

EISTI



Ecole
Internationale
des Sciences
du Traitement
de l'Information

PROBABILITES II

-

CONVERGENCE

ET

INTRODUCTION

AUX CHAINES DE MARKOV

SOMMAIRE

REFERENCES :	2
Chapitre 1 : Transformations de variables aléatoires $X \rightarrow Y=\varphi(X)$	3
1. Transformations	3
1.1. Cas où X et Y sont des variables aléatoires discrètes	3
1.2. Cas où X est une v.a continue et Y est une v.a. discrète	3
1.3. Cas où X et Y sont des variables aléatoires continues	3
2. Espérance d'une variable aléatoire transformée	5
2.1. Cas où X est une variable aléatoire discrète	5
2.2. Cas où X est une variable aléatoire continue	5
Chapitre 2 : Convergence d'une suite de variables aléatoires	6
1. Convergence en loi	6
1.1. Définition	6
1.2. Théorème de la limite centrale	6
1.3. Applications du théorème de la limite centrale	6
2. Convergence en probabilités	7
2.1. Définition	7
2.2. Loi faible des grands nombres	7
3. Convergence presque sûre	8
3.1. Définition	8
3.2. Loi forte des grands nombres	8
4. Convergence en moyenne quadratique	8
5. Liens entre les différents types de convergence	9
6. Quelques inégalités	9
Chapitre 3 : Chaînes de Markov à temps discret	10
1. Définition d'une chaîne de Markov à temps discret	10
1. Matrice de transition	11
2. Loi de probabilité d'état	12
2.1. Loi de probabilité à l'instant n	12
2.2. Probabilité de transition en n étapes	12
2.3. Graphe d'une chaîne de Markov	13
3. Classification des états	14
3.1. Propriétés des états	14
3.2. Regroupement des états par classes d'équivalence	15
4. Comportement asymptotique	16
4.1. Définitions et propriétés	16
4.2. Comportement asymptotique	17
5. Délais et probabilité d'atteinte	18
5.1. Délais et probabilité d'absorption	18
5.2. Délais et probabilité d'atteinte	19
ANNEXE : Matrices stochastiques	21

REFERENCES :

Lefebvre M. (2005). *Processus stochastiques appliqués*. Ed. Hermann

Ruegg A. (1989). *Processus stochastiques*. Ed. Presses polytechniques et universitaires romandes

CHAPITRE 1 : TRANSFORMATIONS DE VARIABLES ALEATOIRES $X \rightarrow Y = \varphi(X)$

1. TRANSFORMATIONS

Soit (Ω, \mathcal{A}, P) un espace probabilisé. Soient X une variable aléatoire sur (Ω, \mathcal{A}) avec $X(\Omega) \subset \mathbb{R}$ et φ une application de \mathbb{R} dans \mathbb{R} . On pose alors $Y = \varphi(X)$ et on suppose que Y est une variable aléatoire.

1.1. Cas où X et Y sont des variables aléatoires discrètes

Si X est une variable aléatoire discrète de support D_X et de fonction de masse p_X , alors Y est une variable aléatoire discrète de support $D_Y = \varphi(D_X)$ et de fonction de masse

$$p_Y(y) = \begin{cases} \sum_{x \in \{x, \varphi(x)=y\} \cap D_X} p_X(x) & \text{si } y \in D_Y \\ 0 & \text{sinon} \end{cases}.$$

1.2. Cas où X est une v.a continue et Y est une v.a. discrète

Si X est une variable aléatoire continue de support D_X et de fonction de densité f_X . Si de plus $\varphi(D_X)$ est un ensemble discret, alors Y est une variable aléatoire discrète de support $D_Y = \varphi(D_X)$ et de fonction de masse

$$p_Y(y) = \begin{cases} \int_{x \in \{x, \varphi(x)=y\} \cap D_X} f_X(x) dx & \text{si } y \in D_Y \\ 0 & \text{sinon} \end{cases}.$$

1.3. Cas où X et Y sont des variables aléatoires continues

Supposons que X soit une variable aléatoire continue de support D_X et de fonction de densité f_X et de fonction de répartition F_X . On suppose de plus que l'application φ est dérivable. On cherche f_Y et F_Y les fonctions de densité et répartition de la variable $Y = \varphi(X)$ de support $D_Y = \varphi(D_X)$.

Cas où φ est strictement monotone

Si φ est strictement monotone, elle est bijective, donc φ^{-1} existe.

- Si φ est strictement croissante, on a $F_X(t) = P(X < t) = P(Y < \varphi(t)) = F_Y(\varphi(t))$, d'où

$$F_Y(t) = F_X(\varphi^{-1}(t)).$$

En dérivant $F_X(x) = F_Y(\varphi(x))$ de chaque coté, on obtient

$f_X(x) = \varphi'(x) \times f_Y(\varphi(x))$, soit

$$f_Y(y) = \frac{f_X(x)}{\varphi'(x)} = \frac{f_X(\varphi^{-1}(y))}{\varphi'(\varphi^{-1}(y))}.$$

- Si φ est strictement décroissante, on a $F_X(x) = P(X < x) = P(Y > \varphi(x)) = 1 - F_Y(\varphi(x))$, d'où

$$F_Y(x) = 1 - F_X(\varphi^{-1}(x)).$$

En dérivant $F_X(t) = 1 - F_Y(\varphi(x))$ de chaque coté, on obtient

$f_X(x) = -\varphi'(x) \times f_Y(\varphi(x))$, soit

$$f_Y(y) = -\frac{f_X(x)}{\varphi'(x)} = -\frac{f_X(\varphi^{-1}(y))}{\varphi'(\varphi^{-1}(y))}.$$

On obtient donc la forme générale,

$$f_Y(y) = \begin{cases} \frac{f_X(\varphi^{-1}(y))}{|\varphi'(\varphi^{-1}(y))|} & \text{si } y \in D_Y \\ 0 & \text{sinon} \end{cases}.$$

Exemple

Soit φ la fonction exponentielle, donc φ est strictement croissante et φ^{-1} est la fonction logarithme. Alors la variable aléatoire $Y = \varphi(X)$ a son support inclus dans \mathbb{R}_+ et pour fonction de densité,

$$f_Y(y) = \frac{f_X(\ln(y))}{\exp(\ln(y))} = \frac{f_X(\ln(y))}{y}$$

Par exemple si X une variable aléatoire continue de loi exponentielle de paramètre 2,

$$f_X(x) = \begin{cases} 2e^{-2x} & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases}$$

alors

$$f_Y(y) = \begin{cases} 2y^{-3} & \text{si } y \geq 1 \\ 0 & \text{si } 0 \leq y < 1 \end{cases}.$$

On peut remarquer que Y est bien une variable aléatoire que $\int_0^{+\infty} f_Y(y) dy = 1$.

Cas où φ est quelconque

Le principe est le même mais il est difficile de le traiter de façon général. On cherche toujours à identifier la fonction de répartition $F_Y(y)$ en cherchant l'antécédent pour X de l'événement $Y < y = \varphi(x)$.

Par exemple, si $Y = X^2$ avec X définie sur \mathbb{R} , alors

$$F_Y(y) = P(Y < y) = P(X^2 < y) = P(-\sqrt{y} < X < \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$$

En dérivant, on obtient

$$f_Y(y) = \frac{1}{2\sqrt{y}} (f_X(\sqrt{y}) + f_X(-\sqrt{y})).$$

2. ESPERANCE D'UNE VARIABLE ALEATOIRE TRANSFORMEE

2.1. Cas où X est une variable aléatoire discrète

Si X est une variable aléatoire discrète de support D_X et de fonction de masse p_X , alors

$$E[\varphi(X)] = \sum_{x \in D_X} \varphi(x)p_X(x).$$

2.2. Cas où X est une variable aléatoire continue

Si X est une variable aléatoire continue de support D_X et de fonction de densité f_X , alors

$$E[\varphi(X)] = \int_{D_X} \varphi(x)f_X(x)dx.$$

Remarque : Dans les deux cas, il faut s'assurer de la convergence.

CHAPITRE 2 : CONVERGENCE D'UNE SUITE DE VARIABLES ALEATOIRES

1. CONVERGENCE EN LOI

1.1. Définition

Soient l'espace probabilisé (Ω, \mathcal{A}, P) et une suite de variables aléatoires X_1, \dots, X_n sur (Ω, \mathcal{A}) . Notons F_1, \dots, F_n les fonctions de répartition correspondantes. On dit que la suite $\{X_1, \dots, X_n\}$ converge *en loi* vers une variable aléatoire X de fonction de répartition F si

$$\lim_{n \rightarrow +\infty} F_n(t) = F(t)$$

pour tout t appartenant au domaine de continuité de F . Ce type de convergence est noté

$$X_n \xrightarrow{L} X \text{ ou bien } \lim_{n \rightarrow +\infty} X_n = X (L).$$

Exemple : Nous avons vu au chapitre 7 que la loi binomiale converge en loi vers une loi de Poisson.

1.2. Théorème de la limite centrale

Théorème (théorème de la limite centrale)

Soit X_1, \dots, X_n une suite de variables aléatoires indépendantes de même loi telle que $E(X_i) = \mu$ et $\text{var}(X_i) = \sigma^2$. Alors la variable aléatoire

$$Y_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{X_i - \mu}{\sigma},$$

converge en loi vers une variable aléatoire Y de la normale centrée réduite,

$$\begin{aligned} Y_n &\xrightarrow{L} N(0,1) \\ \Leftrightarrow \frac{1}{n} \sum_{i=1}^n X_i &\xrightarrow{L} N\left(\mu, \frac{\sigma^2}{n}\right) \end{aligned}$$

Ce théorème joue un rôle capital en statistique car l'étude de somme de variables aléatoires indépendantes est primordial.

1.3. Applications du théorème de la limite centrale

Convergence de la loi binomiale vers la loi normale

Soit X_1, \dots, X_n une suite de variables aléatoires indépendantes de loi binomiale $b(n, p)$, alors

$$\frac{X_n - np}{\sqrt{np(1-p)}} \xrightarrow{L} N(0,1).$$

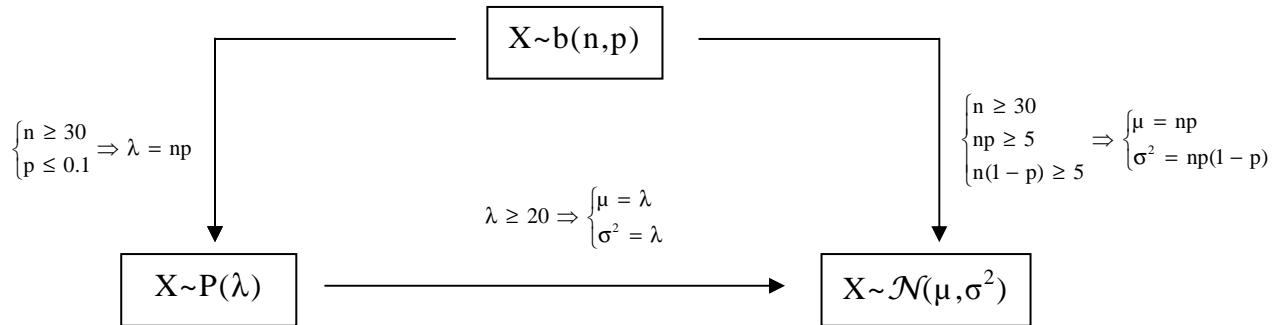
En pratique, on utilise cette approximation lorsque $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$.

Convergence de la loi de Poisson vers la loi normale

Soit X_1, \dots, X_n une suite de variables aléatoires indépendantes de loi de Poisson $P(\lambda)$, alors

$$\frac{X_n - \lambda}{\sqrt{\lambda}} \xrightarrow{L} \mathcal{N}(0,1).$$

En pratique, on utilise cette approximation lorsque $\lambda \geq 20$.



2. CONVERGENCE EN PROBABILITES

2.1. Définition

Soit l'espace probabilisé (Ω, \mathcal{A}, P) . Une suite de variables aléatoires X_1, \dots, X_n sur (Ω, \mathcal{A}) converge *en probabilités* vers une variable aléatoire X si

$$\forall \varepsilon > 0, \lim_{n \rightarrow +\infty} P[|X_n - X| \geq \varepsilon] = 0.$$

Ce type de convergence est noté

$$X_n \xrightarrow{P} X \text{ ou bien } \lim_{n \rightarrow +\infty} X_n = X (P).$$

2.2. Loi faible des grands nombres

Soit X_1, \dots, X_n une suite de variables aléatoires pas nécessairement indépendantes telles que $E(X_i)$ existe. On dit que la suite obéit à la loi faible des grands nombres si la variable aléatoire

$$Y_n = \frac{1}{n} \sum_{i=1}^n (X_i - E(X_i)) = E(\bar{X}) - \frac{1}{n} \sum_{i=1}^n E(X_i),$$

converge en probabilité 0.

Théorème

Si X_1, \dots, X_n une suite de variables aléatoires indépendantes de même loi telle que $E(X_i) = \mu$, alors la suite obéit à la loi faible des grands nombres, *i.e*

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{P} \mu.$$

3. CONVERGENCE PRESQUE SURE

3.1. Définition

Soit l'espace probabilisé $(\Omega, \mathcal{A}, \mathbf{P})$. Une suite de variables aléatoires X_1, \dots, X_n sur (Ω, \mathcal{A}) converge *presque sûrement* vers une variable aléatoire X si

$$\mathbf{P}\left[\lim_{n \rightarrow +\infty} X_n = X\right] = 1.$$

Ce type de convergence est noté

$$X_n \xrightarrow{\text{p.s.}} X \text{ ou bien } \lim_{n \rightarrow +\infty} X_n = X \text{ (p.s.)}.$$

3.2. Loi forte des grands nombres

Soit X_1, \dots, X_n une suite de variables aléatoires pas nécessairement indépendantes telles que $E(X_i)$ et $\text{var}(X_i)$ existent. On dit que la suite obéit à la loi forte des grands nombres si la variable aléatoire

$$Y_n = \frac{1}{n} \sum_{i=1}^n (X_i - E(X_i)) = E(\bar{X}) - \frac{1}{n} \sum_{i=1}^n E(X_i),$$

converge presque sûrement vers 0.

Théorème

Si X_1, \dots, X_n une suite de variables aléatoires indépendantes de même loi telle que $E(X_i) = \mu$ et $\text{var}(X_i) = \sigma^2$, alors la suite obéit à la loi forte des grands nombres, i.e

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\text{p.s.}} \mu.$$

4. CONVERGENCE EN MOYENNE QUADRATIQUE

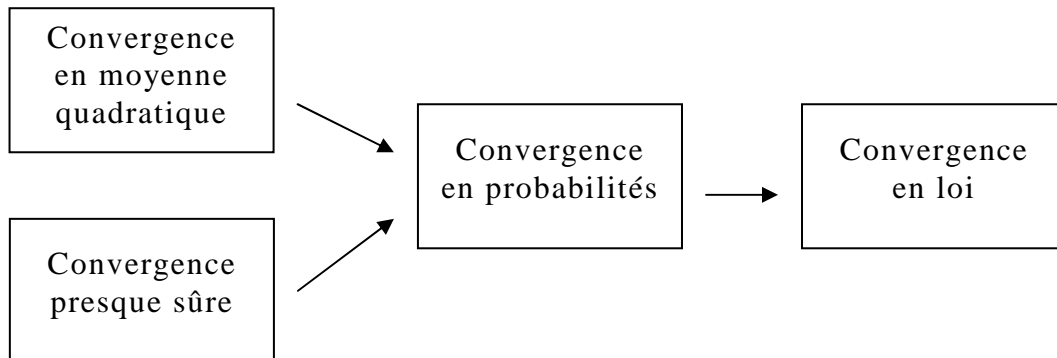
Soit l'espace probabilisé $(\Omega, \mathcal{A}, \mathbf{P})$ et une suite de variables aléatoires X_1, \dots, X_n sur (Ω, \mathcal{A}) telles que $E(X_i)$ et $E(X_i^2)$ existent. On dit que la suite converge *en moyenne quadratique* vers une variable aléatoire X si

$$\lim_{n \rightarrow +\infty} E[(X_n - X)^2] = 0.$$

Ce type de convergence est noté

$$X_n \xrightarrow{\text{m.q.}} X \text{ ou bien } \lim_{n \rightarrow +\infty} X_n = X \text{ (m.q.)}.$$

5. LIENS ENTRE LES DIFFERENTS TYPES DE CONVERGENCE



6. QUELQUES INEGALITES

Théorème : Soit X une v.a. sur $(\Omega, \mathcal{A}, \mathbf{P})$ et g une fonction. Alors pour tout $c > 0$, on a

$$P[g(X) \geq c] \leq \frac{E[g(X)]}{c}$$

- Inégalité de Markov : $g(X) = |X - E(X)|^p$

$$P\left[|X - E(X)|^p \geq c^p\right] \leq \frac{E\left[|X - E(X)|^p\right]}{c^p}$$

- Inégalité de Tchebychev : $c = \delta \sqrt{\text{var}(X)}$

$$P\left[|X - E(X)| \geq \delta \sqrt{\text{var}(X)}\right] \leq \frac{1}{\delta^2}$$

CHAPITRE 3 : CHAINES DE MARKOV A TEMPS DISCRET

Dans ce chapitre, nous allons étudier une suite de variables aléatoires $\{X_n\}_{n \in \mathbb{N}}$ dont la définition est assez élémentaire mais qui permet tout de même une description mathématique de nombreux phénomènes aléatoires rencontrés en pratique.

1. DEFINITION D'UNE CHAÎNE DE MARKOV A TEMPS DISCRET

Soient (Ω, \mathcal{A}, P) un espace probabilisé et (E, \mathcal{B}) un espace mesurable où $E = \{e_i, i \in I\}$ est un ensemble fini ou dénombrable.

Définition

Soit $\{X_n\}_{n \in \mathbb{N}}$ une suite de variables aléatoires définie sur Ω avec \mathbb{N} l'ensemble des temps et E l'espace des états. On dit que $\{X_n\}_{n \in \mathbb{N}}$ est une *chaîne de Markov* si pour tout $(n+1)$ -uplet (e_0, \dots, e_n) d'éléments de E tel que $P(\bigcap_{i=0}^{n-1} [X_i = e_i]) \neq 0$, on a

$$P(X_n = e_n \mid X_0 = e_0, \dots, X_{n-1} = e_{n-1}) = P(X_n = e_n \mid X_{n-1} = e_{n-1}).$$

Interprétation

Cela signifie que la loi conditionnelle de X_n sachant (X_0, \dots, X_{n-1}) est la même que la loi conditionnelle de X_n sachant X_{n-1} . Si on considère que l'indice n représente le temps, alors cette propriété traduit le fait que le futur (n) ne dépend du passé $(0, 1, \dots, n-1)$ qu'à travers le présent ($n-1$).

Exemple 1 (marche aléatoire sur \mathbb{Z})

On pose $E = \mathbb{Z}$. Soit $\{Y_n\}_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles indépendantes de même loi telle que pour tout $n \in \mathbb{N}$, $P[Y_n = 1] = p$ et $P[Y_n = -1] = 1 - p$ où $0 < p < 1$. Pour tout $n \in \mathbb{N}$, on pose $X_n = Y_0 + Y_1 + \dots + Y_n$. Alors $\{X_n\}_{n \in \mathbb{N}}$ est une chaîne de Markov.

Définition

Une chaîne de Markov $\{X_n\}_{n \in \mathbb{N}}$ est dite *homogène* si pour tout $e_j \in E$ et pour tout $e_i \in E$ tel que $P(X_{n-1} = e_i) \neq 0$, la probabilité conditionnelle $P(X_n = e_j \mid X_{n-1} = e_i)$ ne dépend pas de n .

Exemple 2 (Propagation de maladie)

Un individu vit dans un milieu où il est susceptible d'attraper une maladie par piqûre d'insecte. Il peut être dans l'un des trois états suivants : immunisé (1), malade (2), non malade et non immunisé (3). D'un mois à l'autre son état peut changer selon les règles suivantes :

- étant immunisé, il peut le rester avec une probabilité 0,9 ou passer à l'état (3) avec une probabilité 0,1
- étant à l'état (3), il peut le rester avec une probabilité 0,5 ou tomber malade avec une probabilité 0,5
- étant malade, il peut le rester avec une probabilité 0,2 ou devenir immunisé avec une probabilité 0,8.

On remarque que l'état dans lequel sera la personne le mois prochain, ne dépend que de son état actuel mais pas de son passif. D'autre part, les probabilités de transition d'un état à l'autre sont stables au cours des mois. Donc la chaîne décrivant l'état d'une personne est une chaîne de Markov homogène.

1. MATRICE DE TRANSITION

Soit $\{X_n\}_{n \in \mathbb{N}}$ une chaîne de Markov homogène.

Définition

La chaîne étant homogène, on note $p_{ij} = P(X_n = e_j \mid X_{n-1} = e_i)$. Cette probabilité est appelée *probabilité de transition* (de l'état e_i à l'état e_j). On appelle *matrice de transition* de la chaîne $\{X_n\}_{n \in \mathbb{N}}$, la matrice $P=(p_{ij})$.

Remarque

E est un ensemble fini ou dénombrable. Dans le cas où E n'est pas fini, on parle encore de « matrice ».

Exemple 2 (suite)

Pour cet exemple, on obtient la matrice de transition

$$P = \begin{pmatrix} 0.9 & 0 & 0.1 \\ 0.8 & 0.2 & 0 \\ 0 & 0.5 & 0.5 \end{pmatrix}$$

Propriété

La matrice P est une matrice stochastique (cf. annexe).

Exemple 3 (Chaîne d'Ehrenfest *)

Soit $d \in \mathbb{N}^*$, On dispose de d objets numérotés de 1 à d répartis dans deux urnes A et B. On choisit au hasard un entier k dans $\{1, \dots, d\}$ et on change d'urne l'objet numéroté k . On note X_n le nombre d'objet dans l'urne A après n tirages indépendants. $\{X_n\}_{n \in \mathbb{N}}$ est alors un chaîne de Markov homogène à valeurs dans $E = \{0, 1, \dots, d\}$ et on a pour tous $i, j \in \{1, \dots, d+1\}$

$$p_{ij} = P(X_n = j-1 \mid X_{n-1} = i-1) = \begin{cases} 0 & \text{si } |i-j| \geq 2 \\ \frac{d-i+1}{d} & \text{si } j = i+1 \\ \frac{i-1}{d} & \text{si } j = i-1 \end{cases} .$$

La matrice de transition P est de taille $d+1$ définie par

* Ce modèle a été utilisé par Paul et Tatiana Ehrenfest pour étudier le transfert de chaleur entre les molécules dans un gaz.

$$P = \begin{pmatrix} 0 & 1 & 0 & \cdots & \cdots & 0 \\ \frac{1}{d} & 0 & \frac{d-1}{d} & \ddots & & \vdots \\ 0 & \ddots & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & & & \ddots & \frac{d-1}{d} & \frac{1}{d} \\ 0 & \cdots & \cdots & 0 & 1 & 0 \end{pmatrix}.$$

2. LOI DE PROBABILITE D'ETAT

Soit $\{X_n\}_{n \in \mathbb{N}}$ une chaîne de Markov homogène à valeurs dans E et $P=(p_{ij})$ sa matrice de transition.

2.1. Loi de probabilité à l'instant n

Définition

Pour tout entier $n \in \mathbb{N}$ et pour tout $e_i \in E$, $i \in I$, on note la probabilité $\pi_i(n) = P(X_n = e_i)$ et le vecteur ligne $\pi(n) = (\pi_i(n))_{i \in I}$.

Le vecteur $\pi(n)$ définit une probabilité sur (E, \mathcal{B}) appelée *loi à l'instant n* de la chaîne $\{X_n\}_{n \in \mathbb{N}}$.

On appelle *loi initiale* de la chaîne $\{X_n\}_{n \in \mathbb{N}}$ le vecteur $\pi(0)$.

Remarque

$\pi(n)$ est un vecteur de probabilité donc la somme de ses composantes est égale à 1.

Exemple 2 (suite)

Une personne saine arrivant dans ce milieu a une chance sur dix d'être immunisée, c'est-à-dire que $\pi(0) = (0.1, 0.9)$. On calcule alors que $\pi(3) = (0.70, 0.14, 0.16)$. Cela signifie par exemple, qu'après 3 mois la personne a 70% de chance d'être immunisée.

Exemple 3 (suite)

Supposons qu'initialement, il y a un objet dans l'urne A et deux dans l'urne B ($d=3$). Alors $\pi(0) = (0, 1, 0, 0)$ et par exemple $\pi(4) = (0, 0.11, 0.64, 0.25)$. Ce qui signifie entre autre que la probabilité d'avoir aucun objet dans l'urne A après 4 tirages est nulle.

2.2. Probabilité de transition en n étapes

Définition

Pour tout entier $n \in \mathbb{N}$, on définit la matrice des probabilités de transition en n étapes,

$$P^{(n)} = (p_{ij}^{(n)})_{i,j \in I} \text{ où } p_{ij}^{(n)} = P(X_n = e_j \mid X_0 = e_i), \forall e_i, e_j \in E.$$

Propriété

Pour tout entier $n \in \mathbb{N}$, on a

$$\begin{aligned} P^{(n)} &= P^n \\ \pi(n) &= \pi(0)P^n \end{aligned}$$

Remarque (formule de Chapman- Kolmogorov)

La formule matricielle de $P^n = P^m P^{n-m}$ s'écrit

$$p_{ij}^{(n)} = \sum_{k \in I} p_{ik}^{(m)} p_{kj}^{(n-m)},$$

et s'interprète de la façon suivante : aller de e_i à e_j en n pas, c'est aller de e_i à un certain état e_k en m pas, puis de e_k à e_j en $n-m$ pas.

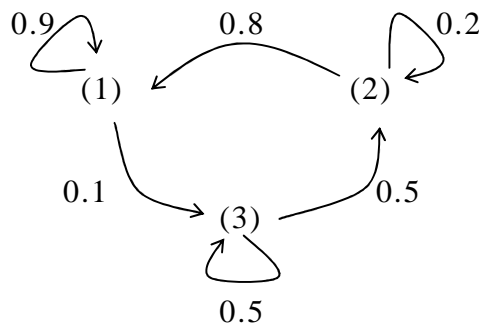
2.3. Graphe d'une chaîne de Markov

La propriété précédente montre qu'une chaîne de Markov homogène est entièrement définie par sa loi initiale $\pi(0)$ et sa matrice de transition P .

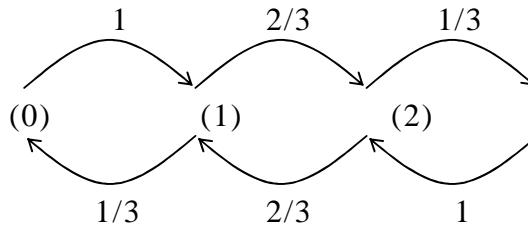
Le graphe associé à la chaîne de Markov $\{X_n\}_{n \in \mathbb{N}}$ est défini par

- ses sommets qui représentent les états du système ($e_i \in E$)
- ses arêtes ($i \rightarrow j$) associées aux probabilités de transition p_{ij}

Exemple 2 (suite)



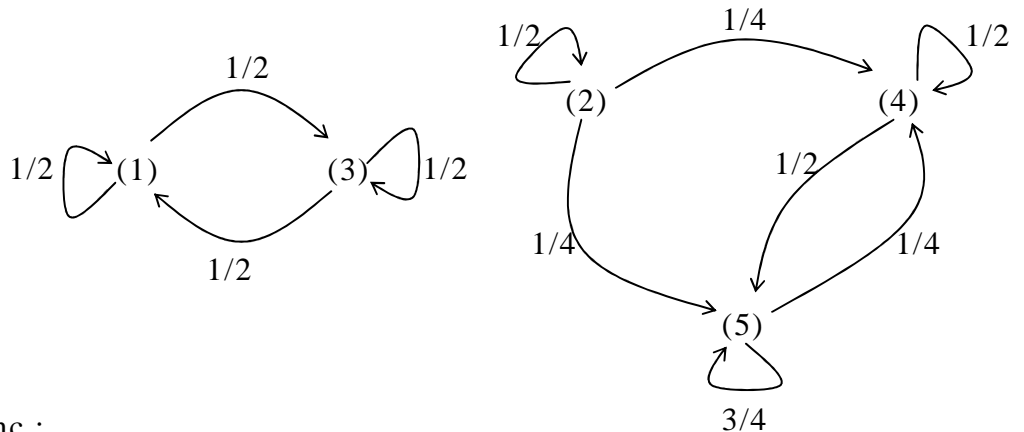
Exemple 3 (suite)



Exemple 4

Soit $\{X_n\}_{n \in \mathbb{N}}$ une chaîne de Markov homogène à valeurs dans $E = \{1, \dots, 5\}$, de matrice de transition

$$P = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{4} & \frac{3}{4} \end{pmatrix}.$$



Un graphe permet donc :

- de voir directement si on peut passer d'un état à l'autre en un temps fini
- d'étudier les propriétés des états du système (cf. § 4)

3. CLASSIFICATION DES ETATS

Soit $\{X_n\}_{n \in \mathbb{N}}$ une chaîne de Markov homogène à valeurs dans E et $P=(p_{ij})$ sa matrice de transition.

3.1. Propriétés des états

Définition

Soit e_i un état de E . On appelle *période* de e_i l'entier d_i défini par le pgcd des longueurs des chemins partant de e_i et arrivant à e_i ,

$$d_i = \text{pgcd}\{n > 0, p_{ii}^{(n)} > 0\}.$$

Si $d_i=1$, on dit que l'état e_i est *apériodique*, sinon l'état e_i est *périodique* de période d_i .

Exemples 2, 3 et 4 (suite)

Pour les exemples 2 et 4, tous les états sont apériodiques puisque pour tout $i \in \{1, \dots, 5\}$, on a $p_{ii}^{(1)} = p_{ii} > 0$. C'est-à-dire que l'on peut revenir sur l'état e_i en une étape. Pour l'exemple 3, tous les états sont de période 2 car il faut un nombre pair d'étapes pour revenir à un état.

Exemple 5

Soit une chaîne de Markov homogène de matrice de transition

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

On a pour tout $n \in \mathbb{N}$,

$$P^{3n} = I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, P^{3n+1} = P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \text{ et } P^{3n+2} = P^2 = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

On remarque que pour $i=1,2,3$, on a $p_{ii}^{(3n+1)} = p_{ii}^{(3n+2)} = 0$ et $p_{ii}^{(3n)} = 1 > 0$, donc les trois états sont de période 3.

Définition

Soit e_i un état de E . On dit que e_i est *récurrent* si,

$$P\left(\bigcup_{n=1}^{\infty} X_n = e_i \mid X_0 = e_i\right) = 1.$$

Dans le cas contraire on dit que e_i est un état *transient* ou *transitoire*.

Cela signifie que si le système est initialement à l'état e_i alors le système est certain de repasser par l'état e_i en un temps fini.

Exemple 4 (suite)

L'état (1) est récurrent et l'état (2) est transitoire.

Propriété

Un état $e_i \in E$ est récurrent (resp. transitoire) si et seulement si la série,

$$\sum_{n \in \mathbb{N}} p_{ii}^{(n)},$$

diverge (resp. converge)

Exemple 5 (suite)

$$\sum_{n \in \mathbb{N}} p_{ii}^{(n)} = \sum_{n \in \mathbb{N}} p_{ii}^{(3n)} + \sum_{n \in \mathbb{N}} p_{ii}^{(3n+1)} + \sum_{n \in \mathbb{N}} p_{ii}^{(3n+2)} = \sum_{n \in \mathbb{N}} 1$$

Cette série diverge donc tous les états sont récurrents. Si on trace le graphe de la chaîne, on voit que cela représente un boucle.

Définition

Soit e_i un état de E . On dit que e_i est *absorbant* si,

$$\forall n \in \mathbb{N}, p_{ii}^{(n)} = 1.$$

Cela signifie que quand le système se trouve à l'état e_i , il ne peut plus en sortir. On remarque qu'un état absorbant est nécessairement apériodique et récurrent.

3.2. Regroupement des états par classes d'équivalence

Définition

Soient e_i et e_j deux états de E . On dit que e_i et e_j *communiquent* si l'on peut passer de l'état e_i à l'état e_j et de l'état e_j à l'état e_i avec des probabilités non nulles,

$$e_i \leftrightarrow e_j \Leftrightarrow \begin{cases} \text{soit } i = j \\ \text{soit } \exists n \geq 1, p_{ij}^{(n)} > 0 \text{ et } \exists m \geq 1, p_{ji}^{(m)} > 0. \end{cases}$$

Cette relation entre deux éléments de E définit une relation d'équivalence sur E . On dit alors que les états e_i et e_j appartiennent à la même *classe d'équivalence*.

Propriété

Les états d'une même classe ont les mêmes propriétés.

On parle alors de classe récurrente, classe transitoire, classe périodique, classe apériodique ou classe absorbante. On remarque que

- si une classe est récurrente, il est impossible d'en sortir,
- si une classe est transitoire alors il est possible d'en sortir mais dans ce cas le processus ne pourra plus y retourner,
- si une classe est absorbante alors elle ne comprend qu'un seul état.

Exemple 4 (suite)

En observant le graphe, on voit que la chaîne admet trois classes : $\{e_1, e_3\}$ classe récurrente, $\{e_2\}$ classe transitoire et $\{e_4, e_5\}$ classe récurrente.

Définition

Une chaîne de Markov est dite *irréductible* si elle n'admet qu'une seule classe.

Exemple 3 (suite)

En observant le graphe, on voit que tous les états communiquent, il n'y a donc qu'une seule classe. Les chaînes d'Ehrenfest sont donc irréductibles.

Remarque

Afin de calculer plus facilement P^n , il peut être intéressant de re-numéroter les états par classes. Dans l'exemple précédent, si on échange e_2 et e_3 alors la matrice de transposition s'écrit,

$$P = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}, \text{ où } A = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \text{ et } B = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{4} & \frac{3}{4} \end{pmatrix},$$

$$\text{d'où } P^n = \begin{pmatrix} A^n & 0 \\ 0 & B^n \end{pmatrix}.$$

4. COMPORTEMENT ASYMPTOTIQUE

Soit $\{X_n\}_{n \in \mathbb{N}}$ une chaîne de Markov homogène à valeurs dans E et $P=(p_{ij})$ sa matrice de transition.

4.1. Définitions et propriétés

Définition

Un vecteur π définissant une probabilité sur E (i.e $\forall i \in I, 0 \leq \pi_i \leq 1$ et $\sum \pi_i = 1$) est appelé *distribution stationnaire* ou *régime stationnaire* de la chaîne de Markov $\{X_n\}_{n \in \mathbb{N}}$ si et seulement si,

$$\pi = \pi P.$$

Cela signifie que π est un vecteur propre de P associé à la valeur propre 1.

On sait que la loi de la chaîne de Markov au temps n est $\pi(n) = \pi(0)P^n$. Si la loi initiale $\pi(0)$ est π , alors la loi au temps n sera aussi π pour tout n . Ce qui justifie le qualificatif de stationnaire.

Exemple 2 (suite)

$$\pi P = \pi \Leftrightarrow (\pi_1 \quad \pi_2 \quad \pi_3) \begin{pmatrix} 0.9 & 0 & 0.1 \\ 0.8 & 0.2 & 0 \\ 0 & 0.5 & 0.5 \end{pmatrix} = (\pi_1 \quad \pi_2 \quad \pi_3) \Leftrightarrow \begin{cases} \pi_2 = 0.125\pi_1 \\ \pi_3 = 0.2\pi_1 \end{cases}$$

En utilisant le fait que $\pi_1 + \pi_2 + \pi_3 = 1$, on obtient la distribution stationnaire $\pi = (0.755, 0.094, 0.151)$.

Exemple 4 (suite)

$$\pi P = \pi \Leftrightarrow (\pi_1 \quad \pi_2 \quad \pi_3 \quad \pi_4 \quad \pi_5) \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{4} & \frac{3}{4} \end{pmatrix} = (\pi_1 \quad \pi_2 \quad \pi_3 \quad \pi_4 \quad \pi_5) \Leftrightarrow \begin{cases} \pi_2 = 0 \\ \pi_1 = \pi_3 \\ \pi_5 = 2\pi_4 \end{cases}$$

En utilisant le fait que $\pi_1 + \pi_2 + \pi_3 + \pi_4 + \pi_5 = 1$, on peut par exemple avoir $\pi = (1/2, 0, 1/2, 0, 0)$ ou $\pi = (0, 0, 0, 1/3, 2/3)$. On voit clairement qu'il n'y a pas unicité de la distribution stationnaire.

Exemple 6

Déterminer une distribution stationnaire à la chaîne de Markov de matrice de transition

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

Propriété

Si E est fini, alors toute chaîne de Markov homogène à valeurs dans E possède au moins une distribution stationnaire.

Propriété

Une chaîne de Markov admet une unique distribution stationnaire si et seulement si elle ne possède qu'une seule classe récurrente.

Exemple 6 (suite)

Si on trace le graphe de la chaîne, on voit qu'il n'y a qu'une seule classe récurrente. Effectivement, il n'y a qu'une seule distribution stationnaire.

Exemple 7

Soit la chaîne de Markov définie par

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

En traçant le graphe de la chaîne, on voit qu'il y a deux classes récurrentes $\{2\}$ et $\{3,4\}$, ce qui exclut l'existence d'une seule distribution stationnaire. En effet, on trouve $\pi = (0, 1-2\alpha, \alpha, \alpha)$ avec $0 \leq \alpha \leq 1/2$.

4.2. Comportement asymptotique

Définition

On appelle *distribution limite* ou *régime permanent* de la chaîne de Markov $\{X_n\}_{n \in \mathbb{N}}$ le vecteur de probabilité π (i.e. $\forall i \in I, 0 \leq \pi_i \leq 1$ et $\sum \pi_i = 1$) tel que

$$\pi = \lim_{n \rightarrow +\infty} \pi(n).$$

Remarque : Si un état e_i est transcient alors $\lim_{n \rightarrow +\infty} \pi_i(n) = 0$.

Etant donné que $\lim_{n \rightarrow +\infty} \pi(n) = \lim_{n \rightarrow +\infty} \pi(0)P^n$, il pourrait y avoir une distribution limite pour chaque loi initiale $\pi(0)$.

Définition

On dit que la chaîne de Markov $\{X_n\}_{n \in \mathbb{N}}$ est *ergodique* si et seulement si elle admet une unique distribution limite indépendante de la loi initiale $\pi(0)$.

Propriété

Si une chaîne admet une distribution limite alors cette distribution est stationnaire. Si la chaîne est ergodique alors la distribution limite est nécessairement l'unique distribution stationnaire de la chaîne.

On peut noter que si la chaîne admet plusieurs distributions stationnaires alors elle ne peut être ergodique. Par exemple, les chaînes d'Ehrenfest ne sont pas ergodiques.

Propriété

Si une chaîne de Markov est apériodique et irréductible alors elle est ergodique.

Propriété

Une chaîne de Markov de matrice de transition P est ergodique si et seulement si la valeur propre 1 de P est simple et si toute autre valeur propre de P est de module strictement inférieure à 1.

Exemple 2 (suite)

Les valeurs propres de la matrice de transition sont $\lambda_1=1$, $\lambda_2=0.3-0.2i$ et $\lambda_3=0.3+0.2i$. On remarque que $|\lambda_2|=|\lambda_3|=0.36 < 1$, donc la chaîne est ergodique (on pouvait aussi remarquer que la chaîne est apériodique et irréductible). On peut donc en déduire que la distribution limite est la distribution stationnaire calculée

précédemment, $\pi=(0.755,0.094,0.151)$. Cela signifie que quel que soit l'état de la personne initialement, à terme elle à 75% de chance d'être immunisée.

5. DELAIS ET PROBABILITE D'ATTEINTE

Soit $\{X_n\}_{n \in \mathbb{N}}$ une chaîne de Markov homogène à valeurs dans E et $P=(p_{ij})$ sa matrice de transition.

5.1. Délais et probabilité d'absorption

Supposons que la chaîne admette des états absorbants, on est généralement intéressé par les deux questions suivantes :

- Combien de temps faudra-t-il pour que le processus soit absorbé étant donné son état initial ?
- S'il existe plusieurs états absorbants, quelle est la probabilité pour le processus d'être absorbé par un état donné ?

Pour répondre à ces questions, il est nécessaire d'introduire les quantités suivantes :

- N_i : le nombre de transitions jusqu'à l'absorption en partant de l'état initial e_i
- $n_i=E(N_i)$: le temps moyen jusqu'à l'absorption en partant de l'état e_i
- b_{ij} : probabilité que le processus soit absorbé dans l'état e_j si son état initial est e_i

Remarque

Si l'état e_i est absorbant, on a $n_i=0$, $b_{ii}=1$ et $b_{ij}=0$ si $j \neq i$.

Exemple 8

Soit la chaîne de Markov définie par

$$P = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{3}{4} & 0 & \frac{1}{4} \\ 0 & 0 & 1 \end{pmatrix}.$$

On considère N_1 le nombre de transitions pour arriver à l'état d'absorption (3) en partant de l'état (1). On trouve

$$P(N_1 = 1) = \frac{1}{2} \quad P(N_1 = 2) = \frac{1}{2} \times \frac{1}{4} \quad P(N_1 = 3) = \frac{1}{2} \times \frac{3}{4} \times \frac{1}{2} \quad P(N_1 = 4) = \frac{1}{2} \times \frac{3}{4} \times \frac{1}{2} \times \frac{1}{4} \dots$$

$$P(N_1 = 2k - 1) = \frac{1}{2} \left(\frac{3}{8}\right)^{k-1} \text{ et } P(N_1 = 2k) = \frac{1}{8} \left(\frac{3}{8}\right)^{k-1}, k=1,2,\dots$$

On en déduit

$$n_1 = E(N_1) = \sum_{k=1}^{\infty} (2k-1) \frac{1}{2} \left(\frac{3}{8}\right)^{k-1} + \sum_{k=1}^{\infty} (2k) \frac{1}{8} \left(\frac{3}{8}\right)^{k-1} = 2.4$$

Si on part de l'état (1), il faudra en moyenne 2.4 unités de temps pour que le processus soit absorbé.

Propriété

Soit S l'ensemble des états non absorbants. Alors pour tout i tel que $e_i \in S$, les quantités n_i et b_{ij} sont les solutions des systèmes d'équations :

$$n_i = 1 + \sum_{k \in S} p_{ik} n_k$$

$$b_{ij} = p_{ij} + \sum_{k \in S} p_{ik} b_{kj}, \quad \forall j \notin S.$$

Exemple 8 (suite)

On a $S=\{(1),(2)\}$. On obtient le système d'équations

$$\begin{cases} n_1 = 1 + p_{11}n_1 + p_{12}n_2 = 1 + \frac{1}{2}n_2 \\ n_2 = 1 + p_{21}n_1 + p_{22}n_2 = 1 + \frac{3}{4}n_1 \end{cases} \Rightarrow n_1=2.4 \text{ et } n_2=2.8$$

Exemple 9

Soit la chaîne de Markov définie par

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{2}{3} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Les états (1) et (4) sont absorbants, donc $S=\{(2),(3)\}$. On obtient le premier système d'équations

$$\begin{cases} n_2 = 1 + p_{22}n_2 + p_{23}n_3 = 1 + \frac{2}{3}n_3 \\ n_3 = 1 + p_{32}n_2 + p_{33}n_3 = 1 + \frac{1}{2}n_2 \end{cases} \Rightarrow n_2=2.25 \text{ et } n_3=2.5$$

Donc que l'on parte de l'état (2) ou (3) le temps d'absorption est à peu près le même. Pour savoir si on a plus de chance que la chaîne soit absorbée par l'état (1) ou l'état (4), on résout les systèmes suivants,

$$\begin{cases} b_{21} = p_{21} + p_{22}b_{21} + p_{23}b_{31} = \frac{1}{3} + \frac{2}{3}b_{31} \\ b_{31} = p_{31} + p_{32}b_{21} + p_{33}b_{31} = \frac{1}{2}b_{21} \end{cases} \Rightarrow b_{21}=0.5 \text{ et } b_{31}=0.25$$

$$\begin{cases} b_{24} = p_{24} + p_{22}b_{22} + p_{23}b_{34} = \frac{2}{3}b_{34} \\ b_{34} = p_{34} + p_{32}b_{24} + p_{33}b_{34} = \frac{1}{2} + \frac{1}{2}b_{24} \end{cases} \Rightarrow b_{24}=0.5 \text{ et } b_{34}=0.75$$

Donc si initialement la chaîne est à l'état (2), il y a autant de chance qu'elle soit absorbée par (1) ou (4), en revanche si l'état initial est (3), il y a trois fois plus de chance que la chaîne soit absorbée par l'état (4).

5.2. Délais et probabilité d'atteinte

Si on considère une chaîne ne comprenant pas d'état absorbant. Soit e_i l'état initial et e_j et e_k deux états différents de e_i . Alors les deux quantités suivantes peuvent présenter un intérêt pratique.

- Le temps moyen nécessaire pour que le processus passe par l'état e_j pour la première fois (*délais moyen d'atteinte* ou *premier passage*)
- La probabilité d'atteindre l'état e_j avant e_k .

Pour les calculer, on utilise les délais et probabilités d'absorption définis précédemment. Pour cela on modifie la matrice de transition P en transformant l'état e_j en état absorbant, si on cherche le délai d'atteinte, ou en transformant les états e_j et e_k en états absorbants, si on cherche la probabilité d'atteinte. Soit P^* la matrice ainsi obtenue. On détermine alors relativement à P^* , le temps d'absorption moyen n_i^* et la probabilité d'absorption b_{ij}^* , et on obtient ainsi les deux quantités cherchées.

Exemple 10

Soit un tétraèdre régulier dont les sommets sont numérotés de 1 à 4. Un robot qui se trouve initialement en 1 met une minute pour parcourir une arête. Arrivé en un sommet il choisit avec la même probabilité l'une des trois arêtes pour continuer sa route. Quel est le temps moyen T nécessaire pour atteindre le sommet 4 ? Quelle

est la probabilité p d'atteindre le sommet 4 sachant qu'il y a de la colle au sommet 2 ?

$$P = \begin{pmatrix} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \end{pmatrix}$$

Il n'y a pas d'état absorbant donc le robot pourrait continuer son voyage indéfiniment. Si on rend l'état (4) absorbant, il faut remplacer la dernière ligne de P par $(0,0,0,1)$ pour avoir P^* . On trouve alors $n_1^* = n_2^* = n_3^* = 3 = T$, donc il lui faudra en moyenne 3 min pour atteindre le sommet 4. Si de plus, on transforme l'état (2) en état absorbant, alors la probabilité qu'il soit absorbé par l'état (4) est $b_{14}^* = p = 0.5$.

Propriétés

Notons n_{ij} le délais moyen d'atteinte de l'état e_j en partant de l'état e_i .

- Si la chaîne est ergodique alors $n_{ii} = \frac{1}{\pi_i}$, où π est le régime permanent de la chaîne.
- Si la chaîne est irréductible alors nous avons le système d'équations
$$n_{ij} = 1 + \sum_{k \neq j} p_{ik} n_{kj}.$$

ANNEXE : Matrices stochastiques

Un vecteur $\pi=(\pi_1,\pi_2,\dots,\pi_n)$ est un *vecteur de probabilité* si ses composantes sont comprises entre 0 et 1 et si leur somme vaut 1.

Une matrice carrée $P=(p_{ij})\in M_n(\mathbb{R})$ est dite *stochastique* si tous ses termes sont compris entre 0 et 1 et si la somme des termes de chaque ligne vaut 1 c'est-à-dire si chaque ligne est un vecteur de probabilité :

$$\forall i,j\in\{1,\dots,n\}, 0\leq p_{ij}\leq 1$$

$$\forall i\in\{1,\dots,n\}, \sum_{j=1}^n p_{ij} = 1 \text{ (somme de la ligne } i)$$

Exemple

La matrice P_1 est stochastique et la matrice P_2 ne l'est pas,

$$P_1 = \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{2} & \frac{1}{6} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} & 0 \end{pmatrix} \text{ et } P_2 = \begin{pmatrix} \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{3}{4} & \frac{1}{2} & -\frac{1}{4} \\ \frac{1}{2} & \frac{2}{3} & \frac{1}{3} \end{pmatrix}.$$

Propriétés générales

Si A et B sont deux matrices stochastiques alors

- AB est une matrice stochastique
- Pour tout entier $m\in\mathbb{N}$, A^m est une matrice stochastique

Propriétés sur les vecteurs et valeurs propres

- Toute matrice stochastique admet 1 comme valeur propre.
- Il existe un vecteur propre π associé à la valeur propre 1 qui définit une distribution de probabilité, c'est-à-dire tel que π est un vecteur de probabilité.
- La valeur propre 1 est la seule à laquelle on peut associer un vecteur de probabilité comme vecteur propre.
- Toute valeur propre a un module inférieur à 1.
- Si le module d'une valeur propre est égale à 1 alors celle-ci est une racine de l'unité.