Simulations des phénomènes aléatoires 3LM246 : PARTIE 2, Chaines de Markov.

Raphaël KRIKORIAN, Irina KOURKOVA
Université Paris 6

Remis à jour pour l'année 2018-2019 par Irina KOURKOVA

Ceci est la dernière (grosse) partie du polycopié du cours LM346 "Processus et Simulations".

Elle porte sur les Chaines de Markov sur l'espace dénombrable mais seulement les Chaines de Markov sur l'espace fini seront portées à l'examen.

Les parties non-obligatoires qui peuvent être omises à la première lecture sont indiquées. Je conseille aussi d'omettre à la première lecture le premier chapitre sur l'espérance conditionnelle (si vous connaissez les probabilités conditionnelles), de passer directement aux Chaines de Markov (page 17) et de consulter ce premier chapitre quand c'est nécessaire.

Le signe!! en début d'une ligne ou d'un paragraphe signifie un point important pour les exercices, il est souhaitable d'y porter une attention particulière.

Cette partie du cours est conçue en accord avec la feuille d'exercices N3. Après chaque section, les exercices qui s'y rapportent sont recommmandés à faire. Pour bien assimiler le contenu du cours et se préparer à l'examen il est souhaitable de suivre ces recommendations .

Le Devoir N3 portera sur tout le contenu jusqu'à la fin de la Section 2.5.1. Le Devoir N4 couvrira tout le contenu en partant de la Section 2.5.2 jusqu'à la fin.

Table des matières

Chapitre 1

Espérance conditionnelle

Ce chapitre ne sera pas demandé à l'examen. Il est écrit pour preparer le chapitre "Chaines de Markov".

Dans tout ce qui suit $(\Omega, \mathcal{B}, \mathbf{P})$ est un espace probabilisé fixé.

1.1 Probabilités conditionnelles

Définition 1.1.1 *Si A et B sont deux événements de la tribu $\mathcal{B}$ et si $\mathbf{P}(B) > 0$ on définit la probabilité de A sachant B comme étant*

$$\mathbf{P}(A|B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}.$$

En fait, tout événement $B \in \mathcal{B}$ définit une nouvelle probabilité sur $(\Omega, \mathcal{B})$:

Proposition 1.1.1 *Si $B \in \mathcal{B}$ est tel que $\mathbf{P}(B) > 0$, l'application $\mathbf{P}_B : \mathcal{B} \to [0,1]$ qui à $A \in \mathcal{B}$ associe $\mathbf{P}_B(A) = \mathbf{P}(A|B)$ est une probabilité.*

La preuve de cette proposition est laissée en exercice au lecteur.

Supposons à présent que nous ayons une partition de Ω en événements $B_1, \ldots, B_r$ (c'est-à-dire que les événements $B_1, \ldots, B_r$ sont disjoints deux à deux et que leur union soit égale à Ω) et que pour tout i, $\mathbf{P}(B_i) > 0$. On peut donc définir r probabilités $\mathbf{P}(\cdot|B_i)$.

Théorème 1.1.1 (Formule des causes) *Sous les hypothèses précédentes*

$$\mathbf{P}(A) = \sum_{i=1}^{r} \mathbf{P}(A|B_i)\mathbf{P}(B_i).$$

Démonstration.— Il suffit de remarquer que A est l'union disjointe des $(A \cap B_i)$ et donc

$$\begin{aligned}\mathbf{P}(A) &= \sum_{k=1}^{r} \mathbf{P}(A \cap B_i) \\ &= \sum_{i=1}^{r} \frac{\mathbf{P}(A \cap B_i)}{\mathbf{P}(B_i)} \cdot \mathbf{P}(B_i)\end{aligned}$$

ce qui est la formule annoncée.

□

Dans la pratique un problème courant est de calculer $\mathbf{P}(B_i|A)$ connaissant les $\mathbf{P}(A|B_j)$.

Théorème 1.1.2 (Formule de Bayes) *Sous les hypothèses précédentes :*

$$\mathbf{P}(B_i|A) = \frac{\mathbf{P}(A|B_i)\mathbf{P}(B_i)}{\sum_{i=1}^{r} \mathbf{P}(A|B_j)\mathbf{P}(B_j)}.$$

Démonstration.— Il suffit d'écrire

$$\mathbf{P}(B_i|A) = \frac{\mathbf{P}(B_i \cap A)}{\mathbf{P}(A)} = \frac{\mathbf{P}(A|B_i)\mathbf{P}(B_i)}{\mathbf{P}(A)},$$

et d'utiliser la formule des causes.

□

1.2 Espérance conditionnelle par rapport à une tribu finie

Espérance sachant un événement Supposons fixé $B \in \mathcal{B}$ un événement tel que $\mathbf{P}(B) > 0$. On sait que $\mathbf{P}(\cdot|B)$ est une probablilité sur $(\Omega, \mathcal{B})$ et on obtient donc un espace probabilisé $(\Omega, \mathcal{B}, \mathbf{P}(\cdot|B))$. Pour toute variable aléatoire $Y : \Omega \to \mathbf{R}$ on peut définir $\mathbf{E}(Y|B)$ comme étant l'espérance de Y sur l'espace probabilisé $(\Omega, \mathcal{B}, \mathbf{P}(\cdot|B))$. C'est un *nombre* (si Y est une v.a.r.).

Dans le cas où par exemple Y prend un nombre fini de valeurs $y_1, \ldots, y_m$ on a

$$\mathbf{E}(Y|B) = \sum_{k=1}^{m} y_k \mathbf{P}(Y = y_k|B).$$

Loi conditionnelle d'une v.a sachant un évènement

Définition 1.2.1 *La loi conditionnelle de la v.a Y sachant B est la loi de Y vue comme v.a sur l'espace probabilisé $(\Omega, \mathcal{B}, \mathbf{P}(\cdot|B))$. C'est donc la mesure de probabilité sur $(\mathbf{R}, Bor(\mathbf{R}))$ définie par $\mu_{Y|B}(\cdot) = \mathbf{P}(Y^{-1}(\cdot)|B)$. Si Y est à valeurs dans un ensemble fini ou dénombrable nous la noterons souvent $p(y|B)$:*

$$p(y|B) = \mathbf{P}(Y = y|B).$$

Espérance sachant une tribu (finie) On suppose fixée une partition $B_1, \ldots, B_r$ de Ω et on fait l'hypothèse que pour tout i, $\mathbf{P}(B_i) > 0$. Considérons $\mathcal{F}$ la plus petite tribu engendrée par les ensembles $B_1, \ldots, B_r$.

Exercice : Montrer que $\mathcal{F}$ est l'ensemble des 2^r unions finies de B_i :

$$\mathcal{F} = \{\bigcup_{i \in J} B_i : J \subset \{1, \ldots, r\}\}.$$

Réciproquement, montrer que toute tribu finie est obtenue de la façon précédente.

Définition 1.2.2 *Si $Y : \Omega \to \mathbf{R}$ est une v.a, on définit l'espérance de Y par rapport à la tribu (finie) $\mathcal{F}$ comme étant la* **variable aléatoire** $\mathbf{E}(Y|\mathcal{F}) : \Omega \to \mathbf{R}$ *valant $\mathbf{E}(Y|B_i)$ sur B_i, $B_i \in \mathcal{F}$:*

$$\mathbf{E}(Y|\mathcal{F}) = \sum_{k=1}^{r} \mathbf{E}(Y|B_k) \cdot \mathbf{1}_{B_k}.$$

1.3 Espérance conditionnelle suivant une v.a, cas discret

Supposons à présent que $X : \Omega \to \mathbf{R}$ soit une v.a ne prenant qu'un nombre fini de valeurs $x_1, \ldots, x_r$ et notons $B_i = X^{-1}(x_i)$, $1 \leq i \leq r$. et $\mathcal{F}$ (ou encore $\mathcal{F}(B_1, \ldots, B_r)$) la tribu engendrée par les B_i. Nous supposerons que pour tout i, $\mathbf{P}(X = x_i) > 0$.

Définition 1.3.1 *On définit l'espérance conditionnelle de Y sachant X comme étant la* **variable aléatoire** $\mathbf{E}(Y|X) : \Omega \to \mathbf{R}$ *définie par $\mathbf{E}(Y|X) = \mathbf{E}(Y|\mathcal{F}(X = x_1, \ldots, X = x_r))$. On a donc*

$$\mathbf{E}(Y|X)(\omega) = \sum_{k=1}^{r} \mathbf{E}(Y|X = x_k) \cdot \mathbf{1}_{X=x_k}.(\omega).$$

(C'est la fonction qui associe à ω $\mathbf{E}(Y|X = x_k)$ si $X(\omega) = x_k$.)

Dans le reste de cette section nous supposerons que X et Y sont deux v.a. prenant un nombre fini ou dénombrable de valeurs.

Mentionnons la propriété suivante évidente mais importante

Proposition 1.3.1 *Si on note ψ la fonction $\psi(x) = \mathbf{E}(Y|X = x)$ on a*

$$\mathbf{E}(Y|X) = \psi(X),$$

(c'est une égalité entre v.a).

Loi conditionnelle de Y sachant X (X, Y à valeurs discrètes) Pour tout $x \in X(\Omega)$ on peut définir la loi conditionnelle de Y sachant l'évènement $(X = x)$.

Définition 1.3.2 *On appelle loi conditionnelle de Y sachant $X = x$ la loi conditionnelle de Y sachant l'évènement $(X = x)$ et nous la noterons $p_{Y|X}(y|x)$:*

$$p_{Y|X}(y|x) = \mathbf{P}(Y = y|X = x) = \frac{p_{X,Y}(x, y)}{p_X(x)},$$

où on a noté $p_{X,Y}$ la loi du couple (X, Y) et p_X la loi de X

Relation entre espérance conditionnelle et loi conditionnelle

Proposition 1.3.2 *On a*

$$\mathbf{E}(Y|X = x) = \sum_{l=1}^{m} y_l p_{Y|X}(y_l|x).$$

Propriété de l'espérance conditionnelle On a alors

Théorème 1.3.1 *On a*

$$\mathbf{E}(\mathbf{E}(Y|X)) = \mathbf{E}(Y).$$

Démonstration.— Il suffit de calculer

$$\begin{aligned}
\mathbf{E}(\mathbf{E}(Y|X)) &= \mathbf{E}(\psi(X)) \\
&= \sum_{k=1}^{r} \mathbf{E}(Y|X = x_k)\mathbf{P}(X = x_k) \\
&= \sum_{k=1}^{r} \left(\sum_{l=1}^{m} y_l \mathbf{P}(Y = y_l|X = x_k) \right) \mathbf{P}(X = x_k) \\
&= \sum_{l=1}^{m} y_l \left(\sum_{k=1}^{r} \mathbf{P}(Y = y_l|X = x_k)\mathbf{P}(X = x_k) \right) \\
&= \sum_{l=1}^{m} y_l \mathbf{P}(Y = y_l) \\
&= \mathbf{E}(Y)
\end{aligned}$$

(on a utilisé la formule des causes).

□

Remarque : L'espérance conditionnelle $\mathbf{E}(Y|X)$ est linéaire par rapport à Y.

Exercice : Démontrer que si X, Y sont des v.a ne prenant qu'un nombre fini de valeurs $\mathbf{E}(f(X,Y)|X) = \psi(X)$ où

$$\psi(x_k) = \sum_{l=1}^{m} f(x_k, y_l) p_{Y|X}(y_l|x_k).$$

Exercice : Soient X et Y deux v.a indépendantes qui suivent des lois de Poisson de paramètres respectifs λ et μ. Soit $S = X + Y$

a) Déterminer la loi de S

b) Calculer l'espérance conditionnelle $\mathbf{E}(X|S)$.

(Réponse : a) S suit une loi de Poisson de paramètre $\lambda + \mu$; b) $\mathbf{E}(X|S)$ est la variable aléatoire $S \cdot \lambda/(\lambda + \mu)$)

1.4 Cas des v.a admettant des densités

Supposons donnée deux v.a Y et X telles que le couple (X, Y) admette une densité $\rho_{X,Y}$. Les v.a X et Y admettent donc des densités respectives ρ_X et ρ_Y égales à :

$$\rho_X(x) = \int_{-\infty}^{\infty} \rho_{X,Y}(x, y)dy, \quad \rho_Y(y) = \int_{-\infty}^{\infty} \rho_{X,Y}(x, y)dx.$$

Nous supposerons dans toute la suite que $\rho_X(x) > 0$ pour tout x.

Nous allons définir l'espérance $\mathbf{E}(Y|X)$ en nous inspirant de la définition 1.3.2, de la proposition 1.3.2 et de la proposition 1.3.1.

Définition 1.4.1 *Nous appelerons densité conditionnelle de Y sachant $X = x$ la fonction $\rho_{Y|X}(y|x)$ définie par*

$$\rho_{Y|X}(y|x) = \frac{\rho_{X,Y}(x,y)}{\rho_X(x)}.$$

Définition 1.4.2 *L'espérance de Y sachant $X = x$ est la fonction de $x \in \mathbf{R}$*

$$\mathbf{E}(Y|X = x) = \int_{\mathbf{R}} y\rho_{Y|X}(y|x)dy.$$

Définition 1.4.3 *L'espérance conditionnelle $\mathbf{E}(Y|X)$ est la variable aléatoire*

$$\mathbf{E}(Y|X) = \psi(X),$$

où $\psi(x) = \mathbf{E}(Y|X = x)$.

Exercice : Soient X, Y deux v.a dont la loi du couple (X, Y) admet une densité égale à $\rho_{X,Y}(x,y) = (1/x) \cdot \mathbf{1}_T(x,y)$ où T est le triangle $T = \{0 < y < x < 1\}$. Calculer $\mathbf{E}(Y|X)$.

Solution : La densité de X se calcule en intégrant $\rho_{X,Y}(x,y)$ par rapport à $y \in \mathbf{R}$. On trouve $\rho_X(x) = \mathbf{1}_{]0,1[}(x)$ et donc pour $x \in]0,1[$

$$\rho_{Y|X}(y|x) = \frac{1}{x}\mathbf{1}_{]0,x[}(y).$$

On trouve $\psi(x) = \mathbf{E}(Y|X = x) = (x/2)$ et donc $\mathbf{E}(Y|X) = \psi(X) = X/2$.

On a encore le théorème important suivant

Théorème 1.4.1

$$\mathbf{E}(\mathbf{E}(Y|X)) = \mathbf{E}(Y).$$

Démonstration.— Notons $\psi(x) = \mathbf{E}(Y|X = x)$. Comme par définition

$\mathbf{E}(Y|X) = \psi(X)$, la formule de transfert donne

$$\begin{aligned}
\mathbf{E}(\mathbf{E}(Y|X)) &= \mathbf{E}(\psi(X)) \\
&= \int_{\mathbf{R}} \psi(x)\rho_X(x)dx \\
&= \int_{\mathbf{R}} \mathbf{E}(Y|X=x)\rho_X(x)dx \\
&= \int_{\mathbf{R}} \left(\int_{\mathbf{R}} y\rho_{Y|X}(x,y)dy \right) \rho_X(x)dx \\
&= \int_{\mathbf{R}} \left(\int_{\mathbf{R}} y\frac{\rho_{X,Y}(x,y)}{\rho_X(x)}dy \right) \rho_X(x)dx
\end{aligned}$$

et comme d'après le théorème de Fubini on peut intervertir l'ordre d'intégration :

$$\begin{aligned}
\mathbf{E}(\mathbf{E}(Y|X)) &= \int_{\mathbf{R}} \left(\int_{\mathbf{R}} \frac{\rho_{X,Y}(x,y)}{\rho_X(x)}\rho_X(x)dx \right) ydy \\
&= \int_{\mathbf{R}} \left(\int_{\mathbf{R}} \rho_{X,Y}(x,y)dx \right) ydy \\
&= \int_{\mathbf{R}} \rho_Y(y)ydy \\
&= \mathbf{E}(Y).
\end{aligned}$$

□

Exercice : Démontrer que $\mathbf{E}(Yg(X)|X) = g(X)\mathbf{E}(Y|X)$.

On a l'analogue suivant du théorème de transfert que nous donnons sans démonstration (comparer avec la définition 1.4.2) :

Théorème 1.4.2 *Si* $Z = f(X,Y)$ *on a* $\mathbf{E}(f(X,Y)|X) = \psi(X)$ *où*

$$\psi(x) = \int_{\mathbf{R}} f(x,y)\rho_{Y|X}(y|x)dy.$$

Remarque : L'espérance conditionnelle $\mathbf{E}(Y|X)$ est linéaire par rapport à Y.

Exercice : Démontrer que si on prend l'énoncé du théorème précédent comme définition de $\mathbf{E}(f(X,Y)|X)$ on a bien $\mathbf{E}(f(X,Y)) = \mathbf{E}(\mathbf{E}(f(X,Y)|Y))$.

1.5 Indépendance

Considérons à présent le cas où les v.a X et Y sont indépendantes. Dans ce cas on sait que

$$p_{X,Y}(x,y) = p_X(x)p_Y(y),$$

dans le cas discret et

$$\rho_{X,Y}(x,y) = \rho_X(x)\rho_Y(y)$$

dans le cas continu. Par conséquent

$$p_{Y|X}(y|x) = p_Y(y)$$

dans le cas discret et

$$\rho_{Y|X}(y|x) = \rho_Y(y)$$

dans le cas continu. On a donc dans les deux cas

Proposition 1.5.1 *Si X et Y sont indépendantes*

$$\mathbf{E}(Y|X) = \mathbf{E}(Y)$$

ou en d'autres termes, la **variable aléatoire** $\mathbf{E}(Y|X)$ *est constante.*

Dans la pratique on a souvent à calculer des espérances de v.a de la forme $Z = f(X,Y)$ où X et Y sont des v.a indépendantes. Il est souvent utile pour cela de *conditionner* Z par rapport à une des variables (disons X) c'est-à-dire de faire le calcul en deux étapes :

i) On introduit la v.a $\mathbf{E}(Z|X) = \mathbf{E}(f(X,Y)|X)$;

ii) On utilise ensuite la relation $\mathbf{E}(Z) = \mathbf{E}(\mathbf{E}(Z|X))$.

Pour l'étape i) on dispose de la proposition suivante :

Proposition 1.5.2 *Soient X et Y deux v.a indépendantes. Notons pour $x \in$ $\mathbf{R}$, Y_x la variable aléatoire $Y_x = f(x,Y)$ (c'est-à-dire l'application de $\Omega \to \mathbf{R}$ telle que $Y_x(\omega) = f(x, Y(\omega))$) et soit $F : \mathbf{R} \to \mathbf{R}$ l'application définie par $F(x) = \mathbf{E}(Y_x)$ ($F(x) = \mathbf{E}(f(x,Y))$). On a*

$$\mathbf{E}(f(X,Y)|X) = F(X).$$

Démonstration.—

Nous faisons la démonstration dans le cas de v.a admettant des densités (le cas discret se traite de la même façon). Notons $Z = f(X,Y)$. On a d'après le théorème 1.4.2

$$\mathbf{E}(Z|X) = \psi(X)$$

où

$$\begin{aligned}\psi(x) &= \int_{\mathbf{R}} f(x,y)\rho_{Y|X}(y|x)dy \\ &= \int_{\mathbf{R}} f(x,y)\rho_Y(y)dy \\ &= \mathbf{E}(Y_x) \\ &= F(x).\end{aligned}$$

On a donc bien la conclusion.

□

Exercice : Soient $N, X_1, \ldots, X_n, \ldots$ une famille indépendante de v.a de même loi, intégrables et à valeurs dans $\mathbf{N}^*$. Soit S la v.a

$$S = \sum_{1 \le i \le N} X_i.$$

Calculer $\mathbf{E}(S)$.

Solution : On calcule d'abord $\mathbf{E}(S|N)$ et pour cela on évalue $\psi(n) = \mathbf{E}(S|N = n)$.

$$\begin{aligned}\mathbf{E}(S|N=n) &= \sum_{1\le i\le n} \mathbf{E}(X_i|N=n) \\ &= \sum_{1\le i\le n} \mathbf{E}(X_i) \\ &= n\mathbf{E}(X_1).\end{aligned}$$

où dans le passage de la première à la deuxième ligne on a utilisé le fait que X_i et N sont indépendantes. On a donc $\mathbf{E}(S|N) = \mathbf{E}(X_1) \cdot N$. On utilise ensuite la formule $\mathbf{E}(S) = \mathbf{E}(\mathbf{E}(S|N))$ et on trouve $\mathbf{E}(S) = \mathbf{E}(X_1)\mathbf{E}(N)$.

1.6 Tribus indépendantes

Définition 1.6.1 *Une famille de tribu $\mathcal{F}_\alpha$, $\alpha \in E$ est dite indépendante si pour toute sous famille finie $\mathcal{F}_{\alpha_k}$, $k = 1, \ldots, r$ et tous $A_{\alpha_k} \in \mathcal{F}_{\alpha_k}$ on a*

$$\mathbf{P}\Big(\bigcap_{k=1}^r A_{\alpha_k}\Big) = \prod_{k=1}^r \mathbf{P}(A_{\alpha_k}).$$

Voyons comment cette notion se rattache à l'indépendance des v.a. Etant donnée une suite de v.a $X_1, \ldots, X_n, \ldots$ on peut toujours définir la tribu $\mathcal{F}(X_i; i \geq 1)$ qui est *la tribu engendrée par les événements* $X_i^{-1}(I)$ où i décrit $i \geq 1$ et I décrit les intervalles de $\mathbf{R}$. Il n'est pas difficile de voir que c'est la plus petite tribu qui rend simultanément mesurables les X_i. La suite de v.a. $X_1, \ldots, X_n, \ldots$ est indépendante si les tribus engendrées par les X_i le sont. Plus généralement on dira qu'une v.a. est *indépendante d'une tribu* $\mathcal{F}$ si la tribu engendrée par X est indépendante de la tribu $\mathcal{F}$.

Proposition 1.6.1 *a) Si les v.a $U_1, \ldots, U_n, X$ sont indépendantes alors X est indépendante de la tribu $\mathcal{F}(U_1, \ldots, U_n)$.*

b) Si $U_1, \ldots, U_n$ sont des v.a. et X est une v.a de la forme $X = F(U_1, \ldots, U_n)$ alors $\mathcal{F}(U_1, \ldots, U_n, X) \subset \mathcal{F}(U_1, \ldots, U_n)$.

Démonstration.— Pour a) il suffit de démontrer

Lemme 1.6.1 *Si A_α, $\alpha \in E$ est une famille d'événements indépendants (i.e les fonctions indicatrices le sont) et si (E_1, E_2) est une partition de E, alors les tribus $\mathcal{F}_1, \mathcal{F}_2$ sont indépendantes où on a noté $\mathcal{F}_i$ la tribu engendrée par les événements $(A_\alpha)_{\alpha \in E_i}$.*

Démonstration.— Notons $\mathcal{C}$ l'ensemble des A pour lesquels, pour tout $B \in \{A_\alpha, \alpha \in E_2\}$ on ait

$$\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B). \tag{1.1}$$

Il est facile de voir que c'est une tribu qui contient $\{A_\alpha, \alpha \in E_1\}$ et donc que l'égalité (1.1) est vraie pour tout $A \in \mathcal{F}_1$ et tout B de la forme A_α, $\alpha \in E_2$. A présent notons $\mathcal{D}$ l'ensemble des B pour lesquels (1.1) soit vraie pour tout $A \in \mathcal{F}_1$: c'est une tribu qui contient $\{A_\alpha, \alpha \in E_2\}$ et donc $\mathcal{F}_2$. Par conséquent, (1.1) est vraie pour tout $A \in \mathcal{F}_1$ et tout $B \in \mathcal{F}_2$.

□

Le point b) est quant à lui clair puisque tout événement de la forme $X^{-1}(I)$ appartient à $\mathcal{F}(U_1, \ldots, U_n)$.

□

On a aussi la proposition simple suivante

Proposition 1.6.2 *La v.a X est indépendante de la tribu $\mathcal{F}$ si et seulement si pour toute fonction continue bornée*

$$\mathbf{E}(\phi(X)|\mathcal{F}) = \mathbf{E}(\phi(X)).$$

1.7 Compléments

Le lecteur aura remarqué que la construction que nous avons donnée de l'espérance conditionnelle dans les cas (des v.a. prenant des valeurs dans un ensemble) discret ou continu n'est pas la mÃªme, ce qui n'est pas très satisfaisant. En fait, une façon unifiée d'introduire cette notion est la suivante. On démontre qu'il est possible d'associer à toute v.a $X \in L^1(\Omega, \mathcal{B}, \mathbf{P})$ et toute tribu $\mathcal{F} \subset \mathcal{B}$, une **v.a.** $\mathbf{E}(X|\mathcal{F}) \in L^1(\Omega, \mathcal{F}, \mathbf{P})$ (donc mesurable par rapport à la tribu $\mathcal{F}$) que l'on appelle l'espérance conditionnelle de X par rapport à la tribu $\mathcal{F}$ qui est définie de façon unique (mod 0) par la propriété suivante : $Y \in L^\infty(\Omega, \mathcal{F}, \mathbf{P})$ (donc $\mathcal{F}$-mesurable) et toute v.a $X \in L^1(\Omega, \mathcal{B}, \mathbf{P})$ on a

$$\mathbf{E}(X \cdot Y) = \mathbf{E}\Big(\mathbf{E}(X|\mathcal{F}) \cdot Y\Big).$$

Si X est de carré intégrable ($X \in L^2(\Omega, \mathcal{B}, \mathbf{P})$) cette espérance conditionnelle est facilement définie comme étant la v.a dans $L^2(\Omega, \mathcal{F}, \mathbf{P})$ qui est la projection orthogonale sur $X \in L^2(\Omega, \mathcal{F}, \mathbf{P}) \subset\in L^2(\Omega, \mathcal{B}, \mathbf{P})$. Dans le cas général, on utilise que toute fonction dans $X \in L^1(\Omega, \mathcal{B}, \mathbf{P})$ se laisse approcher L^1 par des fonctions dans $X \in L^2(\Omega, \mathcal{B}, \mathbf{P})$. Il est alors possible de vérifier que par exemple :

1) Pour toute v.a. $Y \in L^\infty(\Omega, \mathcal{F}, \mathbf{P})$ (donc $\mathcal{F}$-mesurable) et toute v.a $X \in L^1(\Omega, \mathcal{B}, \mathbf{P})$ on a

$$\mathbf{E}(X \cdot Y|\mathcal{F}) = Y \cdot \mathbf{E}(X|\mathcal{F});$$

2) Si $\mathcal{F}_1 \subset \mathcal{F}_2$ et $X \in L^1(\Omega, \mathcal{B}, \mathbf{P})$ on a

$$\mathbf{E}\Big(\mathbf{E}(X|\mathcal{F}_1)|\mathcal{F}_2\Big) = \mathbf{E}(X|\mathcal{F}_2).$$

3) Si $X \in L^1(\Omega, \mathcal{B}, \mathbf{P})$,

$$\mathbf{E}(X|\mathcal{B}) = X \qquad \text{et} \qquad \mathbf{E}(X|\{\emptyset, \Omega\}) = \mathbf{E}(X).$$

4) Si la v.a $X \in\in L^1(\Omega, \mathcal{B}, \mathbf{P})$ est indépendante de la tribu $\mathcal{F}$

$$\mathbf{E}(X|\mathcal{F}) = \mathbf{E}(X).$$

Chapitre 2

Chaînes de Markov

2.1 Chaîne de Markov et processus stochastiques

Un processus stochastique est une collection de variables aléatoires $(X(t)(\cdot), t \in T)$, $X(t)(\cdot) : \Omega \to E$ où la variable t est interprétée comme un temps. L'ensemble E s'appelle l'espace des états du processus. Le temps peut être discret auquel cas $T = \mathbf{N}$ ou $\mathbf{Z}$ (on parle alors de processus à temps discret) ou continu auquel cas T sera un intervalle (fini ou infini) de $\mathbf{R}$ (on parle alors de processus à temps continu). Des exemples de tels processus sont par exemple le temps qu'il fait le jour $t = n$ ($X(t)$ prend ses valeurs dans $\{beau, mauvais\}$) ou le nombre $X(t)$ de clients arrivés à un guichet avant le temps t, ou encore la position au temps t d'une molécule dans un gaz *etc.* Comprendre un tel processus c'est connaître pour tout n et tous $s_1 < \ldots < s_n$ la loi du vecteur $(X_{i_1}, \ldots, X_{i_n})$ [1]

On dit qu'un processus est *stationnaire* si pour tout n et tous temps t, $s_1, \ldots, s_n$ la loi du vecteur aléatoire $(X_{t+s_1}, \ldots, X_{t+s_n})$ est indépendante de t (mais évidemment dépend de $s_1, \ldots, s_n$). En d'autres termes, un processus est stationnaire s'il est invariant par translation temporelle. L'objet de ce chapitre est essentiellement de démontrer qu'un processus de Markov raisonnable se stabilise asymptotiquement, c'est-à-dire converge dans un certain sens vers un processus stationnaire.

Une *chaîne de Markov* est un cas particulier important de processus stochastique. Nous ne parlerons dans la suite que de chaînes de Markov homogène à temps discret. Nous noterons X_n au lieu de $X(t)$ les v.a du processus associé.

Définition 2.1.1 *Une chaîne de Markov à temps discret est un processus*

1. en fait pour connaître la loi de tous ces vecteurs aléatoires, il suffit de connaître la loi des vecteurs de la forme $(X_0, \ldots, X_n)$

stochastique à temps discret $X_0, X_1, \dots, X_n, \dots$ *qui vérifie : pour tout* n *et tous états* $i_0, i_1, \dots, i_{n-1}, i, j \in E$ *la probabilité conditionnelle*

$$\mathbf{P}(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1} \dots X_0 = i_0) = \mathbf{P}(X_{n+1} = j | X_n = i)$$

ne dépend que de i *et de* j *et de* n*. On dit que la chaîne de Markov est homogène si cette quantité ne dépend que de* i, j *et pas de* n*. En d'autres termes pour une chaîne de Markov homogène, l'état du système dans le futur (au temps* $n+1$*) ne dépend que de son état dans le présent (temps* n*) et pas des états antérieurs dans le passé (temps* $0 \leq k \leq n-1$*). On note* $p_{ij} = \mathbf{P}(X_{n+1} = j | X_n = i)$ *et on appelle ce nombre la probabilité de transition de* i *à* j*. La loi de* X_0 *s'appelle la loi initiale de la chaîne.*

N.B. Dans la suite nous ne considérons que des chaînes de Markov discrètes homogènes avec un nombre fini ou dénombrable d'états.

On note indifféremment $\mathcal{F}(X_i, i \geq 0)$ ou $\sigma(X_i, i \geq 0)$ la tribu engendrée par les événements $(X_i \in I_i)$, $i \geq 0$, I_i intervalle ouvert de $\mathbf{R}$. Nous supposerons dans la suite que les X_i prennent un nombre fini r de valeurs.

En fait, une chaîne de Markov est déterminée par la loi μ de X_0 et les probabilités de transitions $p_{ij} = \mathbf{P}(X_{n+1} = j | X_n = i)$ dans la mesure où la connaissance de ces lois permet de connaître la loi du vecteur aléatoire $(X_{s_1}, \dots, X_{s_n})$ pour tout n et tous temps $s_1 < \dots < s_n$. En effet,

Théorème 2.1.1 *a) Notons* $p_j^{(n)} = \mathbf{P}(X_n = j)$ *la loi de* X_n *et* $p^{(n)}$ *le vecteur ligne associé. Si on note* P *la matrice de transitions de la chaîne c'est-à-dire la matrice* $r \times r$ *dont les coefficients sont les* $p_{ij} = \mathbf{P}(X_{n+1} = j | X_n = i)$ *on a pour tout* k

$$p^{(n+k)} = p^{(k)} P^n.$$

Ainsi, la loi de X_0 *et les probabilités de transitions déterminent la loi de* X_n*.*
b) Pour tous $i_1, \dots, i_n$

$$\begin{aligned} &\mathbf{P}(X_{s_n} = i_{s_n}, \dots, X_{s_1} = i_{s_1}) = \\ &\quad \mathbf{P}(X_{s_n} = i_{s_n} | X_{s_{n-1}} = i_{s_{n-1}}) \cdots \mathbf{P}(X_{s_2} = i_{s_2} | X_{s_1} = i_{s_1}) \mathbf{P}(X_{s_1} = i_{s_1}). \end{aligned} \tag{2.1}$$

Par conséquent, (cf. a)) la loi du vecteur aléatoire $(X_{s_1}, \dots, X_{s_n})$ *est déterminée par la loi de* X_0 *et les probabilités de transitions.*

! ! Le corollaire de ce Thm pour faire les exercices est ce que pour calculer la probabilité ci-dessus il faut calculer les puissances de la matrice P. Alors cette probabilité s'exprime en terme d'un produit :

$$\mathbf{P}(X_{s_n} = i_{s_n}, \dots, X_{s_1} = i_{s_1}) = (P^{s_n - s_{n-1}})_{i_{s_{n-1}}, i_{s_n}} \cdots (P^{s_2 - s_1})_{i_{s_1}, i_{s_2}} (\mu P^{s_1})_{i_{s_1}},$$

μ étant le vecteur ligne composé de $\mu_i = \mathbf{P}(X_0 = i)$, $i \in E$.
Démonstration.— a) Il suffit de remarquer que

$$\begin{aligned} p_j^{(n+1)} &= \sum_{k=1}^{r} \mathbf{P}(X_{n+1} = j, X_n = k) \\ &= \sum_{k=1}^{r} \mathbf{P}(X_{n+1} = j | X_n = k)\mathbf{P}(X_n = k) \\ &= \sum_{k=1}^{r} p_k^{(n)} P_{kj} \end{aligned}$$

ce qui s'écrit matriciellement $p^{(n+1)} = p^{(n)}P$. En itérant cette dernière formule on obtient a).
b) Démontrons le lemme suivant :

Lemme 2.1.1 *Si A est un événement dans $\mathcal{F}(X_0, \ldots, X_{n-1})$ alors*

$$\mathbf{P}(X_{n+k} = i_{n+k} | X_n = i_n, A) = \mathbf{P}(X_{n+k} = i_{n+k} | X_n = i_n).$$

Démonstration.— On sait que cette formule est vraie quand $k = 1$ et que A est de la forme $C = \{X_{n-1} = i_{n-1} \ldots X_0 = i_0\}$ (c'est la définition d'une chaîne de Markov). Tout événement de $\mathcal{F}(X_0, \ldots, X_{n-1})$ est une union disjointe finie d'événements C_l de cette forme C : $A = \bigcup_l C_l$. Comme

$$\begin{aligned} \mathbf{P}(X_{n+1} = i_{n+1} | X_n = i_n, A) &= \frac{\mathbf{P}(X_{n+1} = i_{n+1}, X_n = i_n, A)}{\mathbf{P}(X_n = i_n, A)} \\ &= \sum_l \frac{\mathbf{P}(X_{n+1} = i_{n+1}, X_n = i_n, C_l)}{\mathbf{P}(X_n = i_n, A)} \\ &= \sum_l \frac{\mathbf{P}(X_{n+1} = i_{n+1} | X_n = i_n, C_l)}{\mathbf{P}(X_n = i_n, A)} \mathbf{P}(X_n = i_n, C_l) \\ &= \sum_l \frac{\mathbf{P}(X_{n+1} = i_{n+1} | X_n = i_n)}{\mathbf{P}(X_n = i_n, A)} \mathbf{P}(X_n = i_n, C_l) \\ &= \mathbf{P}(X_{n+1} = i_{n+1} | X_n = i_n) \sum_l \frac{\mathbf{P}(X_n = i_n, C_l)}{\mathbf{P}(X_n = i_n, A)} \\ &= \mathbf{P}(X_{n+1} = i_{n+1} | X_n = i_n), \end{aligned}$$

le passage de la troisième à la quatrième ligne étant justifié par le fait évoqué plus haut que le lemme est vrai quand $k = 1$ et pour des événements de la forme C_l. Nous avons donc démontré que le lemme est vrai pour $k = 1$. Pour

k quelconque on raisonne par récurrence : supposons le lemme vrai pour k et montrons qu'il est vrai pour $k+1$. Comme précédemment,

$$\begin{aligned}
&\mathbf{P}(X_{n+(k+1)} = i_{n+k+1} | X_n = i_n, A) \\
&= \sum_{j \in E} \mathbf{P}(X_{n+k+1} = i_{n+k+1} | X_{n+1} = j, X_n = i_n, A) \mathbf{P}(X_{n+1} = j | X_n = i_n, A)
\end{aligned}$$

On applique l'hypothèse de récurrence avec $n+1$ à la place de n et $A' = \{X_n = i_n\} \cap A$ à la place de A ($A' \in \mathcal{F}(X_0, \ldots, X_{n-1}, X_n)$), ce qui donne pour tout $j \in E$

$$\begin{aligned}
\mathbf{P}(X_{n+k+1} = i_{n+k+1} | X_{n+1} = j, X_n = i_n, A) &= \mathbf{P}(X_{n+k+1} = i_{n+k+1} | X_{n+1} = j, A') \\
&= \mathbf{P}(X_{n+k+1} = i_{n+k+1} | X_{n+1} = j)
\end{aligned}$$

et on applique l'hypothèse de récurrence avec $k = 1$, n et A pour obtenir

$$\mathbf{P}(X_{n+1} = j | X_n = i_n, A) = \mathbf{P}(X_{n+1} = j | X_n = i_n).$$

Finalement

$$\begin{aligned}
&\mathbf{P}(X_{(n+1)+k} = i_{(n+1)+k} | X_n = i_n, A) \\
&\qquad = \sum_{j \in E} \mathbf{P}(X_{(n+1)+k} = i_{(n+1)+k} | X_{n+1} = j) \mathbf{P}(X_{n+1} = j | X_n = i_n)
\end{aligned}$$

et par conséquent

$$\mathbf{P}(X_{n+(k+1)} = i_{n+(k+1))} | X_n = i_n, A) = \mathbf{P}(X_{n+(k+1)} = i_{n+(k+1)} | X_n = i_n).$$

□

Pour démontrer b) il suffit d'écrire

$$\begin{aligned}
&\mathbf{P}(X_{s_n} = i_{s_n}, X_{s_{n-1}} = i_{s_{n-1}}, \ldots, X_{s_1} = i_{s_1}) = \\
&\qquad \mathbf{P}(X_{s_n} = i_{s_n} | X_{s_{n-1}} = i_{s_{n-1}}, \ldots, X_{s_1} = i_{s_1}) \\
&\qquad \mathbf{P}(X_{s_{n-1}} = i_{s_{n-1}} | X_{s_{n-2}} = i_{s_{n-2}} \ldots X_{s_1} = i_{s_1}) \cdots \\
&\qquad \cdots \mathbf{P}(X_{s_2} | X_{s_1} = i_{s_1}) \mathbf{P}(X_{s_1} = i_{s_1})
\end{aligned}$$

et d'appliquer le lemme.

□

Il est temps de faire les exercices 3,4,5,6. Si vous avez peu de temps, on peut se borner par les exercices 3 et 4.

2.1.1 Propriétés de Markov

Nous démontrons dans cette section deux théorèmes qui seront d'un usage constant dans la suite dans les demonstrations de résultats.

Théorème 2.1.2 *Soient* $n \geq 1$, $A \in \mathcal{F}(X_0, \ldots, X_{n-1})$ *et* $B \in \mathcal{F}(X_{n+1}, X_{n+2}, \cdots)$. *Pour tout* $i \in E$ *on a*

$$\mathbf{P}(B|X_n = i, A) = \mathbf{P}(B|X_n = i).$$

Cette propriété signifie que le futur ne dépend du passé qu'à travers le présent.

Démonstration.— Le lemme 2.1.1 montre que le théorème est vrai pour tout événement B de la forme $\{X_{n+k} = j\}$. Montrons qu'il est vrai pour toutes leurs intersections finies $\{X_{n+k_1} = j_1\} \cap \{(X_{n+k_2} = j_2\} \cap \cdots \cap \{(X_{n+k_p} = j_p\}$:

$$\begin{aligned}
&\mathbf{P}(X_{n+k_p} = j_p, \ldots, X_{n+k_1} = j_1 | X_n = i, A) \\
&\quad = \mathbf{P}(X_{n+k_p} = j_p | X_{n+k_{p-1}} = j_{p-1} \ldots, X_{n+k_1} = j_1, X_n = i, A) \cdots \\
&\qquad \cdots \mathbf{P}(X_{n+k_2} = j_2 | X_{n+k_1} = j_1, X_n = i, A) \mathbf{P}(X_{n+k_1} = j_1 | X_n = i, A)
\end{aligned}$$

et l'utilisation du lemme 2.1.1 démontre que

$$\begin{aligned}
\mathbf{P}(X_{n+k_p} = j_p, \ldots, X_{n+k_1} = j_1 | X_n = i, A) &= \mathbf{P}(X_{n+k_p} = j_p | X_{n+k_{p-1}} = j_{p-1}) \cdots \\
&\cdots \mathbf{P}(X_{n+k_2} = j_2 | X_{n+k_1} = j_1) \mathbf{P}(X_{n+k_1} = j_1 | X_n = i)
\end{aligned}$$

c'est-à-dire en vertu du lemme 2.1.1

$$\begin{aligned}
&\mathbf{P}(X_{n+k_p} = j_p, \ldots, X_{n+k_1} = j_1 | X_n = i, A) \\
&\qquad\qquad = \mathbf{P}(X_{n+k_p} = j_p, \ldots, X_{n+k_1} = j_1 | X_n = i).
\end{aligned}$$

Ce qui précéde démontre que le théorème est vrai pour les événements B qui sont intersections finies d'événements de la forme $(X_k = i)$, $k \geq n+1$ et donc pour tout événement dans la tribu $\mathcal{F}(X_{n+1}, \ldots, X_{n+p})$. On constate que si B_l est une suite croissante d'événement de $\mathcal{B}$ pour lesquels le théorème est vrai, alors $\bigcup_l B_l$ vérifie encore le théorème. Un argument facile (de classe monotone) montre que tout événement de la tribu engendrée par les $\mathcal{F}(X_{n+1}, \ldots, X_{n+p})$, pour tout $p \geq 0$, vérifie le théorème. Or, cette tribu est $\mathcal{F}(X_k : k \geq n+1)$.

□

Opérateur de translation Ce paragraphe ne sera pas demandé dans les exercices. Notons $\mathcal{F}$ la tribu $\mathcal{F}(X_k : k \geq 0)$.

Proposition 2.1.1 *Pour tout $T \in \mathbf{N}$, il existe une application $\theta_T : (\Omega, \mathcal{F}) \to (\Omega, \mathcal{F})$ ($\mathcal{F}$-mesurable) définie de manière unique de la fa con suivante : pour tous n, $s_1 < \ldots < s_n$ et tous $i_1, \ldots, i_n \in E$*

$$\theta_T^{-1}(\{X_{s_1} = i_1, \ldots, X_{s_n} = i_n\}) = \{X_{s_1+T} = i_1, \ldots, X_{s_n+T} = i_n\}.$$

On appelle θ_T l'application de translation. C'est aussi l'unique application $\theta_T : (\Omega, \mathcal{F}) \to (\Omega, \mathcal{F})$ ($\mathcal{F}$-mesurable) telle que $X_n \circ \theta_T = X_{n+T}$ pour tout n.

Il est facile de voir que $\theta_T = \theta^T := \theta \circ \cdots \circ \theta$ (T fois) où $\theta = \theta_1$. Remarquons que $\theta_T^{-1}\mathcal{F}(X_0, \ldots, X_n) = \mathcal{F}(X_T, \ldots, X_{n+T})$. On peut faire agir cette application de translation sur les v.a : Si Y est une v.a on peut définir la v.a $Y \circ \theta_T$.

Exemple : Si $B = \{X_5 = a, X_3 = b\} \cup \{X_1 = c, X_3 = d\}$ alors $\theta^{-3}B = \{X_8 = a, X_6 = b\} \cup \{X_4 = c, X_6 = d\}$.

Théorème 2.1.3 *Pour tous événements $B \in \mathcal{F}(X_k : k \geq m)$, $i \in E$*

$$\mathbf{P}(\theta_1^{-n}B | X_{m+n} = i\}) = \mathbf{P}(B | X_m = i).$$

2.2 Matrices de transition et lois initiales

Définition 2.2.1 *On dit qu'une matrice $r \times r$ à coefficients réels P est stochastique si pour tous $1 \leq i, j \leq r$ on a*

a) $0 \leq P_{ij} \leq 1$.

b) pour tous i, $\sum_{j=1}^{r} P_{ij} = 1$ (somme sur les lignes).

Proposition 2.2.1 *Soit P une matrice stochastique. Pour toute loi de probabilité μ sur l'espace des états E il existe un espace probabilisé $(\Omega, \mathcal{F}, \mathbf{P}_\mu)$ et une suite de v.a $(X_n)_{n \geq 0}$ telle que (X_n) soit une chaîne de Markov admettant P comme matrice de transitions et μ comme loi initiale.*

Démonstration.— Nous donnons seulement une idée de la démonstration. L'espace des états Ω est l'ensemble des suites $\underline{e} = (e_k)_{k \geq 0} \in E^{\mathbf{N}}$ telles que pour tout $k \geq 0$, $P_{e_k, e_{k+1}} > 0$. L'espace des événements est la tribu engendrée par les cylindres

$$C(j_0, \ldots, j_n) = \{(e_k)_{k \geq 0} \in \Omega : e_0 = j_0, \ldots, e_n = j_n\}$$

où n parcourt $\mathbf{N}$ et $j_1, \ldots, j_n$ parcourt E^n. On peut alors démontrer [2] qu'il existe une *unique* mesure de probabilité $\mathbf{P}_\mu$ définie sur $\mathcal{F}$ telle que pour tout cylindre $C(j_0, \ldots, j_n)$

$$\mathbf{P}(C(j_0, \ldots, j_n)) = \mu_{j_0} P_{j_0 j_1} \cdots P_{j_{n-1} j_n}.$$

Si on définit $X_n : \Omega \to E$ par $X_n(\underline{e}) = e_n$ on voit facilement que les X_n sont des v.a et que la suite $(X_n)_{n \geq 0}$ est une chaîne de Markov de loi initiale μ et de matrice de transitions P.

□

2.3 Simulation d'une Chaine de Markov

Une autre fa con de définir une chaîne de Markov discrète est de la définir comme une suite de variables aléatoires $(X_n)_{n \geq 0}$ définie par

$$X_{n+1} = F(X_n, U_n), \qquad X_0$$

où la suite de variables aléatoires $X_0, U_0, U_1, \ldots, U_n, \ldots$ est indépendante les variables aléatoires U_i étant à valeurs dans un ensemble B et $F : E \times B \to E$ est une application (mesurable). Montrons l'équivalence de ces deux définitions. Nous aurons besoin du lemme suivant dont la preuve est une conséquence immédiate de la proposition 1.6.1

Lemme 2.3.1 *U_n est indépendant de la tribu $\mathcal{F}(X_n, \ldots, X_0) \subset \mathcal{F}(X_0, U_0, U_1, \ldots, U_{n-1})$.*

Notons alors pour $i, j \in E$

$$T_{i \to j} = \{b \in B, F(i, b) = j\}.$$

On a

$$\begin{aligned}
\mathbf{P}(X_{n+1} = j | X_n = i) &= \frac{\mathbf{P}(X_{n+1} = j, X_n = i)}{\mathbf{P}(X_n = i)} \\
&= \frac{\mathbf{P}(U_n \in T_{i \to j}, X_n = i)}{\mathbf{P}(X_n = i)} \\
&= \frac{\mathbf{P}(U_n \in T_{i \to j}) \mathbf{P}(X_n = i)}{\mathbf{P}(X_n = i)} \\
&= \mathbf{P}(U_n \in T_{i \to j})
\end{aligned}$$

2. Il s'agit d'un théorème de Kolmogorov (cf. aussi théorème de prolongement de Carathéodory)

où on a utilisé le fait que U_n et X_n sont indépendants. Calculons à présent

$$\mathbf{P}(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = \frac{\mathbf{P}(X_{n+1} = j, X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0)}{\mathbf{P}(X_n = i, X_{n-1} = i_{n-1} \ldots X_0 = i_0)}.$$

Si on note A l'événement $A = (X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0)$, on peut écrire

$$\begin{aligned}\mathbf{P}(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) &= \frac{\mathbf{P}(X_{n+1} = j, X_n = i, A)}{\mathbf{P}(A)} \\ &= \frac{\mathbf{P}(U_n \in T_{i\to j}, A)}{\mathbf{P}(A)}\end{aligned}$$

et puisque U_n est indépendant de la tribu $\mathcal{F}(X_n, \ldots, X_0)$ on a

$$\mathbf{P}(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = \mathbf{P}(U_n \in T_{i\to j}.$$

On a donc démontré que pour tous $i, j \in E$

$$\mathbf{P}(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = \mathbf{P}(X_{n+1} = j | X_n = i),$$

ce qui signifie que (X_n) est une chaîne de Markov.

Démontrons la réciproque : toute chaîne de Markov (avec un nombre fini d'états) peut se représenter sous la forme d'un système dynamique stochastique. Introduisons les probabilités de transitions

$$p_{ij} = \mathbf{P}(X_{n+1} = j | X_n = i),$$

et la loi $p^{(0)}$ de X_0. Nous assimilerons notre chaîne de Markov à ces données. Nous allons en fait démontrer qu'étant données une telle famille de probabilités de transitions et la loi de X_0, on peut construire un espace B, une fonction F et des v.a. $U_1, \ldots, U_n, \ldots$ telles que : $X_0, U_1, \ldots, U_n, \ldots$ soient indépendantes et $X_{n+1} = F(X_n, U_n)$. Pour cela définissons B comme étant l'ensemble des r-uplets $u = (u^{(1)}, \ldots, u^{(r)})$ $(r = \#E)$ où $u^{(i)} \in E$ et soit $F : E \times B \to E$ définie par $F(i, u) = u^{(i)}$ si $i \in E$ et $u = (u^{(1)}, \ldots, u^{(r)})$. D'après un théorème général de probabilités, on sait qu'étant données des lois de probabilités $p^{(i)}$, il est toujours possible de construire un espace de probabilité $(\Omega, \mathcal{B}, \mathbf{P})$ et des v.a U_i *indépendantes* admettant les $(p_j^{(i)})_{j\in E}$ comme lois (nous supposerons les U_i indépendantes de X_0 également). Définissons par récurrence $X_{n+1} = F(X_n, U_n)$, X_0 étant donné. On a alors,

$$(X_{n+1} = j, X_n = i) = (U_n^{(i)} = j) \cap (X_n = i);$$

on sait que U_n et X_n sont indépendantes (lemme 2.3.1) et donc,

$$\mathbf{P}(X_{n+1} = j | X_n = i) = \frac{\mathbf{P}((X_{n+1} = j, X_n = i)}{\mathbf{P}(X_n = i)} = \mathbf{P}(U_n^{(i)} = j) = p_{ij}.$$

Ce résultat est particulièrement utile en simulation. Supposons qu'on veut simuler une suite de v.a. $X_0, X_1, \ldots, X_n, \ldots$ sur un espace E qui est une chaine de Markov de matrice de transition P. D'abord on simule X_0 de loi initiale donnée. Lorsque $X_0, \ldots, X_n$ sont simulés, on simule X_{n+1} de manière suivante. Si $X_n = i$, on divisie le segment $[0,1]$ en segments Δ_k^i de longueur $p_{i,k}$ pour $k \in E$. La somme des longueurs $\sum_{k \in E} |\Delta_k^i| = \sum_{k \in E} p_{i,k} = 1$. On simule une v.a. U_{n+1} de loi uniforme sur $[0,1]$ indépendante des précédentes $U_1, \ldots, U_n$. On pose alors $X_{n+1} = j$ si $U_{n+1} \in \Delta_j^i$. Alors $P(X_{n+1} = j \mid X_n = i) = P(U_{n+1} \in \Delta_j^i) = p_{i,j}$.

2.4 Classe d'états, Irréductibilité

Définition 2.4.1 *On dit qu'un état i conduit à un état j s'il existe n tel que $(P^n)_{ij} > 0$ c'est-à-dire s'il existe n tel que $\mathbf{P}(X_n = j | X_0 = i) > 0$.*

!! Pour le voir de manière pratique, sans calculer les puissances de P, remarquons que i mène à j ssi ils existent des états $j_1, j_2, \ldots, j_k$ tels que le produit des éléments de la matrice P : $P_{i,j_1} P_{j_1,j_2} \cdots P_{j_k,j} > 0$, cad tous les éléments de ce produit sont > 0.

On dit que i et j sont dans la même classe et on écrit $i\mathcal{R}j$ si i conduit à j et j conduit à i. C'est une relation d'équivalence et il est pertinent de considérer ses classes d'équivalence disjoints de E.

Définition 2.4.2 *Une classe d'états $C \subset E$ est dite fermée si pour tout état j tel que il existe un état $i \in C$ qui conduit à j, on a $j \in C$.*

Autrement dit, une classe C est fermée, si la chaine de Markov quitte cette classe avec probabilité 0.

Définition 2.4.3 *Lorsque l'espace d'états se compose d'une seule classe (forcément fermée), cad tous les états communiquent, la Chaine de Markov est dite irréductible.*

Définition 2.4.4 *Un état i tel que $p_{i,i} = 1$ est dit absorbant.*

Il est temps de faire l'exercice 7A et 7B. Si vous avez peu de temps, vous pouvez juste identifier toutes les classes pour les matrices Q_1, Q_2, Q_3 de l'ex. 7B.

2.5 Récurrence et transience

2.5.1 Temps et nombre de retours

Définition 2.5.1 *Si $A \subset E$, on appelle*

(a) Temps d'entrée dans A la v.a $T_A : \Omega \to \mathbf{N} \cup \{\infty\}$ définie par $T_A := \inf\{n \geq 0 : X_n \in A\}$ (donc $T_A = 0$ si $X_0 \in A$).

(b) Temps de retour dans A la v.a $S_A : \Omega \to \mathbf{N} \cup \{\infty\}$ définie par $S_A := \inf\{n \geq 1 : X_n \in A\}$.

(c) Temps de retours successifs dans A les v.a $S_A^{(0)} = 0$ et par récurrence

$$S_A^{(k+1)} = \inf\{n > S_A^{(k)} : X_n \in A\}.$$

(d) Nombre de visites du processus à A la v.a

$$N_A = \sum_{k=0}^{\infty} \mathbf{1}_A \circ X_k. = \#\{k \geq 0 : X_k \in A\}.$$

Proposition 2.5.1 *Les v.a S_A, T_A et $S_A^{(k)}$ sont des temps d'arrêt c'est-à-dire que pour tout n les événements $\{S_A = n\}$ (resp. $\{T_A = n\}$, $\{S_A^{(k)} = n\}$) appartiennent à $\mathcal{F}_n = \sigma(X_0, \ldots, X_n)$.*

Démonstration.— Démontrons le par exemple pour S_A : l'événement $\{S_A = k\}$ est $\{X_0 \in A, X_1 \notin A, \ldots, X_{k-1} \notin A, X_k \in A\}$ qui est bien dans $\mathcal{F}_k$.

□

On note dans la suite $S(i) = S_{\{i\}}$, $S^{(k)}(i) = S_{\{i\}}^{(k)}$ et $N(i) = N_{\{i\}}$.

Définition 2.5.2 *On dit que l'état $i \in E$ est récurrent si $a(i) = \mathbf{P}_i(S(i) < \infty) = 1$ [3] , transient si $a(i) = \mathbf{P}_i(S(i) < \infty) < 1$.*

Le résultat important est alors

Théorème 2.5.1 *(a) Si $i \in E$ est récurrent alors $\mathbf{P}_i(N(i) = \infty) = 1$ tandis que si i est transient $\mathbf{P}_i(N(i) = \infty) = 0$ (c'est-à-dire $\mathbf{P}_i(N(i) < \infty) = 1$).*

b) Si $i \in E$ est transient alors sous $\mathbf{P}_i$ la v.a $N(i)$ suit une loi géometrique de paramètre $a(i)$: $\mathbf{P}_i(N(i) = k | X_0 = i) = (1 - a(i))a(i)^{k-1}$.

Démonstration.— Soit donc une chaîne de Markov de loi initiale δ_i (ce qui signifie $\mathbf{P}(X_0 = i) = 1$). Notons $R(n, i)$ l'événement $\{S^{(n)}(i) < \infty\}$: "il y a au moins n retours en i" et $R(n-1, i, t)$ l'événement $\{S^{(n-1)}(i) <$

3. la notation $\mathbf{P}_i$ est seulement là pour rappeler que la chaîne de Markov que l'on considère admet δ_i pour loi initiale ; en d'autres termes $\mathbf{P}_i(A) = \mathbf{P}(A | X_0 = i)$

$\infty, S^{(n-1)}(i) = t\}$: "il y a au moins $n-1$ retours en i et le $(n-1)$-ième a lieu au temps t". Ce dernier événement appartient à $\mathcal{F}_t$. Par conséquent, si on note $S_t(i) = \inf\{n \geq t+1 : X_n = i\}$ on a

$$\begin{aligned}\mathbf{P}(R(n,i)) &= \sum_{t=0}^{\infty} \mathbf{P}(R(n-1,i), S^{(n-1)}(i) = t, X_t = i, S_t(i) < \infty) \\ &= \sum_{t=0}^{\infty} \mathbf{P}(S_t(i) < \infty | R(n-1,i,t), X_t = i)\mathbf{P}(R(n-1,i,t), X_t = i)\end{aligned}$$

et d'après les théorèmes 2.1.2 et 2.1.3 (on remarque que $R(n-1,i,t) \subset \{X_t = i\}$, $\{S_t(i) < \infty\} \in \mathcal{F}(X_{t+1}, X_{t+2}, \ldots)$)

$$\begin{aligned}\mathbf{P}(R(n,i)) &= \sum_{t=0}^{\infty} \mathbf{P}(S_t(i) < \infty | X_t = i)\mathbf{P}(R(n-1,i,t), X_t = i) \\ &= \sum_{t=0}^{\infty} \mathbf{P}(S_0(i) < \infty | X_0 = i)\mathbf{P}(R(n-1,i,t)) \qquad \text{(Th. 2.1.3)} \\ &= \mathbf{P}(S(i) < \infty | X_0 = i) \sum_{t=0}^{\infty} \mathbf{P}(R(n-1,i,t)) \\ &= \mathbf{P}(S(i) < \infty | X_0 = i)\mathbf{P}(R(n-1,i))\end{aligned}$$

Si on note $a(i) = \mathbf{P}_i(S(i) < \infty)$ on a donc

$$\mathbf{P}(R(n,i)) = a(i)^n.$$

Par conséquent, si $a(i) = 1$ chaque événement $R(n,i)$ a une probabilité égale à 1 et il en est donc de même de l'intersection dénombrable $\cap_{n\geq 1} R(n,i)$ qui n'est rien d'autre que l'événement "il y a une infinité de retours en i". En revanche, si $a(i) < 1$, on a $\mathbf{P}_i(N(i) > n) = \mathbf{P}_i(R(n,i)) = a(i)^n$ et $\mathbf{P}_i(N(i) = n) = \mathbf{P}_i(N(i) > n-1) - \mathbf{P}_i(N(i) > n) = (1-(a(i))a(i)^{n-1}$. Ainsi, $N(i)$ suit une loi géométrique et $\mathbf{P}_i(N(i) = \infty) = 0$ (car $\sum_{n=1}^{\infty} \mathbf{P}_i(N(i) = n) = 1$).

□

Proposition 2.5.2 *Notons $N_i = N_{\{i\}}$. On a $\mathbf{E}_i(N_i) = V_{ii}$ où V est la matrice $r \times r$*

$$V = I + P + P^2 + \cdots = \sum_{k=0}^{\infty} P^k.$$

Démonstration.— Le nombre N_i de visites à i est la v.a $N_i = \sum_{k\geq 0} \mathbf{1}_{\{i\}} \circ X_k$

et son espérance est

$$\begin{aligned}
\mathbf{E}(N_i) &= \sum_{k\geq 1} \mathbf{E}(\mathbf{1}_{\{i\}} \circ X_k) \\
&= \sum_{k\geq 0} \mathbf{P}(X_k = i) \\
&= \sum_{k\geq 0} \mathbf{P}(X_k = i | X_0 = i) \qquad\qquad (\mathbf{P}(X_0 = i) = 1) \\
&= \sum_{k\geq 0} (P^k)_{ii}.
\end{aligned}$$

Ainsi, si on pose

$$V = I + P + P^2 + \cdots = \sum_{k=0}^{\infty} P^k$$

on a

$$\mathbf{E}_i(N_i) = V_{ii}.$$

□

Corollaire 2.5.1 *L'état $i \in E$ est récurrent ($\mathbf{E}_i(N_i) = \infty$) si et seulement si*

$$\sum_{k\geq 0} (P^k)_{ii} = \infty.$$

Proposition 2.5.3 *Si E est fini, il existe au moins un état récurrent.*

Exercice : Définissons N_{ij} le nombre de visites en j d'une chaîne commen cant en i.

a) Démontrer que

$$\mathbf{E}(N_{ij}) = V_{ij}.$$

b) Prouver que

$$V_{ij} = \mathbf{P}_i(S_j < \infty) V_{jj}.$$

Remarque. Remarquons que une classe d'états qui communiquent contenant un état récurrent i ($\sum_{k=0}^{\infty} (P^k)_{ii} = \infty$) est composé d'états récurrents puisque si $(P^k)_{ij} > 0$, $(P^l)_{ji} > 0$ et $(P^m)_{ii} > 0$

$$(P^{k+l+m})_{jj} \geq (P^l)_{ji} (P^k)_{ii} (P^m)_{ij}.$$

On a donc

$$\sum_{k=0}^{\infty} (P^k)_{jj} \geq \sum_{k=0}^{\infty} (P^{k+l+m})_{jj} \geq (P^l)_{ji} (P^m)_{ij} \sum_{k=0}^{\infty} (P^k)_{ii} = \infty$$

car i est récurrent. D'après le critère de récurrence j est également récurrent.

De même, la classe $C(i)$ d'un état transient i ne contient que des états transients puisque si dans le cas contraire un tel élément $j \in C(i)$ était récurrent sa classe ne contiendrait que des éléments récurrents et en particulier i serait récurrent.

Corollaire. *Dans une classe d'états qui communiquent, tous les états sont récurrents ou tous sont transients.*

Proposition 2.5.4 *L'espace des états E se décompose en deux types de classes : les classes récurrentes et les classes transitoires. Si C est une classe et $i \in C$ alors,*

– si C est récurrente elle est fermée c'est-à-dire que tout élément de i reste dans C $\mathbf{P}$-p.s :

$$\mathbf{P}_i(\exists n \geq 1 : X_n \notin C) = 0.$$

– si C est transiente et finie , $\mathbf{P}$-p.s tout élément quitte C définitivement à partir d'un certain temps :

$$\mathbf{P}_i(\{\exists N,\ \forall n \geq N : X_n \notin C\}) = 1.$$

Démonstration.—

a) Supposons C récurrente et $i \in C$. L'événement $\{X_n \notin C\}$ implique (i.e est inclus dans) l'événement $\bigcap_{k\geq n}\{X_k \notin C\}$ (ce qui signifie qu'une fois sorti de C on n'y revient plus jamais). En effet, dans le cas contraire, il existe $j \notin C$, $i' \in C$, $0 < n < n'$ tels que l'événement $\{X_0 = i, X_n = j, X_{n'} = i'\}$ est de probabilité positive. Ceci implique que i conduit à j et que j conduit à i'. Comme i et i' sont dans la même classe cela signifie que j conduit à i et i conduit à j : mais alors $j \in C$ ce qui est une contradiction. Maintenant, si la classe est récurrente on sait que $\mathbf{P}_i$-p.s. X_n visite une infinité de fois C et par conséquent pour tout n l'événement $\bigcap_{k\geq n}\{X_k \notin C\}$ est de probabilité nulle au même titre que $\{X_n \notin C\}$. L'union sur n de ces événements est encore de probabilité nulle.

b) On a vu que

$$\mathbf{E}(N_i) = \mathbf{E}(\sum_{k=0}^{\infty} \mathbf{1}_{\{i\}} \circ X_k) = V_{ii}\mathbf{P}(X_0 = i).$$

où $V_{ii} = \sum_{k=0}^{\infty}(P^k)_{ii}$. On sait en outre que l'état i est transient si et seulement si V_{ii} est fini Par conséquent, comme $\mathbf{1}_C = \sum_{i\in C} \mathbf{1}_i$

$$\mathbf{E}(N_C) = \sum_{i\in C} V_{ii}\mathbf{P}(X_0 = i),$$

et on voit que si la classe C est transiente (donc composée uniquement d'états transients) cette somme est finie. Ceci démontre que N_C est $\mathbf{P}$-p.s fini.

□

! ! Par conséquent, pour une Chaine de Marlkov finie, une classe d'états fermée est récurrente, une classe non fermée est transiente. Une Chaine de Markov finie(!) irréductible se compose d'une seule classe fermée, elle est donc récurrente.

Il n'est pas difficile de voir que

Corollaire 2.5.2 *Si une chaîne de Markov est finie,* $\mathbf{P}$*-p.s. tout état aboutit à une classe récurrente.*

En effet, tout état transient est visité un nombre fini de fois p.s. Comme dans un espace d'états fini, il y a un nombre fini d'états transients, l'ensemble d'états transients ne peut pas être visité une infinité de fois avec probabilité positive. Par conséquent, la chaine de Markov aboutit à une classe récurrente p.s. Les probabilités avec lesquelles elle aboutit a des différentes classes récurrentes, ainsi que l'espérance de ce temps seront le sujet de la section suivante.

Un autre corollaire est

Corollaire 2.5.3 *(i) La restriction de la chaîne aux seuls états d'une classe récurrente (ou plus généralement close) est encore de Markov, ce qui signifie que la matrice* $\tilde{P}_{ij} = P_{ij},\ i, j \in C$ *est stochastique.*

(ii) Si la Chaine de Markov se retrouve dans cette classe récurrente, elle visite chaque état de cette classe une infinité de fois p.s.

Démonstration.— Pour (i) il suffit de vérifier que si $i \in C$ $\sum_{j \in C} p_{ij} = 1$, ce qui résulte de $\mathbf{P}(X_1 \notin C | X_0 = i) = p_{ij} = 0$ si $i \in C$ et $j \notin C$.

Pour prouver (ii) , comme tout état est récurrent dans cette classe, il suffit de montrer que $\mathbf{P}(T^j < \infty \mid X_0 = i) = 1$ pour tous i et j dans cette classe. On choisit m avec $(P^m)_{i,j} > 0$. On a

$$1 = \mathbf{P}(X_n = j \text{ pour une infinité de } n \mid X_0 = j) \le \mathbf{P}(X_n = j \text{ pour un } n \ge m{+}1 \mid X_0 = j)$$

$$= \sum_{k \in I} \mathbf{P}(X_n = j \text{ pour un } n \ge m + 1 \mid X_m = k, X_0 = j)\mathbf{P}(X_m = k \mid X_0 = j)$$

$$= \sum_{k \in I} \mathbf{P}(T^j < \infty \mid X_0 = k)(P^m)_{j,k}$$

où la dernière égalité vient de la propriété de Markov. Comme $\sum_{k \in I} (P^m)_{j,k} = 1$, $(P^m)_{j,k} \ge 0$ et $(P^m)_{i,j} > 0$, on doit avoir $\mathbf{P}(T_j < \infty \mid X_0 = i) = 1$.

□

Il est temps de faire Exercice 9 a).

2.5.2 Temps et probabilité d'atteinte

Si $A \subset E$ rappelons que le temps d'atteinte (ou d'entrée) de A est la v.a. $T_A := \inf\{n \geq 0 : X_n \in A\}$. Nous définirons également la probabilité d'atteinte de A à partir de $i \in E$ de la facon suivante :

$$\nu_A(i) = \mathbf{P}(\exists n \geq 0 : X_n \in A | X_0 = i).$$

Théorème 2.5.2 *Les fonctions $\nu : E \to \mathbf{R}$ et $(i \mapsto \mathbf{E}_i(T_A)) : E \mapsto \mathbf{R}$ vérifient pour $i \notin A$ les équations*

$$(\Delta \nu)(i) = 0, \qquad \Delta(\mathbf{E}_i(T_A))(i) = -\nu_A(i)$$

où $(\Delta f)(i) = -f(i) + \sum_{j \in E} P_{ij} f(j)$

Démonstration.— Notons pour $m < n$, $A_m^n(i)$ l'événement $\{X_m = i, X_{m+1} \notin A, \ldots, , X_{n-1} \notin A, X_n \in A\}$. On a si $i \notin A$

$$\begin{aligned}
\nu_A(i) &= \sum_{n=1}^{\infty} \mathbf{P}_i(A_0^n(i)) \\
&= \sum_{n=2}^{\infty} \sum_{j \in E-A} \frac{\mathbf{P}(A_1^n(j), X_1 = j, X_0 = i)}{\mathbf{P}(X_0 = i)} + \sum_{j \in A} \mathbf{P}(X_1 = j | X_0 = i) \\
&= \sum_{n=2}^{\infty} \sum_{j \in E-A} \mathbf{P}(A_1^n(j) | X_1 = j, X_0 = i) \mathbf{P}(X_1 = j | X_0 = i) + \sum_{j \in A} P_{ij} \\
&= \sum_{j \in E-A} \sum_{n=2}^{\infty} \mathbf{P}(A_1^n(j) | X_1 = j) P_{ij} + \sum_{j \in A} P_{ij} \\
&= \sum_{j \in E-A} \sum_{n=1}^{\infty} \mathbf{P}(A_0^{n-1}(j) | X_0 = j) P_{ij} + \sum_{j \in A} P_{ij} \\
&= \sum_{j \in E-A} P_{ij} \nu_A(j) + \sum_{j \in A} P_{ij}
\end{aligned}$$

Comme $\sum_{j \in A} P_{ij} \nu_A(j) = \sum_{j \in A} P_{ij}$ (car $\nu_A(j) = 1$ si $j \in A$) on a bien pour $i \notin A$

$$\nu_A(i) = \sum_{j \in E} P_{ij} \nu_A(j).$$

De même si $i \notin A$

$$\begin{aligned}
\mathbf{E}_i(T_A) &= \sum_{k=0}^{\infty} k\mathbf{P}_i(T_A = k) \\
&= \sum_{k=1}^{\infty} \sum_{j \in E-A} k\mathbf{P}(T_A = k, X_1 = j | X_0 = i) + \sum_{j \in A} 1 \cdot \mathbf{P}(T_A = 1, X_1 = j | X_0 = i) \\
&= \sum_{k=1}^{\infty} \sum_{j \in E-A} k\mathbf{P}(T_A = k, X_1 = j, X_0 = i)\mathbf{P}(X_1 = j | X_0 = i) + \sum_{j \in A} P_{ij}
\end{aligned}$$

et donc

$$\begin{aligned}
\mathbf{E}_i(T_A) &= \sum_{k=1}^{\infty} \sum_{j \in E-A} k\mathbf{P}(T_A = k | X_1 = j) P_{ij} + \sum_{j \in A} P_{ij} \\
&= \sum_{k=1}^{\infty} \sum_{j \in E-A} (k - 1 + 1)\mathbf{P}(T_A = k - 1 | X_0 = j) P_{ij} + \sum_{j \in A} P_{ij} \\
&= \sum_{l=0}^{\infty} \sum_{j \in E-A} l\mathbf{P}(T_A = l | X_0 = j) P_{ij} + \sum_{j \in E-A} \sum_{k=1}^{\infty} \mathbf{P}(T_A = k - 1 | X_0 = j) P_{ij} + \sum_{j \in A} P_{ij} \\
&= \sum_{j \in E-A} \mathbf{E}_j(T_A) P_{ij} + \sum_{j \in E-A} P_{ij}\nu_A(j) + \sum_{j \in A} P_{ij}\nu_A(j) \quad (\text{car } \nu_A(j) = 1 \text{ si } j \in A) \\
&= \sum_{j \in E} \mathbf{E}_j(T_A) P_{ij} + \sum_{j \in E} P_{ij}\nu_A(j) \quad (\text{car } T_A = 0 \text{ si } j \in A) \\
&= \sum_{j \in E} \mathbf{E}_j(T_A) P_{ij} + \nu_A(i) \quad (\text{d'après } (\Delta\nu_A)(i) = 0)
\end{aligned}$$

□

Il est temps de faire les exercices 8 et 9 entièrement.

2.5.3 Exemple (non-obligatoire) de la ruine du joueur

Il s'agit de la chaîne de markov sur $E = \{0, 1, \ldots, r\}$ avec barrières absorbantes : les transitions sont $p_{0j} = \delta_{0,j}$, $p_{r,j} = \delta_{r,j}$ et pour $i \notin \{0, r\}$ $p_{i,j} = p\delta_{p,p+1} + (1 - p)\delta_{p,p-1}$. Elle modélise le jeu suivant. Le joueur arrive au casino avec $0 \leq k \leq r$ euros et mise euro par euro ; s'il gagne, la banque lui donne un euro, sinon il perd le sien, la probabilité de gain à chaque mise étant p et celle de perte $1 - p$. Le jeu s'arrête quand le joueur arrive à 0 euro (ruine) ou à une somme de r euros (victoire). La probabilité $\nu_0(k)$ (resp. $\nu_r(k)$) d'atteinte de 0 (resp. r) à partir de k est la probabilité de ruine (resp. victoire) d'un joueur dont la mise initiale est de k. La probabilité $\nu_\infty(k) = 1 - (\nu_0(k) + \nu_r(k))$ est la probabilité qu'il n'y ait ni ruine ni gain, et donc jeu infini : elle est en fait nulle comme nous allons le voir.

D'après le théorème précédent, on a pour tout $k \neq 0, r$ l'équation

$$\nu_0(k) = p\nu_0(k+1) + (1-p)\nu_0(k-1)$$

à laquelle il faut adjoindre les *conditions aux limites* $\nu_0(0) = 1$, $\nu_0(r) = 0$ (cette dernière condition étant claire puisque r est absorbant). On aura de même

$$\nu_r(k) = p\nu_r(k+1) + (1-p)\nu_r(k-1), \qquad \nu_r(0) = 0, \quad \nu_r(r) = 1.$$

Démontrons que la probabilité $\nu_\infty(k)$ d'un jeu infini est nulle pour tout k.

Lemme 2.5.1 *Si une suite de nombre réels vérifie pour tout $k \neq 0, r$ l'équation $(\Delta x)_k := -x_k + px_{k+1} + (1-p)x_{k-1} = 0$ avec les conditions aux limites $x_0 = 0$, $x_r = 0$ alors pour tout $k = 0, \ldots, r$ $x_k = 0$.*

Démonstration.— La preuve repose sur le "principe du maximum" : Si la suite n'est pas nulle, alors le maximum x_{k^*} de x_k, $k \in E$ est non nul et atteint en $k^* \neq 0, r$. On a alors

$$|x_{k^*}| \leq p|x_{k^*+1}| + (1-p)|x_{k^*-1}| \leq \max(|x_{k^*+1}|, |x_{k^*-1}|) \leq |x_{k^*}|$$

si bien que $|x_{k^*}| = |x_{k^*+1}| = |x_{k^*-1}|$ et itérant l'argument avec $k^* \pm 1$ *etc.* on voit que pour tout $k \in E$, $|x_k| = |x_{k^*}|$. En particulier, $0 = |x_0| = |x_{k^*}| \neq 0$: une contradiction.

□

Ainsi, $\nu_\infty(k) = 0$ pour tout k.

Calculons à présent $\nu_0(k)$. Commencons par le cas particulier où $p = 1/2$. On doit dans ce cas résoudre en posant $x_k = \nu_0(k)$

$$x_k = \frac{1}{2}x_{k+1} + \frac{1}{2}x_{k-1}, \ 1 \leq k \leq r-1, \qquad x_0 = 1, \quad x_r = 0.$$

On constate que $y_k = x_k - (r-k)/r$ vérifie $\Delta y = 0$ mais avec conditions aux limites nulles. Le lemme permet de conclure que $y_k = 0$ c'est-à-dire

$$\nu_0(k) = x_k = \frac{r-k}{r}.$$

Si $p \neq 1/2$ la méthode est moins élégante : on voit que x_k satisfait l'équation aux différences secondes

$$x_{k+1} = \frac{1}{p} x_k - \frac{1-p}{p} x_{k-1} \qquad 0 \leq k \leq r-1$$

ce qu'on peut récrire

$$\begin{pmatrix} x_{k+1} \\ x_k \end{pmatrix} = \begin{pmatrix} \frac{1}{p} & -\frac{1-p}{p} \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_k \\ x_{k-1} \end{pmatrix}$$

ou encore si on note $X_k = \begin{pmatrix} x_{k+1} \\ x_k \end{pmatrix}$

$$X_k = A^k X_0, \qquad A = \begin{pmatrix} \frac{1}{p} & -\frac{1-p}{p} \\ 1 & 0 \end{pmatrix}.$$

Les valeurs propres de A sont 1 et $(1-p)/p$. Puisqu'elles sont distinctes A est diagonalisable et il est facile de voir que x_k est de la forme $x_k = a + b[(1-p)/p]^k$ où a, b sont à déterminer. Comme $x_0 = 1$, $x_r = 0$ on en déduit

$$b = 1 - a = \frac{1}{1 - \left(\frac{1-p}{p}\right)^r},$$

soit

$$x_k = \frac{(\frac{p}{1-p})^{r-k} - 1}{(\frac{p}{1-p})^r - 1}.$$

On peut également calculer le temps moyen de jeu avant arrêt (c'est-à-dire ruine ou victoire) $\tau(i) = \mathbf{E}_i(T_{\{0,r\}})$. Comme $\nu_{\{0,r\}}(k) = \nu_0(k) + \nu_r(k) = 1$ pour tout k, on a à présent,

$$\Delta\tau = -1, \qquad \tau(0) = \tau(r) = 0.$$

Si $p = 1/2$ cette équation se résout de la fa con suivante : on constate que $\Delta(kr - k^2) = -1$ et donc en notant $y_k = \tau(k) - k(r-k)$ on voit que $\Delta y = 0$ et $y(0) = y(r) = 0$ si bien que y est identiquement nul :

$$\tau(k) = k(r-k).$$

Le cas $p \neq 1/2$ est plus fastidieux à traiter mais ne pose aucune difficulté.

2.5.4 Marches aléatoires symétriques dans $\mathbf{Z}^d$, exemple non-obligatoire.

L'espace des états est $E = \mathbf{Z}^d$ et la matrice de transition est $P_{ij} = 1/2^d$ si $\sum_{s=1}^d |i_s - j_s| = 1$ (c'est-à-dire si i et j sont voisins). On peut voir une telle marche comme le système dynamique stochastique $X_{n+1} = X_n + U_n$ où $U_1, \ldots, U_n, \ldots$ est une suite de vecteurs aléatoires à valeurs dans $\{1, -1\}^d$ indépendants et de même loi uniforme. Il est clair que tous les états communiquent et il n'y a donc qu'une seule classe. Comme E est infini on ne peut pas conclure que cette classe est récurrente. En fait, on va voir que cela dépend de la dimension d. Puisque tous les états communiquent, pour déterminer si l'unique classe est récurrente, il suffit de déterminer le caractère récurrent ou transient de 0, ce qui est équivalent à la divergence ou la convergence de $\sum_{k=0}^\infty (P^k)_{00}$. Estimons pour cela $(P^k)_{00}$, la probabilité de retour en 0 en exactement n pas ; c'est aussi la probabilité pour que $U_1 + \cdots + U_n = 0$ et c'est donc la probabilité de l'événement

$$\bigcap_{s=1}^d \{U_1^{(s)} + \cdots + U_n^{(s)} = 0\}$$

où $U_l^{(s)}$ est la s-ième composante du vecteur U_l. Comme les composantes des vecteurs $U_1, \cdots, U_n$ sont également indépendantes (**Pourquoi ?**) on peut écrire

$$(P^n)_{00} = \mathbf{P}(U_1^{(1)} + \cdots + U_n^{(1)} = 0)^d.$$

La v.a $Y_n = U_1^{(1)} + \cdots + U_n^{(1)}$ est une somme de n v.a de Bernoulli (à valeurs dans $\{1, -1\}$ indépendantes et par conséquent, $(Y_n + n)/2$ suit une loi Binomiale $(n, 1/2)$. Ainsi, si n est impair $\mathbf{P}(Y_n = 0) = 0$ tandis que si $n = 2m$ est pair

$$\mathbf{P}(Y_{2m} = 0) = \binom{2m}{m} (\frac{1}{4})^m,$$

et la formule de Stirling $n! \sim n^n e^{-n} \sqrt{2\pi n}$ montre que

$$\mathbf{P}(Y_{2m} = 0) \sim \frac{1}{\sqrt{\pi m}}.$$

Par conséquent,

$$(P^{2m})_{00} \sim \left(\frac{1}{\sqrt{\pi m}}\right)^d.$$

Ainsi,

(a) si $d = 1$ ou $d = 2$ on a $d/2 \leq 1$ et la série $\sum_{n\geq 0}(P^n)_{00}$ diverge : 0 est récurrent au même titre que tout point de $\mathbf{Z}^d$

(b) si $d \geq 3$ $(d/2 > 1)$ la série $\sum_{n\geq 0}(P^n)_{00}$ converge et 0 est transient comme tout autre point de $\mathbf{Z}^d$.

2.6 Lois stationnaires

Définition 2.6.1 *On dit qu'une loi μ sur E est stationnaire (ou encore invariante)pour la matrice stochastique P si $\mu P = \mu$.*

Proposition 2.6.1 *La loi μ est stationnaire pour P si et seulement si la chaîne de Markov de loi initiale μ et de matrice de transitions P vérifie pour tous n, j*

$$\mathbf{P}(X_n = j) = \mu(j),$$

(cela signifie que la loi de X_n est également μ).

Proposition 2.6.2 *Soit une chaîne de Markov admettant une loi stationnaire μ. Alors, tout état i tel que $\mu(\{i\}) > 0$ est récurrent. Si en outre la chaîne est irréductible elle est récurrente.*

Démonstration.— On a $\mu P^k = \mu$ pour tout entier $k \geq 0$ et donc $\mu(I + \cdots + P^{n-1}) = n\mu$. Par conséquent $\sum_{j\in E}\mu_j V_{ji} = \infty$ pour i tel que $\mu(i) > 0$. On a vu précédemment (cf. exercice après la proposition 2.5.3) que pour tous i, j on a $V_{ji} \leq V_{ii}$ si bien que $\sum_{j\in E}\mu_j V_{ji} \leq V_{ii}$ car $\sum_{j\in E}\mu(j) = 1$. On a donc $V_{ii} = \infty$ et l'état i est donc récurrent.

□

Théorème 2.6.1 *Si l'espace des états est fini, toute chaîne de Markov sur E (de matrice stochastique P) admet au moins une loi stationnaire.*

Démonstration.— Si μ_0 est une loi quelconque sur E, par exemple la loi δ_i, notons $\mu_n = \mu P^n$ la loi de X_n où (X_n) est la chaîne de Markov de matrice P et de loi initiale μ_0. Notons encore $\tilde{\mu}_n$ la loi $(\mu_0 + \cdots + \mu_{n-1})/n$ si bien que

$$\tilde{\mu}_n = \frac{1}{n}(\mu_0 + \mu_0 P + \cdots + \mu_0 P^{n-1}).$$

Il existe une sous-suite n_k telle que pour tout $i \in E$ $\tilde{\mu}_{n_k}(i)$ converge vers un réel $\pi(i)$. Il est clair que les $\pi(i)$ sont dans $[0, 1]$ et vérifient $\sum_{i\in E}\pi(i) = 1$:

π est bien une loi de probabilité. Elle est stationnaire. En effet,

$$\begin{aligned}
\pi P &= \lim_{n_k\to\infty} \frac{1}{n_k}(\mu_0 + \mu_0 P + \cdots + \mu_0 P^{n_k-1})P \\
&= \lim_{n_k\to\infty} \frac{1}{n_k}(\mu_0 P + \mu_0 P^2 + \cdots + \mu_0 P^{n_k}) \\
&= \lim_{n_k\to\infty} \left(\frac{1}{n_k}(\mu_0 + \mu_0 P + \cdots + \mu_0 P^{n_k-1}) + \frac{1}{n_k}(\mu_0 P^{n_k} - \mu_0)\right) \\
&= \pi
\end{aligned}$$

□

Quand l'espace des états est infini, l'existence d'une telle loi de probabilité invariante n'est pas garantie. L'unicité n'est pas toujours vraie même si E est fini. On peut préciser un peu le théorème précédent :

Proposition 2.6.3 *Si la chaîne de Markov est irréductible (et finie) alors toute mesure de probabilité stationnaire a pour support l'espace des états.*[4] *En particulier, si i est un état récurrent il existe une mesure stationnaire μ dont le support est la classe $C(i)$ de i.*

Démonstration.— Soit l un état où $\mu(l) > 0$ (un tel état existe toujours puisque $\sum_k \mu(k) = 1$.) Si j est un état quelconque on sait qu'il existe n tel que $p_{lj}^{(n)} > 0$. Or, d'après l'invariance

$$\mu_j = \sum_{k\in E} \mu_k (P^n)_{kj} \geq \mu_l (P^n)_{lj} > 0.$$

La dernière partie de la proposition résulte du fait que si i est récurrent il est contenu dans une classe close et on peut se ramener au cas irréductible.

□

Théorème 2.6.2 *Si une chaîne de Markov est finie alors les assertions suivantes sont équivalentes : (a) elle est irréductible ; (b) elle admet une unique mesure de probabilité stationnaire.*

Démonstration.— Supposons la chaîne irréductible et faisons l'hypothèse qu'il existe deux mesures de probabilités μ_1, μ_2 différentes telles que $\mu_1 = \mu_1 P$ et $\mu_2 = \mu_2 P$. Si on appelle x le vecteur ligne $x = \mu_1 - \mu_2$ on a $x = xP$ et $\sum_{i\in E} x_i = 0$; en particulier il existe deux indices i, j tels que x_i et x_j sont de

4. Le support d'une mesure de probabilité μ sur E est l'ensemble de j où $\mu(j) > 0$.

signes opposés. Puisque la matrice P est irréductible il existe un exposant m tel que $(P^m)_{ij} > 0$. Ecrivons $x = xP^m$ puis,

$$\begin{aligned}
\sum_{j\in E} |x_j| &= \sum_{j\in E} |\sum_{k\in E} x_k (P^m)_{kj}| \\
&\leq \sum_{j\in E}\sum_{k\in E} |x_k||(P^m)_{kj}| \\
&\leq \sum_{k\in E}\sum_{j\in E} |x_k|(P^m)_{kj} \\
&\leq \sum_{k\in E} |x_k|
\end{aligned}$$

puisque la matrice P est stochastique ($\sum_{j\in E}(P^m)_{ij} = 1$). Or, cette dernière inégalité est une égalité ; par conséquent dans la suite d'inégalités que nous avons écrites nous avions en fait déjà des égalités. Il en résulte que pour tout j,

$$|x_j| = |\sum_{k\in E} x_k (P^m)_{kj}| = \sum_{k\in E} |x_k|(P^m)_{kj},$$

ce qui n'est possible que si les $|x_k|(P^m)_{kj}$, sont tous de même signe quand k varie dans le sous-ensemble de E constitué des k pour lesquels $|x_k|(P^m)_{kj}$ est non nul. Puisque $x_i \neq 0$ et que par définition de m le coefficient $(P^m)_{ij} > 0$, ceci entraîne que x_j est de même signe que x_i ce qui est une contradiction.

Supposons à présent que la mesure de probabilité stationnaire soit unique et supposons qu'il y ait deux classes C_1, C_2. Puisque les mesures de *probabilité* ne chargent pas les classes transientes (cf. proposition 2.6.2), on peut supposer que C_1 et C_2 sont récurrentes. Mais alors, on sait que si $i \in C_1$ et $j \in C_2$ on peut trouver deux mesures de probabilités stationnaires μ_1, μ_2 qui ont pour supports respectifs C_1 et C_2 (proposition 2.6.3) ; en particulier elles sont différentes.

□

!! Pour chercher une loi stationnaire, il faut chercher le vecteur ligne $\mu = (\mu_i)_{i\in E}$ avec $\mu_i \geq 0$ et $\sum_{i\in E} \mu_i = 1$ qui vérifie $\mu = \mu P$.

Dans le cas d'une chaine de Markov irréductible finie la solution de l'équation $\mu = \mu P$ est unique a une constante près. Pour avoir la loi de probabilité stationnaire, il faut la normaliser par la constante $\sum_{i\in E} \mu_i$.

Dans le cas d'une chaine de Markov finie et non irreductible $\mu_i = 0$ pour tout i transient. Alors le systeme $\mu = \mu P$ va se decomposer en systemes plus petites de type $\mu^C = \mu^c P^C$ où P^C est la restriction de P pour une classe recurrente C. La solution de chaque système est unique à une constante près. Leur combinaison va donner les lois stationnaires, voir les exercices.

Finallement, il est utile de se souvenir du Thm suivant.

Théorème 2.6.3 *Pour une chaine de Markov irréductible finie de loi de probabilité stationnaire μ, on a* $\mathbf{E}(T^i \mid X_0 = i) = (\mu^i)^{-1}$.

Démonstration.— Pour raccourcir les notations, on va écrire ici $\mathbf{E}(\cdot \mid X_0 = i) = \mathbf{E}_i(\cdot)$ et $\mathbf{P}(\cdot \mid X_0 = i) = \mathbf{P}_i(\cdot)$

On va prouver que le vecteurγ^i avec $\gamma^i_j = \mathbf{E}_i \sum_{k=0}^{T^i-1} \mathbf{1}_j \circ X_k$ pour $j \in E$ vérifie l'équation $\gamma^i P = \gamma^i$. Alors

$$\sum_{j\in E} \gamma^i_j = \sum_{j\in E} \mathbf{E}_i \sum_{k=0}^{T^i-1} \mathbf{1}_j \circ X_k = \mathbf{E}_i \sum_{k=0}^{T^i-1} \sum_{j\in E} \mathbf{1}_j \circ X_k = \mathbf{E}_i \sum_{k=0}^{T^i-1} 1 = \mathbf{E}_i T^i.$$

Donc l'unique loi stationnaire $\mu_j = \gamma^i_j / \mathbf{E}_i T^i$. Comme $\gamma^i_i = 1$, le Thm serait prouvé.

Pour prouver l'égalité $\gamma^k = \gamma^k P$, notons d'abord que $T^k < \infty$ p.s. pour tout k. Nous pouvons donc écrire

$$\gamma^k_j = \mathbf{E}_k \sum_{n=1}^{T^k} 1_{\{X_n = j\}} + \mathbf{P}_k(X_0 = j) - \mathbf{P}_k(X_{T^k} = j) = \mathbf{E}_k \sum_{n=1}^{T^k} 1_{\{X_n = j\}} \tag{2.2}$$

car $\mathbf{P}_k(X_0 = j) - \mathbf{P}_k(X_{T^k} = j)$ est $1-1 = 0$ si $j = k$ et $0-0 = 0$ si $j \neq k$. (Si la CM n'avait pas été récurrente, on aurait eu $\mathbf{P}(T^k < \infty) < 1$ et on n'aurait pas pu considérér X_{T^k} !) Alors,

$$\gamma^k_j = \mathbf{E}_k \sum_{n=1}^{\infty} 1_{\{X_n = j, n \leq T^k\}} = \sum_{n=1}^{\infty} \mathbf{P}_k(X_n = j, n \leq T^k). \tag{2.3}$$

Par la propriété de Markov

$$\begin{aligned}\mathbf{P}_k(X_{n-1} = i, X_n = j, T^k \geq n) &= \mathbf{P}_k(X_n = j \mid X_{n-1} = i, T^k \geq n)\mathbf{P}_k(X_{n-1} = i, T^k \geq n)\\ &= \mathbf{P}_k(X_n = j \mid X_{n-1} = i, X_1, \ldots, X_{n-2} \neq k)\mathbf{P}_k(X_{n-1} = i, T^k \geq n)\\ &= \mathbf{P}_k(X_n = j \mid X_{n-1} = i) P_k(X_{n-1} = i, T^k \geq n) = P_{i,j}\mathbf{P}_k(X_{n-1} = i, T^k \geq n).\end{aligned} \tag{2.4}$$

Alors

$$\gamma^k_j = \sum_{n=1}^{\infty} \mathbf{P}_k(X_n = j, n \leq T^k) = \sum_{n=1}^{\infty} \sum_{i\in I} \mathbf{P}_k(X_{n-1} = i, X_n = j, n \leq T^k)$$

$$= \sum_{i\in I} P_{i,j} \sum_{n=1}^{\infty} \mathbf{P}_k(X_{n-1} = i, n \leq T^k) = \sum_{i\in I} P_{i,j} \mathbf{E}_k \sum_{m=0}^{\infty} 1_{\{X_m = i, m \leq T^k - 1\}}$$

$$= \sum_{i \in I} P_{i,j} \mathbf{E}_k \sum_{m=0}^{T^k - 1} 1_{\{X_m = i\}} = \sum_{i \in I} \gamma_i^k P_{i,j}.$$

□

Après la lecture de ce paragraphe, il faut faire les exercices 11 (au moins a),b), e) f)) et 12. Il est souhaitable mais non-indispendable de faire l'exercice 13.

2.6.1 Convergence vers la loi stationnaire

On a vu dans la section précédente que si le nombre d'états de notre chaîne de Markov est finie, tout élément atteint une classe récurrente avec probabilité 1. Cela ne signifie pas que l'état asymptotique de tout état soit constant à partir d'un certain moment car cette classe récurrente peut être constituée de plusieurs états. Pour comprendre la loi asymptotique de notre chaîne, il suffit de comprendre les lois asymptotiques des chaînes irréductibles. Nous entendons par cela la chose suivante : calculer $\lim_{n\to\infty} \mathbf{P}(X_n = j | X_0 = i)$ si la limite existe ou en d'autres termes, calculer $\lim_{n\to\infty} P^n$. Il est facile de voir qu'une telle limite n'existe pas forcément. Par exemple, si on considère une chaîne à deux états $\{1, 2\}$ de matrice de transition $p_{11} = p_{22} = 0$, $p_{12} = p_{21} = 1$ on voit facilement que $P^{2k} = I_2$ et $P^{2k+1} = P$ si bien que la suite P^n admet deux valeurs d'adhérence. En revanche cette suite évaluée sur certaines sous-suites de temps (ici pairs ou impairs) convergent. Nous verrons que c'est un fait général.

On a d'abord besoin de la définition suivante.

Definition. Le pgcd$\{n > 0 \ : P^n_{i,i} > 0\}$ est dit la période de l'état i. Si $d = 1$, l'état i est apériodique.

Théorème 2.6.4 *Si d est la période de i, alors il exsite $N \geq 0$ tel que $P^{nd}_{i,i} > 0$ pour tout $n > N$.*

Soit P irréductible, apériodique. Alors pour tout $i, j \in I$ il existe $N = N(i,j)$ tel que $P^n_{i,j} > 0$ pour tout $n > N$.

Preuve. Soit d la prériode de i et l'ensemble $S = \{n > 0 \ : P^{(nd)}_{i,i} > 0\}$.

Alors, par définition de la période, pgcd$S$=1. Alors, ils existent $n_1, \ldots, n_k \in S$ et $z_1, \ldots, z_k \in \mathbf{Z}$ tels que $\sum_{j=1}^k z_j n_j = 1$. Soit $N = n_1(\sum_{j=1}^k |z_j| n_j)$. Alors, pour tout $n \geq N$ il existe $h \in \mathbf{Z}_+$ et $r \in \{0, 1, \ldots, n_1 - 1\}$ tels que

$$n = N + hn_1 + r = \sum_{j=1}^k n_1 |z_j| n_j + hn_1 + \Big(\sum_{j=1}^k z_j n_j\Big) r = hn_1 + \sum_{j=1}^k (n_1 |z_j| + r z_j) n_j.$$

Comme $n_1, \dots, n_k \in S$, $P_{i,i}^{dn_1 \times h_1} > 0, \dots, p_{i,i}^{dn_k \times h_k} > 0$ pour tout $h_1, \dots, h_k \in \mathbf{Z}_+$. En particulier, pour $n \geq N$

$$P_{i,i}^{nd} = p_{i,i}^{dn_1 \times h} \prod_{j=1}^{k} P_{i,i}^{dn_j \times (n_1|z_j| + rz_j)} > 0,$$

donc $n \in S$ pour tout $n \geq N$.

Si P est irréductible, apériodique, il existe $N(i)$ tel que $P_{i,i}^{(n)} > 0$ pour tout $n > N(i)$ et il existe $k = k(i,j)$ tel que $P_{i,j}^{(k)} > 0$, d'où $P_{i,j}^{(n)} > 0$ pour tout $n \geq N(i) + k(i,j)$. □

Théorème 2.6.5 *Tous les états d'une classe d'états qui communiquent C ont la même période.*

Preuve. Soient $i, j \in C$, d_i, d_j leurs périodes. Si r, s sont tels que $P_{i,j}^r > 0$, $P_{j,i}^s > 0$, alors, $P_{i,i}^{(r+s)} \geq P_{i,j}^{(r)} P_{j,i}^{(s)} > 0$ et d_i donc divise $r + s$. D'autre part, pour tout n tel que $P_{j,j}^{(n)} > 0$

$$P_{i,i}^{(r+n+s)} \geq P_{i,j}^{(r)} P_{j,j}^{(n)} P_{i,j}^{(s)} > 0,$$

donc d_i divise $r + s + n$. Donc d_i divise tout n tel que $P_{j,j}^{(n)} > 0$ et donc d_i divise d_j. Par la symétrie, d_j doit aussi diviser d_i, d'où $d_j = d_i$. □.

Exemple. Une CM irréductible de matrice

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

a la période 3. Une CM de matrice

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \end{pmatrix}$$

est apériodique.

Rémarque utile. Si $P_{i,i} > 0$, alors l'état i a la période 1. En effet, ceci implique $P_{i,i}^n \geq P_{i,i} \cdots P_{i,i} > 0$ pour tout $n > 0$. Donc, si dans une matrice de transition d'une CM irréductible, un élément sur la diagonale est positif, alors, tous les états ont la période 1, la CM est apériodique. L'inverse n'est pas vraie !

Le Thm suivant est central dans le cours.

Théorème 2.6.6 *Soit une chaîne irréductible, finie, apériodique. Alors, pour toute mesure de probabilité ν (non nécessairement stationnaire) on a*

$$\lim_{n\to\infty} \nu P^n = \mu$$

où μ est l'unique mesure stationnaire.

Démonstration.— Notons $\mathcal{M}_0$ l'ensemble des vecteurs lignes x tels que $\sum_{i\in E} x_i = 0$. Définissons pour $x \in \mathcal{M}_0$ la norme suivante :

$$\|x\| = \frac{1}{2}\sum_{i\in E} |x_i| = \sum_{i\in E} (x_i)^+.$$

(Avec la notation $z^+ = \max(0,z)$). L'application $x \mapsto xP$ envoie $\mathcal{M}_0$ dans lui même. On a le lemme suivant

Lemme 2.6.1 *Soit Q une matrice stochastique telle que $\alpha = \min_{i,j\in E} Q_{ij} > 0$. Alors, Q est une $(1-\alpha)$-contraction : pour tout $x \in \mathcal{M}_0$*

$$\|xQ\| \le (1-\alpha)\|x\|.$$

Démonstration.— On a

$$\begin{aligned}
\|xQ\| &= \sum_{i:(xQ)_i>0} (xQ)_i \\
&= \sum_{i:(xQ)_i>0} \sum_{k\in E} x_k (Q)_{ki} \\
&\le \sum_{i:(xQ)_i>0} \sum_{k:x_k>0} |x_k| (Q)_{ki} \\
&\le \sum_{k:x_k>0} \sum_{i:(xQ)_i>0} x_k (Q)_{ki}
\end{aligned}$$

Comme $xQ \in \mathcal{M}_0$ on a $\sum_{i\in E}(xQ)_i = 0$ et l'ensemble des $i \in E$ pour lesquels $(xQ)_i > 0$ n'est pas E tout entier. Par conséquent,

$$\begin{aligned}
\sum_{i:(xQ)_i>0} (Q)_{ki} &= \sum_{i\in E} (Q)_{ki} - \sum_{i:(xQ)_i\le 0} (Q)_{ki} \\
&= 1 - \sum_{i:(xQ)_i\le 0} (Q)_{ki} \\
&\le (1-\alpha).
\end{aligned}$$

Revenant aux inégalités précédentes

$$\|xQ\| \leq (1-\alpha)\|x\|.$$

$\square$

Puisque P est apériodique, pour tout $i, j \in E$ il existe $N(i,j)$ tel que $P^n_{i,j} > 0$ pour tout $n \geq N(i,j)$. Comme l'espace E est fini, alors il existe un entier $m = \max N(i,j)$ tel que $Q = P^m$ soit à coefficients strictement positifs. L'application $x \mapsto xP^m$ de $(\mathcal{M}_0, \|\cdot\|)$ dans lui même est donc une $(1-\alpha)$-contraction et par conséquent pour tout entier l, $\|xP^{lm}\| \leq (1-\alpha)^l\|x\|$. Si μ est l'unique mesure de probabilité stationnaire et ν une mesure de probabilité $\mu - \nu \in \mathcal{M}_0$ et donc pour tous entiers l, c

$$\|(\mu-\nu)P^{lm+c}\| \leq (1-\alpha)^l\|(\mu-\nu)P^c\|$$

ce qui s'écrit (en utilisant $\mu P^n = \mu$)

$$\|\mu - \nu P^{lm+c}\| \leq (1-\alpha)^l\|(\mu-\nu)P^c\|.$$

Comme tout entier n s'écrit de fa con unique $n = lm + c$ avec $0 \leq c < m$, $l \geq n/m$ (division euclidienne de n par m) et comme $0 \leq (1-\alpha) < 1$, la suite $\mu - \nu P^n$ converge vers 0 exponentiellement vite.

$\square$

! ! Ce Thm nous dit que si P est irréductible, apériodique et finie, pour toute loi initiale ν

$$\mathbf{P}(X_n = j) \to \mu_j, \quad n \to \infty.$$

Donc au bout d'un tres grand temps la chaine de Markov sera dans un état aléatoire, cet état est j avec probabilité approximative μ_j. Indépendament de la loi initiale ν !

Notamment si $\nu = \delta_i$, cad l'état initiale est i, on a

Corollaire 2.6.1 *Soit une chaîne irréductible finie apériodique. Alors, pour tout $i, j \in E$ la limite suivante existe et est indépendante de i*

$$\lim_{n\to\infty} (P^n)_{ij} = \mu_j,$$

où μ est l'unique mesure de probabilité stationnaire.

! ! Donc P^n converge vers une matrice dont la $j$eme ligne se compose de mêmes éléments μ_j.

L'hypothèse d'apériodicité est cruciale dans ce Thm. En effet, on peut considérer une CM irréductible de deux états avec $P_{12} = 1$ et $P_{21} = 1$ qui n'est pas apériodique. Elle a une mesure invariante $(1/2, 1/2)$. Néanmoins, si $X_0 = 1$, alors $\mathbf{P}(X_n = 1) = 1$ pour n pair et $\mathbf{P}(X_n = 1) = 0$ pour n impair, donc $\mathbf{P}(X_n = 1)$ ne converge pas vers $1/2$, $n \to \infty$.

Si la Chaine de Markov est périodique, cette limite généralement n'existe pas.

! !Et si la CM est finie mais pas irréductible ? Comment calculer $\lim_{n\to\infty} \mathbf{P}(X_n = j \mid X_0 = i)$?

Premierement, si j est transient, il est visité un nombre fini de fois p.s., et par conséquent cette limite vaut 0.

Supposons que j appartient à une classe de récurrence C. Si i appartient aussi à C, alors la matrice P restreinte a cette classe P^C possède l'unique loi de probabilité stationnaire μ^C. Si de plus C est apériodique (il faut le vérifier !) on a $\mathbf{P}(X_n = j \mid X_0 = i) \to \mu_j^C$ par le Thm.

Si i appartient à une autre classe récurrente $\mathbf{P}(X_n = j \mid X_0 = i) = 0$ pour tout $n \geq 0$.

Finallement, si i est transient et C est apériodique $\mathbf{P}(X_n = j \mid X_0 = i) \to \nu_i(C)\mu_j^C$ avec $\nu_i(C)$ la probabilité d'absorption dans C.

Il est temps de faire l'exercice 14.

2.7 Théorème ergodique

Le Thm suivant est appellé aussi la loi de grands nombres pour les chaines de Markov.

Théorème 2.7.1 *Soit $(X_n)_{n=0}^{\infty}$ une chaine de Markov de matrice de transition P irréductible finie de loi stationnaire μ. Alors*

$$\frac{\sum_{k=0}^{n-1} \mathbf{1}_i \circ X_k}{n} \to \mu_i, \quad n \to \infty, p.s.$$

De plus pour toute fonction $f : E \to \mathbf{R}$

$$\frac{1}{n}\sum_{k=0}^{n-1} f(X_k) \to \int_I f d\pi = \sum_{i\in E} \mu_i f(i), \quad n \to \infty, p.s.$$

Remarque. Ce Thm est vrai pour les CM autant périodiques qu'apériodiques ! Il n'y a pas d'hypothèse d'apériodicité dans l'énoncé.

Démonstration.— On note $\sum_{k=0}^{n-1} 1_{\{X_k=i\}} = N_i(n)$ le nombre de visites dans i avant l'instant n. Soit $\sum_{k=0}^{\infty} 1_{\{X_k=i\}} = N_i$ est le nombre total de visites dans i.

Comme pour une CM irréductible finie $\mathbf{P}(T^i < \infty) = 1$, il suffit de considérer une CM de mesure initiale δ_i. Soient $Q^i_r = S(i)^{(r)} - S(i)^{(r-1)}$ les intervalles de temps entre le $r$ème et le $(r-1)$ème visite dans i. Les variables aléatoires $Q^i_1, Q^i_2, Q^i_3, \dots$ sont indépendantes de même loi et $\mathbf{E}(Q^i_r \mid X_0 = i) = \mathbf{E}(T^i \mid X_0 = i)$. On a

$$Q^i_1 + Q^i_2 + \cdots + Q^i_{N_i(n)-1} \leq n - 1$$

où $Q^i_1 + Q^i_2 + \cdots + Q^i_{N_i(n)-1}$ est l'instant de la dernière visite dans i avant l'instant n. On a aussi

$$Q^i_1 + Q^i_2 + \cdots + Q^i_{N_i(n)} \geq n$$

où $Q^i_1 + Q^i_2 + \cdots + Q^i_{N_i(n)}$ est l'instant de la première visite dans i après l'instant $n-1$. On obtient

$$\frac{Q^i_1 + Q^i_2 + \cdots + Q^i_{N_i(n)-1}}{N_i(n)} \leq \frac{n}{N_i(n)} \leq \frac{Q^i_1 + Q^i_2 + \cdots + Q^i_{N_i(n)}}{N_i(n)}. \tag{2.5}$$

Par la loi des grands nombres

$$\frac{Q^i_1 + Q^i_2 + \cdots + Q^i_n}{n} \to \mathbf{E}(T^i \mid X_0 = i) \quad n \to \infty, \quad p.s.$$

Comme P est récurrente,

$$\mathbf{P}(N_i(n) \to \infty, \quad n \to \infty) = 1,$$

d'où dans (2.5)

$$\frac{n}{N_i(n)} \to \mathbf{E}(T^i \mid X_0 = i), \quad n \to \infty, \quad p.s.$$

Il reste a rappeller que $\mathbf{E}(T^i \mid X_0 = i) = 1/\mu_i$ pour terminer la preuve du premier résultat énoncé dans le Thm.

Toute $f : E \to \mathbf{R}$ s'écrit comme une combinaison linéaire finie $\sum_{i \in E} f(i) \mathbf{1}_i$. Ceci implique tout de suite le résultat énoncé pour f.

$\square$

! ! Ce Thm dit que la proportion du temps qu'une chaine de Markov irréductible finie passe dans un état i est asymptotiquement la loi stationnaire μ_i.

Il n'y a pas d'apériodicité dans les hypothèses ! !

Faisons l'exercice 15.

Si une chaine de Markov finie n'est pas irréductible, et $i \in C$ est dans une classe récurrente C, le Thm reste vraie avec μ_i^C qui est la loi stationnaire de la restriction P^C. Si i est transient, cette limite vaut 0.

Ce Thm peut être utile pour simuler la loi stationnaire d'une chaine de Markov. S'il est difficile de résoudre $\mu = \mu P$ par ordinateur pour P grande, on simule une trajectoire $X_0, X_1, X_2, \ldots$ et on calcule la proportion du temps passé dans chaque état, ceci donne approximativement la loi stationnaire. Et voici le fin du cours !