Chaînes de Markov

résumé : • on aborde ici un outil mathématique utilisé pour représenter les systèmes discrets dont l'état change au cours du temps de manière aléatoire

• on parle ici de tps discret, càd qu'on s'intéresse à la succession des états, pas le moment où ils surviennent

• on cherche à déterminer (1) la probabilité des différents états en fonction du temps

(2) le comportement du système après un temps long.

1. Quelques exemples

(1) "La ruine du joueur"

Un joueur arrive au casino avec i€ en poche et se lance dans des paris successifs et indépendants qui lui rapportent 1€ avec probabilité p et -1€ avec probabilité $q = 1 - p$. Le jeu s'arrête lorsqu'il arrive à amasser N€ (on suppose $1 \leq i < N$) ou lorsqu'il a tout perdu.

Avec quelle probabilité le joueur sera-t-il ruiné ?

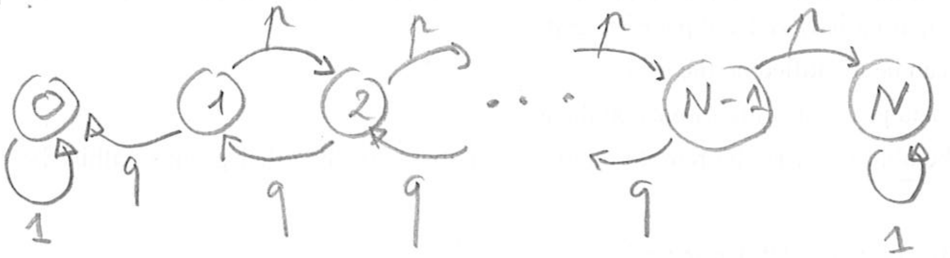
Modélisons le problème

$X_n \in \{0, \dots, N\}$: somme détenue par le joueur à l'étape n

$$\mathbb{P}[X_{n+1} \mid X_n = k] = \begin{cases} 1 & \text{si } X_{n+1} = k = 0 \\ 1 & \text{si } X_{n+1} = k = N \\ p & \text{si } k \neq 0 \text{ et } X_{n+1} = k+1 \\ q & \text{si } k \neq N \text{ et } X_{n+1} = k-1 \\ 0 & \text{sinon} \end{cases}$$

On peut schématiser l'évolution du jeu à l'aide d'un graphe, où les sommets correspondent à toutes les situations possibles (c'est-à-dire les sommes que le joueur peut détenir), et les arcs aux enchaînements possibles entre les situations.

Soit :

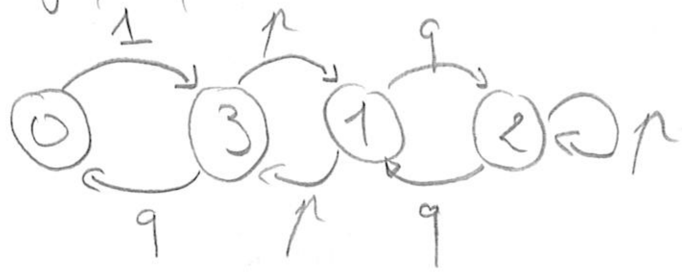


(2) Gestion de parapluies

Je dispose de K parapluies qui peuvent se trouver chez moi ou à mon travail. Je me déplace de mon domicile à mon travail (et vice versa) en emportant un parapluie seulement s'il pleut. En admettant qu'il pleut sur un trajet avec une probabilité p et indépendamment des autres trajets, avec quelle fréquence sur le long terme vais-je devoir marcher sous la pluie ?

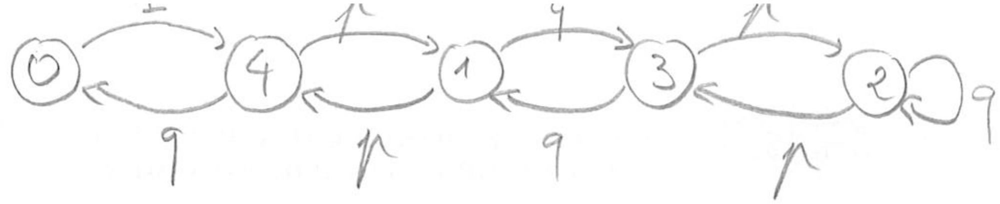
Représentation graphique :

$K = 3$



$q = 1 - p$

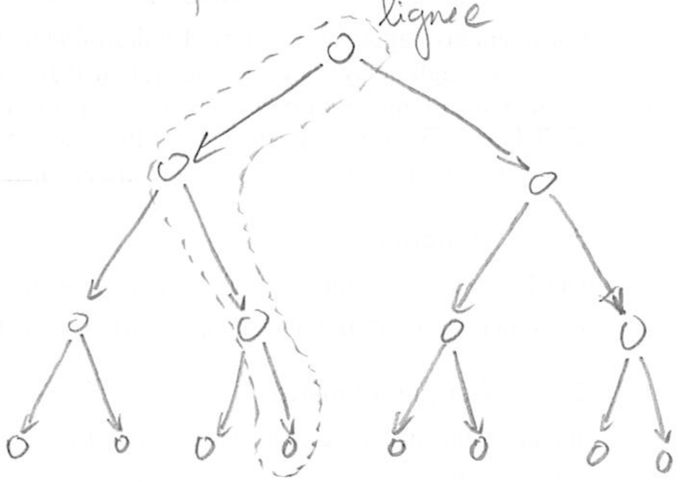
K = 4



$X_n \in \{0, \dots, K\}$: nb de parapluies où je me trouve après n voyages

$$\mathbb{P}[X_{n+1} | X_n] = \begin{cases} 1 & \text{si } X_n = 0 \text{ et } X_{n+1} = 4 \\ 0 & \text{si } X_n = 0 \text{ et } X_{n+1} = 0 \\ \dots & \text{etc} \end{cases}$$

(3) Évolution d'une séquence nucléotidique

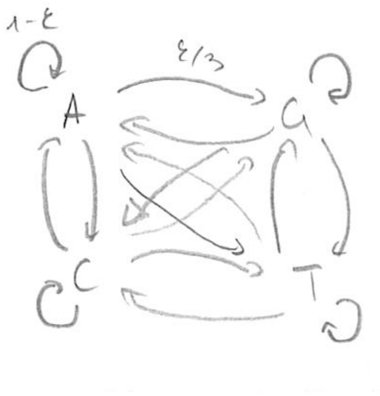


Soit une population de cellules en division (voir ci-contre) dont on considère une lignée. On s'intéresse à l'évolution d'un site donné dans le génome, sous l'effet de mutations supposées neutres. et chaque division, le site a une probabilité ε de muter uniformément vers un autre nucléotide.

Quelle est la distribution de probabilité à long terme de ce site ?

On pose $X_n \in \{A, C, G, T\}$ le nucléotide au site considéré à la génération n

$$\mathbb{P}[X_{n+1} | X_n] = \begin{cases} 1 - \varepsilon & \text{si } X_{n+1} = X_n \\ \varepsilon / 3 & \text{sinon} \end{cases}$$



Ce modèle d'évolution est connu sous le nom de modèle de Jukes et Cantor.

2. Définition

Soit $S = \{s_1, \dots, s_M\}$ un ensemble fini.

On appelle $\boxed{\text{chaîne de Markov à temps discret}}$ une famille de v.a. $(X_n)_{n \in \mathbb{N}}$ tq $X_n \in S$ et

$$\mathbb{P}[X_{n+1} \mid X_0, X_1, \dots, X_n] = \mathbb{P}[X_{n+1} \mid X_n] \qquad (1)$$

S est appelé espace d'état de la chaîne.

La propriété (1) dit que la distribution de X_{n+1} ne dépend que de X_n et pas de ce qui s'est passé avant. Elle est appelée propriété de Markov.

On notera
$$p_{i \to j}(n) = \mathbb{P}[X_{n+1} = s_j \mid X_n = s_i]$$
$$\text{et} \quad p_i(n) = \mathbb{P}[X_n = s_i]$$

Pour i et n fixés, $\{p_{i \to j}(n)\}_j$ est une distribution de probabilité sur S.

Donnons maintenant un exemple où la propriété de Markov n'est pas vérifiée : dans l'exemple 1 on ajoute la règle selon laquelle le joueur peut perdre une fois. On lui redonne alors sa mise initiale et il peut continuer à jouer. La famille $(X_n)_{n \geq 0}$ n'est alors plus une chaîne de Markov (CM). Si en revanche on considère le couple (X_n, Y_n) où Y_n représente le nombre de fois où le joueur a perdu,

celui-ci vérifie bien la propriété de Markov (à démontrer !).
Cette situation est très fréquente : en rajoutant de l'information à l'espace d'état, on peut se permettre "d'oublier" le passé.

Une CM $(X_n)_{n\in\mathbb{N}}$ est dite <u>homogène</u> lorsque $p_{i\to j}(n)$ ne dépend pas de n, c'est-à-dire quand les probabilités de transition entre états sont constantes au cours du temps.

On restera dorénavant dans ce cas particulier, qui permet la représentation des CM sous forme de graphe, comme abordé au paragraphe suivant.

3. Matrice stochastique et graphes

On appelle matrice stochastique toute matrice carrée dont les éléments sont dans $[0;1]$ et telle que chacune de ses lignes somme à 1.

C'est notamment le cas de la matrice $P(n) = \left(p_{i\to j}(n)\right)_{i,j\in S}$ qui est la matrice stochastique associée à une CM. Dans le cas homogène, on notera simplement P.

<u>Exemples</u> : la matrice stochastique dans l'exemple 2 pour $K=3$ vaut

$$P = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & q & p \\ 0 & q & p & 0 \\ q & p & 0 & 0 \end{pmatrix}$$

dans l'exemple 3, la matrice stochastique vaut

$$P = \begin{pmatrix} 1-\varepsilon & \varepsilon/3 & \varepsilon/3 & \varepsilon/3 \\ \varepsilon/3 & 1-\varepsilon & \varepsilon/3 & \varepsilon/3 \\ \varepsilon/3 & \frac{\varepsilon}{3} & 1-\varepsilon & \varepsilon/3 \\ \frac{\varepsilon}{3} & \frac{\varepsilon}{3} & \frac{\varepsilon}{3} & 1-\varepsilon \end{pmatrix}$$

On rappelle maintenant le lien bien connu entre matrices et graphes.

Soit un graphe $G = (V, E)$ où $V = \{v_1, .., v_n\}$ est l'ensemble des sommets, $E \subset V \times V$ l'ensemble des arcs. La matrice d'adjacence de G est la matrice carrée de dimension $|V|$ telle que

$$M_{ij} = \begin{cases} 1 & \text{si } (v_i, v_j) \in E \\ 0 & \text{sinon} \end{cases}$$

Dans la même idée, on définit le graphe associé à une CM homogène comme le graphe $G = (V, E)$ tel que $V = S$ et $E = \{(s_i, s_j) \mid P_{ij} \neq 0\}$. Ainsi on peut voir la matrice d'adjacence comme une version "binarisée" de la matrice stochastique. Sur le graphe associé à une CM, on représente les probabilités de transition sur les arcs.

4. Évolution temporelle

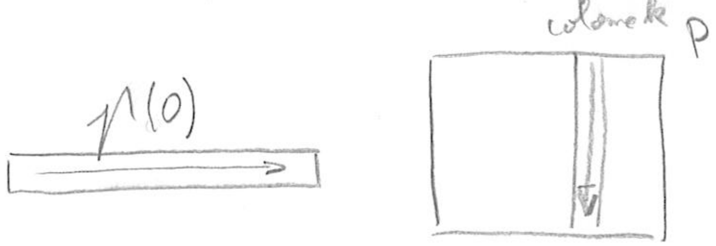
Soit une CM homogène de matrice stochastique P. Au temps 0, on suppose la distribution $p(0)$ sur S. Quelle est la distribution de probabilité après une transition ?

Autrement dit, quelle est la distribution de X_1 ?

$$\underset{\substack{\| \\ p_k(1)}}{P[X_1 = s_k]} = \sum_{s_i \in S} P[X_0 = s_i]\, P[X_1 = s_k \mid X_0 = s_i]$$

$$= \sum_{i=1}^{n} p_i(0)\, P_{ik}$$

On reconnait ici une multiplication vecteur/matrice



Ce qui nous donne : $\quad p(1) = p(0)P$

Qu'en est-il de $p(2)$? Un calcul similaire nous amène

$$p(2) = p(1)P = p(0)P.P = p(0)P^2.$$

On peut démontrer ainsi par récurrence :

> **Proposition** : soit $(X_n)_{n \geq 0}$ une CM homogène de matrice stochastique P. La distribution de probabilité sur l'espace d'état après k transitions vaut
>
> $$\boxed{p(k) = p(0)P^k}$$

Passons maintenant à une question similaire : quelle est la probabilité de passer d'un état s_i à un état s_j en n étapes ? On notera cette grandeur $P_{ij}^{(n)}$. Les $(P_{ij}^{(n)})_{ij}$ forment aussi une matrice stochastique.

Pour $n \geq 1$

$$
\begin{aligned}
p_{ij}^{(n)} &= \mathbb{P}[X_n = j \mid X_0 = i] \\
&= \sum_{s_k \in S} \mathbb{P}[X_n = j \mid X_{n-1} = s_k, X_0 = i] \cdot \mathbb{P}[X_{n-1} = s_k \mid X_0 = i] \\
&= \sum_{s_k \in S} \mathbb{P}[X_n = j \mid X_{n-1} = s_k] \cdot \mathbb{P}[X_{n-1} = s_k \mid X_0 = i]
\end{aligned}
$$

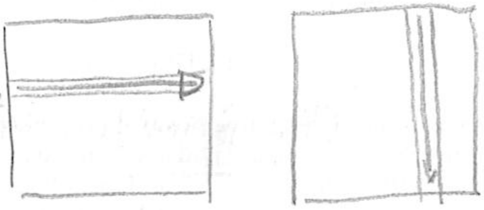
(propriété de Markov)

$$
= \sum_{k=1}^{M} p_{kj}\, p_{ik}^{(n-1)} = \sum_{k=1}^{M} p_{ik}^{(n-1)} p_{kj}
$$

On reconnaît ici le produit matriciel

j^e colonne

i^e ligne



$P^{(n-1)}$ $\quad$ P

c'est-à-dire que $\quad P^{(n)} = P\,P^{(n-1)}$

d'où par récurrence

$$
\boxed{P^{(n)} = P^n}
$$

D'une manière similaire on peut démontrer les équations de Chapman-Kolmogorov

$$
\boxed{P^{(k+l)} = P^{(k)} P^{(l)}}
$$

Exercice : soit la CM homogène de matrice stochastique

$$P = \begin{bmatrix} 0,8 & 0,2 & 0 \\ 0 & 0,5 & 0,5 \\ 0 & 0,4 & 0,6 \end{bmatrix}$$

1) Dessinez le graphe associé

2) Pour $\mu_0 = [1 \ 0 \ 0]$, calculez μ_1, μ_2, μ_3

5. Distribution stationnaire

On appelle distribution stationnaire une distribution π sur S telle que

$$\pi P = \pi$$

C'est-à-dire qu'une distribution stationnaire n'est pas modifiée par le passage du temps.

Exercice : déterminez une distribution stationnaire pour la CM de l'exercice précédent.

Remarque : si π est une distribution stationnaire de la chaîne alors c'est une valeur propre (à gauche) de P.

On retiendra ces deux résultats importants :

Théorème : une CM homogène à espace d'état fini a au moins une distribution stationnaire

Théorème : si le graphe associé à cette CM est fortement connexe, alors cette distribution est unique.

Une CM dont le graphe associé est fortement connexe est appelée irréductible

Sous certaines conditions, une CM converge vers une distribution stationnaire quand le temps passe. Sur ce sujet, le principal résultat est donné pour mémoire par le théorème suivant

Théorème :

- On appelle période d'un état i l'entier d_i défini par

$$d_i = GCD\{n \geqslant 1 \mid P_{ii}^{(n)} > 0\}$$

si $d_i = 1$, l'état est dit apériodique

- Une chaîne irréductible contenant un état apériodique converge vers son unique état stationnaire π

$$\lim_{n \to +\infty} P_{ij}^{(n)} = \pi_j$$

On remarque en particulier que la limite ne dépend pas du point de départ.

Application : modèle de Jukes et Cantor

$$\begin{bmatrix} 1-\varepsilon & \frac{\varepsilon}{3} & \frac{\varepsilon}{3} & \frac{\varepsilon}{3} \\ \frac{\varepsilon}{3} & 1-\varepsilon & \frac{\varepsilon}{3} & \frac{\varepsilon}{3} \\ \frac{\varepsilon}{3} & \frac{\varepsilon}{3} & 1-\varepsilon & \frac{\varepsilon}{3} \\ \frac{\varepsilon}{3} & \frac{\varepsilon}{3} & \frac{\varepsilon}{3} & 1-\varepsilon \end{bmatrix}$$

Le polynome caractéristique est

$$\det(M - \lambda I_4) = \frac{1}{27}(\lambda - 1)(4\varepsilon + 3\lambda - 3)^3$$

On en déduit les 4 valeurs propres

$$\lambda_1 = 1 \qquad \lambda_2 = \lambda_3 = \lambda_4 = \frac{1}{3}(3 - 4\alpha)$$

et les matrices de passage

$$S = \begin{pmatrix} 1 & -1 & -1 & -1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix} \qquad S^{-1} = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{3}{4} \\ -\frac{1}{4} & -\frac{1}{4} & \frac{3}{4} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{3}{4} & -\frac{1}{4} & -\frac{1}{4} \end{pmatrix}$$

d'où l'on peut écrire

$$P = S D S^{-1} \qquad \text{avec} \qquad D = \operatorname{diag}(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$$

On en déduit $P^n = S D^n S^{-1}$

$$\text{et} \quad \lim_{n \to +\infty} P^n = S \left(\lim_{n \to +\infty} D^n \right) S^{-1} = S \begin{pmatrix} 1&0&0&0 \\ 0&0&0&0 \\ 0&0&0&0 \\ 0&0&0&0 \end{pmatrix} S^{-1} = \frac{1}{4} \begin{pmatrix} 1&1&1&1 \\ 1&1&1&1 \\ 1&1&1&1 \\ 1&1&1&1 \end{pmatrix}$$

d'où pour toute distribution initiale $p_0 = \begin{pmatrix} a \\ b \\ c \\ 1-a-b-c \end{pmatrix}^T$

$$\lim_{n \to +\infty} p_0 P^n = \frac{1}{4} (1 \ 1 \ 1 \ 1)$$

On peut également tirer de ce modèle une conséquence utile pour dater la divergence des espèces. La diagonalisation permet un calcul explicite de P^n.

$$P^n = S D^n S^{-1} = \frac{1}{4} \begin{pmatrix} 3\left(1-\frac{4\varepsilon}{3}\right)^n + 1 & 1-\left(1-\frac{4\varepsilon}{3}\right)^n & 1-\left(1-\frac{4\varepsilon}{3}\right)^n & 1-\left(1-\frac{4\varepsilon}{3}\right)^n \\ 1-\left(1-\frac{4\varepsilon}{3}\right)^n & 3\left(1-\frac{4\varepsilon}{3}\right)^n + 1 & 1-\left(1-\frac{4\varepsilon}{3}\right)^n & 1-\left(1-\frac{4\varepsilon}{3}\right)^n \\ 1-\left(1-\frac{4\varepsilon}{3}\right)^n & 1-\left(1-\frac{4\varepsilon}{3}\right)^n & 3\left(1-\frac{4\varepsilon}{3}\right)^n + 1 & 1-\left(1-\frac{4\varepsilon}{3}\right)^n \\ 1-\left(1-\frac{4\varepsilon}{3}\right)^n & 1-\left(1-\frac{4\varepsilon}{3}\right)^n & 1-\left(1-\frac{4\varepsilon}{3}\right)^n & 3\left(1-\frac{4\varepsilon}{3}\right)^n + 1 \end{pmatrix}$$

La probabilité qu'un site mute au bout de n générations est donnée par la somme des termes non-diagonaux sur une ligne, c'est-à-dire :

$$p_{mut}(n) = \frac{3}{4}\left(1 - \left(1 - \frac{4\varepsilon}{3}\right)^n\right)$$

Cette relation peut s'inverser :

$$n = \frac{\ln\left(1 - \frac{4 p_{mut}}{3}\right)}{\ln\left(1 - \frac{4\alpha}{3}\right)}$$

En comptant le nombre de substitutions entre deux séquences homologues, on obtient une estimée $\hat{p}_{mut}$, qu'on peut alors utiliser pour estimer le temps de divergence entre les deux séquences.

6. Chaînes d'ordre supérieur

Une application importante des chaînes de Markov concerne la modélisation des séquences nucléotidiques ou protéiques. On remarque en effet que les variations observées dans des sites homologues ne sont pas indépendantes des sites adjacents. Un exemple célèbre que nous verrons en TP est la sous représentation des di-nucléotide CpG. Les occurrences CG sont hyper-mutables sur la position G. Par conséquent, la probabilité d'observer un G après un C n'est pas la même qu'après un A.

Plus généralement, on peut imaginer que certains 2, 3, 4, 5 mers sont sur- ou sous-représentés dans le génome. Autrement dit l'état X_n du site en position n dépend des états en $n-1$, $n-2$, $n-3$ etc.

S'agit-il encore d'une CM ? Cela semble être incompatible avec la propriété de Markov, mais seulement en apparence. Prenons un cas où X_n dépend de X_{n-1} et X_{n-2}. On peut définir $Y_n = (X_n, X_{n-1})$ dont l'espace d'état est S^2. On note alors que $Y_n = (X_n, X_{n-1})$ ne dépend que de $Y_{n-1} = (X_{n-1}, X_{n-2})$: on se ramène alors à une CM.

On appelle CM d'ordre K, $K \geq 1$, une famille de v.a. $(X_n)_{n \in \mathbb{N}}$ tq

$$\forall n \geq K, \quad \mathbb{P}[X_n \mid X_0, \dots, X_{n-1}] = \mathbb{P}[X_n \mid X_{n-1}, \dots, X_{n-K}]$$

On peut réutiliser tous les résultats précédents sur les CM en utilisant la construction vue plus haut.

7. Vraisemblance

Soit $(X_n)_{n \geq 0}$ une CM homogène de matrice stochastique P inconnue. On observe une réalisation finie $x_0, \dots, x_N$ de la chaîne. Comment inférer P ?

$\rightsquigarrow$ maximum de vraisemblance !

On commence par écrire le modèle

soit $\left| \begin{array}{ll} \pi & \text{un vecteur de probabilité de dimension } |S| \\ & \text{(distribution de proba sur l'état initial)} \\ P & \text{une matrice stochastique} \end{array} \right.$

$$M1 \begin{cases} X_0 \sim \mathcal{C}at(\pi, S) \\ X_{i+1} \sim \mathcal{C}at(P_{X_i \cdot}, S) \quad \forall i \geq 1 \end{cases}$$

On remarque ce nouveau paramètre π qui détermine la probabilité sur l'état initial.

Dans le cours précédent, nous avons vu un modèle plus simple de séquence :

$$M0 \begin{cases} \pi \text{ un vecteur de probabilité de dimension } |S| \\ X_i \sim \mathcal{C}at(\pi, S) \quad \forall i \geq 0 \end{cases}$$

nous en avions déduit des estimations au maximum de vraisemblance de la forme $\hat{\pi}_X = \frac{k_X}{n}$ où k_X est le nombre d'occurences de la lettre X et n la longueur totale de la séquence. L'écriture de la vraisemblance pour M1 amène à des résultats analogues

$$L(P, \pi \mid x_0, \ldots, x_N) = \mathbb{P}[X_0 = x_0, \ldots, X_N = x_N \mid P, \pi]$$

$$= \mathbb{P}[X_0 = x_0 \mid P, \pi] \, \mathbb{P}[X_1 = x_1, \ldots, X_N = x_N \mid P, \pi, X_0 = x_0]$$

$$= \pi_{x_0} \mathbb{P}[X_1 = x_1 \mid P, \pi, X_0 = x_0] \, \mathbb{P}[X_2 = x_2, \ldots, X_N = x_N \mid P, \pi, X_0 = x_0, X_1 = x_1]$$

$$= \pi_{x_0} P_{x_0 x_1} \mathbb{P}[X_2 = x_2 \mid P, \pi, X_0 = x_0, X_1 = x_1] \, \mathbb{P}[X_3 = x_3, \ldots, X_N = x_N \mid P, \pi, X_0 = x_0, X_1 = x_1, X_2 = x_2]$$

$$= \pi_{x_0} P_{x_0 x_1} \mathbb{P}[X_2 = x_2 \mid P, \pi, X_1 = x_1] \mathbb{P}[X_3 = x_3, \dots, X_N = x_N \mid P, \pi, X_0 = x_0, X_1 = x_1, X_2 = x_2]$$

$$= \pi_{x_0} P_{x_0 x_1} P_{x_1 x_2} \mathbb{P}[X_3 = x_3, \dots, X_N = x_N \mid P, \pi, X_0 = x_0, X_1 = x_1, X_2 = x_2]$$

etc

$$= \pi_{x_0} \prod_{i=1}^{N} P_{x_{i-1} x_i}$$

La log vraisemblance est alors.

$$\log L(P, \pi \mid x_0, \dots, x_N) = \log \pi_{x_0} + \sum_{i=1}^{N} \log P_{x_{i-1} x_i}$$

que l'on peut réécrire pour faire apparaître les comptages $k_{u,v}$ du nombre de transitions $u \to v$.

$$\log L(P, \pi \mid x_0, \dots, x_N) = \log \pi_{x_0} + \sum_{u,v \in S} k_{u,v} \log P_{uv}$$

En se rappelant que $\sum_{v \in S} P_{uv} = 1$ pour tout u,

on trouverait (comme pour M_0) que

$$\hat{P}_{u,v} = \frac{k_{u,v}}{\sum_{v' \in S} k_{u,v'}}$$

à essayer !

Note : ce cours s'inspire beaucoup de ceux donnés par Sylvain Mousset et Laurent Guéguen.