Characterization of stationary probability measures for Variable Length Markov Chains

Peggy Cénac, Brigitte Chauvin, Frédéric Paccaut, Nicolas Pouyanne

Characterization of stationary probability measures for Variable Length Markov Chains

Peggy Cénac[*], Brigitte Chauvin[†], Frédéric Paccaut[‡] and Nicolas Pouyanne[†]

Tuesday 3rd July, 2018

Abstract

By introducing a key combinatorial structure for words produced by a Variable Length Markov Chain (VLMC), the *longest internal suffix*, precise characterizations of existence and uniqueness of a stationary probability measure for a VLMC chain are given. These characterizations turn into necessary and sufficient conditions for VLMC associated to a subclass of probabilised context trees: the *shift-stable* context trees. As a by-product, we prove that a VLMC chain whose stabilized context tree is again a context tree has at most one stationary probability measure.

MSC 2010: 60J05, 60C05, 60G10.

Keywords: variable length Markov chains, stationary probability measure.

Contents

[*]Université de Bourgogne, Institut de Mathématiques de Bourgogne, IMB UMR 5584 CNRS, 9 rue Alain Savary - BP 47870, 21078 DIJON CEDEX, France.

[†]Laboratoire de Mathématiques de Versailles, UVSQ, CNRS, Université Paris-Saclay, 78035 Versailles, France

[‡]LAMFA, CNRS, UMR 7352, Université de Picardie Jules Verne, 33 rue Saint-Leu, 80039 Amiens, France.

1 Introduction

Infinite random sequences of letters can be viewed as stochastic chains or as strings produced by a source, in the sense of information theory. In this last frame, context tree models have been introduced by Rissanen [1983] as a parsimonious generalization of Markov models to perform data compression. They have been successfully studied and used since then in many fields of applications of probability, including bioinformatics, universal coding, statistics or linguistics.
Statistical use of context tree models requires the possibility of constructing efficient estimators of context trees. For this matter, various algorithms (for example Rissanen's *Context* algorithm, or the so-called "context tree weighting" Willems et al. [1995]) have been developed. The necessity of considering *infinite depth* of context trees has emerged, even for finite memory Markov processes (see for instance Willems [1998]). Moreover, estimators for not necessarily finite memory processes lead to consider infinite context trees. And estimating context trees is harder (namely the consistency is no more ensured), when the depth of the context tree is infinite. This problem in addressed in Csiszár and Talata [2006] and Talata [2013].
In biology, persistent random walks are one possible model to address the question of anomalous diffusions in cells (see for instance Fedotov et al. [2015]). Actually, such random walks are non Markovian and the displacements and the jumping times are correlated. As pointed in Cénac et al. [2013, 2018], Cénac et al. [2017], persistent random walks can be viewed as VLMC for an infinite context tree.

Variable length Markov chains are also a particular case of processes defined by a g-function (where the g-function is piecewise constant on a countable set of cylinders), also called "chaînes à liaisons complètes" after Doeblin and Fortet [1937] or "chains with infinite order" after Harris [1955]. Stationary probability measures for VLMC are g-measures. The question of uniqueness of g-measures has been adressed by many authors when the function g is continuous (in this case, the existence is straightforward), see Johansson and Öberg [2003], Fernández and Maillard [2005]. Recently, interest raised also for the question of existence and uniqueness when g is not continuous, see Gallo [2011], Gallo and Garcia [2013], De Santis and Piccioni [2012] for a perfect simulation point of view and the more ergodic theory flavoured Gallo and Paccaut [2013].

In this paper we go towards some necessary and sufficient conditions to ensure existence and uniqueness of a stationary probability measure for a general VLMC. In this introduction we give the thread of the story and the main results. Key concepts are precisely defined in Section 2 while Section 3 is devoted to stating and proving the main theorem (Theorem 1). The particular case of so-called stable context trees is detailed in Section 4, giving a NSC (Theorem 2). Several illustrating examples are given in Section 5.

Let us recall briefly the probabilistic presentation of Variable Length Markov Chains (VLMC), following Cénac et al. [2012]. Introduce the set $\mathscr{L}$ of left-infinite words on the alphabet $\mathscr{A} = \{0, 1\}$ and consider a saturated tree $\mathscr{T}$ on this alphabet, *i.e.* a tree such that each node has 0 or 2 children, whose leaves are words (possibly infinite) on $\mathscr{A}$. The set of leaves, denoted by $\mathscr{C}$ is supposed to be at most countable.

To each leaf $c \in \mathscr{C}$, called a context, is attached a probability Bernoulli distribution q_c on $\mathscr{A}$. Endowed with this probabilistic structure, such a tree is named a probabilised context tree. The related VLMC is defined as the Markov chain $(U_n)_{n\geqslant 0}$ on $\mathscr{L}$ whose transitions are given, for $\alpha \in \mathscr{A}$, by

$$\mathbf{P}(U_{n+1} = U_n\alpha | U_n) = q_{\mathrm{pref}(\overline{U_n})}(\alpha), \tag{1}$$

where $\mathrm{pref}(u) \in \mathscr{C}$ is defined as the only prefix of the right-infinite word u appearing as a leaf of the context tree. As usual, the bar denotes mirror word♫ and concatenation of words is written without any symbol.

The heuristic is classically as follows: assume that a stationary measure exists for a given VLMC. Describe all properties this measure should have, then try to prove these properties are sufficient to get the existence of a stationary measure.

Let π be a stationary measure for (U_n). It is entirely defined by its value $\pi(\mathscr{L}w)$ on cylinders $\mathscr{L}w$, for all finite words w. By definition (1) of the VLMC, for any letter $\alpha \in \mathscr{A}$ and w non internal word of the context tree,

$$\pi(\mathscr{L}\overline{w}\alpha) = q_{\mathrm{pref}(w)}(\alpha)\pi(\mathscr{L}\overline{w}). \tag{2}$$

A detailed proof of this formula and the following ones will be given at Lemma 1. This formula applies again for $\pi(\mathscr{L}\overline{w})$, and so on, and so forth, until... it is not possible anymore, which means that the suffix of w is of the form αs where $\alpha \in \mathscr{A}$ and s is an internal word of the context tree. This leads to point out the following decomposition of any finite word w:

$$w = \beta_1\beta_2 \dots \beta_{p_w}\alpha_w s_w,$$

where

- p_w is a nonnegative integer and $\beta_i \in \mathscr{A}$, for all $i = 1, \dots, p_w$,
- s_w is the longest internal suffix of w,
- $\alpha_w \in \mathscr{A}$.

With this decomposition, s_w is called the *lis* of w and $\alpha_w s_w$ the α-*lis* of w. Consequently, for any stationary measure π and for any finite word w, write $w = v\alpha_w s_w$ where v is a finite word and $\alpha_w s_w$ is the α-lis of w so that

$$\pi(\mathscr{L}\overline{w}) = \mathrm{casc}(w)\pi(\mathscr{L}\overline{\alpha_w s_w}), \tag{3}$$

♫ In the whole paper, words are as usual read from left to right. One exception to this rule: since the process (U_n) of left-infinite words grows by adding letters to the right, all words that are used to describe suffixes of U_n are read from right to left, hence are written with a bar. See also Remark 1.

where $\operatorname{casc}(w)$, the *cascade* of w is defined as

$$\operatorname{casc}(w)=\prod_{1\leqslant k\leqslant p_w} q_{\operatorname{pref}\sigma^k(w)}(\beta_k).$$

In the above formula, σ is the shift mapping defined by $\sigma\left(\alpha_0\alpha_1\alpha_2\cdots\right)=\alpha_1\alpha_2\cdots$.

Formula (3) indicates that a stationary measure π is entirely determined by its value $\pi(\mathscr{L}\overline{\alpha v})$ for all αv, where α is a letter and v an internal word of the context tree. Further, notice that for any internal word v, by disjoint union, using Formula (2),

$$\pi\left(\mathscr{L}\overline{\alpha v}\right)=\pi\left(\mathscr{L}\overline{v}\alpha\right)=\sum_{c\in\mathscr{C},\ c=v\cdots}\pi\left(\mathscr{L}\overline{c}\right)q_c\left(\alpha\right). \tag{4}$$

This means that π is in fact determined by the $\pi\left(\mathscr{L}\overline{c}\right)$ where c is a context. Admit for this introduction that only finite contexts matter (Lemma 2) and denote by $\mathscr{C}^f$ the set of finite contexts. Formula (3) applies to $\pi\left(\mathscr{L}\overline{c}\right)$:

$$\pi\left(\mathscr{L}\overline{c}\right)=\operatorname{casc}(c)\pi\left(\mathscr{L}\overline{\alpha_c s_c}\right),$$

so that π is entirely determined by its value $\pi\left(\mathscr{L}\overline{\alpha s}\right)$ on all the αs which are α-lis *of contexts*. Denote by $\mathscr{S}$ the set of all the α-lis of contexts. All these values of π are connected since by Formula (4), for any $\alpha s\in\mathscr{S}$,

$$\pi\left(\mathscr{L}\overline{\alpha s}\right)=\sum_{c\in\mathscr{C}^f,\ c=s\cdots}\operatorname{casc}(\alpha c)\pi\left(\mathscr{L}\overline{\alpha_c s_c}\right)=\sum_{\beta t\in\mathscr{S}}\pi\left(\mathscr{L}\overline{\beta t}\right)\sum_{\substack{c\in\mathscr{C}^f\\ c=s\cdots\\ c=\cdots[\beta t]}}\operatorname{casc}(\alpha c),$$

where the notation $c=\cdots[\beta t]$ means that βt is the α-lis of c.
Introduce the square matrix $Q=\left(Q_{\alpha s,\beta t}\right)_{(\alpha s,\beta t)\in\mathscr{S}^2}$ (at most countable) defined by

$$Q_{\alpha s,\beta t}=\sum_{\substack{c\in\mathscr{C}^f\\ c=t\cdots\\ c=\cdots[\alpha s]}}\operatorname{casc}\left(\beta c\right),$$

so that the above formula on π writes

$$\pi\left(\mathscr{L}\overline{\alpha s}\right)=\sum_{\beta t\in\mathscr{S}}\pi\left(\mathscr{L}\overline{\beta t}\right)Q(\beta t,\alpha s).$$

In otherwords, $\left(\pi\left(\mathscr{L}\overline{\alpha s}\right)\right)_{\alpha s\in\mathscr{S}}$ is a left-fixed vector of the matrix Q. It appears that the study of the matrix Q acting on the α-lis of contexts is the key tool to characterize a stationary measure for the VLMC. It allows us to prove in Section 3 the following theorem.

Theorem. *Let $(\mathscr{T},q)$ be a probabilised context tree and U the associated VLMC. Assume that $\forall\alpha\in\mathscr{A}$, $\forall c\in\mathscr{C}$, $q_c(\alpha)\neq 0$.*

(i) Assume that there exists a finite U-stationary probability measure π on $\mathscr{L}$. Then the cascade series $\displaystyle\sum_{c\in\mathscr{C}^f,\ c=\cdots[\alpha s]}\operatorname{casc}(c)$ converge. Calling $\kappa_{\alpha s}$ its sum,

$$\sum_{\alpha s\in\mathscr{S}}\pi\left(\mathscr{L}\overline{\alpha s}\right)\kappa_{\alpha s}=1. \tag{5}$$

(ii) Assume that the cascade series $\sum_{c\in\mathscr{C}^f,\ c=\cdots[\alpha s]} \operatorname{casc}(c)$ *converge. Then, there is a bijection between the set of U-stationary probability measures on $\mathscr{L}$ and the set of left-fixed vectors $(v_{\alpha s})_{\alpha s\in\mathscr{S}}$ of Q that satisfy*

$$\sum_{\alpha s\in\mathscr{S}} v_{\alpha s}\kappa_{\alpha s} = 1. \tag{6}$$

The characterization given in this theorem is expressed via the probability distributions q_c. Nevertheless, the role of context lis and α-lis suggests that the *shape* of the context tree matters a lot. Actually, in the case of *stable* trees (so called because they are stable by the shift), it turns out that the matrix Q is stochastic and irreducible. Consequently, the theorem above provides in Theorem 2 a NSC in terms of the recurrence of Q. As a corollary, for context trees $(\mathscr{T}, q)$ whose stabilized (the smallest stable context tree that contains $\mathscr{T}$) has again at most countably many infinite branches, the associated chain admits either a unique or no stationary probability measure.
Moreover, in the particular case of a *finite* number of context α-lis, Theorem 3 gives a rather easy condition for the existence and uniqueness of a stationary probability measure: there exists a unique probability measure if and only if the cascade series converge. Section 4, which deals with stable trees, is made complete with the link that can be highlighted with the semi-Markov chains theory.
Section 5 contains a list of examples that illustrate different configurations of infinite branches and context α-lis, and also the respective roles of the geometry of the context tree on the one hand and on the other the asymptotic behaviour of the Bernoulli distributions q_c when the size of the context c (having a given α-lis) grows to infinity. The final Section 6 gives some tracks towards better characterizations in the non-stable case. A conjecture is set: if the set of infinite branches does not contain any shift-stable subset, then there exists a unique stationary probability measure for the associated VLMC.

2 Definitions

2.1 VLMC

In the whole paper, $\mathscr{A} = \{0,1\}$ denotes the set of *letters* (the *alphabet*) and $\mathscr{L}$ and $\mathscr{R}$ respectively denote the left-infinite and the right-infinite $\mathscr{A}$-valued sequences[♫]:

$$\mathscr{L} = \mathscr{A}^{-\mathbb{N}} \text{ and } \mathscr{R} = \mathscr{A}^{\mathbb{N}}.$$

The set of finite words, sometimes denoted by $\mathscr{A}^*$, will be denoted by $\mathscr{W}$:

$$\mathscr{W} = \bigcup_{n\in\mathbb{N}} \mathscr{A}^n,$$

the set $\mathscr{A}^0$ being the emptyset. When $\ell\in\mathscr{L}$, $r\in\mathscr{R}$ and $v,w\in\mathscr{W}$, the concatenations of ℓ and w, v and w, w and r are respectively denoted by ℓw, vw and wr. More over, the finite word w being given,

$$\mathscr{L}w$$

denotes the cylinder made of left-infinite words having w as a suffix. Finally, when $w=\alpha_1\cdots\alpha_d\in\mathscr{W}$, $\ell=\cdots\alpha_{-2}\alpha_{-1}\alpha_0\in\mathscr{L}$ or $r=\alpha_0\alpha_1\alpha_2\cdots\in\mathscr{R}$, the bar denotes the mirror object :

$$\overline{w} = \alpha_d\alpha_{d-1}\cdots\alpha_1\in\mathscr{W},$$
$$\overline{\ell} = \alpha_0\alpha_{-1}\alpha_{-2}\cdots\in\mathscr{R},$$
$$\overline{r} = \cdots\alpha_2\alpha_1\alpha_0\in\mathscr{L}.$$

[♫] $\mathbb{N}$ denotes the set of natural integers $\{0,1,2,\cdots\}$.

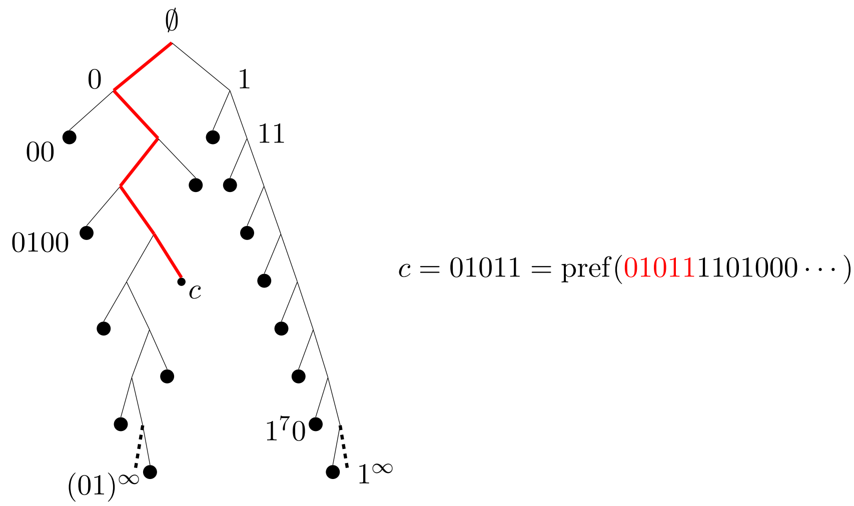


Figure 1: an example of context tree. It has two infinite branches: 1^∞ and $(01)^\infty$.

A VLMC is an $\mathscr{L}$-valued Markov chain, defined by a so-called *probabilised context tree*. We give here a compact description. One can refer to Cénac et al. [2012] for an extensive definition.
A *context tree* is a rooted binary tree $\mathscr{T}$ which has an *at most countable set of infinite branches*; an infinite sequence $r \in \mathscr{R}$ is an *infinite branch* of $\mathscr{T}$ whenever all its finite prefixes belong to $\mathscr{T}$. As usual, the nodes of the tree are canonically labelled by words on $\mathscr{A}$. In the example of Figure 1, the tree has two infinite branches: $(01)^\infty$ and 1^∞. A node of a context tree $\mathscr{T}$ will be called a *context* when it is a finite leaf or an infinite branch of $\mathscr{T}$. The sets of all contexts, finite leaves and infinite branches are respectively denoted by

$$\mathscr{C},\ \mathscr{C}^f \text{ and } \mathscr{C}^i.$$

They are all at most countable sets. A finite word $w \in \mathscr{W}$ will be called an *internal node* when it is *strictly* internal as a node of $\mathscr{T}$; it will be called *nonexternal* whenever it is internal or a context. In the same vein a finite word or a right-infinite sequence will be said *external* when it is strictly external and *noninternal* when it is external or a context. The sets of internal words is denoted by

$$\mathscr{I}.$$

Definition 1 (pref of a noninternal word)**.** *Let $\mathscr{T}$ be a context tree and w be a noninternal finite or right-infinite word. Then, we denote by* $\mathrm{pref}(w)$ *the unique prefix of w which is a context of $\mathscr{T}$.*

Definition 2 (shift mapping)**.** *The shift mapping $\sigma : \mathscr{R} \to \mathscr{R}$ is defined by $\sigma(\alpha_0\alpha_1\alpha_2\cdots) = \alpha_1\alpha_2\cdots$. The definition is extended to finite words (with $\sigma(\emptyset) = \emptyset$).*

A *probablised context tree* is a context tree $\mathscr{T}$ endowed with a family of probability measures $(q_c)_{c\in\mathscr{C}}$ on $\mathscr{A}$ indexed by the (finite or infinite) contexts of $\mathscr{T}$. To any probabilised context tree, one can associate a VLMC, which is the $\mathscr{L}$-valued Markov chain $(U_n)_{n\geqslant 0}$ defined by its transition probabilities given by

$$\forall n \geqslant 0,\ \mathbf{P}\left(U_{n+1} = U_n\alpha | U_n\right) = q_{\mathrm{pref}\left(\overline{U_n}\right)}(\alpha). \tag{7}$$

Note that if one denotes by X_n the rightmost letter of the sequence $U_n \in \mathscr{L}$ so that

$$\forall n \geqslant 0,\ U_{n+1} = U_n X_{n+1},$$

then, when the context tree has at least one infinite context, the final letter process $(X_n)_{n\geqslant 0}$ is generally *not* a Markov process. When the tree is finite, $(X_n)_{n\geqslant 0}$ is a usual $\mathscr{A}$-valued Markov chain whose order is the height of the tree, *i.e.* the length of its longest branch. The vocable VLMC is somehow confusing but commonly used.

Definition 3 (non-nullness)**.** *A probabilised context tree* $(\mathscr{T}, q)$ *is* non-null *if for all* $c \in \mathscr{C}$ *and all* $\alpha \in \mathscr{A}$, $q_c(\alpha) \neq 0$.

Remark 1. It would have been probably simpler to define a VLMC as an $\mathscr{R}$-valued Markov chain $(U_n)_n$ taking as transition probabilities: $\mathbf{P}\left(U_{n+1} = \alpha U_n | U_n\right) = q_{\mathrm{pref}(U_n)}(\alpha)$. With this definition, words would have grown by adding a letter to the left, and everything would have been read from left to right, avoiding the emergence of bars in the text ($\overline{w}$, $\overline{U_n}$, …). We decided to adopt our convention (U_n is $\mathscr{L}$-valued) in order to fit the already existing literature on VLMC and more generally on stochastic processes.

2.2 Cascades, lis and α-lis

Definition 4 (lis and α-lis)**.** *Let* $\mathscr{T}$ *be a context tree. If* $w \in \mathscr{W}$ *is a non empty finite word,* w *can be uniquely written as*

$$w = \beta_1\beta_2 \ldots \beta_{p_w}\alpha_w s_w,$$

where
- $p_w \geqslant 0$ *and* $\beta_i \in \mathscr{A}$*, for all* $i \in \{1, \ldots, p_w\}$,
- $\alpha_w \in \mathscr{A}$,
- s_w *is the longest internal strict suffix of* w.

Note that s_w *may be empty. When* $p_w = 0$*, there are no* β*'s and* $w = \alpha_w s_w$.
Vocabulary: the longest internal suffix s_w *is abbreviated as the* lis *of* w*; the noninternal suffix* $\alpha_w s_w$ *is the* α-lis *of* w.

Any word has an α-lis, but we will be mainly interested by the α-lis of *contexts*. The set of α-lis *of the finite contexts* of $\mathscr{T}$ will be denoted by $\mathscr{S}(\mathscr{C})$, or more shortly by $\mathscr{S}$:

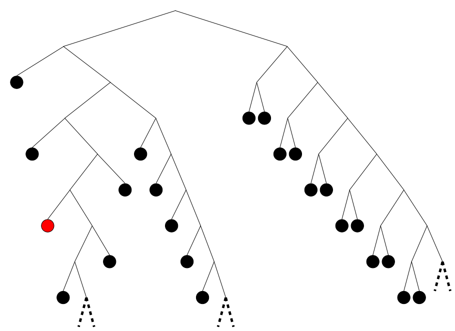
$$\mathscr{S} = \left\{\alpha_c s_c,\ c \in \mathscr{C}^f\right\};$$

this is an at most countable set (like $\mathscr{C}$). For any $u, v, w \in \mathscr{W}$, the notations

$$v = u \cdots \quad \text{and } w = \cdots [u] \tag{8}$$

stand respectively for "u is a prefix of v" and "u is the α-lis of w".

Example 1 (computation of a lis)**.**



In this example, the finite contexts of the context tree are the following ones (they completely define the context tree): $(01)^p00$, $(01)^r1$, 01^r0, 1^q00, 1^q01, $p \geqslant 0$, $q \geqslant 1$, $r \geqslant 2$.

Take for example the context 010100. *Remove successively letters from the left until you get an internal word:* 10100 *is external,* 0100 *is noninternal,* 100 *is noninternal,* 00 *is noninternal. The suffix* 0*, for the first time, is internal: this is the lis of* 010100. *The last removed letter is* $\alpha = 0$ *so that the* α*-lis is* 00.

In the array hereunder, the left side column consists in the list of all α*-lis of the context tree. For any* $\alpha s \in \mathscr{S}$*, the list of all contexts having* αs *as an* α*-lis is given in the right side column.*

$\alpha s \in \mathscr{S}$	*contexts having* αs *as an* α*-lis*
00	1^q00, $(01)^p00$, $p \geqslant 0$, $q \geqslant 1$
101	$1^q01, q \geqslant 1$
01011	$(01)^r1$, $r \geqslant 2$
01^r0, $r \geqslant 2$	01^r0

Definition 5 (cascade)**.** *Let* $(\mathscr{T}, q)$ *be a probabilised context tree. If* $w \in \mathscr{W}$ *writes* $w = \beta_1\beta_2 \ldots \beta_p \alpha s$ *where* $p \geqslant 0$ *and where* αs *is the* α*-lis of* w*, the* cascade *of* w *is defined as*

$$\operatorname{casc}(w) = \prod_{1 \leqslant k \leqslant p} q_{\operatorname{pref} \sigma^k(w)}(\beta_k),$$

where an empty product equals 1*, which occurs if, and only if* w *equals its own* α*-lis. The cascade of* $\emptyset$ *is defined as being* 1*. Note that* $\operatorname{casc}(\alpha s) = 1$ *for any* $\alpha s \in \mathscr{S}$*.*

In Example 1, $\operatorname{casc}(010100) = q_{101}(0)q_{0100}(1)q_{100}(0)q_{00}(1)$.

Remark 2. For any $w \in \mathscr{W}$, $\operatorname{casc}(w) = \operatorname{casc}(0w) + \operatorname{casc}(1w)$ if, and only if w is noninternal. Indeed, if w is internal, the sum equals 2 whereas $\operatorname{casc}(w) \leqslant 1$.

Definition 6 (cascade series)**.** *For every* $\alpha s \in \mathscr{S}$*, the* cascade series *of* αs *(related to* $(\mathscr{T}, q)$*) is the at most countable family of cascades of the finite contexts having* αs *as their* α*-lis. In other words, with notations (8), it is the family*

$$\left(\operatorname{casc}(c)\right)_{c \in \mathscr{C}^f,\ c = \cdots[\alpha s]}.$$

Since the cascades are positive numbers, the summability of a family of cascades of a probabilised context tree is equivalent to the convergence of the series associated to any total order on the set of contexts indexing the family. The assertion

$$\forall \alpha s \in \mathscr{S}, \quad \sum_{c \in \mathscr{C}^f,\ c = \cdots[\alpha s]} \operatorname{casc}(c) < +\infty \tag{9}$$

will be called *convergence of the cascade series.* When the cascade series converge, $\kappa_{\alpha s}$ denotes the sum of the cascade series relative to $\alpha s \in \mathscr{S}$:

$$\kappa_{\alpha s} = \sum_{c \in \mathscr{C}^f,\ c = \cdots[\alpha s]} \operatorname{casc}(c). \tag{10}$$

Remark 3. The finiteness of the set $\mathscr{C}^i$ of infinite branches on one side, and of the set $\mathscr{S}$ of context α-lis on the other side are not related. In Section 5.1, one finds an example of context tree for which $\mathscr{S}$ is finite while $\mathscr{C}^i$ is infinite. In the example of Section 5.3, $\mathscr{S}$ is infinite while $\mathscr{C}^i$ is finite. The left-comb of left-combs has infinite $\mathscr{C}^i$ and $\mathscr{S}$ (see Section 5.4). Finally, the double bamboo has finite $\mathscr{C}^i$ and $\mathscr{S}$, see Example 2.

2.3 α-lis matrix Q

For any $(\alpha s, \beta t) \in \mathscr{S}^2$, with notations (8), define

$$Q_{\alpha s,\beta t} = \sum_{\substack{c\in\mathscr{C}^f \\ c=t\cdots \\ c=\cdots[\alpha s]}} \operatorname{casc}(\beta c) \in [0, +\infty].$$

As the set $\mathscr{S}$ is at most countable, the family $Q = (Q_{\alpha s,\beta t})_{(\alpha s,\beta t)\in\mathscr{S}^2}$ will be considered as a matrix, finite or countable, for an arbitrary order. The convergence of the cascade series of $(\mathscr{T}, q)$ is sufficient to ensure the finiteness of the coefficients of Q.

3 Results for a general context tree

In this section no assumption is made on the shape of the context tree. After two key lemmas, we state and prove the main theorem that establishes precise connections between stationary probability measures of the VLMC and left-fixed vectors of the matrix Q defined in Subsection 2.3.

3.1 Two key lemmas

Lemma 1. (Cascade formulae)
Let $(\mathscr{T}, q)$ be a probabilised context tree and π be a stationary probability measure for the corresponding VLMC.
(i) For every noninternal finite word w and for every $\alpha \in \mathscr{A}$,

$$\pi\left(\mathscr{L}\overline{w}\alpha\right) = q_{\mathrm{pref}(w)}(\alpha)\pi\left(\mathscr{L}\overline{w}\right). \tag{11}$$

(ii) For every right-infinite word $r \in \mathscr{R}$ and for every $\alpha \in \mathscr{A}$,

$$\pi\left(\overline{r}\alpha\right) = q_{\mathrm{pref}(r)}(\alpha)\pi\left(\overline{r}\right). \tag{12}$$

(iii) For every finite non empty word w, if one denotes by $\alpha_w s_w$ the α-lis of w, then

$$\pi\left(\mathscr{L}\overline{w}\right) = \operatorname{casc}(w)\pi\left(\mathscr{L}\overline{\alpha_w s_w}\right). \tag{13}$$

Proof of lemma 1. (i) Assume first that $\pi\left(\mathscr{L}\overline{w}\right) \neq 0$. Then, since w is noninternal, $\mathrm{pref}(w)$ is well defined so that, by stationarity,

$$\begin{aligned}
\pi\left(\mathscr{L}\overline{w}\alpha\right) &= \mathbf{P}_\pi\left(U_1 \in \mathscr{L}\overline{w}\alpha\right) \\
&= \mathbf{P}_\pi\left(U_1 \in \mathscr{L}\overline{w}\alpha | U_0 \in \mathscr{L}\overline{w}\right)\mathbf{P}_\pi\left(U_0 \in \mathscr{L}\overline{w}\right) \\
&= q_{\mathrm{pref}(w)}(\alpha)\mathbf{P}_\pi\left(U_0 \in \mathscr{L}\overline{w}\right) = q_{\mathrm{pref}(w)}(\alpha)\pi\left(\mathscr{L}\overline{w}\right)
\end{aligned}$$

proving (11). If $\pi\left(\mathscr{L}\overline{w}\right) = 0$, then, by stationarity, $\pi\left(\mathscr{L}\overline{w}\alpha\right) = \mathbf{P}_\pi\left(U_1 \in \mathscr{L}\overline{w}\alpha\right) \leqslant \mathbf{P}_\pi\left(U_0 \in \mathscr{L}\overline{w}\right) = 0$ so that (11) remains true.

(ii) Since r is infinite, the context $\mathrm{pref}(r)$ is always defined (it may be finite or infinite). Consequently,

$$\begin{aligned}
\pi\left(\overline{r}\alpha\right) &= \mathbf{P}_\pi\left(U_1 = \overline{r}\alpha\right) \\
&= \mathbf{E}_\pi\left(\mathbf{E}_\pi\left(\mathbf{1}_{\{U_1=\overline{r}\alpha\}}|U_0\right)\right) \\
&= \mathbf{E}_\pi\left(\mathbf{1}_{\{U_0=\overline{r}\}}\mathbf{P}_\pi\left(U_1 = U_0\,\alpha|U_0\right)\right) = q_{\mathrm{pref}(r)}(\alpha)\pi\left(\overline{r}\right).
\end{aligned}$$

(iii) Direct induction from Formula (11). □

The following lemma ensures that a stationary probability measure weights finite words and only finite words.

Lemma 2. *Let $(\mathscr{T}, q)$ be a non-null probabilised context tree. Assume that π is a stationary probability measure for the associated VLMC. Then*
(i) $\forall w \in \mathscr{W}$, $\pi\left(\mathscr{L}\overline{w}\right) \neq 0$.
(ii) $\forall r \in \mathscr{R}$, $\pi(\overline{r}) = 0$.

Proof of lemma 2. (i) We prove that if w is a finite word and if $\alpha \in \mathscr{A}$, then $[\pi\left(\mathscr{L}\overline{w}\alpha\right) = 0] \Rightarrow [\pi\left(\mathscr{L}\overline{w}\right) = 0]$. An induction on the length of w is then sufficient to prove the result since $\pi(\mathscr{L}) = 1$.

1. Assume that $w \notin \mathscr{I}$ and that $\pi\left(\mathscr{L}\overline{w}\alpha\right) = 0$. Then, as a consequence of the cascade formula (11), $0 = \pi\left(\mathscr{L}\overline{w}\right)q_{\mathrm{pref}(w)}(\alpha)$. As no q_c vanishes, $\pi\left(\mathscr{L}\overline{w}\right) = 0$.

2. Assume now $w \in \mathscr{I}$ and $\pi\left(\mathscr{L}\overline{w}\alpha\right) = 0$. Then, by disjoint union and stationarity of π,

$$0 = \sum_{\substack{c\in\mathscr{C}^f\\ c=w\ldots}} \pi\left(\mathscr{L}\overline{c}\alpha\right) + \sum_{\substack{c\in\mathscr{C}^i\\ c=w\ldots}} \pi\left(\overline{c}\alpha\right) = \sum_{\substack{c\in\mathscr{C}^f\\ c=w\ldots}} \pi\left(\mathscr{L}\overline{c}\right)q_c(\alpha) + \sum_{\substack{c\in\mathscr{C}^i\\ c=w\ldots}} \pi\left(\overline{c}\right)q_c(\alpha).$$

As no q_c vanishes, all the $\pi\left(\mathscr{L}\overline{c}\right)$ and the $\pi(\overline{c})$ necessarily vanish so that, by disjoint union,

$$\pi\left(\mathscr{L}\overline{w}\right) = \sum_{\substack{c\in\mathscr{C}^f\\ c=w\ldots}} \pi\left(\mathscr{L}\overline{c}\right) + \sum_{\substack{c\in\mathscr{C}^i\\ c=w\ldots}} \pi\left(\overline{c}\right) = 0.$$

(ii) Denote $r = \alpha_1\alpha_2\cdots$ and $r_n = \alpha_n\alpha_{n+1}\cdots$ its n-th suffix, for every $n \geqslant 1$. Since π is stationary, an elementary induction from Formula (12) implies that, for every $m \geqslant 1$,

$$\pi\left(\overline{r}\right) = \left(\prod_{k=1}^{m} q_{\mathrm{pref}(r_{k+1})}\left(\alpha_k\right)\right)\pi\left(\overline{r_{m+1}}\right). \tag{14}$$

1. Assume first that $r = st^\infty$ is ultimately periodic, where s and t are finite words, $t = \beta_1\cdots\beta_T$ being nonempty. Then, because of (14), $\pi(\overline{r}) \leqslant \pi\left(\overline{t}^\infty\right)$ and $\pi\left(\overline{t}^\infty\right) = \rho\pi\left(\overline{t}^\infty\right)$ where

$$\rho = \prod_{k=1}^{T} q_{\mathrm{pref}(\beta_{k+1}\cdots\beta_T t^\infty)}\left(\beta_k\right).$$

Since the probability measures q_c are all assumed to be nontrivial, then $0 < \rho < 1$, which implies that $\pi\left(\overline{t}^\infty\right) = 0$.

2. Assume on the contrary that r is aperiodic. Then, $m \neq n \Longrightarrow r_n \neq r_m$ for all $n, m \geqslant 1$: the r_n are all distinct among the infinite branches of the context tree. Thus, by disjoint union,

$$\sum_{n\geqslant 1} \pi\left(\overline{r_n}\right) \leqslant \sum_{c\in\mathscr{C}^i} \pi\left(\overline{c}\right) \leqslant \pi\left(\mathscr{L}\right) = 1,$$

which implies in particular that $\pi\left(\overline{r_n}\right)$ tends to 0 when n tends to infinity. Since $\pi\left(\overline{r}\right) \leqslant \pi\left(\overline{r_n}\right)$ because of Formula (14), this leads directly to the result.

□

3.2 Main theorem

Definition 7. *Let $A = (a_{\ell c})_{(\ell,c)\in\mathscr{E}^2}$ be a matrix with real entries, indexed by a totally ordered set $\mathscr{E}$ supposed to be finite or denumerable. A* left-fixed vector *of A is a row-vector $X = (x_k)_{k\in\mathscr{E}} \in \mathbb{R}^{\mathscr{E}}$, indexed by $\mathscr{E}$, such that $XA = X$. In particular, this implies that the matrix product XA is well defined, which means that for any $c \in \mathscr{E}$, the series $\sum_\ell x_\ell a_{\ell,c}$ is convergent. Note that, whenever X and A are infinite dimensional and have nonnegative entries, this summability does not depend on the chosen order on the index set $\mathscr{E}$.*

Denote by $\mathscr{M}_1\left(\mathscr{L}\right)$ the set of probability measures on $\mathscr{L}$. Define the mapping f as follows:

$$\begin{array}{rcl} f : \mathscr{M}_1\left(\mathscr{L}\right) & \longrightarrow & [0,1]^{\mathscr{S}} \\ \pi & \longmapsto & \Big(\pi\left(\mathscr{L}\overline{\alpha s}\right)\Big)_{\alpha s\in\mathscr{S}}. \end{array}$$

Theorem 1. *Let $(\mathscr{T}, q)$ be a non-null probabilised context tree and U the associated VLMC.*
(i) Assume that there exists a finite U-stationary probability measure π on $\mathscr{L}$. Then the cascade series (9) converge. Furthermore, using notation (10),

$$\sum_{\alpha s\in\mathscr{S}} \pi\left(\mathscr{L}\overline{\alpha s}\right)\kappa_{\alpha s} = 1. \tag{15}$$

(ii) Assume that the cascade series (9) converge. Then, f induces a bijection between the set of U-stationary probability measures on $\mathscr{L}$ and the set of left-fixed vectors $(v_{\alpha s})_{\alpha s\in\mathscr{S}}$ of Q which satisfy

$$\sum_{\alpha s\in\mathscr{S}} v_{\alpha s}\kappa_{\alpha s} = 1. \tag{16}$$

For an example of application of this theorem, see Section 5.1.

Proof of theorem 1.
Proof of *(i)*. If π is a stationary probability measure, disjoint union, Lemma 2(i) and the cascade formula (13) imply that

$$1 = \sum_{c\in\mathscr{C}^f} \pi\left(\mathscr{L}\overline{c}\right) = \sum_{c\in\mathscr{C}^f} \mathrm{casc}(c)\pi\left(\mathscr{L}\overline{\alpha_c s_c}\right).$$

Gathering together all the contexts that have the same α-lis leads to

$$1 = \sum_{\alpha s\in\mathscr{S}} \pi\left(\mathscr{L}\overline{\alpha s}\right)\left(\sum_{c\in\mathscr{C}^f,\ c=\cdots[\alpha s]} \mathrm{casc}(c)\right).$$

Now, by Lemma 2(ii), $\pi\left(\mathscr{L}\overline{\alpha s}\right)\neq 0$ for all $\alpha s\in\mathscr{S}$. This forces the sums of cascades to be finite.

Proof of *(ii)*.
1) *Injectivity.* Let π be a stationary probability measure on $\mathscr{L}$. As the cylinders based on finite words generate the whole σ-algebra, π is determined by the $\pi\left(\mathscr{L}\overline{w}\right)$, $w\in\mathscr{W}$. Now write any $w\in\mathscr{W}\setminus\{\emptyset\}$ as $w=p\alpha s$ where αs is the α-lis of w and $p\in\mathscr{W}$ (beware, αs may not be the α-lis of a context). As π is stationary, the cascade formula (13) entails $\pi\left(\mathscr{L}\overline{w}\right)=\mathrm{casc}(w)\pi\left(\mathscr{L}\overline{\alpha s}\right)$. As a consequence π is determined by its values on the words $\overline{\alpha s}$ where $s\in\mathscr{I}$ is internal and $\alpha\in\mathscr{A}$. Now, as $s\in\mathscr{I}$, by disjoint union, cascade formula (11) and Lemma 2(i),

$$\pi\left(\mathscr{L}\overline{\alpha s}\right)=\pi\left(\mathscr{L}\overline{s}\alpha\right)=\sum_{c\in\mathscr{C}^f,\ c=s\cdots}\pi\left(\mathscr{L}\overline{c}\right)q_c\left(\alpha\right).$$

This means that π is in fact determined by the $\pi\left(\mathscr{L}\overline{c}\right)$ where c is a finite context. Lastly, as above, the stationarity of π, the cascade formula (13) and the decomposition of any context c into $c=p_c\alpha_c s_c$ where $\alpha_c s_c$ is the α-lis of c together imply that π is determined by the $\pi\left(\mathscr{L}\overline{\alpha s}\right)$ where $s\in\mathscr{S}$ (remember, $\mathscr{S}$ denotes the set of all α-lis *of contexts*). We have proved that the restriction of f to stationary measures is one-to-one.
2) *Image of a stationary probability measure.* Let $\pi\in\mathscr{M}_1\left(\mathscr{L}\right)$ be stationary. By disjoint union, as above, if $\alpha s\in\mathscr{S}$,

$$\pi\left(\mathscr{L}\overline{\alpha s}\right)=\sum_{c\in\mathscr{C}^f,\ c=s\cdots}\pi\left(\mathscr{L}\overline{c}\right)q_c\left(\alpha\right).$$

Applying the cascade formula (13) to all contexts in the sum and noting that $\mathrm{casc}(\alpha c)=q_c(\alpha)\,\mathrm{casc}(c)$, one gets

$$\pi\left(\mathscr{L}\overline{\alpha s}\right)=\sum_{c\in\mathscr{C}^f,\ c=s\cdots}\mathrm{casc}(\alpha c)\pi\left(\mathscr{L}\overline{\alpha_c s_c}\right).$$

Gathering together all the contexts that have the same α-lis entails

$$\pi\left(\mathscr{L}\overline{\alpha s}\right)=\sum_{\beta t\in\mathscr{S}}\pi\left(\mathscr{L}\overline{\beta t}\right)\left(\sum_{c\in\mathscr{C}^f,\ c=s\cdots=\cdots[\beta t]}\mathrm{casc}(\alpha c)\right).$$

This means that the row vector $\left(\pi\left(\overline{\alpha s}\right)\right)_{\alpha s\in\mathscr{S}}$ is a left-fixed vector for the matrix Q. We have shown that f sends a stationary probability measure to a left-fixed vector for Q. Moreover, as in the proof of (i), Equality (15) holds.
3) *Surjectivity.*
Let $(v_{\alpha s})_{\alpha s\in\mathscr{S}}\in[0,1]^{\mathscr{S}}$ be a row vector, left-fixed by Q, that satisfies $\sum_{\alpha s\in\mathscr{S}}v_{\alpha s}\kappa_{\alpha s}=1$. Let μ be the function defined on $\mathscr{S}$ by $\mu\left(\alpha s\right)=v_{\alpha s}$. Denoting by $\alpha_c s_c$ the α-lis of a context c, μ extends to any finite nonempty word in the following way:

$$\forall w\in\mathscr{W}\setminus\{\emptyset\},\ \mu(w)=\mathrm{casc}(w)\sum_{c\in\mathscr{C}^f,\ c=s_w\cdots}\mathrm{casc}(\alpha_w c)\mu\left(\alpha_c s_c\right)\in[0,+\infty]. \tag{17}$$

Notice that this definition actually extends μ because of the fixed vector property, and that, at this moment of the proof, $\mu(w)$ might be infinite. Notice also that this implies $\mu(w)=\mathrm{casc}(w)\mu(\alpha_w s_w)$ for any $w\in\mathscr{W}$, $w\neq\emptyset$.
For every $n\geqslant 1$ and for all $w\in\mathscr{W}$ such that $|w|=n$, define $\pi_n\left(\overline{w}\right)=\mu\left(w\right)$; this clearly defines a $[0,+\infty]$-valued measure π_n on $\mathscr{A}_{-n}=\prod_{-n\leqslant k\leqslant -1}\mathscr{A}$. Besides, π_1 is a probability measure. Indeed, because of Definition (17) and Remark(2),

$$\mu(0)+\mu(1)=\sum_{c\in\mathscr{C}^f}\left(\mathrm{casc}(0c)+\mathrm{casc}(1c)\right)\mu\left(\alpha_c s_c\right)=\sum_{c\in\mathscr{C}^f}\mathrm{casc}(c)\mu\left(\alpha_c s_c\right)$$

which can be written

$$\mu(0)+\mu(1)=\sum_{\alpha s\in\mathscr{S}}\mu\left(\alpha s\right)\sum_{c\in\mathscr{C}^f,\ c=\cdots[\alpha s]}\mathrm{casc}(c)=\sum_{\alpha s\in\mathscr{S}}v_{\alpha s}\kappa_{\alpha s}=1$$

the last equality coming from the assumption on $(v_{\alpha s})_{\alpha s}$.
In view of applying Kolmogorov extension theorem, the consistency condition states as follows : $\pi_{n+1}(\mathscr{A}\overline{w})=\pi_n(\overline{w})$ for any $w\in\mathscr{W}$ of length n. This is true because

$$\mu(w0)+\mu(w1)=\mu(w). \tag{18}$$

Indeed, for any $a\in\mathscr{A}$, since s_w is internal, s_wa is either internal or a context. Furthermore,

- if $s_wa\in\mathscr{I}$ then $s_{wa}=s_wa$, $\alpha_{wa}=\alpha_w$ and $\mathrm{casc}(wa)=\mathrm{casc}(w)$ so that

$$\mu(wa)=\mathrm{casc}(w)\sum_{c\in\mathscr{C}^f,\ c=s_wa\cdots}\mathrm{casc}\left(\alpha_wc\right)\mu\left(\alpha_cs_c\right); \tag{19}$$

- if $s_wa\in\mathscr{C}$ then denote $\kappa=s_wa$ so that $\mathrm{casc}(wa)=\mathrm{casc}(w)\,\mathrm{casc}\left(\alpha_w\kappa\right)$, $\alpha_{wa}=\alpha_\kappa$ and $s_{wa}=s_\kappa$. Thus, $\mu(wa)=\mathrm{casc}(wa)\sum_{c=s_\kappa\cdots}\mathrm{casc}\left(\alpha_\kappa c\right)\mu\left(\alpha_cs_c\right)=\mathrm{casc}(w)\,\mathrm{casc}\left(\alpha_w\kappa\right)\mu\left(\alpha_\kappa s_\kappa\right)$, which implies that (19) still holds, the sum being reduced to one single term since s_wa is itself a context.

Valid in all cases, Formula (19) easily implies Claim (18). Consequently all the π_n are probability measures. By Kolmogorov extension theorem, there exists a unique probability measure π on $\mathscr{L}$ such that $\pi_{|\mathscr{A}_{-n}}=\pi_n$ for every n.
Furthermore, $\pi(\overline{c})=0$ for any infinite context c. Indeed, one has successively,

$$\begin{aligned}1&=\sum_{c\in\mathscr{C}^f}\pi\left(\mathscr{L}\overline{c}\right)+\sum_{c\in\mathscr{C}^i}\pi\left(\overline{c}\right)\\&=\sum_{c\in\mathscr{C}^f}\mu(c)+\sum_{c\in\mathscr{C}^i}\pi\left(\overline{c}\right).\end{aligned}$$

Besides,

$$\sum_{c\in\mathscr{C}^f}\mu(c)=\sum_{c\in\mathscr{C}^f}\mathrm{casc}(c)\mu\left(\alpha_cs_c\right)=\sum_{\alpha s\in\mathscr{S}}v_{\alpha s}\kappa_{\alpha s}=1$$

so that $\sum_{c\in\mathscr{C}^i}\pi\left(\overline{c}\right)=0$.
The stationarity of π follows from the identity $\mu(0w)+\mu(1w)=\mu(w)$ for any finite word w. Namely :

- if $w\notin\mathscr{I}$, then for $a\in\mathscr{A}$, $s_{aw}=s_w$, $\alpha_{aw}=\alpha_w$ hence

$$\mu(0w)+\mu(1w)=\left(\mathrm{casc}(0w)+\mathrm{casc}(1w)\right)\sum_{c\in\mathscr{C}^f,\ c=s_w\cdots}\mathrm{casc}\left(\alpha_wc\right)\mu\left(\alpha_cs_c\right).$$

  Now, Remark 2 entails the claim.

- if $w\in\mathscr{I}$, then for $a\in\mathscr{A}$, $s_{aw}=w$, $\alpha_{aw}=a$ and $\mathrm{casc}(aw)=1$ thus

$$\mu(aw)=\sum_{c\in\mathscr{C}^f,\ c=w\cdots}\mathrm{casc}(ac)\mu\left(\alpha_cs_c\right).$$

Using again Remark 2, it comes

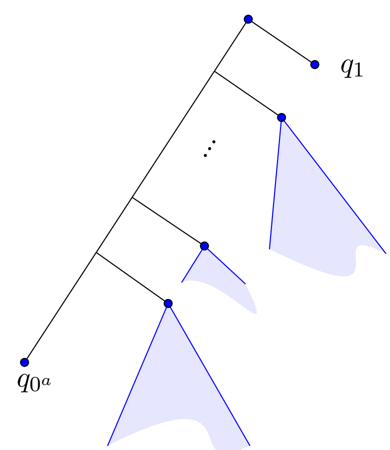
$$
\begin{aligned}
\mu(0w)+\mu(1w) &= \sum_{c\in\mathscr{C}^f,\ c=w\cdots} (\mathrm{casc}(0c)+\mathrm{casc}(1c))\,\mu(\alpha_c s_c)\\
&= \sum_{c\in\mathscr{C}^f,\ c=w\cdots} \mathrm{casc}(c)\mu(\alpha_c s_c)\\
&= \sum_{c\in\mathscr{C}^f,\ c=w\cdots} \mu(c)\\
&= \sum_{c\in\mathscr{C}^f,\ c=w\cdots} \pi\left(\mathscr{L}\overline{c}\right)\\
&= \pi\left(\mathscr{L}\overline{w}\right) - \sum_{c\in\mathscr{C}^i,\ c=w\cdots} \pi\left(\overline{c}\right) = \mu(w).
\end{aligned}
$$

Since $f(\pi)=(v_{\alpha s})_{\alpha s}$, this concludes the proof. □

3.3 Two particular cases

3.3.1 Finite hat, 1 or 0 as a context

Proposition 1. *Let $(\mathscr{T},q)$ be a non-null probabilised context tree and U the associated VLMC. Assume that 1 and 0^a are contexts, for some integer $a\geqslant 1$ (or symmetrically that $0\in\mathscr{C}$ and $1^a\in\mathscr{C}$). Then, U admits a unique invariant probability measure.*



The proof is based on Kac's theorem (see Theorem 10.2.2 together with Theorem 10.0.1 in Meyn and Tweedie [2009]). Indeed, for any context c, the cylinder $\mathscr{L}\overline{c}$ is an *atom* (see [Meyn and Tweedie, 2009, Chap 5, p. 96]). Besides, the non triviality of the q_c yields the λ-irreductibility of $(U_n)_{n\geqslant 0}$ (see [Meyn and Tweedie, 2009, Chap 5]), where λ denotes the Lebesgue measure on $[0,1]$. Due to Kac's theorem, $(U_n)_n$ is positive recurrent if and only if $\mathbf{E}\left[\tau_1|U_0\in\mathscr{L}1\right]<\infty$, where

$$
\tau_1=\inf\{n\geqslant 1,\quad U_n\in\mathscr{L}1\}
$$

is the first return time in $\mathscr{L}1$. Since

$$
\begin{aligned}
\mathbf{E}\left[\tau_1|U_0\in\mathscr{L}1\right] &= \sum_{n=1}^{\infty} n\,\mathbf{P}\left(U_n\in\mathscr{L}10^{n-1}1|U_0\in\mathscr{L}1\right)\\
&\leqslant \sum_{n=1}^{a} n\,\mathbf{P}\left(U_n\in\mathscr{L}10^{n-1}1|U_0\in\mathscr{L}1\right)+\sum_{n=a+1}^{\infty} nq_{0^a}(0)^{n-a-1}q_{0^a}(1),
\end{aligned}
$$

and since $q_{0^a}(0)\neq 1$, we have $\mathbf{E}\left[\tau_1|U_0\in\mathscr{L}1\right]<\infty$. Thus, the non triviality of the q_c implies that U is positive recurrent and hence U admits a unique invariant probability measure.

3.3.2 Finite number of infinite branches and uniformly bounded q_c

Proposition 2. *Let $(\mathscr{T},q)$ be a probabilised context tree and U the associated VLMC. Assume that*

(i) $\exists\varepsilon > 0$, $\forall c \in \mathscr{C}$, $\forall \alpha \in \mathscr{A}$, $\varepsilon < q_c(\alpha) < 1-\varepsilon$ *(strong non-nullness);*
(ii) $\mathscr{C}^i$ *is a finite set.*
Then, U admits at least one invariant probability measure.

This result is a consequence of the theorem proved in Gallo and Paccaut [2013], stated in the framework of g-measures, which contains the VLMC processes. Assuming regularity conditions on the g-function (which plays the role of the q_c), a uniqueness result is also obtained.
The proof relies on a careful study of the so-called transfer operator associated to the g-function. The infinite contexts in the VLMC framework play the role of discontinuities of the g-function. When the g-function is continuous, it is straightforward to find a fixed point for the dual of the transfer operator. This fixed point is an invariant probability measure for the process. The result extends to the case when the g-function has a finite number of discontinuities.

4 Shift-stable context trees

This section deals with a subclass of context trees defined by a hypothesis put on their shape (see definition and characterizations of stable trees in Section 4.1). For this subclass, using the notion of descent tree (see Section 4.2), the matrix Q is proved to be irreducible and stochastic. In Section 4.4, using Theorem 1, a necessary and sufficient condition is given for a stable tree VLMC to admit a (unique) stationary probability measure (Theorem 2). In particular, when $\mathscr{S}$ is finite, this NSC reduces to the convergence of cascade series (Theorem 3), a rather easy to handle condition. Finally, Section 4.6 is dedicated to the link that can be made with the semi-Markov chains theory.

4.1 Definitions, examples

Proposition 3. *Let $\mathscr{T}$ be a context tree. The following conditions are equivalent.*

(i) The unlabelled tree $\mathscr{T}$ is invariant by the shift σ : $\forall\alpha \in \mathscr{A}$, $\forall w \in \mathscr{W}$, $\alpha w \in \mathscr{T} \implies w \in \mathscr{T}$. In an equivalent manner, $\sigma(\mathscr{T}) \subseteq \mathscr{T}$.

(ii) If c is a finite context and $\alpha \in \mathscr{A}$, then αc is noninternal.

(iii) $\forall\alpha \in \mathscr{A}$, $\mathscr{T} \subset \alpha\mathscr{T}$, where $\alpha\mathscr{T} = \{\alpha w, w \in \mathscr{T}\}$.

(iv) For any VLMC $(U_n)_n$ associated with $\mathscr{T}$, the process $\left(\mathrm{pref}\left(\overline{U_n}\right)\right)_{n\in\mathbb{N}}$ defines a Markov chain with state space $\mathscr{C}$.

Proof.
$(i) \implies (ii)$. Take $c \in \mathscr{C}$ and $\alpha \in \mathscr{A}$. If $\alpha c \in \mathscr{I}$ then $\alpha c0 \in \mathscr{T}$. The item (i) implies $c0 \in \mathscr{T}$, which contradicts $c \in \mathscr{C}$.
$(ii) \implies (i)$. Take $\alpha \in \mathscr{A}$ and w such that $\alpha w \in \mathscr{T}$. If $w \notin \mathscr{T}$ then there exists a finite context c such that $w = cw'$ with $w' \neq \emptyset$. It comes $\alpha cw' \in \mathscr{T}$, which implies $\alpha c \in \mathscr{I}$ and this contradicts (ii).
$(i) \iff (iii)$ is easy.
$(ii) \implies (iv)$ What needs to be proved is that $\mathrm{pref}\left(\overline{U_{n+1}}\right)$ only depends on U_n through $\mathrm{pref}\left(\overline{U_n}\right)$. In other words, we shall prove that for all $s \in \mathscr{L}, \alpha \in \mathscr{A}$, $\mathrm{pref}\left(\alpha\overline{s}\right)$ only depends on s through $\mathrm{pref}(\overline{s})$. This is clear because

$$\mathrm{pref}(\alpha\overline{s}) = \mathrm{pref}\left(\alpha\,\mathrm{pref}(\overline{s})\right).$$

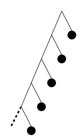
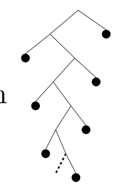
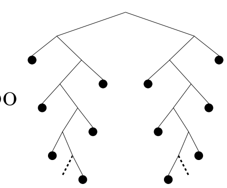
Indeed, $\mathrm{pref}(\overline{s}) \in \mathscr{C}$ and (ii) implies $\alpha\,\mathrm{pref}(\overline{s}) \notin \mathscr{I}$. Therefore $\alpha\,\mathrm{pref}(\overline{s})$ writes cw with $c \in \mathscr{C}$ and $w \in \mathscr{W}$. On one hand, this entails $\mathrm{pref}\left(\alpha\,\mathrm{pref}(\overline{s})\right) = c$. On the other hand, this means that cw is a prefix of $\alpha\overline{s}$ thus $\mathrm{pref}(\alpha\overline{s}) = c$.

$(iv) \implies (ii)$ We shall prove the contrapositive. Assume there exists $c \in \mathscr{C}^f$ and $\alpha \in \mathscr{A}$ such that $\alpha c \in \mathscr{I}$. Let $s \in \mathscr{L}$ such that $\mathrm{pref}(\overline{s}) = c$. As $\alpha c \in \mathscr{I}$, $\mathrm{pref}(\alpha\overline{s})$ is a context which has αc as a strict prefix. Therefore, $\mathrm{pref}(\alpha\overline{s})$ does not only depend on $\mathrm{pref}(\overline{s})$, but going further in the past of s is needed. □

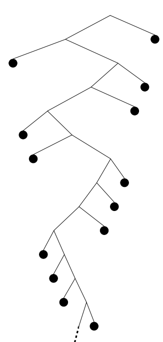
Definition 8 (shift stable tree)**.** *A context tree is* shift stable*, shortened in the sequel as* stable *when one of the four equivalent conditions of Proposition 3 is satisfied.*

Remark 4. If $\mathscr{T}$ is stable then $\sigma(\mathscr{I}) \subseteq \mathscr{I}$. Namely, if $v \in \mathscr{I}$, $v \neq \emptyset$, then $v\alpha \in \mathscr{T}$ for any $\alpha \in \mathscr{A}$. As $\mathscr{T}$ is stable, $\sigma(v\alpha) = \sigma(v)\alpha \in \mathscr{T}$, which implies $\sigma(v) \in \mathscr{I}$.

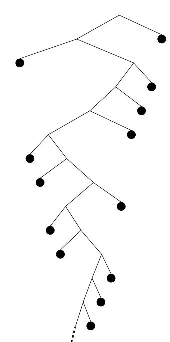
Definition 9 (stabilizable tree, stabilized of a tree)**.** *A context tree is* stabilizable *whenever the stable tree* $\bigcup_{n\in\mathbb{N}} \sigma^n(\mathscr{T})$ *has at most countably many infinite branches (*i.e. *when the latter is again a context tree). When this occurs,* $\bigcup_{n\in\mathbb{N}} \sigma^n(\mathscr{T})$ *is called the* stabilized *of* $\mathscr{T}$*; it is the smallest stable context tree containing* $\mathscr{T}$*.*

For example, the left-comb  is stable. On the contrary, the bamboo blossom  is non-stable; it is stabilizable, its stabilized being the double bamboo  .

Remark 5. A context tree is not necessarily stabilizable as the following examples show.



This context tree consists in saturating the infinite word $010011\ldots0^k1^k\cdots$ by adding hairs. This filament tree is stabilizable, its stabilized being the context tree having the $\{0^l1^k0^{k+1}1^{k+1}\cdots\}$ and the $\{1^l0^k1^{k+1}0^{k+1}\cdots\}$, $k \geqslant 1, 0 \leqslant l \leqslant k-1$ as internal nodes. Its countably many infinite branches are the 0^k1^∞ and the 1^k0^∞, $k \geqslant 0$.

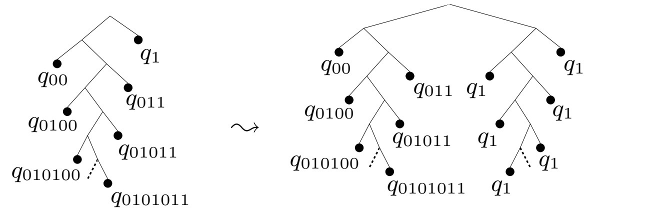


This context tree, denoted by $\mathscr{T}$ for a while, consists in saturating the infinite word $0100011011000001\cdots$ made of the concatenation of all finite words taken in length-alphabetical order. It is not stabilizable. Indeed, any finite word belongs to the smallest stable tree that contains $\mathscr{T}$, the latter having thus has uncountably many infinite branches.

Proposition 4. *Let $(\mathscr{T}, q)$ be a stabilizable probabilised context tree and $\widehat{\mathscr{T}}$ its stabilized. For every context c of $\widehat{\mathscr{T}}$, define $\widehat{q}_c = q_{\text{pref}(c)}$ where the function* pref *is relative to $\mathscr{T}$. Then $(\mathscr{T}, q)$ and $(\widehat{\mathscr{T}}, \widehat{q})$ define the same VLMC.*

Proof. Both VLMC, as Markov processes on $\mathscr{L}$, have the same transition probabilities. □

The example of the opposite figure illustrates the Proposition for the bamboo blossom and its stabilized tree, the double bamboo.



$(\mathscr{T}, q)$ $\left(\widehat{\mathscr{T}}, \widehat{q}\right)$

4.2 Descent tree of a stable context tree

In the stable case, the finite contexts organize in a remarkable way: all the contexts that have the same α-lis may be seen as the nodes of a tree called *descent tree*. The labels of these nodes are read from right to left (unlike the usual case where labels are read from left to right). This situation is precised by the following lemmas. In particular, Lemma 4 explains how the various node types of a descent tree (with two children, one or no child) correspond to different context types having two, only one or the empty lis as a prefix.

Lemma 3. *Let $(\mathscr{T}, q)$ be a stable context tree.*
(i) Any context α-lis is a context. In otherwords, $\mathscr{S} \subseteq \mathscr{C}$.
(ii) Assume that $c = \cdots[\alpha s] \in \mathscr{C}^f$. Then all $\sigma^k(c)$, $0 \leqslant k \leqslant |c| - |\alpha s|$ are also contexts.
(iii) For any $\alpha s \in \mathscr{S}$, the set $\{c \in \mathscr{C}^f, c = \cdots[\alpha s]\}$ constitutes the nodes of a tree, defined below as a descent tree.

Proof. Let $\alpha s \in \mathscr{S}$ and let $c \in \mathscr{C}^f$ such that $c = \cdots[\alpha s]$. Since $\mathscr{T}$ is stable, for any $k \in \mathbb{N}$, the node $\sigma^k(c)$ is internal or a context. By maximality of s, this implies that the $\sigma^k(c)$, for $0 \leqslant k \leqslant |c| - |\alpha s|$, have αs as a suffix and are noninternal, thus contexts. This proves (ii), thus (i) and (iii). □

Definition 10. *Let $\mathscr{T}$ be a context tree and $\alpha s \in \mathscr{S}$ be a context α-lis. The* descent tree associated with αs *is the tree whose nodes are the finite contexts of $\mathscr{T}$ having αs as an α-lis. The nodes are labelled reading from right to left (instead of the more common other labelling that reads words from left to right). The root is αs. This tree is denoted by $\mathscr{D}_{\alpha s}$:*

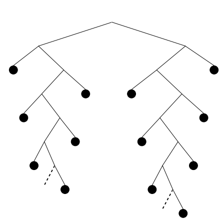
$$\mathscr{D}_{\alpha s} = \{\sigma^n(c),\ c \in \mathscr{C},\ c = \cdots[\alpha s],\ 0 \leqslant n \leqslant |c| - |\alpha s|\}\,.$$

The saturated descent tree associated with αs *is the descent tree of αs completed by the daughters of all its nodes:*

$$\overline{\mathscr{D}_{\alpha s}} = \mathscr{D}_{\alpha s} \cup \{\beta c,\ \beta \in \mathscr{A},\ c \in \mathscr{D}_{\alpha s}\}\,.$$

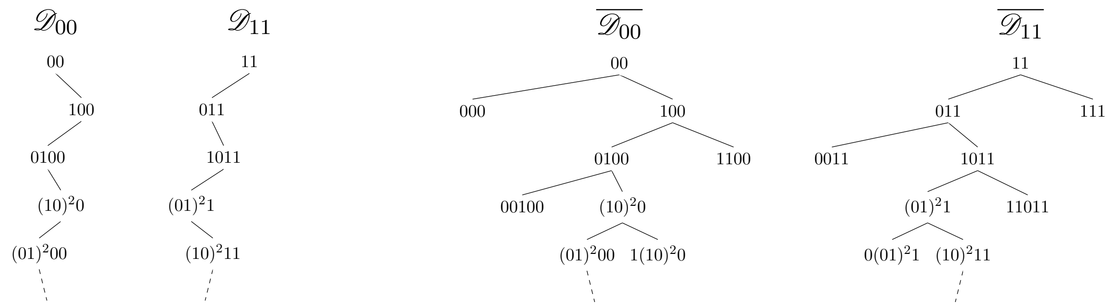
As an illustration, see the double bamboo in Example 2. For the left-comb of left-comb of Section 5.4, the descent tree that corresponds to the α-lis 10^n1 is simply an infinite left-comb.

Example 2 **(double bamboo).**



α-lis αs	contexts having αs as an α-lis
00	$(01)^k00, k \geqslant 0$; $(10)^k0, k \geqslant 1$
11	$(10)^k11, k \geqslant 0$; $(01)^k1, k \geqslant 1$

The descent trees: The saturated descent trees:



Lemma 4. *Let $(\mathscr{T}, q)$ be a stable context tree. Let $c \in \mathscr{C}$ be a context of $\mathscr{T}$.*
(i) Are equivalent:
(i.1) $0c \in \mathscr{C}$ and $1c \in \mathscr{C}$;
(i.2) c does not admit any context lis as a prefix.
(ii) Let $\alpha \in \mathscr{A}$. Are equivalent:
(ii.1) $\alpha c \notin \mathscr{C}$ and $\overline{\alpha} c \in \mathscr{C}$;
(ii.2) there exists a unique context lis t such that $c = t \cdots$, $\alpha t \in \mathscr{C}$ and $\overline{\alpha} t \notin \mathscr{C}$.
(iii) Are equivalent:
(iii.1) $0c \notin \mathscr{C}$ and $1c \notin \mathscr{C}$;
(iii.2) there exists a unique context lis t_0 and a unique context lis t_1 (that might be equal) such that: $c = t_0 \cdots = t_1 \cdots$, $0t_0 \in \mathscr{C}$ and $1t_1 \in \mathscr{C}$.

Proof. •$(i.1 \Rightarrow i.2)$ Assume that t is a context lis such that $c = t \cdots$. Then $0t$ or $1t$ is a context α-lis. Since $\mathscr{T}$ is stable, Lemma 3 implies that $0t$ or $1t$ is thus a context. Consequently, $0c \notin \mathscr{C}$ or $1c \notin \mathscr{C}$ because two different contexts cannot be prefix of one another.
•$(ii.1 \Rightarrow ii.2)$ Assume (say) that $\alpha = 0$, *i.e.* that $0c \notin \mathscr{C}$ and $1c \in \mathscr{C}$.
(a) Existence of t. Since $\mathscr{T}$ is stable, $\mathscr{T}$ is a subtree of $\mathscr{A}\mathscr{T}$ so that $0c$, which is thus noninternal, is an external node. Let $c' = \mathrm{pref}(0c)$. The context c' is a strict prefix of $0c$ that satisfies $c' = 0 \cdots$. Since $\mathscr{T}$ is stable, $\sigma(c')$ is nonexternal. But $\sigma(c')$ cannot be a context because it is a prefix of c which is a context. Thus $\sigma(c')$ is internal. This implies that $t := \sigma(c')$ is the lis of c' and a prefix of c as well, and that $0t = c' \in \mathscr{C}$. Finally, since $c' = 0t = \mathrm{pref}(0c)$, t is a prefix of c, so that $1t$ is a prefix of the context $1c$. Consequently, $1t \notin \mathscr{C}$ because two different contexts cannot be prefix of one another.
(b) uniqueness of t. Assume that $c = t \cdots = t' \cdots$ where t and t' are different context lis's such that $0t \in \mathscr{C}$ and $0t' \in \mathscr{C}$. Then $0t$ and $0t'$ are different contexts that are both prefixes of $0c$, thus prefix of one another, which is not possible.
•$(iii.1 \Rightarrow iii.2)$ Assume that $0c \notin \mathscr{C}$ and $1c \notin \mathscr{C}$. As in the proof of $(ii.1 \Rightarrow ii.2)$, there is a unique context lis t_0 such that $c = t_0 \cdots$ and $0t_0 \in \mathscr{C}$. By the same argument, there is a unique context lis t_1 such that $c = t_1 \cdots$ and $1t_1 \in \mathscr{C}$.
• End of the proof.
On one side, (i.1), (ii.1) and (iii.1) are disjoint cases that cover all possible situations. On the other side, the same can be said about (i.2), (ii.2) and (iii.2), because of what follows. A context cannot be written $c = s \cdots = t \cdots$ where, for a same $\alpha \in \mathscr{A}$, $\alpha s \in \mathscr{S}$, $\alpha t \in \mathscr{S}$ and $s \neq t$. Indeed, once again, αs and αt would be different contexts that are both prefixes of αc, thus prefix of one another, which is not possible. Thus, the three equivalences are proven. □

Remark 6. In terms of descent trees, Lemma 4 can be seen the following way. A context that belongs to case (i) has valence 3 in its descent tree (one parent, two children, it is called a bifurcation). A context that belongs to case (ii) has valence 2 in its descent tree (one parent, one child, it is called monoparental). A context that belongs to case (iii) has valence 1 in its descent tree (it is a leaf: one parent, no child).

4.3 Properties of Q

Definition 11. *A (finite or denumerable) matrix $(a_{r,c})_{r,c}$ is said* row-stochastic *whenever all its rows (are summable and) sum to* 1, i.e.

$$\forall r,\ \sum_c a_{r,c} = 1.$$

Proposition 5. *Let $(\mathscr{T}, q)$ be a stable probabilised context tree. Assume that the cascade series (9) converge. Then, the matrix Q is row-stochastic.*

The row-stochasticity of Q writes

$$\forall \alpha s \in \mathscr{S},\ \sum_{\beta t \in \mathscr{S}} Q_{\alpha s, \beta t} = 1.$$

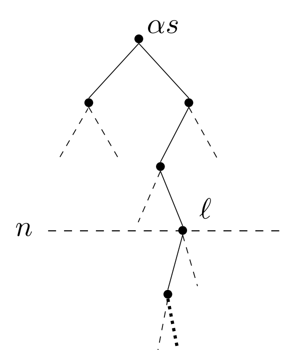
Lemma 5. *Let $\mathscr{D}$ be the descent tree associated with some α-lis of a stable probabilised context tree and let $\overline{\mathscr{D}}$ be the associated saturated descent tree.*
(i) For any $n \in \mathbb{N}$, if $\overline{\mathscr{D}}_n = \{w \in \overline{\mathscr{D}},\ |w| \leqslant n\}$ denotes the n-th truncated tree of $\overline{\mathscr{D}}$ and $\mathscr{L}\left(\overline{\mathscr{D}}_n\right)$ the set of its leaves, then

$$1 = \sum_{\ell \in \mathscr{L}\left(\overline{\mathscr{D}}_n\right)} \operatorname{casc}(\ell).$$

(ii) Denote by $\mathscr{I}(\mathscr{D})$ the set of internal nodes of $\mathscr{D}$ and $\mathscr{L}\left(\overline{\mathscr{D}}\right)$ the set of finite leaves of $\overline{\mathscr{D}}$. If $\displaystyle\sum_{\ell \in \mathscr{I}(\mathscr{D})} \operatorname{casc}(\ell)$ is summable, then

$$1 = \sum_{\ell \in \mathscr{L}\left(\overline{\mathscr{D}}\right)} \operatorname{casc}(\ell).$$

Proof of Lemma 5. Denote by $\mathscr{T}$ the probabilised context tree and by αs the root of the descent tree $\mathscr{D}$.



(i) Let ℓ be a leaf of $\overline{\mathscr{D}}_n$ and assume that ℓ is an internal node of $\overline{\mathscr{D}}_{n+1}$. In particular, $\ell \in \mathscr{D}$ which means that ℓ is a suffix of some context c with α-lis αs. Then, either $\ell = c$ is a context, or ℓ writes $\ell = \cdots \alpha s$ and c writes $c = \cdots \beta \ell = \cdots [\alpha s]$, $\beta \in \mathscr{A}$, which prevents ℓ to be internal (in $\mathscr{T}$) by maximality of s. In any case, ℓ is a noninternal node of $\mathscr{T}$. Consequently, by Remark 2, the daughters 0ℓ and 1ℓ are leaves of $\overline{\mathscr{D}}_{n+1}$ that satisfy $\operatorname{casc}(\ell) = \operatorname{casc}(0\ell) + \operatorname{casc}(1\ell)$. This proves (i) by induction on n (note that the cascade of an α-lis is always 1).

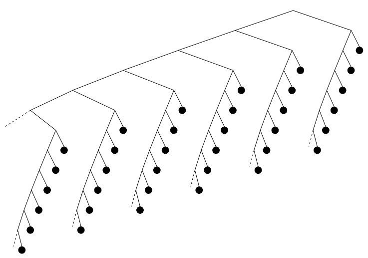
(ii) Because of (i), for any n,

$$\sum_{\substack{\ell \in \mathscr{L}\left(\overline{\mathscr{D}}\right) \\ |\ell| \leqslant n}} \operatorname{casc}(\ell) = \sum_{\ell \in \mathscr{L}\left(\overline{\mathscr{D}}_n\right)} \operatorname{casc}(\ell) - \sum_{\substack{\ell \in \mathscr{I}\left(\overline{\mathscr{D}}\right) \\ |\ell| = n}} \operatorname{casc}(\ell) = 1 - \sum_{\substack{\ell \in \mathscr{I}(\mathscr{D}) \\ |\ell| = n}} \operatorname{casc}(\ell).$$

Because of the summability assumption, the last sum indexed by ℓ tends to 0 when n tends to infinity. Since the cascades are nonnegative numbers, this shows that $\sum_{\ell \in \mathscr{L}\left(\overline{\mathscr{D}}\right)} \operatorname{casc}(\ell)$ is summable and proves the result. ☐

Proof of Proposition 5. Let $\alpha s \in \mathscr{S}$. Name $\mathscr{D} = \mathscr{D}_{\alpha s}$ its descent tree and $\overline{\mathscr{D}} = \overline{\mathscr{D}_{\alpha s}}$ its saturated descent tree. The assumption guarantees that the family $(\mathrm{casc}(c))_{c\in\mathscr{D}}$ is summable, or equivalently that the family $(\mathrm{casc}(\ell))_{\ell\in\mathscr{I}(\mathscr{D})}$ is summable (notations of Lemma 5). Remember that $Q_{\alpha s,\beta t}$ is the sum of $\mathrm{casc}\,(\beta c)$ where c runs over all contexts such that $c = \cdots[\alpha s] = t\cdots$. Take $c \in \mathscr{D}$, which means that c is a context with α-lis αs. The proof is now based on Lemma 4. If c belongs to case (i), it admits no context-lis as a prefix so that it never appears as a term in some $Q_{\alpha s,\beta t}$. If c belongs to case (ii), then the only $\beta t \in \mathscr{S}$ such that $c = t\cdots$ is a leaf of $\overline{\mathscr{D}}$ whose sister is an internal node of $\overline{\mathscr{D}}$. If finally c belongs to case (iii), then the contexts $0t_0 \in \mathscr{S}$ and $1t_1 \in \mathscr{S}$ such that $c = t_0\cdots = t_1\cdots$ give rise to two sister leaves of $\overline{\mathscr{D}}$, $0c$ and $1c$. Putting the three cases together one sees that any word of the form βc where $c = \cdots[\alpha s] = t\ldots$ and $\beta t \in \mathscr{S}$ appears once and only once as a term of some $Q_{\alpha s,\beta t}$. Consequently, the sum of all $Q_{\alpha s,\beta t}$ where βt runs over $\mathscr{S}$ can be written as the sum of the cascades of all leaves of $\overline{\mathscr{D}}$. By Lemma 5, this leads to the result. □

Remark 7. Any stochastic matrix with strictly positive coefficients $A = (a_{ij})_{i\geqslant 0,j\geqslant 0}$ is a α-lis transition matrix of a non-null stable context tree. It may be realised with a left-comb of left-combs (see Example 5.4).



The contexts are 0^i10^j1, $i,j \geqslant 0$, the α-lis of 0^i10^j1 being 10^j1. One can check that

$$Q_{i,j} := Q_{10^i1,10^j1} = \mathrm{casc}(10^j10^i1).$$

A calculation shows that if

$$q_{0^i10^j1}(1) = \frac{a_{ji}}{1-\sum_{k=0}^{i-1}(1-a_{jk})},$$

then $Q_{ij} = a_{ij}$. The question whether any stochastic matrix (with some zero coefficients) can be realised seems to be more difficult. Namely, zero coefficients in Q assuming non-zero $q_c(\alpha)$ constraint the shape of the context tree.

Proposition 6. *Let $(\mathscr{T}, q)$ be a non-null stable probabilised context tree. Then the matrix Q is irreducible.*

Proof. Let αs and βt be two α-lis. We shall show that there is a path in the matrix Q from αs to βt. Recall that there is a path of length one from αs to $\alpha' s'$ if there exists a context the α-lis of which is αs and which begins with the lis s', $\alpha' s'$ being a α-lis. In this case, the non-nullness of $(\mathscr{T}, q)$ implies that the transition from αs to $\alpha' s'$ is not zero.
To illustrate the transition from αs to βt, let us use the concatenated word

$$\beta t\alpha s = \beta t_q t_{q-1}\ldots t_1\alpha s.$$

Let us start from αs and add letters from t to the left. As $\alpha s \in \mathscr{C}$ and $\mathscr{T}$ is stable, when we add letters to the left, we get either contexts or strictly external words. Let us distinguish the two following cases:

- Either we always get contexts. In this case, since no internal node is obtained by adding the letters of t, $t\alpha s$ is a context with α-lis αs . Moreover, as βt is a context, $\beta t\alpha s \notin \mathscr{T}$ which implies by Lemma 4 that the context $t\alpha s$ begins with the unique lis t. As βt is an α-lis, this is the case of a one-step transition form αs to βt.

- Or let k be the least integer such that $t_k \dots t_1 \alpha s \notin \mathscr{T}$. Firstly $t_{k-1} \dots t_1 \alpha s$ is a context with α-lis αs (same argument as above). Secondly, by Lemma 3, the context $t_{k-1} \dots t_1 \alpha s$ begins with a unique lis, denoted by u and $t_k u$ is an α-lis. Now the aim is to have a transition with the whole of t. Therefore, one needs to check that $t_k u$ contains the whole end of t. If it were not the case, u would write $t_{k-1} \dots t_l$ with $l > 1$ and $t_k \dots t_l$ would be a context and a suffix of the internal node $\beta t_q \dots t_l$. This can't happen in a stable tree. We have found a transition from αs to $t_k u = t_k \dots t_1 u'$. Now we go on adding letters from t to the left and we again have the dichotomy described in the two items of this proof. By repeating this a finite number of times, we get the result.

□

4.4 Stationary measure for the VLMC vs recurrence of Q

The following result relates the existence and the uniqueness of a stationary probability measure of a VLMC to the recurrence of Q. Let us recall the definition of recurrence and state a necessary and sufficient condition to get a (unique) invariant probability measure for stable trees.

Definition 12. *Let $A = (a_{ij})$ be a stochastic irreducible countable matrix. Denote by $a_{ij}^{(k)}$ the (i,j)-th entry of the (stochastic) matrix A^k. The matrix A is* recurrent *whenever there exists i such that*

$$\sum_{k=1}^{\infty} a_{ii}^{(k)} = 1.$$

Any stochastic irreducible countable matrix may be viewed as the transition matrix of an irreducible Markov chain with countable state space. The recurrence means that there is a state i (and this is true for every state because of irreducibility) for which the first return time is a.s. finite. When in addition the expectation of the return times are finite, the matrix is classically called *positive recurrent*.

Theorem 2. *Let $(\mathscr{T}, q)$ be a non-null probabilised context tree. Assume that $\mathscr{T}$ is stable. Then, the following assertions are equivalent.*

1. *The VLMC associated to $(\mathscr{T}, q)$ has a unique stationary probability measure*
2. *The VLMC associated to $(\mathscr{T}, q)$ has at least a stationary probability measure*
3. *The three following conditions are satisfied:*
   - (c_1) *the cascade series (9) converge*
   - (c_2) *Q is recurrent*
   - (c_3) *$\sum_{\alpha s \in \mathscr{S}} v_{\alpha s} \kappa_{\alpha s} < \infty$ where $\mathbb{R}\left(v_{\alpha s}\right)_{\alpha s}$ is the unique line of left-fixed vectors of Q.*

Proof.
(*3.* $\Longrightarrow$ *1.*) Since Q is recurrent and irreducible, there exists a unique line $\mathbb{R}v$ of left-fixed vectors for Q, where $v = \left(v_{\alpha s}\right)_{\alpha s \in \mathscr{S}}$ (see for example [Seneta, 2006, Theorem 5.4]). Theorem 1(ii) coupled with the assumption on the series $\sum_{\alpha s \in \mathscr{S}} v_{\alpha s} \kappa_{\alpha s}$ entails directly the existence and uniqueness of a stationary probability measure.
(*2.* $\Longrightarrow$ *3.*) If there exists a stationary probability measure, then Theorem 1(i) asserts that the cascade series converge and that Q admits at least one direction of left-fixed vectors $\mathbb{R}v$ such that $\sum_{\alpha s \in \mathscr{S}} v_{\alpha s} \kappa_{\alpha s} < \infty$. Besides, every $\kappa_{\alpha s}$ is greater than 1. Indeed, the cascade of any α-lis is 1 and, in the stable case, any α-lis is a context. Thus, v is summable and Q is positive recurrent (see for instance [Seneta, 2006, Corollary of Theorem 5.5]). Since it is irreducible (Proposition 6), it admits a unique direction of left-fixed vectors $\mathbb{R}v$, proving (*3.*) by Theorem 1. □

Remark 8. Actually, as shown in the end of the proof, when Q is recurrent and when the series $\sum_{\alpha s\in\mathscr{S}} v_{\alpha s}\kappa_{\alpha s}$ converge, then Q is positive recurrent.

Remark 9. One can find a VLMC defined by a stable tree such that the cascade series converge and the matrix Q is transient.
To build such an example, recall that, by Remark 7, any stochastic matrix with strictly positive coefficients can be realized as the matrix Q of a stable tree (take for example a left-comb of left-combs). The matrix $A = (a_{ij})_{i\geqslant 1, j\geqslant 1}$ defined by

- $a_{i,i+1} = 1 - \frac{1}{(i+1)^2}$ for all $i \geqslant 1$,
- $a_{ij} = \frac{1}{(i+1)^2 2^{j-1}}$ if $j \geqslant i+2$,
- $a_{ij} = \frac{1}{(i+1)^2 2^{i+1-j}}$ if $j \leqslant i$.

is stochastic and transient. Indeed, if one associates a Markov chain to the stochastic matrix A and if one denotes by T_1 the return time to the first state,

$$\mathbf{P}(T_1 = \infty) \geqslant \prod_{i\geqslant 1} a_{i,i+1} \geqslant \prod_{i\geqslant 2}\left(1 - \frac{1}{i^2}\right) = \frac{1}{2}.$$

Consequently, A does not have any nonzero left-fixed vector.
Consider now the VLMC defined by a left-comb of left-combs (see 5.4) probabilised in the unique way such that $Q_{10^q1,10^p1} = a_{q,p}$ for every (p, q). A simple computation shows that the series of cascade converge (geometrically). Simultaneously, since Q is transient, Theorem 2 shows that the VLMC admits no stationary probability measure.

Remark 10. Assume that the context tree is stable and that the cascade series (9) converge.
If the VLMC admits a stationary probability measure, then, as already seen, every left-fixed vector of Q is summable (the family of its coordinates is summable).
The reciprocal implication is false: the left-fixed vectors of Q may be summable while no finite measure is stationary for the VLMC, because condition (c_3) in Theorem 2 is not satisfied. One can find in Section 5.4.2 such an example with a left-comb of left-combs.

Remark 11. If one removes the stability assumption, the $\kappa_{\alpha s}$ may not be bounded below, so that the argument that shows the summability of left-fixed vectors of Q fails. Indeed, with a non-stable context tree having an infinite $\mathscr{S}$, one may have

$$\inf_{\alpha s\in\mathscr{S}} \kappa_{\alpha s} = 0.$$

Such an example is developed in 5.5.

Notice that Theorem 2 also provides results for non-stable trees as the following corollary shows, using notations of Proposition 4.

Corollary 1. *Let $(\mathscr{T}, q)$ be a non-null probabilised context tree. Suppose that $\mathscr{T}$ is stabilizable and denote by $\widehat{\mathscr{T}}$ its stabilized. If $(\widehat{\mathscr{T}}, \widehat{q})$ satisfies the conditions of Theorem 2, then the VLMC associated with $(\mathscr{T}, q)$ admits a unique invariant probability measure. If not, it does not admit any invariant probability measure. In particular, a VLMC associated to such a context tree $(\mathscr{T}, q)$ never admits several stationary probability measures.*

Proof. This is a consequence of Proposition 4 and Theorem 2. $\square$

4.5 Existence and uniqueness when $\mathscr{S}$ is finite and $\mathscr{T}$ stable

When the matrix Q is finite, stochasticity and irreducibility are sufficient to get a unique left-fixed vector, therefore

Theorem 3 (finite number of α-lis)**.** *Let $(\mathscr{T}, q)$ be a non-null probabilised context tree. Assume that $\mathscr{T}$ is stable and that $\#\mathscr{S} < \infty$. Then (i) and (ii) are equivalent.*
(i) The VLMC associated to $(\mathscr{T}, q)$ has a unique stationary probability measure.
(ii) The cascade series (9) converge.

Proof. (i) $\Longrightarrow$ (ii) is contained in Theorem 1(i). Assume reciprocally that the cascade series converge. Since Q is stochastic, irreducible and finite dimensional, it admits a unique direction of left-fixed vectors, so that Theorem 1(ii) allows us to conclude. $\square$

To see how this theorem applies, see both examples in Section 5.6. The first one – the left-comb of right-combs – is spectacularly simple ($\#\mathscr{S} = 1$). The second one is a straightforward application of Theorem 3.

4.6 The α-lis process as a semi-Markov chain

For this section, semi-Markov chains can be comprehended thanks to Barbu and Limnios [2008]. Remember that for a VLMC (U_n) defined on a *stable* context tree, by (iv) in Definition 4.1, the process $(C_n)_{n\geqslant 0}$ defined by

$$C_n = \text{pref}\left(\overline{U_n}\right)$$

defines a Markov chain with state space $\mathscr{C}$, the set of contexts. Moreover, the process $(Z_n)_{n\geqslant 0}$ of α-lis of contexts, where

$$Z_n = \alpha_{C_n} s_{C_n},$$

is a semi-Markov chain with state space $\mathscr{S}$, as detailed in the following.

Definition 13. *Let $(J_n, S_n)_{n\geqslant 0}$ be a Markov chain with state space $\mathscr{S} \times \mathbb{N}$ such that $S_0 = 0$ and (S_n) is increasing. The semi-Markov chain associated with (J, S) is the $\mathscr{S}$-valued process $(Z_n)_{n\geqslant 0}$ defined by $Z_0 = J_0$ and*

$$\forall k \text{ such that } S_n \leqslant k < S_{n+1}, \quad Z_k = J_n.$$

In otherwords, the S_n are jump times and Z_k stagnates at a same state between two successive jump times.

In a VLMC (U_n) defined on a *stable* context tree, both processes (C_n) and (Z_n) evolve as follows: start with $Z_0 = \alpha s$ which is the α-lis of C_0. As time goes by, U_n grows by addition of letters and C_n goes down in the descent tree of αs (the length of C_n increases and Z_n remains equal to αs) until there is $n_0 \geqslant 0$ and $\beta \in \mathscr{A}$ such that $\beta C_{n_0} \notin \mathscr{C}$. Since a descent tree is not the complete tree (if not, the context tree would also be the complete tree, which is not allowed for a VLMC), n_0 is finite and lemma 4 implies the existence of a context lis t, prefix of C_{n_0} such that $\beta t \in \mathscr{S}$. Therefore, at the next time step, with nonzero probability $q_{C_{n_0}}(\beta)$, one has $C_{n_0+1} = Z_{n_0+1} = \beta t$, the process C_n has gone up to the root of the descent tree $\mathscr{D}_{\beta t}$. With probability $q_{C_{n_0}}(\beta)$, at time $n_0 + 1$, the process (Z_n) "changes" state; in this case, $n_0 + 1$ is a jump time. Remark that one may have that $n_0 + 1$ is a jump time *and* $Z_{n_0+1} = Z_{n_0}$ (it is the case when the context process goes back to the root of the descent tree it belongs to).
More generally, the evolution of C_n and Z_n shows that jump times may be defined as follows (with $S_0 = 0$) : for $n \geqslant 1$,

$$S_n = \min\{k > S_{n-1}, C_k = Z_k\}.$$

In particular, $C_{S_n} = Z_{S_n}$. At all the other times k, Z_k is a strict suffix of C_k. The jump times may be equivalently defined as

$$S_n = \min\{k > S_{n-1}, |C_k| \leqslant |C_{k-1}|\}.$$

For $n \geqslant 0$, let

$$J_n = Z_{S_n}$$

be the state of the α-lis process at the n-th jump, so that $S_{n+1} - S_n$ is the sojourn time in the state J_n. For $k \geqslant 1$, if $N(k) = \max\{n, S_n \leqslant k\}$ is the number of jump times in the interval $[1, k]$, one also have $J_{N(k)} = Z_k$.

Proposition 7. *Let (U_n) be a VLMC on a stable context tree and assume that the cascade series (9) converge. Let (J_n, S_n) be as defined above. Then the sojourn times $S_{n+1} - S_n$ are a.s. finite, (J_n, S_n) is a Markov chain on $\mathscr{S} \times \mathbb{N}$, (S_n) is increasing and (Z_n) is a semi-Markov chain associated with (J, S).*

Proof. Let

$$\begin{aligned}
q_{\alpha s, \beta t}(k) &= \mathbf{P}(J_{n+1} = \beta t, S_{n+1} = S_n + k | J_0, \ldots, J_{n-1}, J_n = \alpha s, S_0, \ldots, S_n) \\
&= \mathbf{P}(C_{S_n+1} \in \mathscr{D}_{\alpha s} \setminus \{\alpha s\}, \ldots, C_{S_n+k-1} \in \mathscr{D}_{\alpha s} \setminus \{\alpha s\}, C_{S_n+k} = \beta t | C_{S_n} = \alpha s) \\
&= \sum_{\substack{c = t \cdots [\alpha s] \\ |c| - |\alpha s| = k}} \operatorname{casc}(\beta c),
\end{aligned}$$

so that $q_{\alpha s, \beta t}(k) = \mathbf{P}(J_{n+1} = \beta t, S_{n+1} = S_n + k | J_n = \alpha s)$. This computation is illustrated in the example below, Remark 12.

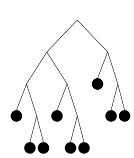
Notice also that $\sum_{k \geqslant 1} q_{\alpha s, \beta t}(k) = Q_{\alpha s, \beta t}$. Therefore, the semi-Markov kernel property of $q_{\alpha s, \beta t}(k)$, namely $\sum_{k \geqslant 1} \sum_{\beta t \in \mathscr{S}} q_{\alpha s, \beta t}(k) = 1$, follows straightforwardly from the stochasticity of Q, see Proposition 5.

Moreover, the stochasticity of Q provides the a.s. finiteness of $S_{n+1} - S_n$. Indeed, for any $\alpha s \in \mathscr{S}$,

$$\begin{aligned}
\sum_{k \geqslant 1} \mathbf{P}(S_{n+1} - S_n = k | J_n = \alpha s) &= \sum_{\beta t \in \mathscr{S}} \sum_{k \geqslant 1} \mathbf{P}(J_{n+1} = \beta t, S_{n+1} - S_n = k | J_n = \alpha s) \\
&= \sum_{\beta t \in \mathscr{S}} \sum_{k \geqslant 1} \sum_{\substack{c = t \cdots [\alpha s] \\ |c| - |\alpha s| = k}} \operatorname{casc}(\beta c) = \sum_{\beta t \in \mathscr{S}} Q_{\alpha s, \beta t} = 1
\end{aligned}$$

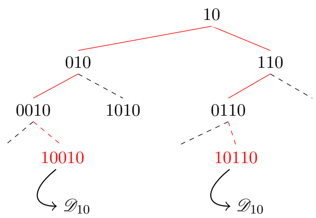
□

Remark 12. The semi-Markov chain contains less information than the chain (U_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

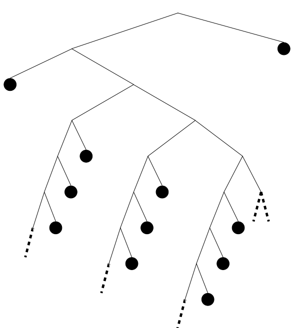
In this example, 0010 and 0110 are two contexts of same length, with the same α-lis 10 and beginning by the same lis 0. Hence if we know that $J_n = 10$, $S_{n+1} - S_n = 3$ and $J_{n+1} = 10$, there are two possibilities to reconstruct the VLMC (U_n). With the notations of the proof above, there are two cascade terms in $q_{10,10}(3)$:



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

5 Miscellaneous examples, bestiary

5.1 Example with $\mathscr{S}$ finite and $\mathscr{C}^i$ infinite



The finite contexts of this context tree are 00, 1, and the 01^p0^q1, $p, q \geqslant 1$. The infinite branches are the 01^p0^∞, $p \geqslant 1$, so that $\mathscr{C}^i$ is infinite. There are four context α-lis: 1, 00, 001 and 101, as the following array shows. Note that there are only three context lis: $\emptyset$, 0 and 01. This tree is non-stable.

$\alpha s \in \mathscr{S}$	contexts having αs as an α-lis
1	1
00	00
001	01^p0^q1, $p \geqslant 1$, $q \geqslant 2$
101	01^p01, $p \geqslant 1$

The cascade series converge as soon as $q_1(1) \neq 1$ and $q_{00}(0) \neq 1$, since in this case,

$$\kappa_1 = \kappa_{00} = \kappa_{001} = \kappa_{101} = 1.$$

Let

$$A = \sum_{k \geqslant 0} q_{01^k01}(0)q_1(0)q_1(1)^k$$

$$B = \sum_{k \geqslant 0, l \geqslant 0} q_{01^k0^l1}(1)q_1(0)q_{00}(1)q_1(1)^k q_{00}(0)^l.$$

The matrix Q writes as follows:

$$Q = \begin{pmatrix} q_1(1) & 0 & 0 & 0 \\ q_{00}(1) & q_{00}(0) & 0 & 0 \\ B & 1-B & 1-B & B \\ 1-A & A & A & 1-A \end{pmatrix}.$$

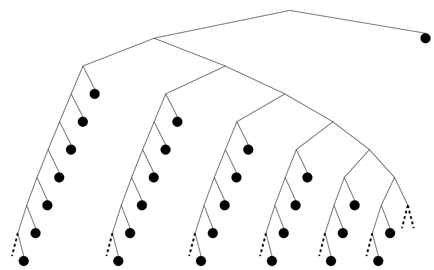
A simple computation leads to the unique direction of left-fixed vectors of Q

$$((A+B)q_{00}(1), Aq_1(0), Aq_1(0)q_{00}(1), Bq_1(0)q_{00}(1))$$

and Theorem 1*(ii)* gives existence and uniqueness of a probability stationary measure for the VLMC associated to this probabilised context tree. This application of Theorem 1 is an alternative argument to Section 3.3.1 this tree is an special case of.

5.2 A non-stable tree: the brush

This example provides an application of Theorem 1 that is not covered by particular cases of Sections 3.3.1 and 3.3.2.



The finite contexts of this non-stable tree are 1 and the 01^p0^q1, $p \geqslant 0$, $q \geqslant 1$. There are infinitely many infinite branches, namely the 01^p0^∞, $p \geqslant 0$. There are only three α-lis, as summed up in the following array.

α-lis αs	contexts having αs as an α-lis
1	1
001	0^q1 and 01^p0^q1, $p \geqslant 1$, $q \geqslant 2$
101	01^p01, $p \geqslant 1$

Compute the cascade series: $\kappa_1 = 1$, and $\kappa_{101} = 1$ as soon as $q_1(1) \neq 1$. If one denotes

$$c_q = \operatorname{casc}(0^q1) = \prod_{k=2}^{q-1} q_{0^k1}(0)$$

for every $q \geqslant 2$ (with $c_2 = 1$), then $\kappa_{001} = 1 + \sum_{q\geqslant 2} c_q$ as soon as this series converges. Finally, under the above hypothesis of the convergence of cascade series, let

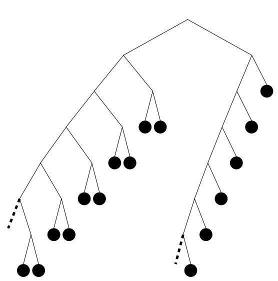
$$A = \sum_{p\geqslant 1,\ q\geqslant 2} \operatorname{casc}(101^p0^q1) \quad \text{and} \quad B = \sum_{p\geqslant 1} \operatorname{casc}(101^p01)$$

– these numbers are easily expressed in terms of the q_c. With these notations, one gets

$$Q = \begin{pmatrix} q_1(1) & 0 & 0 \\ 1+A & 1-A & A \\ B & 1-B & B \end{pmatrix}.$$

A simple glance to this matrix shows that 1 is its Perron-Frobenius eigenvalue, and that all its left-fixed vectors are proportional to $(A+1-B, (1-B)q_1(0), Aq_1(0))$. Under the convergence of cascade series, the corresponding VLMC admits a unique stationary probability measure.

5.3 Example with $\mathscr{S}$ infinite and $\mathscr{C}^i$ finite

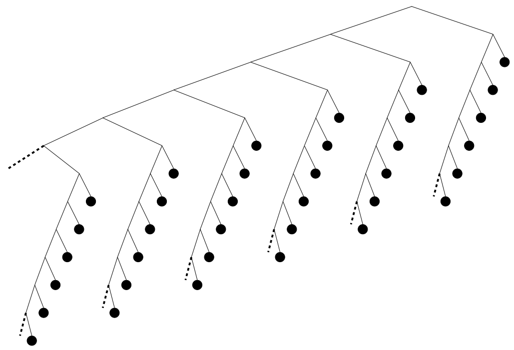


The finite contexts of this tree are the 0^q10, 0^q11 and 10^p1, $p \geqslant 0$, $q \geqslant 1$ while the infinite ones are $\mathscr{C}^i = \{0^\infty, 10^\infty\}$. There are infinitely many context α-lis, as made precise by the array. This tree is stable.

$\alpha s \in \mathscr{S}$	contexts having αs as an α-lis
11	0^q11, $q \geqslant 1$
010	0^q10, $q \geqslant 1$
10^p1, $p \geqslant 1$	10^p1

5.4 The left-comb of left-combs

5.4.1 Definition and notations



The *left-comb of left-combs* is the context tree as drawn on the left: the finite contexts are the 0^p10^q1, $p,q \geqslant 0$. Remark in passing that, for any corresponding VLMC, the transition probabilities of the Markov process on $\mathscr{L}$ depend only on the largest suffix of the form 0^q10^p of the current left-infinite sequence $U_n = \cdots 0^q10^p$ (p being possibly infinite).

A left-comb of left-combs is a stable context tree. Its has infinitely many infinite branches, namely 0^∞ and the 0^p10^∞, $p \geqslant 0$. For any $p, q \geqslant 0$, the α-lis of 0^p10^q1 is 10^q1. In particular, the set $\mathscr{S}$ of α-lis of contexts is countably infinite. In this case, for any $q \geqslant 0$, the set of contexts having 10^q1 as an α-lis is also countably infinite.
Probabilise this context tree by a family $(q_c)_c$ of probability measures on $\{0,1\}$ and denote, for every $q, p \geqslant 0$,

$$c_{q,p} = \operatorname{casc}(0^p10^q1) = \prod_{0 \leqslant k \leqslant p-1} q_{0^k10^q1}(0).$$

The convergence of cascade series is equivalent to the finiteness of

$$\kappa_{10^q1} = \sum_{p \geqslant 0} c_{q,p}, \ \ \forall q \geqslant 0.$$

The square matrix Q is infinite, defined by $Q_{10^q1,10^p1} = \operatorname{casc}(10^p10^q1) = c_{q,p} - c_{q,p+1}$ for all $p, q \geqslant 0$. It has always finite entries, even if one cascade series diverges. One sees immediately that Q is line-stochastic if, and only if $c_{q,p}$ tends to 0 as p tends to infinity, for all $q \geqslant 0$.

5.4.2 Stationarity and summability of left-fixed vectors of Q

This paragraph is devoted to an example of (stable) VLMC that satisfies that the following properties:
- the cascade series converge;
- the VLCM admits no stationary probability measure;
- every left-fixed vector of Q is summable.

The context tree of the example is a left-comb of left-combs with the above notations.

Let $v_p = \frac{1}{p+1} - \frac{1}{p+2}$ and $R_p = \sum_{q \geqslant p} v_q = \frac{1}{p+1}$ for every $p \geqslant 0$ (more generally, on can build a similar counter-example based on any positive sequence $(v_p)_p$ such that $\sum_{p \geqslant 0} v_p = 1$ and $\sum p v_p$ diverge). Define S by

$$S(x) = \sum_{q \geqslant 0} v_q x^{\frac{1}{q+1}}.$$

The series is normally convergent on the real interval $[0,1]$ so that S is continuous on $[0,1]$ and satisfies $S(0) = 0$ and $S(1) = 1$. Furthermore, S is derivable and increasing on $[0,1]$ since the derived series converges normally on any compact subset of $]0,1]$. Finally, $S(x) \geqslant v_q x^{\frac{1}{q+1}}$ on $[0,1]$ for every $q \geqslant 0$. Consequently, for every $t > 0$, there exists $C_t > 0$ such that

$$S^{-1}(x) \leqslant C_t x^t \tag{20}$$

for every $x \in [0,1]$.

Take now the probabilised left-comb of left-combs defined by the relations

$$\forall q, p \geqslant 0,\ c_{q,p} = S^{-1}\left(R_p\right)^{\frac{1}{q+1}}.$$

Note that these equations fully define the corresponding VLCM because the probabilities $q_{0^p10^q1}$ are characterized by these $c_{q,p}$ *via* the equalities $q_{0^p10^q1}(0) = c_{q,p+1}/c_{q,p}$. The definition of S implies that $\sum_{q\geqslant 0} v_q c_{q,p} = R_p$ for every $p \geqslant 0$, which precisely means that $v = vQ$ (the row-vector v is a left-fixed vector for Q). Besides, for any $q \geqslant 0$, applying (20) for $t = 2(q+1)$ leads to inequalities

$$\forall p \geqslant 0,\ c_{q,p} \leqslant C_{2(q+1)}\left(\frac{1}{p+1}\right)^2.$$

Thus, the sequences (v_q) and $(c_{q,p})$ satisfy the following properties.

1. $\forall q \geqslant 0, \sum_p c_{q,p} < \infty$,
2. $\forall p \geqslant 0, \sum_{q\geqslant 0} v_q c_{q,p} = \sum_{q\geqslant p} v_q$,
3. $\sum_q v_q < \infty$,
4. $\sum_{q,p\geqslant 0} v_q c_{q,p} = +\infty$.

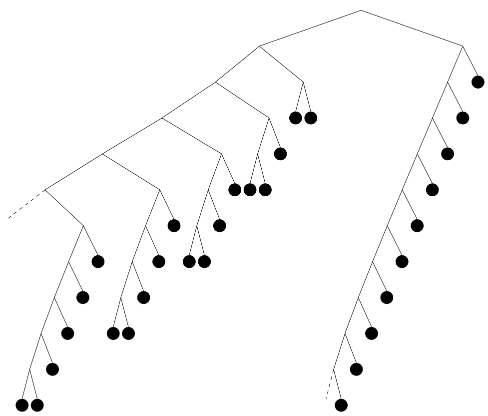
In terms of the VLMC, with general notations of Section 3, these properties translate into:

1. the cascade series converge,
2. $(v_{\alpha s})_{\alpha s\in\mathscr{S}}$ is a left-fixed vector for Q,
3. $\sum_{\alpha s\in\mathscr{S}} v_{\alpha s} < \infty$,
4. there exists a unique stationary positive measure π on $\mathscr{L}$ such that $\pi\left(\mathscr{L}\overline{\alpha s}\right) = v_{\alpha s}$ for every $\alpha s \in \mathscr{S}$. The measure π is not finite.

The existence and uniqueness of the measure π in item 4. can be shown by simple adaptation of the proof of Theorem 1, the total mass of π being

$$\pi(\mathscr{L}) = \sum_{\alpha s\in\mathscr{S}} v_{\alpha s}\kappa_{\alpha s} = \sum_{\substack{\alpha s\in\mathscr{S}\\ c\in\mathscr{C},\ c=\cdots[\alpha s]}} \mathrm{casc}(c) v_{\alpha s} = \sum_{q,p\geqslant 0} v_q c_{q,p}.$$

5.5 Tree of small kappas



The tree of small kappas is the context tree as drawn on the left. Its finite contexts are the following ones:

⋆ 0^m10^k1, $m \geqslant 1$, $0 \leqslant k \leqslant m-1$;

⋆ 0^m10^m, $m \geqslant 1$;

⋆ 10^m1, $m \geqslant 0$.

This context tree gets two infinite branches, namely 0^∞ and 10^∞. It is non-stable. There are infinitely many context α-lis, as summed up in the following array.

α-lis αs	contexts having αs as an α-lis
11	$0^m11,\ m \geqslant 0$
$10^k1,\ k \geqslant 1$	10^k1 and $0^m10^k1,\ m \geqslant k+1$
$10^m,\ m \geqslant 1$	0^m10^m

When the context tree is probabilised, the convergence of cascade series is equivalent to the convergence of the series

$$\kappa_{11} = \sum_{m \geqslant 0} \prod_{j=0}^{m-1} q_{0^j11}(0) \quad \text{and} \quad \kappa_{10^k1} = 1 + \sum_{m \geqslant k+1} \prod_{j=0}^{m-1} q_{0^j10^k1}(0), \ \forall k \geqslant 1.$$

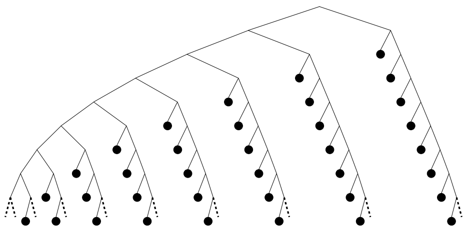
The remaining $\kappa_{\alpha s}$ are defined by sums of one sole term, namely, for all $m \geqslant 1$,

$$\kappa_{10^m} = \operatorname{casc}(0^m10^m) = \prod_{j=1}^{m-1} q_{0^j10^j}(0).$$

In particular, the sequence $(\kappa_{10^m})_m$ is not bounded below by any positive number as soon as the infinite product diverges to 0.

5.6 Variations on the left-comb of right-combs

This Section produces two examples of stable context trees that give rise to a direct application of Theorem 3. The first one, named left-comb of right-combs, is particularly simple because if has only one α-lis of contexts. The left-comb of right-combs augmented by a cherry stem, a variation of the former one, gets four α-lis of contexts. Because of Theorem 3, both corresponding VLMC have a (unique) stationary probability measure if, and only if their cascade series converge.



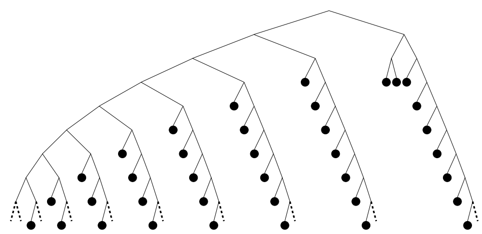
The finite contexts of the left-comb of right-combs (drawn on the left) are the 0^p1^q0, $p \geqslant 0$, $q \geqslant 1$. It has infinitely many infinite branches, namely the 0^p1^∞, $p \geqslant 0$. This context tree is stable and all finite contexts have 01 as an α-lis. The matrix Q, which is thus 1-dimensional, is reduced to (1). The convergence of the unique cascade series consists in the summability of the double sum

$$\sum_{p\geqslant 0, q\geqslant 1} \prod_{j=0}^{p-1} q_{0^j1^q0}(0) \prod_{k=1}^{q-1} q_{1^k0}(1).$$

Note that the transition probabilities of this Markov chain depend only of the largest suffix of the form 1^q0^p of the current left-infinite sequence $U_n = \cdots 1^q0^p$ (q being possibly infinite).

The left-comb of right-combs with a cherry stem consists in simply replacing the context 01 of the preceding tree by the cherries 100 and 101. The tree is still stable and it has four context α-lis, as resumed in the array.

α-lis αs	contexts having αs as an α-lis
100	100
101	101
010	0^p10, $p \geqslant 1$
110	0^p1^q0, $p \geqslant 0$, $q \geqslant 2$



In this last example, the convergence of cascade series is equivalent to the finiteness of both sums

$$\kappa_{010} = \sum_{p\geqslant 1} \prod_{k=1}^{p-1} q_{0^k10}(0) \quad \text{and} \quad \kappa_{110} = \sum_{p\geqslant 1} \prod_{j=0}^{p-1} q_{0^j1^q0}(0) \prod_{k=2}^{q-1} q_{0^k1^q0}(1).$$

6 More about the non-stable case

Staying in the framework of non-nullness of the $q_c(\alpha)$ for all $c \in \mathscr{C}$ and $\alpha \in \mathscr{A}$ in order to avoid degenerate situations, we think that the non-stable case may be much more investigated, thanks to the following tracks.
Namely, for *totally non-stable* context trees, as defined below, we claim the following conjecture.

Definition 14. *A context tree is* totally non-stable*, when the set of infinite branches $\mathscr{C}^i$ has no shift-invariant subset.*

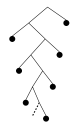
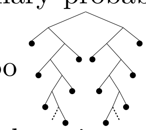
Among examples in Section 5, Example 5.1, is totally non-stable, though Examples 5.2 and 5.5 are non-stable but not totally non-stable (they have an infinite comb among their infinite branches).

Conjecture 1. *Let $(\mathscr{T}, q)$ be a non-null probabilised context tree. If $\mathscr{T}$ is totally non-stable, then there exists a unique probability stationary measure for the VLMC associated to $(\mathscr{T}, q)$.*

Using Proposition 4 as in Corollary 1, a first step in this direction consists in proving the following weaker conjecture.

Conjecture 2. *Let $(\mathscr{T}, q)$ be a non-null probabilised context tree. If $\mathscr{T}$ is totally non-stable and stabilizable, then its stabilized satifies the conditions of Theorem 2.*

In particular, we believe that the conditions of convergence of cascade series that are required by the existence of a stationary probability measure (see (i) in Theorem 1) are due to the existence of shift-stable subsets of $\mathscr{C}^i$. In otherwords, for a totally non-stable stabilizable context tree, the convergence of cascade series of the stabilized automatically holds because the general terms of these series decay exponentially fast. The very simple bamboo blossom example below comforts this impression.

The bamboo blossom is the context tree  . One can refer to Cénac et al. [2012] to retrieve a necessary and sufficient condition for existence and uniqueness of a stationary probability measure. The bamboo blossom is stabilizable, its stabilized being the double bamboo  . The double bamboo bossom has two α-lis: 00 and 11. The corresponding sums of cascade series are respectively

$$\kappa_{00} = \sum_{n\geqslant 0}\prod_{k=1}^{n-1} q_{(10)^k0}(0)\prod_{k=0}^{n-1} q_{(01)^k00(1)} + \sum_{n\geqslant 1}\prod_{k=1}^{n-2} q_{(10)^k0}(0)\prod_{k=0}^{n-1} q_{(01)^k00(1)}$$
$$\kappa_{11} = \sum_{n\geqslant 0}\prod_{k=0}^{n-1} q_{(10)^k11}(0)\prod_{k=1}^{n-1} q_{(01)^k1(1)} + \sum_{n\geqslant 1}\prod_{k=0}^{n-1} q_{(10)^k11}(0)\prod_{k=1}^{n-2} q_{(01)^k1(1)}.$$

If one probabilises the double bamboo in the sense of Proposition 4, then $q_{(10)^k0} = q_1$ for all $k \geqslant 1$ and $q_{(10)^k11} = q_1$ for all $k \geqslant 0$. The sums of cascade series become thus

$$\kappa_{00} = \sum_{n\geqslant 0} q_1(0)^{n-1}\prod_{k=0}^{n-1} q_{(01)^k00(1)} + \sum_{n\geqslant 1} q_1(0)^{n-2}\prod_{k=0}^{n-1} q_{(01)^k00(1)}$$
$$\kappa_{11} = \sum_{n\geqslant 0} q_1(0)^{n}\prod_{k=1}^{n-1} q_{(01)^k1(1)} + \sum_{n\geqslant 1} q_1(0)^{n}\prod_{k=1}^{n-1} q_{(01)^k1(1)},$$

converging as soon as $q_1(0) < 1$. This is the only condition to ensure existence and uniqueness of the stationary probability measure, as it was stated in Cénac et al. [2012]. Note that the set of infinite branches of the bamboo blossom has no shift-invariant subset.

References

V.S. Barbu and N. Limnios. *Semi-Markov chains and hidden semi-Markov models toward applications*, volume 191 of *Lecture Notes in Statistics*. Springer, New York, 2008. ISBN 978-0-387-73171-1. Their use in reliability and DNA analysis. 23

P. Cénac, B. Chauvin, F. Paccaut, and N. Pouyanne. Variable length Markov chains and dynamical sources. *Séminaire de Probabilités XLIV, Lecture Notes in Math.*, 2046:1–39, 2012. 3, 6, 31

P. Cénac, B. Chauvin, S. Herrmann, and P. Vallois. Persistent random walks, variable length markov chains and piecewise deterministic markov processes. *Markov Processes and Related Fields*, 19(1): 1–50, 2013. 2

P. Cénac, B. De Loynes, Y. Offret, and A. Rousselle. Recurrence of Multidimensional Persistent Random Walks. Fourier and Series Criteria. working paper or preprint, December 2017. URL https://hal.archives-ouvertes.fr/hal-01658494. 2

P. Cénac, A. Le Ny, B. De Loynes, and Y. Offret. Persistent random walks. I. Recurrence versus transience. *J. Theoret. Probab.*, 31(1):232–243, 2018. 2

I. Csiszár and Z. Talata. Context tree estimation for not necessarily finite memory processes, via bic and mdl. *IEEE Transactions on Information Theory*, 52(3):1007–1016, 2006. 2

E. De Santis and M. Piccioni. Backward coalescence times for perfect simulation of chains with infinite memory. *J. Appl. Probab.*, 49(2):319–337, 2012. 3

W. Doeblin and R. Fortet. Sur des chaînes à liaisons complètes. *Bull. Soc. Math. France*, 65:132–148, 1937. 2

S. Fedotov, A. Tan, and A. Zubarev. Persistent random walk of cells involving anomalous effects and random death. *Phys. Rev. E*, 91:042124, 2015. 2

R. Fernández and G. Maillard. Chains with complete connections: general theory, uniqueness, loss of memory and mixing properties. *J. Stat. Phys.*, 118(3-4):555–588, 2005. doi: 10.1007/s10955-004-8821-5. 2

S. Gallo. Chains with unbounded variable length memory: perfect simulation and a visible regeneration scheme. *Adv. in Appl. Probab.*, 43(3):735–759, 2011. doi: 10.1239/aap/1316792668. 2

S. Gallo and N. L. Garcia. Perfect simulation for locally continuous chains of infinite order. *Stochastic Process. Appl.*, 123(11):3877–3902, 2013. doi: 10.1016/j.spa.2013.05.010. 3

S. Gallo and F. Paccaut. On non-regular g-measures. *Nonlinearity*, 26:763–776, 2013. doi: 10.1088/0951-7715/26/3/763. 3, 15

T. E. Harris. On chains of infinite order. *Pacific J. Math.*, 5:707–724, 1955. 2

A. Johansson and A. Öberg. Square summability of variations of g-functions and uniqueness of g-measures. *Math. Res. Lett.*, 10(5-6):587–601, 2003. 2

S. Meyn and R.L. Tweedie. *Markov chains and stochastic stability*. Cambridge University Press, Cambridge, second edition, 2009. 14

J. Rissanen. A universal data compression system. *IEEE Trans. Inform. Theory*, 29(5):656–664, 1983. 2

E Seneta. *Non-negative matrices and Markov chains; rev. version*. Springer series in statistics. Springer, New York, NY, 2006. 21

Z. Talata. Divergence rates of markov order estimators and their application to statistical estimation of stationary ergodic processes. *Bernoulli*, 19 (3):846885, 2013. 2

F.M.J. Willems. The context-tree weighting method: Extensions. *IEEE Trans. Inform. Theory*, 44 (2):792–798, 1998. 2

F.M.J. Willems, Y.M. Shtarkov, and T.J. Tjalkens. Context tree weighting: Basic properties. *IEEE Trans. Inform. Theory*, 41:653–664, 1995. 2