

Chaînes de Markov homogènes: Théorie et applications

Travail de Maturité réalisé au Lycée Denis-de-Rougemont
sous la direction de Mme Anne Gertsch

Ott Männik

1 Introduction

Depuis des millénaires, les hommes ont joué à des jeux de hasard; les plus anciens objets ressemblant à des dés ont précédé la Grèce antique, selon l'*Encyclopedia Britannica*. Ce n'est toutefois qu'au XVII^{ème} siècle que l'étude mathématique des probabilités a été entreprise par Pierre de Fermat et Blaise Pascal. Pour estimer au mieux le comportement des systèmes imprévisibles, on a commencé à étudier les *processus stochastiques*, c'est-à-dire des processus dont l'évolution est aléatoire.

Au début du XX^{ème} siècle, le mathématicien russe Andreï Andreïevitch Markov (1856-1922) a découvert et étudié une famille particulière de processus stochastiques qui feront l'objet de ce travail: **les chaînes de Markov**.

Les chaînes de Markov, processus dont l'évolution n'est pas influencée par le passé, mais seulement par le présent, trouvent aujourd'hui de nombreuses applications dans quasiment tous les domaines en lien avec la statistique. Parmi beaucoup d'autres, nous trouvons l'exemple de *PageRank*, un des algorithmes utilisés par Google pour déterminer la popularité d'un site web et qui sera traité plus tard.

Ce travail portera sur les chaînes de Markov homogènes et finies, nous développerons dans un premier temps les notions de base nécessaires pour les définir, les décrire et les employer à la résolution de problèmes mathématiques. Ensuite, nous traiterons quelques propriétés et chaînes spécifiques intéressantes, comme les chaînes absorbantes ou régulières, tout en fournissant des exemples d'application à des problèmes théoriques ou tirés de la vie réelle.

2 Définition et représentation schématique

2.1 Processus stochastiques

Avant d'entamer l'étude des processus stochastiques, il nous faut d'abord les définir. Nous nous contenterons d'ailleurs ici de n'étudier que des processus stochastiques **discrets**, c'est-à-dire des processus où les unités de temps appartiennent à un ensemble dénombrable, le plus souvent \mathbb{N} .

Nous définissons donc un **processus stochastique discret** comme une suite indexée de variables aléatoires X_t ayant une certaine probabilité de se retrouver dans un état i à un temps t donné. Nous définissons aussi l'**espace des états** d'un processus stochastique comme un ensemble dénombrable comportant tous les états possibles dans lesquels le processus stochastique peut se trouver. En général, nous donnerons simplement un numéro à chaque état. Pour un processus stochastique X_1, X_2, X_3, \dots avec un espace des états $S = \{1, 2, 3, \dots\}$, nous notons $P(X_t = i)$ la probabilité de se retrouver dans un état $i \in S$ au temps t et $P(X_1 = j \mid X_0 = i)$ la probabilité de passer de l'état $i \in S$ à l'état $j \in S$ entre les temps 0 et 1, aussi appelée la **probabilité de transition**.

Un exemple de processus stochastique serait le jeu de l'oie, où l'on jette un ou deux dés pour avancer d'un nombre aléatoire de pas. Certains états ou cases ont des propriétés particulières, ils peuvent par exemple renvoyer le joueur au départ ou alors lui faire perdre un tour.

2.2 Propriété de Markov, chaînes de Markov

Nous avons mentionné dans l'introduction qu'il existe des processus stochastiques dont l'évolution ne dépend que de l'état où ils se trouvent à présent et est donc indépendante du passé. Cette propriété est appelé "absence de mémoire" ou **propriété de Markov**.

Plus formellement, un processus stochastique $X_1, X_2, \dots, X_t, \dots$ avec des états correspondants $i_1, i_2, \dots, i_t, \dots \in S = \{1, 2, 3, \dots\}$ satisfait la propriété de Markov si

$$P(X_{t+1} = i_{t+1} \mid X_t = i_t, X_{t-1} = i_{t-1}, \dots, X_1 = i_1) = P(X_{t+1} = i_{t+1} \mid X_t = i_t).$$

Nous définissons alors **une chaîne de Markov** comme un processus stochastique discret qui satisfait la propriété de Markov. Nous disons encore qu'une chaîne de Markov est **finie** si son espace des états est une ensemble fini, et **homogène** si les probabilités de transition restent constantes en fonction du temps t . Les processus stochastiques continus satisfaisant la propriété de Markov sont appelés processus de Markov, mais ne seront pas étudiés dans ce texte.

Les processus stochastiques que nous pouvons supposer markoviens sont nombreux; parmi les plus simples se trouvent les lancements de pièces ou certaines marches aléatoires, mais les chaînes de Markov ont aussi des applications bien plus sophistiquées, comme le complètement de texte dans les messageries des natels.

Revenons à notre exemple du jeu de l'oie. Ce processus stochastique *n'est pas* une chaîne de Markov à cause des cases qui font perdre un tour au joueur X_t . Soit une telle case i et supposons que le joueur se trouve à un certain temps t en i . La probabilité de pouvoir avancer, peu importe le nombre de pas, dépend alors des valeurs précédentes de X_t . En effet, si nous savons que le joueur a déjà passé un tour sur i , il est sûr de pouvoir avancer. En revanche, si nous savons qu'il vient d'y arriver, il n'avancera certainement pas:

$$P(X_{t+1} > i \mid X_t = i, X_{t-1} = i) = 1 \neq P(X_{t+1} > i \mid X_t = i, X_{t-1} \neq i) = 0.$$

Evidemment, $P(X_{t+1} > i \mid X_t = i)$ n'est égale à aucune de ces deux probabilités et donc la propriété de Markov n'est pas respectée.

2.3 Diagrammes de transition

Afin de représenter clairement et de manière visuelle les chaînes de Markov, nous construisons des **diagrammes de transition**, dans lesquels les probabilités de transition de la chaîne de Markov en question sont représentés par des flèches reliant les états correspondants et comportant chacune la probabilité qui lui est associée.

Prenons comme exemple un internaute connecté à un réseau avec 4 sites, numérotés 1, 2, 3 et 4 respectivement. A chaque unité de temps, il clique au hasard sur un lien qui l'envoie sur un autre site du réseau. Si le site ne présente aucun lien, l'internaute y reste. L'état dans lequel il se trouve peut donc être représenté par la variable aléatoire X_t . Nous choisissons que le site 1 contient des liens pour 2, 3, et 4, le site 2 pour 1 et 3 et le site 3 pour 1 et 4. Le site 4 ne contient aucun lien. Nous avons donc une liste des probabilités de transition.

- $P(X_1 = 4 \mid X_0 = 4) = 1$
- $P(X_1 = 3 \mid X_0 = 2) = \frac{1}{2}$
- $P(X_1 = 1 \mid X_0 = 3) = \frac{1}{2}$
- $P(X_1 = 2 \mid X_0 = 1) = \frac{1}{3}$
- $P(X_1 = 4 \mid X_0 = 3) = \frac{1}{2}$
- $P(X_1 = 3 \mid X_0 = 1) = \frac{1}{3}$
- $P(X_1 = 1 \mid X_0 = 2) = \frac{1}{2}$
- $P(X_1 = 4 \mid X_0 = 1) = \frac{1}{3}$

Les probabilités de transition restantes sont nulles. Pour des raisons de lisibilité, les transitions de probabilité nulle sont presque toujours omises et souvent les transitions d'un état vers lui-même aussi car il est facile de déduire leur probabilité à partir des autres. Nous pouvons maintenant construire le diagramme de transition de cette marche aléatoire:

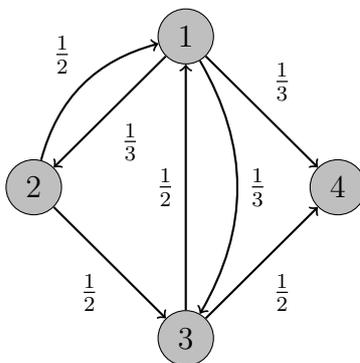


Figure 2.1

Il devient alors facile de suivre les changements d'états possibles de X_t et comprendre le fonctionnement de ce système.

3 Représentation matricielle

Les diagrammes de transition nous permettent de visualiser et de développer une connaissance intuitive des chaînes de Markov. Ils ne sont cependant que peu utiles pour des calculs ou des prédictions. Il faut donc trouver une représentation mathématiquement rigoureuse grâce auxquels nous pouvons travailler de manière efficace. Nous découvrirons

dans ce chapitre un lien entre les chaînes de Markov et l'algèbre linéaire qui comble les faiblesses de la représentation par diagrammes.

3.1 Vecteurs de probabilité

Pour représenter la loi de probabilité d'une chaîne de Markov ou d'autres processus stochastiques discrets, nous utilisons des vecteurs appelés **vecteurs de probabilité** dont chaque composante associe à chaque état la probabilité de s'y trouver à un moment donné. Par exemple, pour un dé simple avec un espace des états $\{1, 2, 3, 4, 5, 6\}$ le vecteur de probabilité est à tout moment $(\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6})$ car la probabilité de lancer un nombre choisi vaut toujours $\frac{1}{6}$.

Nous savons que toute valeur de probabilité doit être comprise entre 0 et 1 et que la somme des probabilités de toutes les issues de l'univers est égale à 1. Nous pouvons donc définir un vecteur de probabilité comme un vecteur $\vec{p} = (p_1, p_2, \dots, p_n)$ avec $\sum_{i=1}^n p_i = 1$ et $0 \leq p_i \leq 1$ pour tout i .

Pour une chaîne de Markov avec un espace des états $\{1, 2, \dots, n\}$, chaque composante de \vec{p} est liée par son indice à l'état correspondant. Nous avons donc $p_i = P(X = i)$.

Etant donné que les chaînes de Markov sont des processus dont la loi de probabilité change en fonction du temps, il est utile de considérer des vecteurs de probabilité correspondant à la même chaîne de Markov, mais à des moments différents dans le temps. Pour faire cela, nous noterons $\vec{p}^{(t)}$ un vecteur de probabilité \vec{p} associé au temps t . Nous pouvons aussi utiliser cette notation pour les composantes de $\vec{p}^{(t)}$. Ainsi, chaque composante $p_i^{(t)} = P(X_t = i)$ de $\vec{p}^{(t)}$ correspond à la probabilité que la chaîne de Markov se trouve dans l'état i au temps t .

3.2 Matrices stochastiques

Définissons maintenant un type de matrice qui, comme nous le verrons, est essentiel dans la représentation des chaînes de Markov: la **matrice stochastique**. La définition d'une matrice stochastique découle directement de celle du vecteur de probabilité; elle est définie comme une matrice carrée dont chaque ligne est un vecteur de probabilité. Il s'agit donc d'une matrice de la forme

$$\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix}$$

avec $\sum_{j=1}^n a_{ij} = 1$ pour tout i et $0 \leq a_{ij} \leq 1$ pour tout i et pour tout j .

Il faut remarquer que cette définition est valable uniquement si nous travaillons avec des vecteurs-ligne, ce qui est le cas ici. En effet, si nous utilisions des vecteurs-colonne, il faudrait redéfinir une matrice stochastique comme une matrice dont chaque colonne est un vecteur de probabilité. La propriété $(AB)^T = B^T A^T$ nous montre cependant que pour un vecteur-ligne p et une matrice A , $(pA)^T = A^T p^T$. Cela revient à dire que travailler avec des vecteurs-ligne ou des vecteurs-colonnes ne

fait aucune différence, tant que l'on ne confond pas les deux notations. Pour des raisons de mise en page et de lisibilité, nous utiliserons dans ce travail des vecteurs-ligne.

Nous démontrons maintenant le premier théorème important de ce travail:

Théorème: Le produit d'un vecteur de probabilité et d'une matrice stochastique est également un vecteur de probabilité.

Preuve:

Soit un vecteur de probabilité $\vec{p} = (p_1, p_2, \dots, p_n)$ et une matrice stochastique

$$A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix}.$$

Vu que le nombre de composantes de \vec{p} est égal au nombre de lignes de A , nous pouvons les multiplier, le produit $\vec{p}A$ étant alors un vecteur à n composantes. Il nous reste donc à démontrer deux propriétés: que toutes les composantes de $\vec{p}A$ sont comprises entre 0 et 1, et que la somme de ces composantes est égale à 1.

Nous savons que chaque composante de $\vec{p}A$ est donné par la somme $\sum_{i=1}^n p_i a_{ij}$, où j est le rang de cette composante. Mais, par définition, tous les p_i et a_{ij} sont des nombres compris entre 0 et 1, ce qui implique que $p_i a_{ij} \leq p_i$ pour tout i et pour tout j , donc

$$0 \leq \sum_{i=1}^n p_i a_{ij} \leq \sum_{i=1}^n p_i = 1$$

pour tout j , ce qui démontre la première propriété.

Ensuite, la somme de toutes les composantes de $\vec{p}A$ est donnée par la relation

$$\sum_{j=1}^n \sum_{i=1}^n p_i a_{ij} = \sum_{i=1}^n p_i \underbrace{\sum_{j=1}^n a_{ij}}_{=1} = \sum_{i=1}^n p_i = 1$$

La deuxième propriété est donc démontrée: le vecteur $\vec{p}A$ est un vecteur de probabilité. \square

Ce résultat nous montre que la multiplication d'un vecteur de probabilité et d'une matrice stochastique correspond à une application linéaire, qui, à partir d'une loi de probabilité initiale, nous donne une nouvelle loi de probabilité.

Soient maintenant deux matrices stochastiques A et B . Chaque ligne de la matrice AB est donnée par la multiplication de la ligne correspondante de A et de la matrice B . Or, par définition, chaque ligne de A est un vecteur de probabilité. Chaque ligne de AB est alors le produit d'un vecteur de probabilité et d'une matrice stochastique, donc aussi un vecteur de probabilité. Il en résulte que AB est une matrice stochastique.

Remarquons encore que, par ce résultat, la multiplication d'une matrice stochastique A par elle-même est une matrice stochastique. La matrice $A^n = \underbrace{AA \dots A}_{n \text{ fois}}$ est donc aussi stochastique.

3.3 Matrice de transition d'une chaîne de Markov

Nous avons vu que la multiplication entre un vecteur de probabilité et une matrice stochastique décrit une application linéaire sur une loi de probabilité. Pour aller plus loin, nous verrons maintenant qu'il s'agit plus précisément de la transition entre deux états d'une chaîne de Markov.

Théorème: Le vecteur de probabilité d'une chaîne de Markov au temps $t + 1$ est donné par le produit de son vecteur de probabilité au temps t et d'une matrice stochastique contenant toutes les probabilités de transition.

Preuve:

Soient une chaîne de Markov, son espace des états $\{1, 2, \dots, n\}$, son vecteur de probabilité $\vec{p}^{(t)} = (p_1^{(t)}, p_2^{(t)}, \dots, p_n^{(t)})$ au temps t et les probabilités de transition entre deux états quelconques $P(X_{t+1} = j \mid X_t = i)$, que nous écrirons a_{ij} .

Nous cherchons le vecteur de probabilité $\vec{p}^{(t+1)} = (p_1^{(t+1)}, p_2^{(t+1)}, \dots, p_n^{(t+1)})$ correspondant à la loi de probabilité après une transition. Vu que les événements $X = i$ et $X = j$ sont disjoints pour tout i et tout j (en effet, notre variable aléatoire ne peut pas admettre deux valeurs à la fois), nous pouvons trouver chaque composante de $\vec{p}^{(t+1)}$ en utilisant le théorème des probabilités totales:

$$p_j^{(t+1)} = P(X_{t+1} = j) = \sum_{i=1}^n P(X_{t+1} = j \mid X_t = i)P(X_t = i) = \sum_{i=1}^n a_{ij}p_i^{(t)}.$$

Or, ce résultat peut être représenté sous la forme d'un produit scalaire entre deux vecteurs:

$$p_j^{(t+1)} = (p_1^{(t)}, p_2^{(t)}, \dots, p_n^{(t)}) \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{nj} \end{pmatrix}.$$

Pour trouver le vecteur $\vec{p}^{(t+1)}$ dans son intégralité, il suffit d'arranger les vecteurs-colonnes \vec{a}_j en une matrice que nous noterons A :

$$\vec{p}^{(t+1)} = (p_1^{(t)}, \dots, p_n^{(t)}) \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} = \vec{p}^{(t)} A.$$

En sommant tous les éléments sur une ligne de A , nous trouvons

$$\sum_{j=1}^n a_{ij} = \sum_{j=1}^n P(X_{t+1} = j \mid X_t = i) = 1.$$

Le vecteur $\vec{p}^{(t+1)}$ est donc bien donné par le produit de $\vec{p}^{(t)}$ et de la matrice stochastique A contenant toutes les probabilités de transition. \square

Nous appelons A la **matrice de transition** de la chaîne de Markov, que nous définissons comme une matrice stochastique dont chaque élément a_{ij} correspond à

la probabilité de passer de l'état i à l'état j dans une unité de temps.

Considérons maintenant l'évolution de cette chaîne de Markov pendant une durée de k unités de temps. Le vecteur de probabilité $p^{(t+k)}$ au temps $t+k$ est donné par la relation suivante:

$$p^{(t+k)} = p^{(t+k-1)}A = p^{(t+k-2)}AA = p^{(t)} \underbrace{AA \dots A}_{k \text{ fois}} = p^{(t)}A^k$$

Les probabilités de transition après k unités de temps sont donc données par la k -ième puissance de la matrice de transition.

Exemple: Loi de Mendel

Dans les années récentes, les chaînes de Markov sont devenues des outils largement utilisés en biologie (et dans bien d'autres disciplines) et donc, ici, nous utiliserons une chaîne de Markov pour modéliser un problème d'hérédité typique des cours de biologie. Cette approche de la génétique est cependant très simplifiée et a plutôt une valeur symbolique et illustrative.

Soit un individu et sa descendance, dont nous étudierons la couleur des yeux. Représentons ces personnes par une variable aléatoire X_t où t représente la génération dans laquelle nous nous trouvons, c'est-à-dire que X_1 est l'enfant de X_0 , X_2 celui de X_1 et ainsi de suite. Pour simplifier, ceci n'étant pas un travail de biologie, nous considérons que la couleur des yeux, limitée à brune ou bleue, ne dépend que d'un gène, où l'allèle **B** dominant correspond aux yeux bruns et l'allèle **b** récessif aux yeux bleus. La variable aléatoire peut donc se trouver dans trois états, numérotés 1, 2 et 3 respectivement: **BB** (homozygote aux yeux bruns), **Bb** (hétérozygote aux yeux bruns) ou **bb** (homozygote aux yeux bleus). Finalement, il faut considérer les probabilités d'avoir un enfant avec une personne **BB**, **Bb** ou **bb**, que nous notons respectivement c_1 , c_2 et c_3 , et qui, dans notre cas simplifié d'une chaîne de Markov homogène, restent constants. Nous pouvons alors calculer toutes les probabilités de transition en utilisant des tableaux de Mendel:

Pour $X_0 = 1(\mathbf{BB})$:

		B		B		et		B		B		et		B		B		:	
B		BB		BB		et	B		BB		BB		et	b		Bb		Bb	
B		BB		BB			b		Bb		Bb			b		Bb		Bb	

$$a_{11} = c_1 + \frac{1}{2}c_2,$$

$$a_{12} = \frac{1}{2}c_2 + c_3,$$

$$a_{13} = 0$$

Pour $X_0 = 2(\mathbf{Bb})$:

		B		b		et		B		b		et		B		b		:	
B		BB		Bb		et	B		BB		Bb		et	b		Bb		bb	
B		BB		Bb			b		Bb		bb			b		Bb		bb	

$$a_{21} = \frac{1}{2}c_1 + \frac{1}{4}c_2,$$

$$a_{22} = \frac{1}{2}c_1 + \frac{1}{2}c_2 + \frac{1}{2}c_3,$$

$$a_{23} = \frac{1}{4}c_2 + \frac{1}{2}c_3$$

Pour $X_0 = 3(\mathbf{bb})$:

$$\begin{array}{|c|c|c|} \hline & \mathbf{b} & \mathbf{b} \\ \hline \mathbf{B} & \mathbf{Bb} & \mathbf{Bb} \\ \hline \mathbf{B} & \mathbf{Bb} & \mathbf{Bb} \\ \hline \end{array} \text{ et } \begin{array}{|c|c|c|} \hline & \mathbf{b} & \mathbf{b} \\ \hline \mathbf{B} & \mathbf{Bb} & \mathbf{Bb} \\ \hline \mathbf{b} & \mathbf{bb} & \mathbf{bb} \\ \hline \end{array} \text{ et } \begin{array}{|c|c|c|} \hline & \mathbf{b} & \mathbf{b} \\ \hline \mathbf{b} & \mathbf{bb} & \mathbf{bb} \\ \hline \mathbf{b} & \mathbf{bb} & \mathbf{bb} \\ \hline \end{array} : \begin{array}{l} a_{31} = 0, \\ a_{32} = c_1 + \frac{1}{2}c_2, \\ a_{33} = \frac{1}{2}c_2 + c_3 \end{array}$$

En choisissant arbitrairement $c_1 = c_2 = c_3 = \frac{1}{3}$ (nous supposons que les génotypes **BB**, **Bb** et **bb** sont équitablement abondants), nous trouvons la matrice de transition suivante:

$$A = \frac{1}{4} \begin{pmatrix} 2 & 2 & 0 \\ 1 & 2 & 1 \\ 0 & 2 & 2 \end{pmatrix}.$$

A partir de la matrice de transition nous construisons aussi le diagramme de transition de cette chaîne.

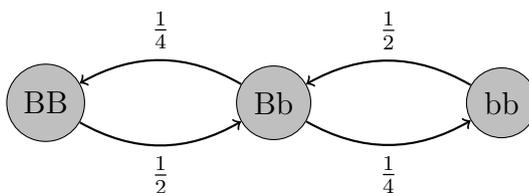


Figure 3.1

Nous pouvons choisir que la première personne a les yeux bleus, donc $X_0 = 3$, ce qui nous donne le vecteur de probabilité $\vec{p}^{(0)} = \vec{p} = (0, 0, 1)$. Il est alors possible de calculer $\vec{p}^{(t)}$ pour tout $t \geq 0$, par exemple

$$\vec{p}^{(2)} = \vec{p}A^2 = \left(\frac{1}{8}, \frac{1}{2}, \frac{3}{8}\right).$$

La probabilité que le petit-fils ou la petite-fille X_2 de X_0 (yeux bleus) ait des yeux bruns est alors de $\frac{5}{8}$.

4 Classification des états

Nous avons désormais les outils nécessaires pour représenter mathématiquement les chaînes de Markov, les visualiser et prédire leur évolution. Dans ce chapitre nous classerons les états en fonction de différentes propriétés, ce qui nous permet de faire un premier pas vers des applications plus concrètes des chaînes de Markov.

4.1 Classes communicantes et réductibilité

Nous établirons ici une relation d'équivalence qui nous permet de partitionner les chaînes de Markov dans des parties que l'on appelle **classes d'états**. Rappelons d'abord que toute relation d'équivalence doit satisfaire trois critères: elle doit être réflexive, symétrique et transitive.

Nous définissons ensuite qu'un état j est **accessible** depuis un état i , que nous écrirons aussi $i \rightarrow j$, si une transition entre i et j est possible dans un certain intervalle de temps. Il existe donc un nombre entier $k \geq 0$ tel que $a_{ij}^{(k)} > 0$. Deux états qui sont accessibles l'un depuis l'autre sont alors appelés **communicants**, nous notons cela $i \leftrightarrow j$. Il s'agit de la relation d'équivalence en question. Pour comprendre pourquoi nous permettons que k soit nul, considérons la chaîne suivante:

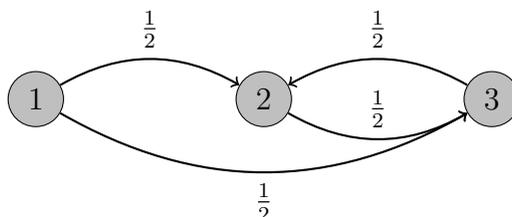


Figure 4.1

Nous voyons clairement que $a_{11}^{(k)} = 0$ pour tout $k > 0$, on ne reste donc jamais dans cet état, et, une fois sorti, on n'y retourne jamais. Or, pour que la relation de communication soit réflexive ($i \leftrightarrow i$ pour tout i), tout état doit communiquer avec lui-même. La seule solution est alors d'admettre que $a_{ii}^{(0)} = (A^0)_{ii} = I_{ii} = 1$ pour n'importe quel état i .

Nous remarquons aussi que la relation de communication est par définition symétrique car $i \leftrightarrow j$ implique $i \rightarrow j$ et $j \rightarrow i$, donc $i \leftrightarrow j \Leftrightarrow j \leftrightarrow i$.

Finalement, nous devons montrer que cette relation est transitive. Considérons trois états i, j et k tels que $i \leftrightarrow j$ et $j \leftrightarrow k$. Cela implique que nous pouvons passer de l'état i à j dans un temps fini, puis de j à k dans un temps également fini. Il en résulte qu'il existe un entier $l \geq 0$ tel que $a_{ik}^{(l)} > 0$. Par un raisonnement identique nous trouvons encore que $a_{ki}^{(m)} > 0$, pour un entier $m \geq 0$. L'état i communique alors avec l'état k .

Il s'agit donc bien d'une relation d'équivalence, ce qui nous permet de partitionner l'espace des états de n'importe quelle chaîne de Markov en ce que nous appelons des **classes d'états** ou **classes communicantes**. Les classes communicantes forment une partition de l'espace des états S , chaque classe regroupant tous les états qui communiquent entre eux. Prenons comme exemple la chaîne suivante (les probabilités de transition ont été supprimées pour des raisons de lisibilité):

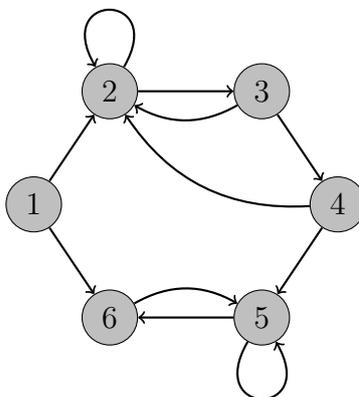


Figure 4.2

Cette chaîne de Markov comporte trois classes d'états: $\{1\}$, $\{2, 3, 4\}$, et $\{5, 6\}$. Nous les appellerons respectivement C_1 , C_2 et C_3 . Nous pouvons d'ailleurs remarquer que l'état 1 n'est accessible que depuis lui-même pour $n = 0$. Regardons maintenant chaque classe séparément.

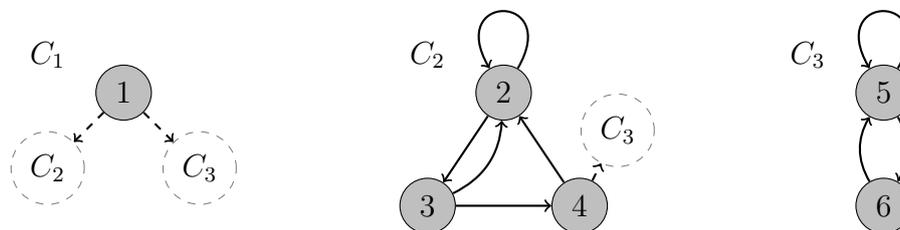


Figure 4.3

Nous voyons que la chaîne peut être divisée en trois parties plus réduites que nous pouvons dans une certaine mesure considérer séparément, ce qui nous facilite l'analyse des chaînes plus complexes. C'est ici qu'apparaît la notion de **réductibilité** qui est indispensable pour l'étude du comportement à long terme des chaînes de Markov. On dit qu'une chaîne de Markov est **réductible** si elle comporte plusieurs classes d'états. Une chaîne **irréductible** est alors une chaîne comportant une seule classe, c'est-à-dire dont tous les états communiquent. Par exemple, la chaîne de la figure 4.2 est réductible car elle comporte plusieurs classes et donc il existe des couples d'états qui ne communiquent pas.

4.2 Etats absorbants et probabilités d'absorption

Nous présenterons ici un type d'état qui nous permet de représenter les conditions sous lesquelles se termine une chaîne de Markov, et une méthode qui permet de calculer la probabilité d'atteindre un état depuis un autre.

Considérons un processus modélisé par une chaîne de Markov qui peut se terminer dans certaines circonstances, par exemple une série d'épreuves avec des conditions de perte ou de victoire. Ces conditions peuvent être représentées par des états dont il est impossible de ressortir, appelés états **absorbants**. Formellement, nous pouvons définir qu'un état i est absorbant si $P(X_{t+1} = i \mid X_t = i) = 1$. Il en découle directement que $P(X_{t+n} = i \mid X_t = i) = 1$ pour tout $n \geq 0$. En plus, tout état absorbant constitue une classe à lui seul car il ne peut communiquer qu'avec lui-même.

Nous avons déjà vu un état absorbant dans ce travail; il s'agit de l'état 4 de l'exemple de l'internaute. En effet, la probabilité d'y sortir est 0 et la chaîne y reste donc pour toujours.

Pour des processus avec plusieurs conditions finales, il est intéressant de connaître quel état absorbant sera atteint avec quelle probabilité, si nous connaissons l'état de départ. La méthode proposée ici est cependant tout aussi applicable aux états non-absorbants.

Soit une variable aléatoire entière $H_i = \min\{t \geq 0 : X_t = i\}$, le plus petit entier non négatif tel que $X_t = i$. C'est le temps du premier passage en i . Si nous considérons qu'un temps de premier passage infini correspond au fait de ne jamais atteindre l'état, la probabilité d'atteindre l'état i au moins une fois est donné par

$P(H_i < \infty)$. Nous définissons alors $h_{ij} = P(H_j < \infty \mid X_0 = i)$, la probabilité d'atteindre l'état j depuis i . Il est clair que si $i = j$, $h_{ij} = 1$ car $P(H_j < \infty \mid X_0 = j) = P(H_j = 0 \mid X_0 = j) = 1$. Dans les autres cas, la chaîne peut emprunter plusieurs chemins différents pour atteindre j et, grâce à la propriété de Markov et au fait que $H_j \geq 1$, nous pouvons exprimer les différentes h_{ij} pour un même j les unes en fonction des autres:

$$\begin{aligned} h_{ij} &= P(H_j < \infty \mid X_0 = i) = \sum_{k=1}^n P(H_j < \infty \mid X_1 = k, X_0 = i) = \\ &= \sum_{k=1}^n P(X_1 = k \mid X_0 = i)P(H_j < \infty \mid X_1 = k) = \\ &= \sum_{k=1}^n P(X_1 = k \mid X_0 = i)P(H_j < \infty \mid X_0 = k) = \sum_{k=1}^n a_{ik}h_{kj} \end{aligned}$$

Nous trouvons donc tous les h_{ij} en résolvant le système d'équations linéaires suivant:

$$\begin{cases} h_{ij} = 1, \text{ si } i = j \\ h_{ij} = \sum_{k=1}^n a_{ik}h_{kj}, \text{ si } i \neq j \end{cases}$$

Exemple: La ruine du joueur

La ruine du joueur est un problème classique des probabilités, modélisé ici par une marche aléatoire sur une partie \mathbb{N} . Un joueur avec une fortune de i francs décide de tenter sa chance dans un jeu de hasard où il parie à chaque tour 1 franc. La probabilité de gagner un pari reste la même pendant tout le jeu, nous la notons p . La probabilité de perdre est alors $q = 1 - p$. Le jeu se termine quand le joueur perd tout son argent (ruine), ou quand il accumule une fortune de N francs (victoire). Le jeu peut être modélisé par une marche aléatoire sur $\{0, 1, 2, \dots, N\}$ avec une probabilité p d'avancer et une probabilité q de reculer, d'un pas dans les deux cas. Les conditions finales 0 et N seront alors représentées par des états absorbants.

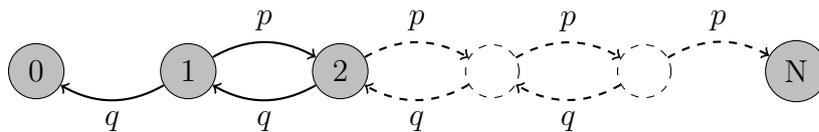


Figure 4.4

Le joueur s'intéresse évidemment à h_{iN} , la probabilité de gagner en commençant avec une fortune de i francs. La matrice de transition de la chaîne est

$$\begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ q & 0 & p & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & q & 0 & p \\ 0 & \dots & 0 & 0 & 1 \end{pmatrix}.$$

Nous connaissons donc déjà $h_{0N} = 0$ et $h_{NN} = 1$. Pour les autres états, tout h_{iN} tel que $0 < i < N$ est donné par $h_{iN} = qh_{i-1,N} + ph_{i+1,N}$. De plus, vu que $p + q = 1$, nous avons

$$h_{iN} = ph_{iN} + qh_{iN} = qh_{i-1,N} + ph_{i+1,N},$$

$$\text{d'où } h_{i+1,N} - h_{iN} = \frac{q}{p}(h_{iN} - h_{i-1,N}).$$

Si nous prenons $i = 1$, nous avons

$$h_{2N} - h_{1N} = \frac{q}{p}(h_{1N} - h_{0N}) = \frac{q}{p}h_{1N}.$$

Nous pouvons alors exprimer $h_{i+1,N} - h_{iN}$ récursivement en fonction de h_{1N} .

$$\begin{aligned} h_{3N} - h_{2N} &= \frac{q}{p}(h_{2N} - h_{1N}) = \left(\frac{q}{p}\right)^2 h_{1N} \\ h_{4N} - h_{3N} &= \frac{q}{p}(h_{3N} - h_{2N}) = \left(\frac{q}{p}\right)^3 h_{1N} \\ &\vdots \\ h_{i+1,N} - h_{iN} &= \left(\frac{q}{p}\right)^i h_{1N} \end{aligned}$$

Si nous écrivons $h_{i+1,N} - h_{1N}$ comme une somme des $h_{k+1,N} - h_{kN}$ avec k compris entre 1 et i , nous pouvons exprimer $h_{i+1,N}$ en fonction de h_{1N} :

$$\begin{aligned} h_{i+1,N} - h_{1N} &= h_{i+1,N} - h_{iN} + h_{iN} - h_{i-1,N} + h_{i-1,N} \cdots - h_{2N} + h_{2N} - h_{1N} = \\ &= \sum_{k=1}^i (h_{k+1,N} - h_{kN}) = \sum_{k=1}^i \left(\frac{q}{p}\right)^k h_{1N}, \\ \text{donc } h_{i+1,N} &= h_{1N} + \sum_{k=1}^i \left(\frac{q}{p}\right)^k h_{1N} = \sum_{k=0}^i \left(\frac{q}{p}\right)^k h_{1N}. \end{aligned}$$

Cette somme vaut $h_{1N} \frac{1 - \left(\frac{q}{p}\right)^{i+1}}{1 - \frac{q}{p}}$ si $p \neq q$ et $h_{1N}(i+1)$ si $p = q$. Mais nous ne connaissons donc toujours pas h_{1N} . Pour le trouver, nous exploitons le fait que $h_{NN} = 1$. Si $p \neq q$, nous trouvons

$$h_{1N} \frac{1 - \left(\frac{q}{p}\right)^N}{1 - \frac{q}{p}} = h_{NN} = 1 \Leftrightarrow h_{1N} = \frac{1 - \frac{q}{p}}{1 - \left(\frac{q}{p}\right)^N}.$$

Si $p = q$, $Nh_{1N} = h_{NN} = 1$ et donc $h_{1N} = \frac{1}{N}$. Avec une simple substitution, nous trouvons h_{iN} pour tout i entre 1 et $N-1$.

$$\left\{ \begin{array}{l} \text{si } p \neq q, h_{iN} = \frac{1 - \frac{q}{p}}{1 - \left(\frac{q}{p}\right)^N} \frac{1 - \left(\frac{q}{p}\right)^i}{1 - \frac{q}{p}} = \frac{1 - \left(\frac{q}{p}\right)^i}{1 - \left(\frac{q}{p}\right)^N} \\ \text{si } p = q, h_{iN} = i \frac{1}{N} = \frac{i}{N} \end{array} \right.$$

Il est clair que le cas le plus simple et le plus souvent considéré est celui où $p = q = \frac{1}{2}$. Dans ce cas, la probabilité de gagner devient un simple rapport entre la fortune de départ et l'argent en jeu. Il est tout de même certain qu'un tel jeu ne nous aiderait pas à sortir des difficultés financières.

4.3 Chaînes de Markov périodiques

Quand nous avons défini la notion d'accessibilité, nous ne nous sommes pas occupés de l'entier k pour lequel la probabilité de transition soit non nulle, tant qu'un tel k soit présent. Il existe cependant des chaînes pour lesquelles certaines probabilités de transition prennent des valeurs strictement positives seulement pour des valeurs spécifiques de k . Considérons la chaîne suivante.

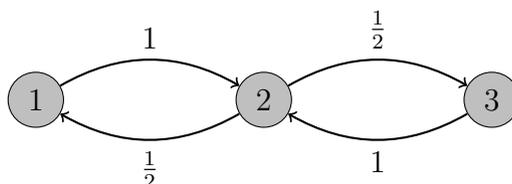


Figure 5.1

Nous voyons que cette chaîne retourne toujours dans l'état 2 après exactement 2 pas. Un calcul matriciel simple affirme que la probabilité $a_{22}^{(k)}$ est non nulle et, dans ce cas, égale à 1 seulement si k est pair, donc un multiple de 2. Nous disons alors que l'état 2 a une **période** de 2. Si nous prenons un autre état de la chaîne, par exemple 1, nous voyons que nous ne sommes pas sûrs d'y retourner dans exactement 2 pas car la chaîne peut tourner dans le boucle 2-3-2 une ou plusieurs fois, sans limite déterminée. Nous remarquons cependant que chaque tour dans ce cycle prend 2 unités de temps et donc la probabilité $a_{11}^{(k)}$ est de nouveau non nulle seulement si t est un multiple de 2. En fait, 2 est le plus grand diviseur commun des entiers k pour lesquels les probabilités de transition d'ordre k de l'état 1 vers lui-même sont non nulles. C'est ainsi que nous définissons la périodicité, la période d_i d'un état i vaut

$$d_i = \text{pgdc}\{k > 0 : a_{ii}^{(k)} > 0\}.$$

Si la période de i est 1, c'est-à-dire s'il existe des k premiers entre eux tels que $a_{ii}^{(k)} > 0$, nous disons que l'état i est **apériodique**. Il est évident que tout état i comportant une auto-transition, donc telle que $a_{ii} > 0$, est apériodique.

Il faut remarquer que dans la chaîne de la figure 5.1, tous les états ont la même période. Cela est aussi vrai en général pour les états d'une classe communicante quelconque. Pour le démontrer, nous utilisons la relation de divisibilité. Nous écrivons alors $a \mid b$ avec $a, b \in \mathbb{Z}$ si a est un diviseur de b . Il est facile de démontrer que $a \mid b$ et $a \mid c$ impliquent $a \mid b + c$, ce qui nous permet de prouver le théorème suivant.

Théorème: Les états d'une classe communicante ont tous la même période.

Preuve:

Soient deux états i et j appartenant à une classe d'états. Il existe alors des entiers m et n strictement positifs tels que $a_{ij}^{(m)} > 0$ et $a_{ji}^{(n)} > 0$. Nous utilisons le fait que la probabilité de retourner à i en passant par j est inférieure ou égale à la probabilité de retourner à i tout court:

$$a_{ii}^{(m+n)} \geq a_{ij}^{(m)} a_{ji}^{(n)} > 0.$$

De même, nous avons

$$a_{jj}^{(m+n)} \geq a_{ji}^{(n)} a_{ij}^{(m)} > 0.$$

Par définition donc, d_i et d_j sont des diviseurs de $m+n$.

Nous savons qu'il existe des entiers k_i et k_j tels que $a_{ii}^{(k_i)} > 0$ et $a_{jj}^{(k_j)} > 0$ qui satisfont de plus $d_i = \text{pgdc}\{k_i\}$ et $d_j = \text{pgdc}\{k_j\}$. Comme avant nous trouvons alors

$$a_{ii}^{(m+n+k_j)} \geq a_{ij}^{(m)} a_{jj}^{(k_j)} a_{ji}^{(n)} > 0$$

$$\text{et } a_{jj}^{(m+n+k_i)} \geq a_{ji}^{(m)} a_{ii}^{(k_i)} a_{ij}^{(n)} > 0.$$

Il s'ensuit que $d_i \mid m+n+k_j$ et $d_j \mid m+n+k_i$, ce qui implique $d_i \mid k_j$ et $d_j \mid k_i$. La période d_i est alors un diviseur commun des k_j et donc, vu que d_j est le plus grand diviseur commun des k_j , nous avons $d_i \mid d_j$. Par un argument similaire, nous trouvons que $d_j \mid d_i$. Or, cela est seulement possible si $d_i = d_j$. Les états i et j ont donc la même période. \square

Nous pouvons ainsi parler des classes, ou même des chaînes périodiques ou apériodiques. Les chaînes apériodiques sont définies comme des chaînes de Markov dont tous les états ont une période de 1. Dans ce travail, la notion de période nous intéresse surtout parce qu'elle rend souvent le comportement asymptotique d'une chaîne de Markov plus difficile à étudier. En effet, les chaînes périodiques comportent par leur définition des probabilités de transition qui seront nulles même pour des lapses de temps t très grands. Comme nous le verrons, cela complique passablement leur comportement à long terme.

Néanmoins, certaines chaînes périodiques, comme celle de l'exemple 5.1, ont un comportement très simple. En construisant la matrice de transition à partir de la figure 5.1 et en calculant ses puissances successives, nous trouvons:

$$A = \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{pmatrix}, A^2 = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}, A^3 = \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{pmatrix} = A$$

Il est facile d'en déduire que la k -ième puissance de la matrice de transition de cette chaîne est donc A pour tout k impair, et A^2 pour tout k pair.

5 Chaînes de Markov régulières

Une grande partie des applications pratiques des chaînes de Markov fait appel à un type de chaînes très utiles grâce à leur comportement asymptotique; c'est les chaînes **régulières**, le dernier type de chaînes que nous traitons dans ce travail.

5.1 Définition et conditions nécessaires

Nous définissons une **chaîne de Markov régulière** comme une chaîne dont la matrice de transition A admet une puissance A^k telle que $a_{ij}^{(k)} > 0$ pour tout i et tout j . Nous parlons alors d'une **matrice stochastique régulière**. A cause de leur utilité

considérable, il est judicieux d'établir quelques conditions nécessaires afin de repérer ces chaînes sans devoir examiner puissance après puissance la matrice de transition. Il est évident que si tous les éléments de $A^{(k)}$ sont strictement positifs, alors ceux de $A^{(k+1)}$, $A^{(k+2)}$ et ainsi de suite le seront aussi. Il suffit donc de trouver un temps k pour lequel cela est vrai.

Considérons maintenant les classes d'états de la chaîne. Si la chaîne comporte plusieurs classes communicantes, il existe forcément des états qui ne sont pas atteignables depuis certains autres. Des probabilités de transition nulles seront donc présentes pour tout k , ce qui est en contradiction avec la définition d'une chaîne régulière. Par conséquent, une chaîne régulière comporte une seule classe; elle est irréductible. De plus, comme mentionné avant, les puissances des matrices de transition des chaînes périodiques comportent toujours des éléments nuls, peu importe la durée de temps. Une chaîne régulière doit donc aussi être aperiodique.

Considérons ensuite un état i quelconque d'une chaîne aperiodique et irréductible. Il existe par définition des entiers m et n (pas en lien avec la taille de la matrice) premiers entre eux tels que $a_{ii}^{(m)} > 0$ et $a_{ii}^{(n)} > 0$. De plus, toute somme $mk + nl$ des multiples positifs de m et de n satisfait aussi la propriété $a_{ii}^{(mk+nl)} > 0$:

$$a_{ii}^{(mk+nl)} \geq \left(a_{ii}^{(m)}\right)^k \left(a_{ii}^{(n)}\right)^l > 0.$$

Il est intuitif et, en utilisant la théorie des nombres, relativement facile de démontrer que tout entier $N \in \mathbb{N}$ suffisamment grand peut être exprimé comme la somme $mk + nl$ si m et n sont premiers entre eux. Ceci nécessite néanmoins l'introduction par exemple de la notion de congruence et dépasse donc le cadre de ce travail. Il en résulte que les éléments diagonaux de la matrice A^N sont strictement positifs pour tout N suffisamment grand. Etant donné que tous les états d'une chaîne irréductible communiquent, il existe des entiers r tels que $a_{ij}^{(r)} > 0$ pour tout j et tout i . Par la relation $a_{ij}^{(N+r)} \geq a_{ii}^{(N)} a_{ij}^{(r)} > 0$ nous trouvons que tous les éléments de la matrice A^{N+r} sont strictement positifs, ce qui est la définition d'une matrice stochastique régulière. Pour résumer, nous pouvons déterminer si une chaîne de Markov est régulière ou pas en vérifiant si elle est irréductible et aperiodique. Cette tâche est encore considérablement facilitée par le fait que tous les états d'une chaîne irréductible ont la même période.

Nous pouvons reprendre notre étude de la loi de Mendel comme un exemple de chaîne régulière. En effet, elle comporte une seule classe et des auto-transitions qui la rendent aperiodique. Sinon, il suffit de remarquer que le carré de sa matrice de transition ne comporte que des éléments positifs:

$$A^2 = \frac{1}{16} \begin{pmatrix} 6 & 8 & 2 \\ 4 & 8 & 4 \\ 2 & 8 & 6 \end{pmatrix}.$$

5.2 Vecteurs stationnaires

Les propriétés utiles des chaînes régulières sont étroitement liées à leurs valeurs et vecteurs propres qui seront le sujet de cette section. Avant de commencer, il nous faut néanmoins nous assurer que la notation de ce travail soit cohérente.

Ainsi, pour éviter les confusions, nous mettons en valeur la distinction entre les vecteurs propres **à droite** et **à gauche** d'une matrice. Si un vecteur(-colonne) propre à droite satisfait $A\vec{v} = \lambda\vec{v}$, un vecteur(-ligne) propre à gauche satisfait $\vec{v}A = \lambda\vec{v}$. Nous remarquons aussi que la transposée d'un vecteur propre à gauche de A est un vecteur propre à droite de A^\top , et inversement. En accord avec le reste du travail, nous continuerons d'employer les vecteurs-ligne à gauche de la matrice.

Tout d'abord, multiplions la transposée A^\top de la matrice stochastique A (régulière ou pas) par le vecteur ligne $c\vec{\mathbf{1}} = (c, c, \dots, c)$ avec $c \in \mathbb{R}^*$. Vu que la somme des éléments d'une colonne de A^\top est 1, tout élément de son produit avec $c\vec{\mathbf{1}}$ est donné par $(c\vec{\mathbf{1}}A^\top)_i = \sum_{j=1}^n cA_{ji}^\top = c \sum_{j=1}^n A_{ji} = c$ et donc $c\vec{\mathbf{1}}A^\top = c\vec{\mathbf{1}}$. Par conséquent, la matrice A^\top admet la valeur propre à gauche 1 et donc A aussi:

$$\det(A - I) = \det((A - I)^\top) = \det(A^\top - I^\top) = \det(A^\top - I) = 0.$$

Ceci nous permet de nous passer de préciser si la valeur propre 1 correspond à un vecteur propre à droite ou à gauche car l'un implique l'autre. Il existe donc pour toute matrice stochastique A au moins un **vecteur stationnaire** ou **invariant**, défini comme un vecteur de probabilité \vec{s} 1-propre à gauche de A et qui satisfait donc $\vec{s}A = \vec{s}$. Il est facile de vérifier que $\vec{s}A^k = \vec{s}$ également.

Les chaînes de Markov avec plusieurs classes communicantes peuvent admettre plusieurs vecteurs stationnaires, mais nous nous intéresserons ici seulement aux chaînes régulières qui en admettent un seul. Nous démontrerons dans un premier temps qu'une chaîne régulière admet au plus un vecteur stationnaire en utilisant deux théorèmes. Nous fournissons la preuve seulement pour le premier qui affirme que l'espace engendré par les vecteurs-ligne d'une matrice a la même dimension que celui engendré par ses vecteurs-colonne. Nous parlons alors du **rang** d'une matrice.

Théorème: Le rang $\text{rg}(A)$ d'une matrice A est égale au rang $\text{rg}(A^\top)$ de sa transposée.

Preuve:

Soient une matrice (pas nécessairement stochastique, mais non nulle) A de taille $m \times n$ et deux espaces vectoriels V_l et V_c , le premier engendré par les vecteurs-ligne de A et le deuxième par ses vecteurs-colonne. Nous notons r la dimension de V_l .

Considérons alors une base $\mathcal{B}(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_r)$ de V_l . Nous commençons par montrer que les vecteurs-colonne $A\vec{x}_1^\top, A\vec{x}_2^\top, \dots, A\vec{x}_r^\top$ sont linéairement indépendants. Pour faire cela, nous posons

$$\vec{0} = \sum_{i=1}^r c_i A\vec{x}_i^\top = A \sum_{i=1}^r c_i \vec{x}_i^\top$$

que nous voulons résoudre pour $c_i \in \mathbb{R}$. Soit alors un vecteur-colonne \vec{v} donné par la combinaison linéaire $\vec{v} = \sum_{i=1}^r c_i \vec{x}_i$. Pour que la relation posée soit respectée, il faut que $A\vec{v}^\top = \vec{0}$ et donc que \vec{v} soit orthogonal à tous les vecteurs-ligne de A . Par extension, \vec{v} doit être orthogonal à tous les vecteurs de la base $\mathcal{B}(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_r)$, ce qui est par sa définition possible seulement si $\vec{v} = \vec{0}$. La seule solution est alors

$c_1 = c_2 = \dots = c_r = 0$ et donc les vecteurs $A\vec{x}_1^\top, A\vec{x}_2^\top, \dots, A\vec{x}_r^\top$ sont bien linéairement indépendants.

Ensuite nous remarquons que tout vecteur $A\vec{x}_i^\top$ appartient à V_c :

$$A\vec{x}_i^\top = \begin{pmatrix} a_{11}(\vec{x}_i)_1 + a_{12}(\vec{x}_i)_2 + \dots + a_{1n}(\vec{x}_i)_n \\ a_{21}(\vec{x}_i)_1 + a_{22}(\vec{x}_i)_2 + \dots + a_{2n}(\vec{x}_i)_n \\ \vdots \\ a_{m1}(\vec{x}_i)_1 + a_{m2}(\vec{x}_i)_2 + \dots + a_{mn}(\vec{x}_i)_n \end{pmatrix} = \sum_{k=1}^n (\vec{x}_i)_k \begin{pmatrix} a_{1k} \\ a_{2k} \\ \vdots \\ a_{mk} \end{pmatrix} \in V_c.$$

Cela veut dire que l'espace V_c admet au moins $r = \dim(V_l)$ vecteurs linéairement indépendants et donc nous avons $\dim(V_c) \geq \dim(V_l)$. En appliquant à A^\top tout le raisonnement précédent, nous trouvons $\dim(V_l) \geq \dim(V_c)$ et donc $\dim(V_l) = \dim(V_c)$. \square

Le deuxième théorème affirme que le rang $\text{rg}(A)$ d'une matrice A de taille $n \times m$ et la dimension de son noyau sont liés par la relation $\text{rg}(A) + \dim(\ker(A)) = n$. Pour une matrice carrée de taille $n \times n$ nous obtenons l'égalité $\text{rg}(A) + \dim(\ker(A)) = \text{rg}(A^\top) + \dim(\ker(A^\top))$ et, par le résultat $\text{rg}(A) = \text{rg}(A^\top)$, nous trouvons $\dim(\ker(A)) = \dim(\ker(A^\top))$. Par conséquent, trouver la dimension de l'espace 1-propre de la matrice stochastique régulière A revient à trouver la dimension du noyau de $A^\top - I$. En travaillant avec les matrices stochastiques régulières, nous considérons que tous les éléments de A sont déjà positifs, sinon il suffit de considérer une puissance A^k de A qui nous convient. Cherchons donc tous les vecteurs 1-propres à gauche de A^\top . Nous savons déjà que $c\vec{1}$ en est un. Supposons alors par l'absurde qu'il existe un vecteur $\vec{v} = (v_1, v_2, \dots, v_n)$ 1-propre linéairement indépendant de $c\vec{1}$. Définissons alors v_M comme la plus grande composante de \vec{v} . Le fait qu'il existe des $v_k < v_M$ et la relation $\vec{v}A^\top = \vec{v}$ nous donne alors l'inégalité suivante:

$$v_M = \sum_{k=1}^n (A^\top)_{Mk} v_k = \sum_{k=1}^n a_{kM} v_k < \sum_{k=1}^n a_{kM} v_M = v_M$$

L'affirmation $v_M < v_M$ étant contradictoire, nous avons démontré que A^\top admet comme vecteurs 1-propres uniquement les multiples du vecteur $\vec{1} = (1, 1, \dots, 1)$. Il en découle que la dimension du noyau de $A^\top - I$, et donc de $A - I$, vaut 1. L'espace 1-propre de A est donc aussi de dimension 1 et ses vecteurs 1-propres sont les multiples les uns des autres. Il existe alors un seul vecteur 1-propre à gauche de A dont la somme des composantes vaut 1. Nous ne savons pas encore s'il s'agit d'un vecteur de probabilité, cela apparaîtra dans la section suivante.

5.3 Convergence des chaînes régulières

Nous commençons cette fois-ci en reprenant l'exemple de la loi de Mendel. La matrice de transition de cette chaîne est facilement diagonalisable et donc nous pouvons trouver une formule générale de A^k . Nous nous contentons ici de la montrer sans présenter les calculs:

$$A^k = \frac{1}{4} \begin{pmatrix} 1 + \frac{1}{2^{k-1}} & 2 & 1 - \frac{1}{2^{k-1}} \\ 1 & 2 & 1 \\ 1 - \frac{1}{2^{k-1}} & 2 & 1 + \frac{1}{2^{k-1}} \end{pmatrix}$$

Nous remarquons que le terme $\frac{1}{2^{k-1}}$ tend vers 0 quand k tend vers l'infini et il nous reste une matrice dont chaque ligne est formée par le vecteur $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$. En le multipliant par A , nous retrouvons le même vecteur: il s'agit de l'unique vecteur stationnaire de A !

Dans cette dernière partie de théorie nous traitons cette propriété qui est probablement la plus intéressante des chaînes régulières: leur convergence vers la loi stationnaire.

Premièrement, nous démontrons que si la matrice stochastique A est régulière, le produit de A^k et d'un vecteur-colonne quelconque tend vers un vecteur à composantes égales quand k tend vers l'infini.

Théorème: Soit A une matrice stochastique régulière et \vec{v} un vecteur colonne à coefficients non négatifs. Alors $\lim_{k \rightarrow \infty} A^k \vec{v} = (u, u, \dots, u)^T$.

Preuve:

Comme avant, nous considérons que tous les éléments de A sont déjà strictement positifs. La preuve est triviale si la matrice est de taille 1 ou si toutes les composantes de \vec{v} sont égales.

Nous considérons donc que $n \geq 2$ et qu'il existe une composante maximale v_M et une composante minimale v_m de \vec{v} de telle manière que $v_M > v_m$. Soit alors a_m le plus petit élément de A . Comme $na_m \leq 1$ et 2 est la valeur minimale de n , nous avons $0 < a_m \leq \frac{1}{2}$.

Nous continuons en cherchant les bornes supérieure (v_{M1}) et inférieure (v_{m1}) des composantes de $\vec{v}_1 = A\vec{v}$. Il est évident que plus l'élément minimal d'une ligne de A est petit, plus les autres éléments seront grands. Pour maximiser v_{M1} , il faut alors considérer que a_m se trouve sur chaque ligne de A de sorte qu'il est multiplié par v_m , puis que tous les composantes restantes de \vec{v} valent v_M et sont multipliées par des composantes dont la somme vaut $1 - a_m$. Nous trouvons alors la valeur maximale possible:

$$v_{M1} = a_m v_m + (1 - a_m) v_M \leq a_m v_M + (1 - a_m) v_M = v_M.$$

De même, nous trouvons la plus petite v_{m1} possible si tous les composants sauf 1 valent v_m et si a_m coïncide avec v_M :

$$v_{m1} = a_m v_M + (1 - a_m) v_m \geq a_m v_m + (1 - a_m) v_m = v_m.$$

Toute composante de \vec{v}_1 est donc bornée par v_{m1} et v_{M1} . Nous pouvons récursivement appliquer le même procédé à $\vec{v}_2 = A\vec{v}_1 = A^2\vec{v}$ et à $\vec{v}_k = A^k\vec{v}$. Nous trouvons alors l'inégalité

$$v_m \leq v_{m1} \leq \dots \leq v_{mk} \leq v_{Mk} \leq \dots \leq v_{M1} \leq v_M$$

qui nous indique déjà que les composantes de \vec{v}_t convergent vers une même limite si par hasard $v_{Mk} - v_{mk}$ tend vers 0 quand t s'approche de l'infini. Pour confirmer cela, nous exprimons $v_{Mk} - v_{mk}$ en fonction de v_M et de v_m :

$$\begin{aligned} v_{M1} - v_{m1} &= a_m v_m + (1 - a_m) v_M - (a_m v_M + (1 - a_m) v_m) = (1 - 2a_m)(v_M - v_m) \\ v_{M2} - v_{m1} &= (1 - 2a_m)(v_{M1} - v_{m1}) = (1 - 2a_m)^2(v_M - v_m) \\ &\vdots \\ v_{Mk} - v_{mk} &= (1 - 2a_m)^k(v_M - v_m). \end{aligned}$$

Or, nous avons posé avant que $0 < a_m \leq \frac{1}{2}$, ce qui implique que $0 \leq 1 - 2a_m < 1$ et donc v_{Mk} et v_{mk} s'approchent bien de la même valeur que nous notons u :

$$\lim_{k \rightarrow \infty} (v_{Mk} - v_{mk}) = \lim_{t \rightarrow \infty} (1 - 2a_m)^k (v_M - v_m) = 0$$

Nous avons donc bien $\lim_{k \rightarrow \infty} A^k \vec{v} = (u, u, \dots, u)^T$. \square

Il est alors facile de trouver la matrice $S = \lim_{t \rightarrow \infty} A^t$ en multipliant $\lim_{t \rightarrow \infty} A^t$ par la matrice identité I . En effet, la i -ème colonne de S correspond à $\lim_{t \rightarrow \infty} A^t \vec{e}_i = (s_i, s_i, \dots, s_i)^T$. Reprenons la relation $v_{m1} = a_m v_M + (1 - a_m) v_m$. Dans notre cas $v_m = 0$ et $v_M = 1$ et donc $s_i = \lim_{t \rightarrow \infty} v_{mt} \geq v_{m1} = a_m > 0$. Nous nous retrouvons finalement avec une matrice stochastique à coefficients strictement positifs:

$$S = \lim_{t \rightarrow \infty} A^t = \begin{pmatrix} s_1 & s_2 & \dots & s_n \\ s_1 & s_2 & \dots & s_n \\ \vdots & \vdots & & \vdots \\ s_1 & s_2 & \dots & s_n \end{pmatrix}$$

Ce n'est pas par hasard que nous avons nommé S et ses éléments ainsi. La relation $SA = \lim_{t \rightarrow \infty} A^{t+1} = S$ et le fait que chaque ligne correspond à un même vecteur $\vec{s} = (s_1, s_2, \dots, s_n)$ prouvent que ces lignes sont chacune le vecteur stationnaire de A , comme nous avons vu avant.

Multiplions maintenant S par un vecteur de probabilité \vec{p} quelconque.

$$\begin{aligned} \vec{p}S &= (p_1, p_2, \dots, p_n) \begin{pmatrix} s_1 & s_2 & \dots & s_n \\ s_1 & s_2 & \dots & s_n \\ \vdots & \vdots & & \vdots \\ s_1 & s_2 & \dots & s_n \end{pmatrix} = \left(\sum_{i=1}^n s_1 p_i, \sum_{i=1}^n s_2 p_i, \dots, \sum_{i=1}^n s_n p_i \right) = \\ &= \sum_{i=1}^n p_i (s_1, s_2, \dots, s_n) = (s_1, s_2, \dots, s_n) = \vec{s} \end{aligned}$$

Ce résultat pourtant évident est de grande importance; à partir de n'importe quelle loi initiale, une chaîne régulière converge vers son unique loi stationnaire, éliminant la nécessité d'effectuer des produits entre vecteurs et matrices. Il suffit de trouver la loi stationnaire en utilisant des procédés que nous connaissons bien.

Nous le prouverons pas rigoureusement, mais la loi stationnaire peut aussi être interprétée comme la proportion de temps que la chaîne finira par passer dans chaque état. Intuitivement, il est facile de l'accepter; sur des durées très longues, le système tend vers un état stable et les effets du déséquilibre initial finissent par devenir insignifiants.

Exemple: Google PageRank

Nous terminons avec un exemple d'application des chaînes de Markov régulières employée auparavant par l'un des géants du Web: Google. Pour améliorer les résultats de leur moteur de recherche, les informaticiens de Google ont conçu une méthode de classement global pour estimer la popularité des sites. Il s'agit d'un internaute aléatoire, déjà traité brièvement dans l'exemple de la figure 2.1 tout au début du travail. Nous pouvons donc le reprendre et construire sa matrice de transition A :

$$A = \frac{1}{6} \begin{pmatrix} 0 & 2 & 2 & 2 \\ 3 & 0 & 3 & 0 \\ 3 & 0 & 0 & 3 \\ 0 & 0 & 0 & 6 \end{pmatrix}$$

La méthode consisterait à examiner la proportion de temps passé dans chaque site en trouvant le vecteur stationnaire de A . Malheureusement, l'état 4 est absorbant et donc la chaîne n'est pas régulière. Pour combler cette difficulté, les ingénieurs de PageRank ont décidé qu'à chaque transition, il y a une probabilité choisie préalablement que l'internaute "s'ennuie" et choisit un site au hasard de manière équiprobable.

Plus précisément, il nous faut considérer une nouvelle chaîne de Markov avec la matrice de transition $G = \alpha A + (1 - \alpha)R$, où α est un constant compris entre 0 et 1 et R est une matrice dont tous les éléments sont donnés par $\frac{1}{n}$, ici $\frac{1}{4}$. La matrice stochastique G , que l'on a nommé **matrice de Google**, est alors régulière tant que $\alpha < 1$. Néanmoins, il vaut mieux donner à α une valeur assez proche de 1 afin de garder la matrice G relativement semblable à A . Il paraît que la valeur utilisée typiquement est $0.85 = \frac{17}{20}$ et donc c'est celle que nous utiliserons. Nous pouvons alors calculer G :

$$G = \frac{17}{20}A + \frac{3}{20}R = \frac{17}{120} \begin{pmatrix} 0 & 2 & 2 & 2 \\ 3 & 0 & 3 & 0 \\ 3 & 0 & 0 & 3 \\ 0 & 0 & 0 & 6 \end{pmatrix} + \frac{3}{80} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} =$$

$$\frac{1}{240} \begin{pmatrix} 9 & 77 & 77 & 77 \\ 111 & 9 & 111 & 9 \\ 111 & 9 & 9 & 111 \\ 9 & 9 & 9 & 213 \end{pmatrix}.$$

Pour avoir une estimation relativement bonne de la popularité de chaque site, il suffit alors de trouver le vecteur stationnaire $\vec{s} = (s_1, s_2, s_3, s_4)$ de G . La relation

$$\vec{s}(G - I) = \frac{1}{240}(s_1, s_2, s_3, s_4) \begin{pmatrix} -231 & 77 & 77 & 77 \\ 111 & -231 & 111 & 9 \\ 111 & 9 & -231 & 111 \\ 9 & 9 & 9 & -27 \end{pmatrix} = \vec{0}$$

nous fournit un système d'équations linéaires à 4 inconnus:

$$\begin{cases} -231s_1 + 111s_2 + 111s_3 + 9s_4 = 0 \\ 77s_1 - 231s_2 + 9s_3 + 9s_4 = 0 \\ 77s_1 + 111s_2 - 231s_3 + 9s_4 = 0 \\ 77s_1 + 9s_2 + 111s_3 - 27s_4 = 0 \end{cases}.$$

Nous n'entrerons pas dans les détails de la résolution de ce système qui est passablement laborieuse et donnons simplement les composantes de \vec{s} :

- $s_1 = \frac{29241}{271868} \approx 0,108$
- $s_2 = \frac{4620}{67967} \approx 0,068$
- $s_3 = \frac{13167}{135934} \approx 0,097$
- $s_4 = \frac{197813}{271868} \approx 0,728$.

Nous voyons alors que la chaîne passe la majorité de temps sur le site 4 qui est référé par 2 autres sites et qui ne présente aucun lien vers les autres. Les sites 1 et 3, bien qu'ils sont eux aussi référés par 2 pages, ne peuvent pas rivaliser le 4. En tant que propriétaire d'un site Web, il paraît donc rentable de supprimer tous les liens vers les pages qui ne répondent pas avec un lien de leur part. Cela s'est en effet produit dans la vie réelle et n'est pas le seul moyen de tromper le système. Par exemple, il est aussi possible d'augmenter le PageRank d'un site Web en créant des sites dont le but est d'établir autant de liens que possible avec le site authentique.

A cause de tels abus, Google ne publie plus les données du PageRank depuis 2016, mais continue probablement de les utiliser en parallèle avec d'autres algorithmes. Cependant, en contraste avec notre petit réseau de 5 pages, le géant du Web doit continuellement effectuer ces calculs sur des réseaux comportant des centaines de millions, voire des milliards de sites!

6 Conclusion

Nous arrivons enfin au bout de ce travail où nous n'avons toutefois que commencé à explorer le monde divers des chaînes de Markov. En effet, s'il y a un aspect des chaînes de Markov que nous voulons mettre en valeur, la chose à retenir de tout ce texte, c'est leur diversité et leur polyvalence.

Rien qu'en étudiant superficiellement le cas restreint des chaînes homogènes et finies, nous avons employé dans la partie théorique du travail des notions de probabilité, d'algèbre linéaire et même un peu de la théorie des nombres. Si nous décidions d'élargir le cadre de ce travail et d'approfondir les recherches, bien d'autres branches des mathématiques s'ajouteraient à cette liste.

Il faut aussi remarquer la variété des domaines auxquels les chaînes de Markov peuvent être appliquées. Ainsi, nous avons traité, bien qu'à titre d'exemple et de manière assez ou très simplifiée, des problèmes de génétique et d'informatique, dont le développement a contribué à une hausse de popularité des chaînes de Markov dans les années récentes.

En vue de leur importance de plus en plus grande, se pourrait-il alors que les chaînes de Markov figureront un jour dans le program des gymnases comme le Lycée Denis-de-Rougemont? Probablement pas, mais même si cela devrait arriver, il resterait encore bien assez de matière pour les futurs rédacteurs de travail de maturité sur

les chaînes de Markov.

Enfin, pour terminer ce travail, nous avons inclus une phrase générée à partir des mots de ce travail par une application Web employant les chaînes de Markov¹:

*CHAINES HOMOGENES ET GRACE AUXQUELS NOUS N'IMPORTE QUEL
ETAT QUI*

¹<https://www.dcode.fr/chaine-markov-texte>

7 Bibliographie

- BERGLUND Nils, *Chaînes de Markov*
<https://www.univ-orleans.fr/mapmo/membres/berglund/markov.pdf>,
consultée le 21. février 2018, cours de l'université d'Orléans.
- LIPSCHUTZ Seymour, *Probabilités: cours et problèmes*, New York, McGraw-Hill Education, 1968.
- RUCH Jean-Jacques *et al.*, *Chaînes de Markov*,
<https://www.math.u-bordeaux.fr/~mchabano/Agreg/ProbaAgreg1314-COURS5-CM.pdf>, consultée le 22. mars 2018, cours de l'Institut de Mathématiques de Bordeaux.
- WATTENHOFER Roger, *Markov Chains & PageRank*,
<https://disco.ethz.ch/courses/fs16/ti2/lecture/chapter11.pdf>,
consultée le 24. mars 2018, cours de l'Ecole Polytechnique Fédérale de Zürich
- WEBER Richard, *Markov Chains*,
<http://www.statslab.cam.ac.uk/~rrw1/markov/M.pdf>,
consultée le 27. mars 2018, cours de l'Université de Cambridge.
- Massachusetts Institute of Technology, *Fundamentals of probability. 6.436/15.085*,
https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-436j-fundamentals-of-probability-fall-2008/lecture-notes/MIT6_436JF08_lec24.pdf,
consultée le 14. mars 2018, cours de l'Institut de technologie du Massachusetts
- Dice, <https://www.britannica.com/topic/dice>, consultée le 24. mars 2018
- Classification of States,
https://www.probabilitycourse.com/chapter11/11_2_4_classification_of_states.php,
consultée le 27. mars 2018
- Markov chain - Wikipedia, https://en.wikipedia.org/wiki/Markov_chain,
consultée le 27. mars 2018.
- Rank (linear algebra) - Wikipedia, [https://en.wikipedia.org/wiki/Rank_\(linear_algebra\)](https://en.wikipedia.org/wiki/Rank_(linear_algebra)),
consultée le 22. mars 2018.
- Chaîne de Markov - Wikipédia, https://fr.wikipedia.org/wiki/Chaîne_de_Markov,
consultée le 27. mars 2018.
- Probabilité - Wikipédia, <https://fr.wikipedia.org/wiki/Probabilité#Historique>,
consultée le 27. mars 2018