } \tau_1^d \\ \text{are integrable} \end{pmatrix} \\
\Updownarrow & & \\
\begin{pmatrix} \text{The VLMC admits} \\ \text{a unique probability} \\ \text{invariant measure} \end{pmatrix} & &
\end{array}$$

The link between recurrence or transience of the PRW and the behavior of the VLMC is only partial. For instance, the PRW may be recurrent while there is no invariant probability measure for the VLMC. The PRW may even be transient while the VLMC admits an invariant probability measure (see Table 1.1).

Dimension two

The PRW in dimension two is driven by a VLMC based on the so-called quadruple comb, as it is defined in example 1.2. Here, the contexts are the $\alpha^k\beta$, where $\alpha, \beta \in \mathcal{A} = \{\mathtt{n}, \mathtt{e}, \mathtt{w}, \mathtt{s}\}$, $\alpha \neq \beta$, $k \geq 1$. The α-LIS of the context $\alpha^k\beta$ is $\alpha\beta$, and its cascade is written as

$$\text{casc}\left(\alpha^k\beta\right) = \prod_{i=1}^{k-1} q_{\alpha^i\beta}(\alpha).$$

Therefore, there are 12 cascade series, namely

$$\kappa_{\alpha\beta} = \sum_{k=1}^{\infty} \text{casc}\left(\alpha^k\beta\right),\ \alpha, \beta \in \mathcal{A},\ \alpha \neq \beta. \qquad [1.27]$$

As in dimension one, since the quadruple comb is a stable context tree having a finite set of context α-LIS, the non-null VLMC that drives the two-dimensional PRW admits a unique stationary probability measure if and only if the 12 cascade series [1.27] converge. This is a consequence of theorem 1.3 and, here again, due to the simplicity of the quadruple comb, one can directly check that *a non-null quadruple-comb VLMC admits a σ-finite stationary measure if and only if* casc $\left(\alpha^k\beta\right) \to 0$ *when k tends to infinity for every $\alpha, \beta \in \mathcal{A}$, $\alpha \neq \beta$.*

The transition matrix of the Markov process $(J_n)_n$ of the PRW bends, denoted by P in formula [1.13], is also written as

$$P(\beta\alpha, \alpha\gamma) = \sum_{n\geq 1} \text{casc}\left(\gamma\alpha^n\beta\right)$$

and all other entries vanish. Relating this expression to definition [1.21] of the Q-matrix of the VLMC leads to the following:

$$P(\beta\alpha, \alpha\gamma) = Q_{\alpha\beta,\gamma\alpha} \qquad [1.28]$$

so that, up to the re-ordering that consists of reversing the indices $\alpha\beta \rightsquigarrow \beta\alpha$, the stochastic matrices P and Q are the same ones. Note that, since the quadruple comb is stable, the process of the α-LIS of the VLMC is Markovian and Q is its transition matrix. Referring to the commutative diagram [1.26], formula [1.28] amounts to saying that the Markov chains $\left(J_n^V\right)_n$ and $\left(\overline{J_n^W}\right)_n$ are identical.

In terms of persistence times of the PRW versus stationary measures for the VLMC, the properties stated in section 1.3.2 show that the following equivalences hold:

$$\begin{array}{ccc}
\begin{pmatrix} \text{for all } \alpha, \beta \in \mathcal{A}, \alpha \neq \beta, \\ \operatorname{casc}\left(\alpha^k \beta\right) \underset{k\to\infty}{\longrightarrow} 0 \end{pmatrix} & \Longleftrightarrow & \text{assumption 1.2} \\
\Updownarrow & & \Updownarrow \\
\begin{pmatrix} \text{The VLMC admits} \\ \text{a } \sigma\text{−finite} \\ \text{invariant measure} \end{pmatrix} & \Longleftrightarrow & \left(\forall n,\ T_n \text{ is a.s. finite}\right)
\end{array}$$

and

$$\begin{array}{ccc}
\begin{pmatrix} \text{for all } \alpha, \beta \in \mathcal{A}, \alpha \neq \beta, \\ \displaystyle\sum_{k\geq 1} \operatorname{casc}\left(\alpha^k \beta\right) < \infty \end{pmatrix} & \Longleftrightarrow & (\forall n,\ T_n \text{ is integrable}) \\
\Updownarrow & & \\
\begin{pmatrix} \text{The VLMC admits} \\ \text{a unique probability} \\ \text{invariant measure} \end{pmatrix} & &
\end{array}$$

The counterexample cited in theorem 1.2 is supported by these equivalences: an example of recurrent two-dimensional PRW having a transient skeleton $(M_n)_n$ cannot be found without assuming that the T_n are a.s. finite but non-integrable, as shown in Cénac *et al*. (2020). Reading the above equivalences shows that such a PRW must be driven by a VLMC the series of cascades of which diverge, while their general terms tend to zero at infinity.

1.6. References

Barbu, V.S., and Limnios, N. (2008). *Semi-Markov Chains and Hidden Semi-Markov Models Toward Applications*. Springer, New York.

Bejerano, G., and Yona, G. (2001). Variations on probabilistic suffix trees: Statistical modeling and prediction of protein families. *Bioinformatics*, 17(1), 23–43.

Busch, J., Ferrari, P., Flesia, A., Fraiman, R., Grynberg, S., and Leonardi, F. (2009). Testing statistical hypothesis on random trees and applications to the protein classification problem. *Ann. Appl. Stat.*, 3(2), 542–563.

Çinlar, E. (1972). Markov additive processes I, II, *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 24(85–93), 95–121.

Cénac, P., Chauvin, B., Paccaut, F., and Pouyanne, N. (2012). Variable length Markov chains and dynamical sources. *Séminaire de Probabilités XLIV, Lecture Notes in Mathematics*, 2046, 1–39.

Cénac, P., Chauvin, B., Herrmann, S., and Vallois, P. (2013). Persistent random walks, variable length Markov chains and piecewise deterministic Markov processes. *Markov Processes and Related Fields*, 19(1), 1–50.

Cénac, P., Chauvin, B., Paccaut, F., and Pouyanne, N. (2018a). Characterization of stationary probability measures for Variable Length Markov Chains [Online]. Available at: https://hal.archives-ouvertes.fr/hal-01829562

Cénac, P., Le Ny, A., De Loynes, B., and Offret, Y. (2018b). Persistent random walks I: Recurrence versus transience. *J. Theoret. Probab.*, 31(1), 232–243.

Cénac, P., Chauvin, B., Noûs, C., Paccaut, F., Pouyanne, N. (2020). Variable Length Memory Chains: Characterization of stationary probability measures [Online]. Available at: https://arxiv.org/abs/004.07893.

Csiszár, I., and Talata, Z. (2006). Context tree estimation for not necessarily finite memory processes, via BIC and MDL. *IEEE Transactions on Information Theory*, 52(3), 1007–1016.

Doeblin, W., and Fortet, R. (1937). Sur des chaînes à liaisons complètes. *Bull. Soc. Math. France*, 65, 132–148.

Erickson, K.B. (1973). The strong law of large numbers when the mean is undefined. *Trans. Amer. Math. Soc.*, 185, 371–381.

Galves, A., and Leonardi, F. (2008). Exponential inequalities for empirical unbounded context trees. *In and Out of Equilibrium 2: Progress in Probability*, 60, 257–269.

Garivier, A., and Leonardi, F. (2011). Context tree selection: A unifying view. *Stochastic Proc. Appl.*, 121, 2488–2506.

Harris, T.E. (1955). On chains of infinite order. *Pacific J. Math.*, 5, 707–724.

Mauldin, R.D., Monticino, M., and von Weizsäcker, H. (1996). Directionally reinforced random walks. *Adv. Math.*, 117(2), 239–252.

Rissanen, J. (1983). A universal data compression system. *IEEE Trans. Inform. Theory*, 29(5), 656–664.