CHAÎNES DE MARKOV

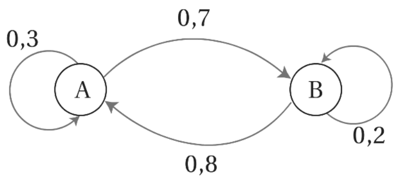
Les chaînes de Markov constituent un des exemples les plus simples de suites de variables aléatoires.

DÉFINITIONS

• Un ***graphe pondéré*** est un graphe dans lequel chaque arête est affectée d'un nombre réel positif appelé ***poids*** de cette arête.

• Un ***graphe probabiliste*** est un graphe orienté pondéré par des réels compris entre 0 et 1 et dans lequel la somme des poids des arêtes issues de chaque sommet est égale à 1.

graphe probabiliste ⟶



DÉFINITIONS

• Soit (X_n) une suite de variables aléatoires à valeurs dans un ensemble $\{a_1; a_2; \ldots; a_N\}$, noté E.

La suite (X_n) est une ***chaîne de Markov à N états*** lorsque :

▼ $\forall k \in \mathbb{N}, \forall (x_0; x_1; \ldots; x_{k+1}) \in E^{k+2}$, $p_{\{X_0=x_0\}\cap\{X_1=x_1\}\cap\ldots\{X_k=x_k\}}(X_{k+1}=x_{k+1}) = p_{X_k=x_k}(X_{k+1}=x_{k+1})$.

Cette probabilité est appelée ***probabilité de transition*** de l'état x_k à l'état x_{k+1}.

Les X_n sont à valeurs dans E : cet ensemble est appelé ***espace des états***.

• On dit qu'une chaîne de Markov (X_n) est ***homogène*** si la probabilité de passer de l'état a_i à l'état a_j ne dépend pas de l'instant k : $p_{X_k=a_i}(X_{k+1}=a_j) = p_{X_0=a_i}(X_1=a_j)$.

La probabilité de transition de l'état a_i à l'état a_j est alors : $p_{X_k=a_i}(X_{k+1}=a_j) = p_{X_0=a_i}(X_1=a_j)$.

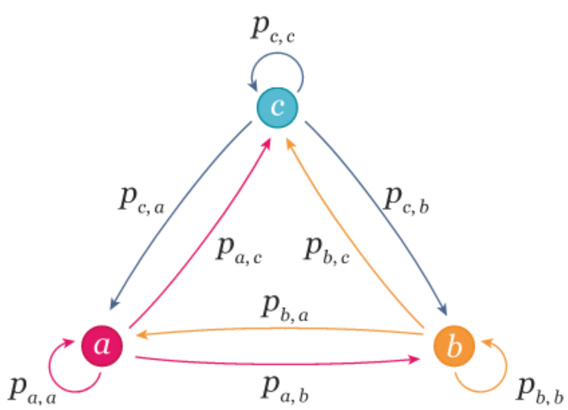
Autrement dit : l'information utile pour la « prédiction » du futur est entièrement contenue dans l'état présent du processus et n'est pas dépendante des états antérieurs (on dit parfois que le système est « sans mémoire »).

<u>Remarques</u> :

• il faudrait supposer que les événements par lesquels on conditionne ne sont pas de probabilités nulles, ce qui sera le cas en pratique ;

• les variables aléatoires (X_n) ne sont pas nécessairement à valeurs réelles ;

• on désigne aussi parfois une chaîne de Markov homogène par l'expression « ***marche aléatoire*** ».



← par exemple, sur ce graphe d'une chaîne de Markov à trois états, on a :

$$p_{X_k=c}(X_{k+1}=b) = p_{c,b}.$$

Andreï Andreïevitch Markov[1] (1856-1922) était un mathématicien russe →

Markov a publié les premiers résultats sur les chaînes de Markov à espace d'états fini en 1906. Une généralisation à un espace d'états infini dénombrable a été publiée par Kolmogorov en 1936. Les processus de Markov sont liés au *mouvement brownien* et à l'*hypothèse ergodique*, deux sujets de physique statistique qui ont été très importants au début du XX^e^ siècle.

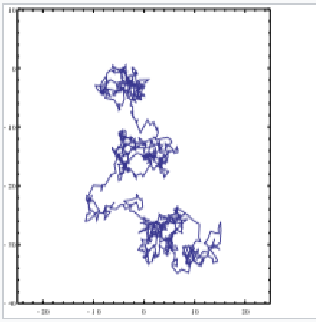
Extraits des pages Wikipédia sur le mouvement brownien et l'hypothèse ergodique :

Le **mouvement brownien**, ou processus de Wiener, est une description mathématique du mouvement aléatoire d'une « grosse » particule immergée dans un fluide et qui n'est soumise à aucune autre interaction que des chocs avec les « petites » molécules du fluide environnant. Il en résulte un mouvement très irrégulier de la grosse particule, qui a été décrit pour la première fois en 1827 par le botaniste Robert Brown en observant des mouvements de particules à l'intérieur de grains de pollen de *Clarkia pulchella* (une espèce de fleur sauvage nord-américaine), puis de diverses autres plantes[1].

La description physique la plus élémentaire du phénomène est la suivante :

- entre deux chocs, la grosse particule se déplace en ligne droite avec une vitesse constante ;
- la grosse particule est accélérée lorsqu'elle rencontre une molécule de fluide ou une paroi.

Ce mouvement permet de décrire avec succès le comportement thermodynamique des gaz (théorie cinétique des gaz), ainsi que le phénomène de diffusion. Il est aussi très utilisé dans des modèles de mathématiques financières.



Mouvement brownien d'une particule.

L'**hypothèse ergodique**, ou **hypothèse d'ergodicité**, est une hypothèse fondamentale de la physique statistique. Elle fut formulée initialement par Ludwig Boltzmann en 1871 pour les besoins de sa théorie cinétique des gaz. Elle s'appliquait alors aux systèmes composés d'un très grand nombre de particules, et affirmait qu'à l'équilibre, la valeur moyenne d'une grandeur calculée de manière statistique est égale à la moyenne d'un très grand nombre de mesures prises dans le temps. La première valeur est celle que permet de calculer la physique statistique, la seconde est proche de ce qu'on peut expérimentalement mesurer. L'hypothèse ergodique est donc fondamentale pour un bon rapprochement entre la théorie et l'expérience.

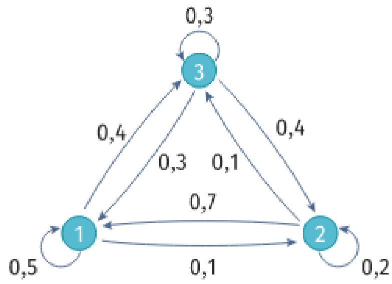
Un système pour lequel l'hypothèse ergodique est vérifiée sera qualifié de *système ergodique*. Dans la plupart des cas, il est très difficile de démontrer rigoureusement qu'un système est ergodique ou non. L'analyse mathématique de ce problème a donné naissance à la théorie ergodique qui précise la nature mathématique de l'hypothèse et donne des résultats sur ses conditions de validité. Mais l'hypothèse ergodique reste souvent une simple hypothèse, jugée vraisemblable *a posteriori* quand elle permet de faire des prédictions correctes. En ce sens, elle constitue un point faible de la physique statistique.

L'hypothèse d'ergodicité intervient également en traitement du signal, où elle consiste à admettre que l'évolution d'un signal aléatoire au cours du temps apporte la même information qu'un ensemble de réalisations. Elle est importante dans l'étude des chaînes de Markov, les processus stationnaires et pour l'apprentissage automatique.

DÉFINITIONS

On considère une chaîne de Markov (X_n) à N états et on note $E=\{1;...;N\}$ l'espace des états.

Cette chaîne de Markov est dite ***homogène*** si pour tout (i,j) de $E\times E$, la probabilité $p_{X_k=i}(X_{k+1}=j)$ ne dépend pas de *k*. La ***matrice de transition*** $(m_{i,j})$ associée à cette chaîne de Markov est la matrice carrée d'ordre N telle que, pour tout (i,j) de $E\times E$, le coefficient $m_{i,j}$ correspond à la probabilité de transition de l'état *i* vers l'état *j*.



← par exemple, la matrice de transition associée à la chaîne de Markov dont le graphe est ci-contre est :

$$\begin{pmatrix} 0,5 & 0,1 & 0,4 \\ 0,7 & 0,2 & 0,1 \\ 0,3 & 0,4 & 0,3 \end{pmatrix}.$$

1 Anecdote : il est cité par le rappeur Freeze Corleone dans son morceau *Tarkov* sur l'album LMF : « J'suis là pour les sous, à propos des chiffres, s/o Andreï Markov » (« s/o » veut dire « shout out » soit "dédicace" en anglais).

Par les définitions précédentes, on a :

> **PROPRIÉTÉ**
>
> La somme des coefficients d'une ligne donnée de la matrice de transition associée à une chaîne de Markov est égale à 1.

C'est pourquoi on définit souvent une chaîne de Markov par un graphe probabiliste.

> **PROPRIÉTÉ**
>
> Soit (X_k) une chaîne de Markov homogène dont l'espace des états est $\{1;2;...;N\}$.
>
> On note M sa matrice de transition. Soit $n \in \mathbb{N}^*$.
>
> La probabilité $p_{X_0=i}(X_n=j)$ est égale au coefficient de la matrice M^n d'indice (i,j).

Autrement dit, le coefficient d'indice (i,j) de la matrice M^n correspond à la probabilité de passer de l'état i à l'état j en n transitions.

Démonstration : par récurrence sur n.

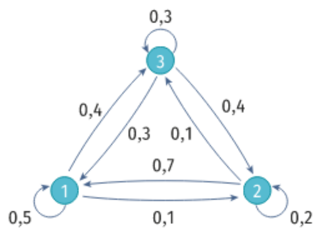
Remarquons d'abord[2] que si A, B et C sont trois événements : $p_A(B\cap C)=p_A(B)\times p_{A\cap B}(C)$.

• Initialisation : $M^1=M$ et par définition, $p_{X_0=i}(X_1=j)=(M)_{i,j}$.

• Hérédité : soit $n \in \mathbb{N}^*$. Supposons que $p_{X_0=0}(X_n=j)=(M^n)_{i,j}$.

$$
\begin{aligned}
p_{X_0=i}(X_{n+1}=j) &= p_{X_0=i}\left(\bigcup_{k=1}^{N}(\{X_n=k\}\cap\{X_{n+1}=j\})\right)\\
&=\sum_{k=1}^{N} p_{X_0=i}(\{X_n=k\}\cap\{X_{n+1}=j\}) && \text{voir ci-dessus}\\
&=\sum_{k=1}^{N} p_{X_0=i}(X_n=k)\times p_{\{X_0=i\}\cap\{X_n=k\}}(X_{n+1}=j) && \text{conséquence* de la définition d'une chaîne de Markov}\\
&=\sum_{k=1}^{N} (M^n)_{i,k}\times p_{X_n=k}(X_{n+1}=j) && \text{+ hypothèse de récurrence} \quad \text{* admise}\\
&=\sum_{k=1}^{N} (M^n)_{i,k}\times p_{X_0=k}(X_1=j) && \text{chaîne de Markov homogène}\\
&=\sum_{k=1}^{N} (M^n)_{i,k}\times (M)_{k,j}\\
&=(M^n M)_{i,j} && \text{produit matriciel}\\
&=(M^{n+1})_{i,j}
\end{aligned}
$$

EXEMPLE C1



On considère une marche aléatoire (X_n) à trois états, modélisée par le graphe probabiliste suivant. On note M la matrice de transition associée. Déterminer une valeur approchée à 10^{-2} de la probabilité de passer de l'état 1 à l'état 2 en 5 étapes.

2 Très facile à démontrer.

DÉFINITIONS

Soit (X_n) une chaîne de Markov dont l'espace des états est $\{1;2;...;N\}$.

On appelle ***distribution initiale*** de cette chaîne de Markov la loi de probabilité de X_0, que l'on note souvent π_0 : $\pi_0 = \begin{pmatrix} p(X_0=1) & p(X_0=2) & \dots & p(X_0=N) \end{pmatrix}$.

On appelle ***distributions*** de la chaîne de Markov les matrices lignes π_k donnant la loi des variables aléatoires X_k : $\pi_k = \begin{pmatrix} p(X_k=1) & p(X_k=2) & \dots & p(X_k=N) \end{pmatrix}$.

THÉORÈME

Soit (X_n) une chaîne de Markov dont l'espace des états est noté $\{1;2;...;N\}$.

On note M sa matrice de transition. Les distributions π_k de cette chaîne de Markov vérifient la relation de récurrence $\pi_{k+1} = \pi_k M$ et on a alors : $\forall n \in \mathbb{N},\ \pi_n = \pi_0 M^n$.

<u>Démonstration</u> : • Soit $j \in \{1;2;...;N\}$. D'après la formule des probabilités totales :

$$\begin{aligned}(\pi_{k+1})_{1,j} = p(X_{k+1}=j) = \sum_{i=1}^{N} p(\{X_{k+1}=j\} \cap \{X_k=i\}) &= \sum_{i=1}^{N} p(X_k=i) \times p_{X_k=i}(X_{k+1}=j) \\ &= \sum_{i=1}^{N} (\pi_k)_{1,i} \times M_{i,j} \\ &= (\pi_k M)_{1,j}.\end{aligned}$$

Conclusion : $\pi_{k+1} = \pi_k M$.

• Par récurrence sur n, on arrive alors facilement à : $\pi_n = \pi_0 M^n$.

EXEMPLE C2

On considère la marche aléatoire (X_n) à trois états de l'exemple C1.

1. On part de l'état 3 : $p(X_0=3)=1$. Déterminer la distribution π_5 donnant la loi de X_5.

2. En déduire une valeur approchée à 10^{-2} de la probabilité d'être à l'état 2 à l'étape 5.

3. Compléter le tableau suivant (arrondir à 10^{-2}) :

k	0	2	4	6	8	10	12	14	16
$p(X_k=1)$									
$p(X_k=2)$									
$p(X_k=3)$									

DÉFINITION

Soit (X_n) une chaîne de Markov homogène à N états, de matrice de transition M.

Une matrice ligne π (dont la somme des coefficients est égale à 1) est une ***distribution invariante*** si : $\pi M = \pi$. On dit aussi que la chaîne de Markov admet alors un ***état stable***.

<u>REMARQUE</u> : si $(M - I_n)$ est inversible, alors la chaîne de Markov n'admet pas de distribution invariante, puisqu'alors π serait la matrice nulle.

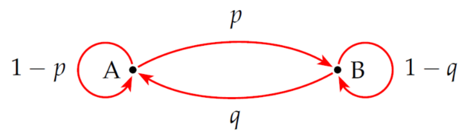
THÉORÈMES (ADMIS)

Soit (X_n) une chaîne de Markov et (π_n) la suite de ses distributions.

On note M sa matrice de transition.

- Si (π_n) est convergente, alors elle converge vers une distribution invariante.
- Si (X_n) est à deux états (avec des coefficients dans $]0;1[$), alors (π_n) converge.
- S'il existe un entier n tel que M^n ne contient aucun 0 (sauf éventuellement sur sa diagonale principale), alors (π_k) converge vers l'unique distribution invariante.

<u>Démonstration du deuxième point</u> :



Soient p et q deux réels de $]0;1[$. $M = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$

1. Démontrer que l'unique distribution invariante est $\begin{pmatrix} \dfrac{q}{p+q} & \dfrac{p}{p+q} \end{pmatrix}$.

2. Démontrer qu'en notant $\pi_n = \begin{pmatrix} x_n & 1-x_n \end{pmatrix}$, on a : $x_{n+1} = (1-p-q)x_n + q$.

3. En déduire (étude d'une suite arithmético-géométrique) que : $\left(x_n - \dfrac{q}{p+q}\right)$ est une suite géométrique de raison $1-p-q$.

4. En déduire que cette suite converge vers 0, puis que (x_n) converge. Conclure.

• Fiche bilan → p.227

• QCM 6 questions corrigées → p.228

• Exercices corrigés → 86 à 88 p.229

• Exercices types corrigés → méthodes 7 et 8 p.216/217

EXEMPLE C3

D'après Bac S, juin 2016 (Amérique du Nord)

On dispose de deux urnes U et V contenant chacune deux boules. Au départ, l'urne U contient deux boules blanches et l'urne V contient deux boules noires.
On effectue des tirages successifs dans ces urnes de la façon suivante : chaque tirage consiste à prendre au hasard, de manière simultanée, une boule dans chaque urne et à la mettre dans l'autre urne.
Pour tout entier naturel n non nul, on note X_n la variable aléatoire égale au nombre de boules blanches que contient l'urne U à la fin du n-ième tirage.

1. **a.** Traduire par une phrase la probabilité $P_{(X_n=1)}(X_{n+1}=1)$ puis déterminer les probabilités conditionnelles suivantes :

$$P_{(X_n=0)}(X_{n+1}=1)\,,\, P_{(X_n=1)}(X_{n+1}=1) \text{ et } P_{(X_n=2)}(X_{n+1}=1).$$

   b. Exprimer $P(X_{n+1}=1)$ en fonction de $P(X_n=0)$, $P(X_n=1)$ et $P(X_n=2)$.

2. Pour tout entier naturel n non nul, on note R_n la matrice ligne définie par :

$$R_n = \begin{pmatrix} P(X_n=0) & P(X_n=1) & P(X_n=2) \end{pmatrix}$$

   et on considère M la matrice $\begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 1 & 0 \end{pmatrix}$.

   On note R_0 la matrice ligne $\begin{pmatrix} 0 & 0 & 1 \end{pmatrix}$.
   On admettra par la suite que, pour tout entier naturel n, $R_{n+1} = R_n \times M$.
   Déterminer R_1 et justifier que, pour tout entier naturel n, $R_n = R_0 \times M^n$.

3. On admet que $M = P \times D \times P^{-1}$ avec :

$$P = \frac{1}{6}\begin{pmatrix} 2 & 3 & 1 \\ -1 & 0 & 1 \\ 2 & -3 & 1 \end{pmatrix},\ D = \begin{pmatrix} -\frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ et } P^{-1} = \begin{pmatrix} 1 & -2 & 1 \\ 1 & 0 & -1 \\ 1 & 4 & 1 \end{pmatrix}.$$

   Établir que, pour tout entier naturel n, $M^n = P \times D^n \times P^{-1}$.

   On admettra que, pour tout entier naturel n, $D^n = \begin{pmatrix} \left(-\frac{1}{2}\right)^n & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

4. **a.** Calculer $D^n \times P^{-1}$ en fonction de n.

   b. Sachant que $R_0 P = \begin{pmatrix} \frac{1}{3} & -\frac{1}{2} & \frac{1}{6} \end{pmatrix}$, déterminer les coefficients de R_n en fonction de n.

5. Déterminer $\lim\limits_{n\to+\infty} P(X_n=0)$, $\lim\limits_{n\to+\infty} P(X_n=1)$ et $\lim\limits_{n\to+\infty} P(X_n=2)$.
   Interpréter ces résultats.

