VLMC
(Variable Length Markov Chains)
Riemannian Comb and Suffix Tries

Brigitte CHAUVIN

(Versailles Univ.)

with Peggy CÉNAC (Dijon),
Frédéric PACCAUT (Amiens), Nicolas POUYANNE (Versailles)

AofA, June 2011

Outline

- Variable Length Markov Chains defined by a probabilized context tree;
- A VLMC defines a probabilistic dynamical source
- NSC existence and unicity of a stationary measure for the infinite comb;
- Mixing properties for the infinite comb
- Suffix tries
- Open questions

Biblio

Rissanen (IEEE, 1983), “A universal data compression system”

Bühlmann et Wyner (Annals of Stat., 1999), “Variable length Markov chains”

Galves, Löcherbach (TICSP Series, 2008), “Stochastic chains with memory of variable length.”

Fernandez, Ferrari, Galves , 2001, “Coupling, renewal and perfect simulation of chains of infinite order”

Comets, Fernandez, Ferrari (AAP, 2002) : “Processes with long memory”

Gallo-Garcia (arxiv, 2011) “A general context-tree-based approach to perfect simulation for chains of infinite order.”

Clement-Flajolet-Vallee (Algorithmica, 2001) “Dynamical sources in Information Theory: Analysis of general tries”

Alphabet $\mathcal{A} = \{0, 1\}$ $\qquad \mathcal{L} = \mathcal{A}^{-\mathbb{N}} =$ left-infinite words

$$\begin{array}{rcl} U_0 & = & \ldots \; X_{-1}X_0 \\ U_1 & = & \ldots \; X_{-1}X_0X_1 \\ \ldots & & \\ U_n & = & \ldots \; X_{-1}X_0X_1 \ldots X_n \\ U_{n+1} & = & \ldots \; X_{-1}X_0X_1 \ldots X_nX_{n+1} = U_nX_{n+1} \end{array}$$

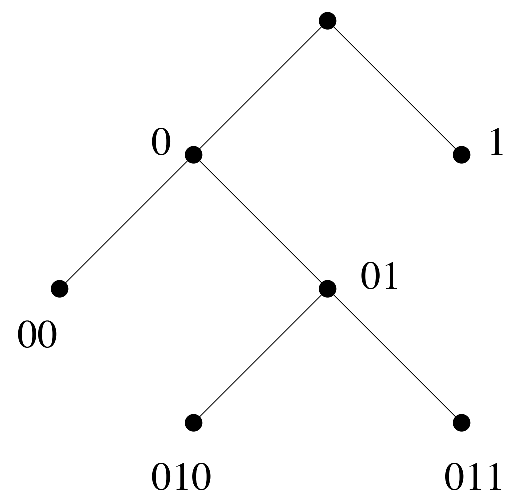
$(U_n)_{n \geq 0}$ is an $\mathcal{L}$-valued Markov chain

Start from U_0 chosen according to an initial probability measure.
From U_n to U_{n+1} add a letter X_{n+1} on the right of U_n.

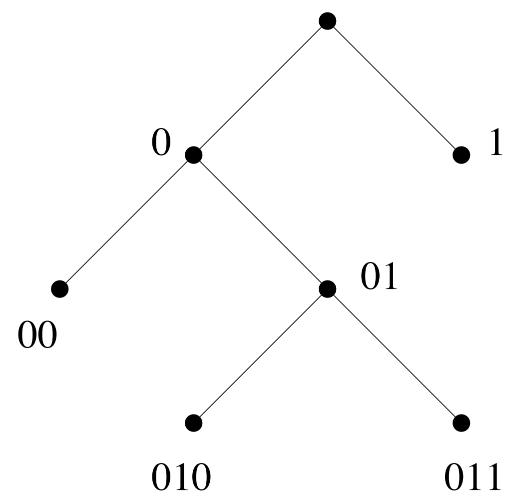
The transition from U_n to U_{n+1} is

$$\mathbb{P}(U_{n+1} = U_n\alpha | U_n)$$

where $\alpha \in \mathcal{A}$ is a letter,
it depends on a suffix of U_n. Which suffix?

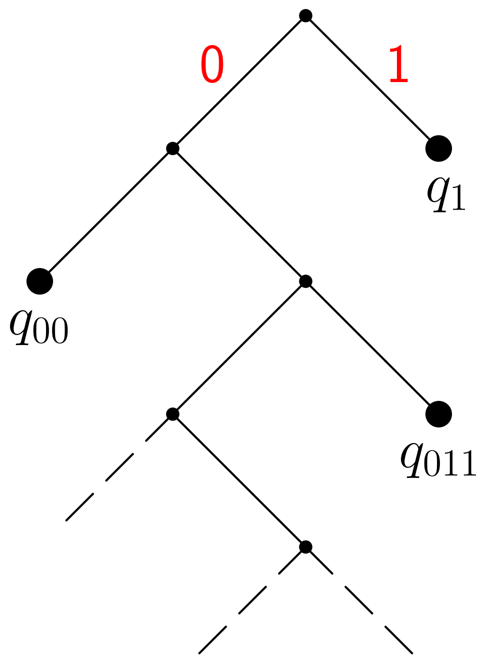


$$U_n = \ldots 1000110$$



$$U_n = \ldots 1000110$$

context : $011 =: \overleftarrow{\text{pref}}(U_n)$



$U_n = \ldots 1000110$

context: $011 =: \overleftarrow{\text{pref}}\,(U_n)$

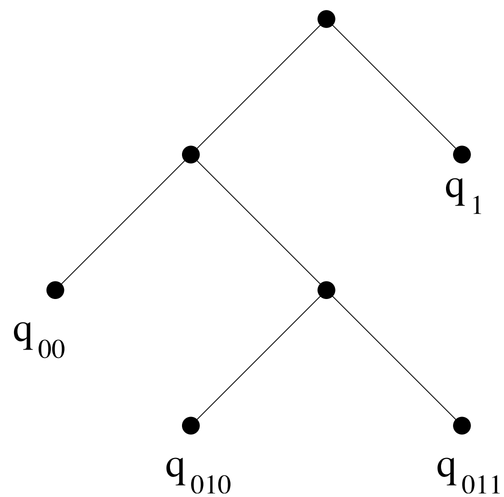
q_{011} = Bernoulli distribution on $\mathcal{A}$

$$\mathbb{P}(U_{n+1} = U_n\alpha | U_n) = q_{011}(\alpha)$$

Definition of the Markov chain:

$$\mathbb{P}(U_{n+1} = U_n\alpha | U_n) = q_{\overleftarrow{\text{pref}}\,(U_n)}(\alpha).$$

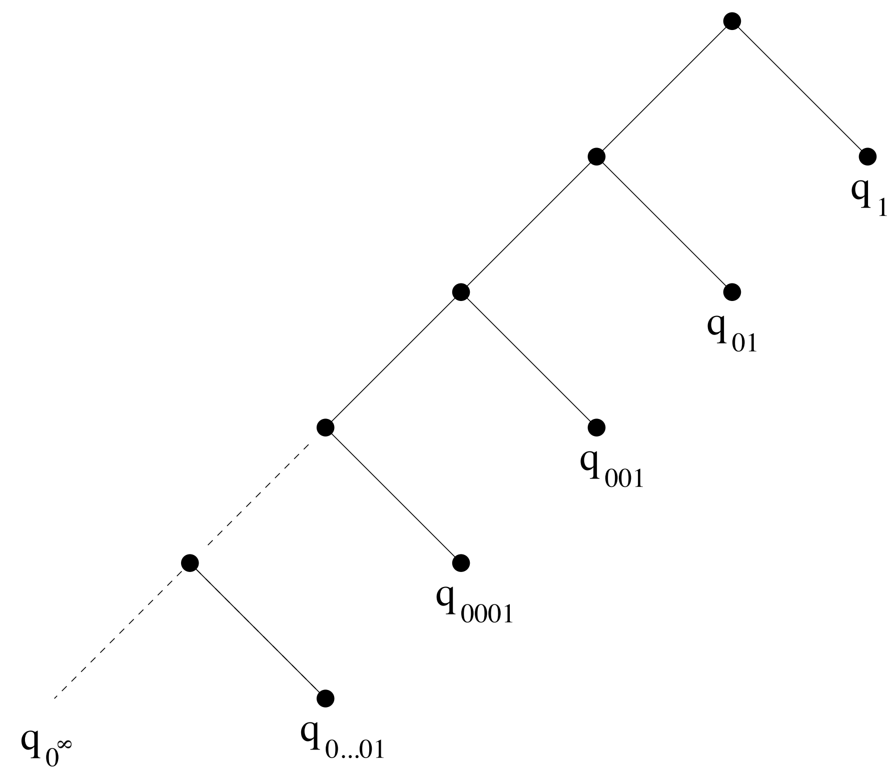
Data: probabilized context tree, with at most countably many infinite branches



finite tree $\rightarrow \quad (X_n)$ Markov chain

infinite tree $\rightarrow \quad (X_n)$ non Markov chain

Question: stationary law for (U_n)?



(X_n) is not Markov.

Stationary measure for the infinite comb

Let (U_n) be a VLMC associated with an infinite comb, probabilized by Bernoulli distributions $q_{0^k 1}, \infty \geq k \geq 0$ on the leaves. Let

$$c_n := \prod_{k=0}^{n-1} q_{0^k 1}(0)$$

Theorem 0.1 *(irreducible case)* $q_{0^\infty}(0) \neq 1$
The Markov process (U_n) admits a stationary measure on $\mathcal{L}$ if and only if the series $\sum c_n$ converges.
In this case, the stationary measure is unique, it is characterized by explicit formulae depending on the $q_{0^k 1}$.

Mixing properties

$\rightarrow$ of a VLMC $(U_n)_{n\in\mathbb{N}}$ defined by an infinite comb where

$$U_n = \ldots X_0 X_1 \ldots X_{n-1} X_n$$

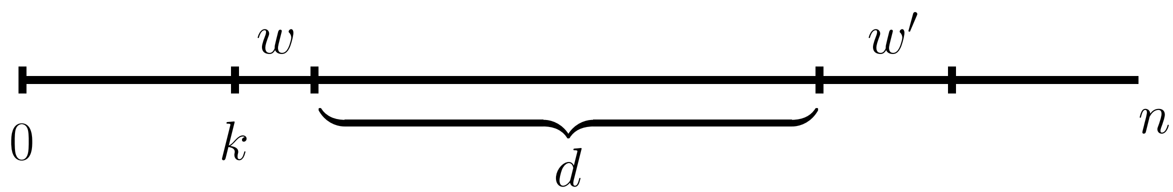
$\rightarrow$ under the stationary distribution π and let

$$\boxed{p_w = \pi(\mathcal{L}w) = \mathbb{P}(X_0 X_1 \ldots X_{|w|-1} = w)}$$

for any finite word w.

Mixing properties of a process

The question: in a random sequence $(X_n)_{n\in\mathbb{N}}$, measure the independence between the occurrences of two finite words w and w' when d letters separate them.



For any $k \geq 0$, compare

$$\mathbb{P}(X_k X_{k+1} \cdots \in w\mathcal{A}^d w' \dots)$$

and

$$\mathbb{P}(X_k \dots X_{k+|w|-1} = w) \times \mathbb{P}(X_{k+|w|+d} \dots X_{k+|w|+d+|w'|-1} = w').$$

In general, take a $\sup_{k\geq 0}$, which is not required when the sequence $(X_n)_n$ is stationary.

For any finite words w and w', consider the function

$$\Phi(d, w, w') = \frac{\left(\sum_{|v|=d} p_{wvw'}\right) - p_w p_{w'}}{p_w p_{w'}}.$$

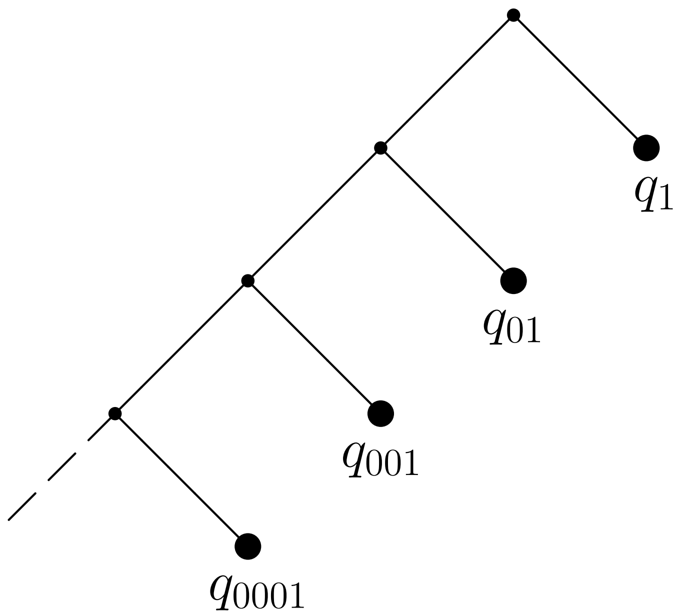
[$\Phi \equiv 0$ when the process $(X_n)_n$ is a sequence of i.i.d. variables]

The mixing property of the process is the decreasing rate to 0 of the function

$$\Phi(d) = \sup_{w,w' \in \mathcal{W}} \Phi(d, w, w').$$

Nice cases: Φ is exponentially decreasing to 0.

Stationary Riemannian infinite comb



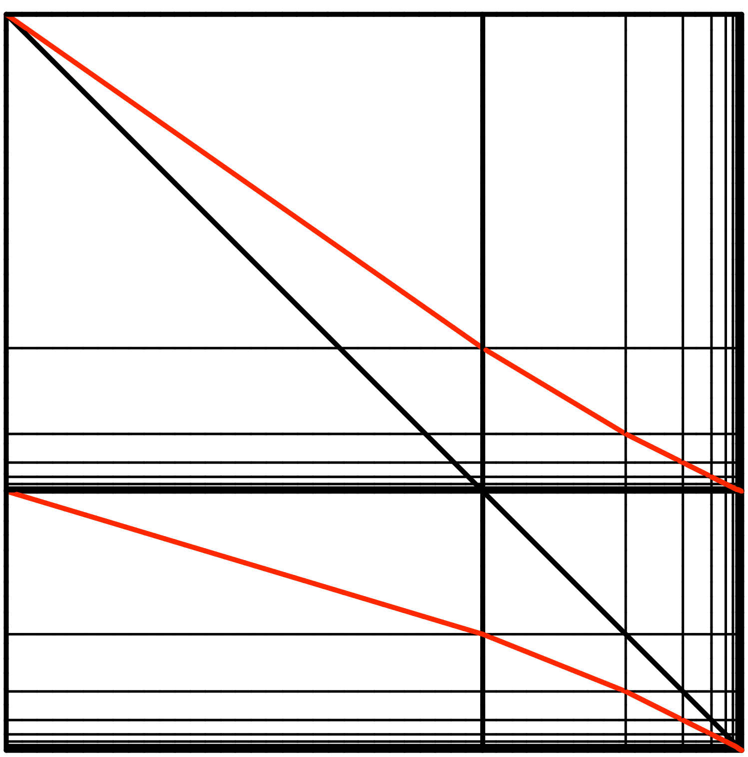
Context probabilities:

$$q_{0^n1}(0) = \left(\frac{n+1}{n+2}\right)^{\alpha}, \ \alpha > 1.$$

$$\boxed{c_n = (1+n)^{-\alpha}}$$

There exists a unique stationary measure π on left-infinite words [CCPP2011].

Consider the VLMC $(U_n)_{n\in\mathbb{N}}$ defined by
- this probabilized context tree;
- the initial distribution $U_0 \sim \pi$.



Mixing of the stationary Riemannian infinite comb

If w and w' are finite words, one gets the explicit formula

$$\Phi(d, w1, 1w') = \zeta(\alpha) u_{d+1} - 1$$

where

$$\sum_{n\geq 0} u_n z^n = \frac{z}{(1-z)\,\mathrm{Li}_\alpha(z)},$$

$\mathrm{Li}_\alpha(z) = \sum_{k\geq 1} \frac{z^k}{k^\alpha}$

Analytic combinatorics!

$$u_n = \frac{1}{\zeta(\alpha)} - \frac{1}{(1-\alpha)\zeta(\alpha)^2}\, n^{1-\alpha} + \mathcal{O}(n^{-\alpha})$$

$$\left|\Phi(d, w1, 1w')\right| \underset{d\to\infty}{\sim} c_\alpha \left(\frac{1}{d}\right)^{\alpha-1}.$$

Suffix Trie of the stationary Riemannian infinite comb

An infinite word $X_0X_1X_2\ldots$ is produced by a stationary Riemannian infinite comb.

Insert in a trie the sequence of the suffixes

$$\begin{aligned} X_0X_1X_2X_3X_4\ldots \\ X_1X_2X_3X_4\ldots \\ X_2X_3X_4\ldots \\ etc\ldots \end{aligned}$$

The height of the trie after the insertion of the n-th suffix is of order $\log n$ under an exponential mixing assumption (Szpankowski, 1983).

Height of a Suffix Trie for a weakly mixing infinite comb

$\rightarrow$ for a "logarithmic" comb (of order 4):

$$\forall \eta > \frac{1}{36}, \qquad a.s. \qquad \frac{L_n}{n^{\frac{1}{4}-\eta}} \longrightarrow +\infty, \qquad n \to +\infty$$

$\rightarrow$ simulations (due to M. Guesdon, INRIA Rocquencourt).

$\rightarrow$ a conjecture:

$$\frac{L_n}{n^{\frac{1}{\alpha}}} \longrightarrow c_\alpha \quad a.s. \text{ or } P$$

Open questions

• get a dictionary between the mixing type of a VLMC and the height of the associated suffix trie; other parameters of the trie?

• NSC of existence and unicity of a stationary measure in more general cases than the infinite comb, depending on the $shape$ of the context tree;

• get a relationship between the renewal properties of some VLMC and the spectral properties of the transfer operator associated with the dynamical system;

• approximate a VLMC by a sequence of Markov chains of increasing order;

• is it possible to see a VLMC as a process on the contexts (automaton)?