Sur les mesures stationnaires des VLMC

P. Cénac B. Chauvin F. Paccaut N. Pouyanne

ALEA, 13 mars 2018

Dijon Versailles Amiens Versailles

[On nous dit qu'il faudrait mettre ici des logos et tout et tout, mais zut]

“Variable Length Markov Chain”

Un alphabet : $\{0, 1\}$

"Variable Length Markov Chain"

Un alphabet : $\{0, 1\}$

Une VLMC est une source probabiliste $X_0, X_1, X_2 \cdots$ de lettres de $\{0, 1\}$ qui sont les lettres finales d'une suite $(U_n)_{n \geqslant 0}$ de mots infinis à gauche qui poussent :

$$\begin{array}{rcl} & U_0 & = \cdots X_0 \\ U_0 X_1 = & U_1 & = \cdots X_0 X_1 \\ U_1 X_2 = & U_2 & = \cdots X_0 X_1 X_2 \\ & \vdots & \\ U_n X_{n+1} = & U_{n+1} & = \cdots X_0 X_1 \cdots X_n X_{n+1} \\ & \vdots & \end{array}$$

“Variable Length Markov Chain”

- Les mots infinis à gauche U_n poussent “Markov” (VLMC) :

$$\mathbf{P}\left(U_{n+1} = U_n\alpha | U_n\right)$$

est une fonction de U_n.

"Variable Length Markov Chain"

• Les mots infinis à gauche U_n poussent "Markov" (VLMC) :

$$\mathbf{P}\left(U_{n+1} = U_n\alpha | U_n\right)$$

est une fonction de U_n.

• Enfin, cette probabilité conditionnelle est fonction d'un *suffixe fini* (mais non borné) de U_n (VLMC).

“Variable Length Markov Chain”

• Les mots infinis à gauche U_n poussent “Markov” (VLMC) :

$$\mathbf{P}\left(U_{n+1} = U_n\alpha | U_n\right)$$

est une fonction de U_n.

• Enfin, cette probabilité conditionnelle est fonction d'un *suffixe fini* (mais non borné) de U_n (VLMC).

Exemples : $\mathbf{P}\left(U_{n+1} = U_n\alpha | U_n\right)$ peut être fonction

(i) de la longueur du dernier run de 0 de U_n

(ii) du dernier run (de 0 ou de 1) de U_n

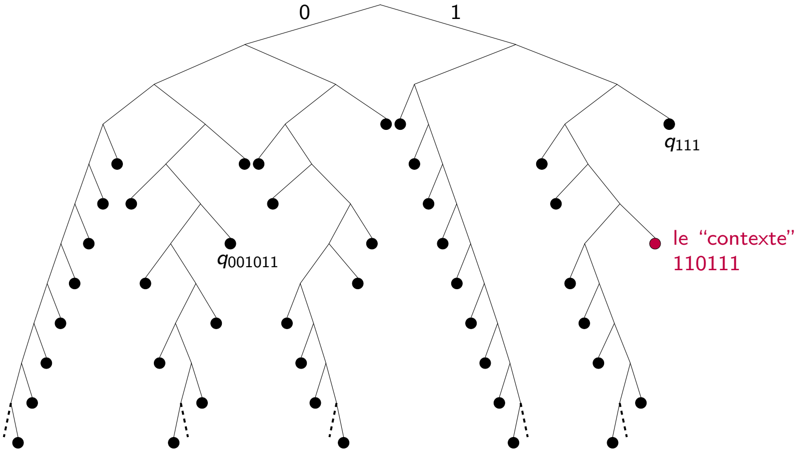
(iii) du dernier motif $1^{\geqslant 0}0^{\geqslant 0}$ de U_n

Les données pour définir une VLMC

(i) Un arbre binaire saturé (*arbre des contextes*)

(ii) Sur chaque feuille c, une mesure de probabilité q_c sur $\{0, 1\}$

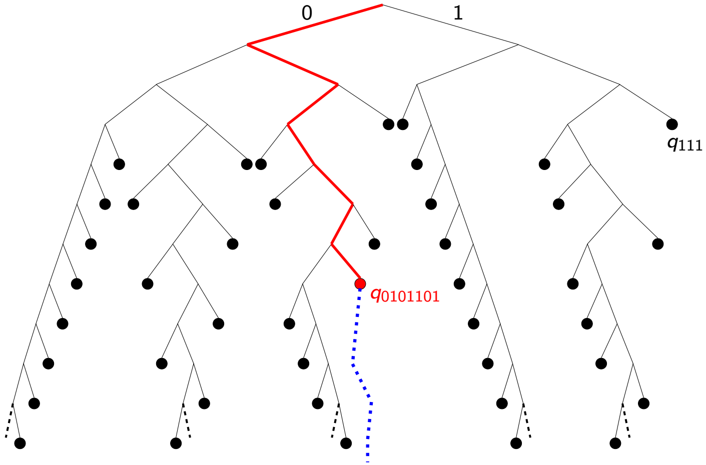
[(iii) Pour U_0, une loi de probabilité sur les mots infinis à gauche]



Définition de la VLMC associée

Pour tout n, suspends le mot infini à gauche U_n par sa dernière lettre, insère-le dans l'arbre et regarde par quelle feuille il sort. Note-la $\overleftarrow{\mathrm{pref}}\,(U_n)$. Alors, pour toute $\alpha \in \{0,1\}$,

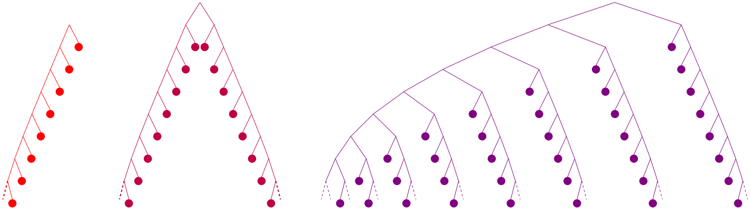
$$\mathbf{P}\left(U_{n+1} = U_n\alpha | U_n\right) = q_{\overleftarrow{\mathrm{pref}}\,(U_n)}\left(\alpha\right)$$



Insère $U_n = \cdots 1001011010$ $\quad$ $\overleftarrow{\text{pref}}\,(U_n) = 0101101$

Exemples

Dernier run de 0 : peigne gauche
Dernier run : double peigne
Dernier motif $1^{\geqslant 0}0^{\geqslant 0}$: peigne gauche de peignes droits



La question

Une telle chaîne de Markov admet-elle une mesure de probabilité stationnaire ?
Combien (quel convexe) ?

Des hypothèses

(i) l'arbre des contextes a un ensemble dénombrable de branches infinies (l'arbre complet est interdit, par exemple)

(ii) $q > 0$

[Remarque : si l'arbre est fini de hauteur h, RAS : $(U_n)_n$ est une chaîne de Markov d'ordre h (PF).]

A quoi ressemble une mesure stationnaire (1)

Suppose que π est une probabilité stationnaire sur $\mathscr{L}$ ♫.

① La tribu est engendrée par les cylindres : π est déterminée par les $\pi(\mathscr{L}w)$, w mot fini

♫ $\mathscr{L}$, ce sont les mots infinis à gauche, $\{0,1\}^{\mathbb{Z}_{\leq 0}}$, quoi.

A quoi ressemble une mesure stationnaire (1)

Suppose que π est une probabilité stationnaire sur $\mathscr{L}$[♫].

① La tribu est engendrée par les cylindres : π est déterminée par les $\pi(\mathscr{L}w)$, w mot fini

② Si $w = \alpha e$ où $\alpha \in \{0,1\}$ et e est un mot fini non interne, puisque π est stationnaire,

$$\pi(\mathscr{L}\overline{w}) = \pi(\mathscr{L}\overline{e}\alpha) = q_{\overrightarrow{\text{pref}}(e)}(\alpha) \times \pi(\mathscr{L}\overline{e})$$

[♫] $\mathscr{L}$, ce sont les mots infinis à gauche, $\{0,1\}^{\mathbb{Z}_{\leq 0}}$, quoi.

Le GIS d'un mot fini

Pour itérer le plus possible : décompose le mot fini non vide w en

$$w = p\alpha s = \beta_1\beta_2\beta_3\cdots\beta_\ell\alpha s$$

où
(i) p est un préfixe de longueur $\geqslant 0$
(ii) α est une lettre
(iii) s est *le plus grand suffixe interne* de w (on dira le *gis*, il peut être vide).

[On appellera aussi αs l'*alpha-gis* du mot w.]

La formule des cascades

Si $w = p\alpha s = \beta_1\beta_2\beta_3\cdots\beta_\ell\alpha s$ où s est le gis de w, appelle

$$\mathrm{casc}(w) = \begin{array}{l} q_{\overrightarrow{\text{pref}}\,(\beta_2\cdots\beta_\ell\alpha s)}(\beta_1) \\ \times q_{\overrightarrow{\text{pref}}\,(\beta_3\cdots\beta_\ell\alpha s)}(\beta_2) \\ \vdots \\ \times q_{\overrightarrow{\text{pref}}\,(\alpha s)}(\beta_\ell) \end{array}$$

la *cascade* de w. Alors,

$$\pi\left(\mathscr{L}\overline{w}\right) = \mathrm{casc}(w) \times \pi\left(\mathscr{L}\overline{\alpha s}\right)$$

A quoi ressemble une mesure stationnaire (2)

③ Grâce à la formule des cascades, π est déterminée par les

$$\pi\left(\mathscr{L}\overline{\alpha s}\right),\ \alpha \text{ lettre, } s \text{ mot interne}$$

A quoi ressemble une mesure stationnaire (2)

③ Grâce à la formule des cascades, π est déterminée par les

$$\pi\left(\mathscr{L}\overline{\alpha s}\right),\ \alpha \text{ lettre, } s \text{ mot interne}$$

④ Union disjointe : si s est interne,

$$\pi\left(\mathscr{L}\overline{\alpha s}\right) = \sum_{\substack{c \text{ contexte} \\ c = s\cdots}} \pi\left(\mathscr{L}\overline{\alpha c}\right)$$

ou encore

$$\pi\left(\mathscr{L}\overline{\alpha s}\right) = \sum_{\substack{c \text{ contexte} \\ c = s\cdots}} q_c(\alpha)\pi\left(\mathscr{L}\overline{c}\right)$$

Donc π est déterminée par les $\pi\left(\mathscr{L}\overline{c}\right)$, c contexte

A quoi ressemble une mesure stationnaire (3)

⑤ On continue : si c est contexte, on décompose $c = p\alpha s$ où s est le gis de c. Alors,

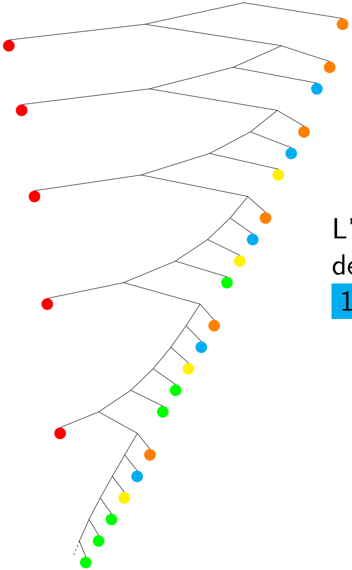
$$\pi\left(\mathscr{L}\overline{c}\right) = \mathrm{casc}(c)\pi\left(\mathscr{L}\overline{\alpha s}\right)$$

Résultat :

$$\boxed{\text{toute mesure stationnaire } \pi \text{ est déterminée par les } \pi\left(\mathscr{L}\overline{\alpha s}\right), \ \alpha s \text{ alpha-gis de contextes.}}$$

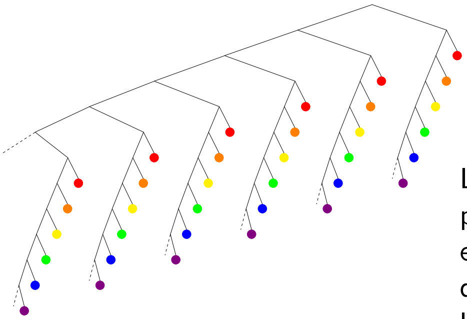
NB : les alpha-gis de contextes sont des données de l'arbre qui définit la VLMC.

Exemples d'arbres et de leurs alpha-gis de contextes (1)



L'arbre 1-arithmétique a 5 alpha-gis de contextes : 00, 001, 001001, 101 et 1

Exemples d'arbres et de leurs alpha-gis de contextes (2)



Le peigne gauche de peignes gauches a un ensemble infini d'alpha-gis de contextes :
les 10^n1, $n \geqslant 0$.

Mesure stationnaire : équation aux gis

⑥ On re-continue : suppose que π est une mesure stationnaire. Par union disjointe et formule des cascades, pour tout alpha-gis de contexte αs :

$$\pi\left(\mathscr{L}\overline{\alpha s}\right) = \sum_{\beta t} \pi\left(\mathscr{L}\overline{\beta t}\right) \underbrace{\left(\sum_{c=s\cdots=\cdots[\beta t]} \mathrm{casc}\left(\alpha c\right)\right)}_{Q_{\beta t,\alpha s}}$$

$\rightsquigarrow$ $\boxed{\text{Matrice } Q,}$ carrée, indexée par les alpha-gis de contextes (dimension finie ou non), définie par les données de la VLMC (l'arbre et les q_c).

Les séries de cascades

Si αs est un alpha-gis de contexte, on dit que *la série des cascades de αs converge* lorsque

$$\sum_{c = \cdots [\alpha s]} \operatorname{casc}(c) < \infty.$$

Lorsque toutes les séries de cascades convergent, on dit que *les séries de cascades convergent*.

Un énoncé, enfin

Théorème (CCPP, 2018)
(i) Si la VLCM a une probabilité stationnaire π, alors
- les séries de cascades convergent,
- $\left(\pi\left(\mathscr{L}\overline{\alpha s}\right)\right)_{\alpha s}$ est un vecteur fixe à gauche de Q, sommable.

(ii) Si les séries de cascades convergent, l'application

$$\pi \mapsto \left(\pi\left(\mathscr{L}\overline{\alpha s}\right)\right)_{\alpha s}$$

établit une bijection entre :
- les mesures sigma-finies sur $\mathscr{L}$, stationnaires, finies sur les alpha-gis de contextes ;
- les vecteurs fixes à gauche de Q.

Remarque : non, non, la bijection n'échange pas les mesures stationnaires finies et les vecteurs fixes sommables. Dommage, hein ?

Le cas sympathique : celui où Q est de dimension finie.

Le cas des arbres stables

L'arbre des contextes est dit *stable* lorsqu'il est invariant par étêtage.

Théorème (CCPP, 2018)
(i) Si l'arbre de la VLMC est stable, la matrice Q est stochastique (par lignes) et irréductible

(ii) Suppose que l'arbre de la VLMC est stable et a un ensemble fini d'alpha-gis. Alors,

$$\exists! \textit{ probabilité stationnaire} \iff \textit{les séries de cascades convergent}$$

Merci !