Université Paris-Dauphine
M1MMD

Année 2017/2018

Processus discrets

Joseph Lehec

Table des matières

Première partie
L'espérance conditionnelle

1 Le cas discret

Soient X, Y deux variables discrètes, par exemple à valeurs dans $\mathbb{N}$. Pour $i \in \mathbb{N}$ vérifiant $\mathbb{P}(X = i) \neq 0$ on a par définition

$$\mathbb{P}(Y = j|X = i) = \frac{\mathbb{P}(Y = j, X = i)}{\mathbb{P}(X = i)}.$$

En supposant Y intégrable on définit également l'espérance conditionnelle de Y sachant $X = i$ par

$$\mathbb{E}[Y \mid X = i] = \frac{\mathbb{E}\left[Y \mathbb{1}_{\{X=i\}}\right]}{\mathbb{P}(X = i)}.$$

On peut voir cette quantité comme une fonction de i et poser

$$u(i) = \mathbb{E}[Y \mid X = i].$$

L'espérance conditionnelle de Y sachant X, notée $\mathbb{E}[Y \mid X]$ est alors la variable définie par la formule

$$\mathbb{E}[Y \mid X] = u(X).$$

Noter que $\mathbb{E}[Y \mid X = i]$ est un réel tandis que $\mathbb{E}[Y \mid X]$ est une variable aléatoire. Posons $A_i = \{X = i\}$ et supposons pour simplifier que $\mathbb{P}(A_i) > 0$ pour tout i. La définition de l'espérance conditionnelle de Y sachant X peut alors se réécrire

$$\mathbb{E}[Y \mid X] = \sum_{i\in\mathbb{N}} \mathbb{E}[Y \mid X = i]\, \mathbb{1}_{\{X=i\}} = \sum_{i\in\mathbb{N}} \mathbb{E}[Y \mid A_i]\, \mathbb{1}_{A_i}.$$

Cette formule montre qu'en fait $\mathbb{E}[Y \mid X]$ ne dépend que de la tribu $\sigma(X)$ engendrée par X. Plus précisément, si X' est une autre variable aléatoire vérifiant $\sigma(X') = \sigma(X)$ alors $\mathbb{E}[Y \mid X] = \mathbb{E}[Y \mid X']$. Remarquons aussi qu'on a l'égalité

$$\mathbb{E}\left[\mathbb{E}[Y \mid X]\, \mathbb{1}_A\right] = \mathbb{E}[Y\, \mathbb{1}_A]$$

pour tout $A \in \sigma(X)$. En effet soit $\mathcal{A} = \sigma(X)$ la tribu engendrée par X, qui est aussi la tribu engendrée par les (A_i). Comme les (A_i) forment une partition dénombrable de Ω, il est facile de voir (exercice) que

$$\mathcal{A} = \left\{ \bigcup_{j\in I} A_j,\ I \subset \mathbb{N} \right\}.$$

Soit $I \subset \mathbb{N}$ et soit $A = \bigcup_{j\in I} A_j$ un élément de $\mathcal{A}$. Comme les (A_i) forment une partition de Ω on a $A_i \cap A = A_i$ si $i \in I$ et $A_i \cap A = \emptyset$ sinon. On a alors le calcul suivant

$$\begin{aligned}
\mathbb{E}\left[\mathbb{E}[Y \mid X]\, \mathbb{1}_A\right] &= \mathbb{E}\left[\sum_{i\in\mathbb{N}} \mathbb{E}[Y \mid A_i]\, \mathbb{1}_{A_i} \mathbb{1}_A\right] \\
&= \mathbb{E}\left[\sum_{i\in I} \mathbb{E}[Y \mid A_i]\, \mathbb{1}_{A_i}\right].
\end{aligned}$$

Et en utilisant Fubini deux fois (exercice : vérifier que Fubini s'applique) on obtient

$$\mathbb{E}\left[\mathbb{E}[Y \mid X]\, \mathbb{1}_A\right] = \sum_{i\in I} \mathbb{E}[Y \mid A_i]\, \mathbb{P}(A_i) = \sum_{i\in I} \mathbb{E}[Y\, \mathbb{1}_{A_i}] = \mathbb{E}[Y\, \mathbb{1}_A],$$

ce qu'il fallait démontrer. On va s'appuyer sur ces remarques pour définir l'espérance conditionnelle dans le cas général.

2 La définition générale

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et soit $\mathcal{G}$ une sous-tribu de $\mathcal{F}$, c'est-à-dire une tribu vérifiant $\mathcal{G} \subset \mathcal{F}$. Étant donnée une variable aléatoire $X : \Omega \to \mathbb{R}$ on va définir dans cette section l'espérance conditionnelle de X sachant $\mathcal{G}$.

Proposition I.1. *Soit X une variable aléatoire intégrable. Il existe une variable Z intégrable vérifiant*

(i) Z est $\mathcal{G}$-mesurable ;

(ii) Pour tout ensemble $A \in \mathcal{G}$ on a $\mathbb{E}[Z\, \mathbb{1}_A] = \mathbb{E}[X\, \mathbb{1}_A]$.

De plus Z est unique à un ensemble négligeable près, c'est-à-dire que si Z' est une autre variable vérifiant les deux propriétés précédentes alors $Z' = Z$ presque surement.

La démonstration est donnée à la fin de cette section.

Definition I.2. La variable Z définie dans la proposition précédente est appelée espérance conditionnelle de X sachant $\mathcal{G}$, notée $\mathbb{E}[X \mid \mathcal{G}]$.

Remarquons tout de suite que :

(i) L'espérance conditionnelle de X sachant $\mathcal{G}$ est une variable aléatoire qui n'est définie qu'à un ensemble de mesure nulle près.

(ii) En appliquant la deuxième propriété définissant l'espérance conditionnelle à l'ensemble $A = \Omega$ on obtient $\mathbb{E}\left[\mathbb{E}[X \mid \mathcal{G}]\right] = \mathbb{E}[X]$.

(iii) Si X est elle-même $\mathcal{G}$-mesurable, en particulier si X est une variable constante, alors $\mathbb{E}[X \mid \mathcal{G}] = X$. En effet, dans ce cas la variable X vérifie trivialement les deux propriétés caractérisant l'espérance conditionnelle.

(iv) À l'inverse, si la variable X est indépendante de la tribu $\mathcal{G}$, en particulier si $\mathcal{G}$ est la tribu triviale $\{\emptyset, \Omega\}$, alors $\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[X]$ p.s. En effet la variable constante égale à $\mathbb{E}[X]$ est évidemment $\mathcal{G}$-mesurable et on a bien par indépendance

$$\mathbb{E}[X\mathbb{1}_A] = \mathbb{E}[X]\,\mathbb{E}[\mathbb{1}_A] = \mathbb{E}[\mathbb{E}[X]\mathbb{1}_A]$$

pour tout $A \in \mathcal{G}$.

Par abus de langage, étant donnée une variable Y, l'espérance conditionnelle de X sachant la tribu engendrée par Y est simplement appelée espérance conditionnelle de X sachant Y. On a donc par définition $\mathbb{E}[X \mid Y] = \mathbb{E}[X \mid \sigma(Y)]$. Rappelons qu'une variable Z est mesurable pour la tribu $\sigma(Y)$ si et seulement s'il existe une fonction f Borélienne telle que $Z = f(Y)$. Par conséquent on a $\mathbb{E}[X \mid Y] = f(Y)$ où f est n'importe quelle fonction Borélienne vérifiant

$$\mathbb{E}[X\mathbb{1}_{\{Y\in B\}}] = \mathbb{E}[f(Y)\mathbb{1}_{\{Y\in B\}}]$$

pour tout Borélien B.

Il faut interpréter l'espérance conditionnelle comme une projection orthogonale. En effet soit F un espace Euclidien, soit G un sous-espace Euclidien de E et soit $x \in E$. Il est alors facile de voir que le projeté orthogonal de x sur G est l'unique élément $z \in G$ vérifiant

(i) $z \in G$;

(ii) Pour tout $y \in G$ on a $\langle z, y\rangle = \langle x, y\rangle$.

Dans le cas où la variable X est de carré intégrable, cette analogie est parfaitement rigoureuse.

Proposition I.3. *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et soit $\mathcal{G}$ une sous-tribu de $\mathcal{F}$. On considère l'espace de Hilbert $L^2(\Omega, \mathcal{F}, \mathbb{P})$. L'ensemble G des variables $X \in F$ qui soient $\mathcal{G}$-mesurables est un sous-espace fermé de F. De plus pour toute variable $X \in F$, l'espérance conditionnelle de X sachant $\mathcal{G}$ est la projection orthogonale de X sur G.*

Démonstration. On a $G = L^2(\Omega, \mathcal{G}, \mathbb{P})$. Donc G est un espace de Hilbert, donc G est complet, donc G est fermé. Soit $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ et soit Z le projeté orthogonal de X sur G. Alors Z est $\mathcal{G}$-mesurable par définition et

$$\mathbb{E}[ZT] = \mathbb{E}[XT], \quad \forall T \in G.$$

En particulier pour tout $A \in \mathcal{G}$, comme $\mathbb{1}_A$ est de carré intégrable et $\mathcal{G}$-mesurable on a bien $\mathbb{1}_A \in G$. On obtient donc

$$\mathbb{E}[Z\mathbb{1}_A] = \mathbb{E}[X\mathbb{1}_A], \quad \forall A \in \mathcal{G},$$

ce qui montre que $Z = \mathbb{E}[X \mid \mathcal{G}]$. □

Démonstration de la Proposition I.1. Commençons par montrer la propriété de monotonie suivante : soient X, X' deux variables intégrables vérifiant $X \leq X'$ p.s. et soient Z, Z' deux variables vérifiant les propriétés (i) et (ii) relatives à X et X' respectivement. Alors

$$Z \leq Z', \quad \text{p.s.}$$

En effet, comme Z et Z' sont $\mathcal{G}$-mesurables, on a $\{Z \geq Z'\} \in \mathcal{G}$. En utilisant la propriété (ii) et l'hypothèse $X \leq X'$ on obtient donc

$$\mathbb{E}[Z\mathbb{1}_{\{Z\geq Z'\}}] = \mathbb{E}[X\mathbb{1}_{\{Z\geq Z'\}}] \leq \mathbb{E}[X'\mathbb{1}_{\{Z\geq Z'\}}] = \mathbb{E}[Z'\mathbb{1}_{\{Z\geq Z'\}}].$$

Donc

$$\mathbb{E}[(Z - Z')\mathbb{1}_{\{Z\geq Z'\}}] \leq 0.$$

Comme une variable positive d'espérance négative est nulle p.s. on en déduit

$$(Z - Z')\mathbb{1}_{\{Z\geq Z'\}} = 0, \quad \text{p.s.}$$

c'est-à-dire $Z \leq Z'$ p.s., ce qu'il fallait démontrer. En appliquant cette propriété de monotonie à $X' = X$ on obtient immédiatement l'unicité de l'espérance conditionnelle.
Montrons maintenant l'existence. On a déjà vu que si X est de carré intégrable, alors $\mathbb{E}[X \mid \mathcal{G}]$ existe et est égale au projeté orthogonal de X sur $L^2(\Omega, \mathcal{G}, \mathbb{P})$. Supposons maintenant que X est une variable intégrable (mais pas forcément de carré intégrable) et positive et posons $X_n = X \wedge n$ pour tout entier n. Alors X_n est une variable bornée, donc de carré intégrable. Soit Z_n son espérance conditionnelle (qui existe d'après le cas précédent). On a $X_n \geq 0$ et $X_n \leq X_{n+1}$ pour tout n. D'après la propriété de monotonie vue plus haut, on en déduit que $Z_n \geq 0$ et $Z_n \leq Z_{n+1}$ pour tout n. En particulier (Z_n) admet une limite presque sure, qu'on appelle Z. Alors Z est $\mathcal{G}$-mesurable comme limite de variables $\mathcal{G}$-mesurables. Soit $A \in \mathcal{G}$. On a par définition de Z_n

$$\mathbb{E}[Z_n\mathbb{1}_A] = \mathbb{E}[X_n\mathbb{1}_A],$$

pour tout n. De plus comme $X_n \nearrow X$ et $Z_n \nearrow Z$ p.s., en passant à la limite dans l'égalité précédente, et en appliquant le théorème de convergence monotone deux fois on obtient

$$\mathbb{E}[Z\mathbb{1}_A] = \mathbb{E}[X\mathbb{1}_A],$$

ce qui montre que Z vérifie (ii).
Enfin, si X est une variable intégrable quelconque, alors les parties positive X_+ et négative X_- de X sont des variables positives et intégrables. D'après ce qui précède $\mathbb{E}[X_+ \mid \mathcal{G}]$ et $\mathbb{E}[X_- \mid \mathcal{G}]$ sont bien définies. On montre alors facilement que

$$Z = \mathbb{E}[X_+ \mid \mathcal{G}] - \mathbb{E}[X_- \mid \mathcal{G}]$$

vérifie bien les propriétés (i) et (ii). □

3 Propriétés de l'espérance conditionnelle

Soit X une variable intégrable, définie sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ et soit $\mathcal{G}$ est une sous-tribu de $\mathcal{F}$.

Lemme I.4. *Pour tout variable Y qui soit $\mathcal{G}$-mesurable et telle que XY soit intégrable (en particulier si Y est bornée) on a*

$$\mathbb{E}\left[\mathbb{E}[X \mid \mathcal{G}]Y\right] = \mathbb{E}[XY].$$

Démonstration. Par définition l'égalité est vraie si Y est une indicatrice. Comme l'égalité est linéaire en Y elle reste vraie si Y est étagée. Si X et Y sont positives, le lemme s'obtient ensuite en approchant de manière monotone Y par des variables étagées et en utilisant la convergence monotone. Si X et Y changent de signe on obtient le résultat en décomposant X et Y en parties positives et négatives. □

Proposition I.5. *L'espérance conditionnelle est*
- *Linéaire :* $\mathbb{E}[\lambda_1 X_1 + \lambda_2 X_2 \mid \mathcal{G}] = \lambda_1 \mathbb{E}[X_1 \mid \mathcal{G}] + \lambda_2 \mathbb{E}[X_2 \mid \mathcal{G}]$ *;*
- *Monotone : si* $X_1 \leq X_2$ *p.s. alors* $\mathbb{E}[X_1 \mid \mathcal{G}] \leq \mathbb{E}[X_2 \mid \mathcal{G}]$ *p.s.*

La linéarité est laissée en exercice, il suffit de vérifier que le membre de droite vérifie les propriétés caractérisant le membre de gauche. La monotonie a déjà été prouvée dans la section précédente.

Proposition I.6 (Convergence monotone)**.** *Soit* (X_n) *une suite de variables intégrables positives. Si* $X_n \nearrow X$ *p.s. alors* $\mathbb{E}[X_n|\mathcal{G}] \nearrow \mathbb{E}[X|\mathcal{G}]$ *p.s.*

Démonstration. Posons $Y_n = \mathbb{E}[X_n|\mathcal{G}]$. D'après la monotonie (Y_n) est une suite croissante. Posons $Y = \lim_n Y_n$. Alors Y est $\mathcal{G}$-mesurable et le théorème de convergence monotone nous permet de passer à la limite dans l'égalité $\mathbb{E}[X_n \mathbb{1}_A] = \mathbb{E}[Y_n \mathbb{1}_A]$ pour obtenir $\mathbb{E}[X \mathbb{1}_A] = \mathbb{E}[Y \mathbb{1}_A]$ pour tout $A \in \mathcal{G}$. Donc $Y = \mathbb{E}[X|\mathcal{G}]$ p.s. □

Lemme I.7 (Inégalité de Jensen)**.** *Soit* $\varphi \colon \mathbb{R} \to \mathbb{R}$ *une fonction convexe telle que* $\varphi(X)$ *soit intégrable. Alors*

$$\mathbb{E}[\varphi(X) \mid \mathcal{G}] \geq \varphi\left(\mathbb{E}[X|\mathcal{G}]\right).$$

La démonstration est laissée en exercice, il suffit d'adapter la démonstration de l'inégalité de Jensen usuelle.

Théorème I.8. *Soit* $p \in [1, +\infty]$*. L'espérance conditionnelle est une contraction dans* L^p *: si* $X \in L^p$ *alors* $\mathbb{E}[X \mid \mathcal{G}] \in L^p$ *et*

$$\|\mathbb{E}[X \mid \mathcal{G}]\|_p \leq \|X\|_p.$$

Démonstration. L'application $x \mapsto |x|^p$ est convexe. Donc par Jensen

$$|\mathbb{E}[X \mid \mathcal{G}]|^p \leq \mathbb{E}[|X|^p \mid \mathcal{G}].$$

En prenant l'espérance de cette inégalité on obtient

$$\mathbb{E}\left[|\mathbb{E}[X \mid \mathcal{G}]|^p\right] \leq \mathbb{E}\left[\mathbb{E}[|X|^p \mid \mathcal{G}]\right] = \mathbb{E}[|X|^p],$$

ce qu'il fallait démontrer. □

Les deux dernières propriétés seront utilisées constamment par la suite.

Proposition I.9. *Si* X, Y *sont telles que* XY *est intégrable et* Y *est* $\mathcal{G}$*-mesurable alors*

$$\mathbb{E}[XY \mid \mathcal{G}] = \mathbb{E}[X \mid \mathcal{G}]\, Y.$$

On peut donc “sortir” de l'espérance conditionnelle les facteurs $\mathcal{G}$-mesurables.

Démonstration. Y est $\mathcal{G}$-mesurable par hypothèse et $\mathbb{E}[X \mid \mathcal{G}]$ est $\mathcal{G}$-mesurable par définition, donc $\mathbb{E}[X \mid \mathcal{G}]\, Y$ est $\mathcal{G}$-mesurable. Soit $A \in \mathcal{G}$ et posons $Z = Y\, \mathbb{1}_A$. Alors Z est $\mathcal{G}$-mesurable pour les mêmes raisons et comme XY est intégrable, XZ l'est aussi. D'après le lemme I.4 on a donc $\mathbb{E}[XZ] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]Z]$. C'est-à-dire

$$\mathbb{E}[XY \mathbb{1}_A] = \mathbb{E}\left[\mathbb{E}[X \mid \mathcal{G}] Y \mathbb{1}_A\right],$$

ce qui permet de conclure. □

Proposition I.10 (Propriété d'emboitement)**.** *Soit* $\mathcal{H}$ *une sous-tribu de* $\mathcal{G}$*. On a donc* $\mathcal{H} \subset \mathcal{G} \subset \mathcal{F}$*. Alors*

$$\mathbb{E}\left[\mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{H}\right] = \mathbb{E}\left[\mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{H}\right] = \mathbb{E}[X \mid \mathcal{H}].$$

Pour retenir cette proposition, il faut penser à l'analogie avec la projection orthogonale et faire un dessin. La démonstration est laissée en exercice.

4 Quelques cas pratiques

Commençons par faire le lien avec le cas avec le cas discret.

Lemme I.11. *Si $\mathcal{G}$ est engendrée par une partition dénombrable $(A_i)_{i\in\mathbb{N}}$, alors*

$$\mathbb{E}[X \mid \mathcal{G}] = \sum_{i\in\mathbb{N}} \mathbb{E}[X \mid A_i]\, \mathbb{1}_{A_i}.$$

En particulier si Y est à une variable aléatoire discrète, disons à valeurs dans $\mathbb{N}$, alors

$$\mathbb{E}[X \mid Y] = \sum_{i\in\mathbb{N}} \mathbb{E}[X \mid Y = i]\, \mathbb{1}_{\{Y=i\}}.$$

La démonstration est laissée en exercice, encore une fois il suffit de vérifier que le membre de droite vérifie les deux propriétés caractérisant $\mathbb{E}[X \mid \mathcal{G}]$.

Donnons maintenant une manière concrète de calculer $\mathbb{E}[X \mid Y]$ quand (X, Y) est couple continu de variables réelles. On suppose donc que (X, Y) possède une densité jointe, notée $f_{X,Y}$ et on note f_X et f_Y les densité marginales respectives de X et Y. Pour y tel que $f_Y(y) > 0$ on pose

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x, y)}{f_X(x)},$$

appelée densité conditionnelle de X sachant Y.

Lemme I.12. *Si X est intégrable alors en posant*

$$u(y) = \int_{\mathbb{R}} x f_{X|Y=y}(x)\, dx$$

on a $\mathbb{E}[X \mid Y] = u(Y)$. Plus généralement, pour toute fonction h telle que $h(X)$ soit intégrable, on pose

$$H(y) = \int_{\mathbb{R}} h(x) f_{X|Y=y}(x)\, dx.$$

On a alors $\mathbb{E}[h(X) \mid Y] = H(Y)$.

Encore une fois on laisse la démonstration en exercice. Notons aussi que le lemme reste valable si (X, Y) est à valeurs dans $\mathbb{R}^n \times \mathbb{R}^m$ (au lieu de $\mathbb{R} \times \mathbb{R}$).

Une situation qui revient très souvent est la suivante.

Proposition I.13. *Soit X, Y deux variables indépendantes, et $f : \mathbb{R}^2 \to \mathbb{R}$ une fonction mesurable telle que $f(X, Y)$ soit intégrable. Alors*

$$\mathbb{E}\left[f(X, Y) \mid Y\right] = F(Y),$$

où F est la fonction donnée par $F(y) = \mathbb{E}[f(X, y)]$ pour tout $y \in \mathbb{R}$.

Démonstration. La variable $F(Y)$ est $\sigma(Y)$-mesurable. Soient $\mathbb{P}_X$ et $\mathbb{P}_Y$ les lois respectives de X et Y. Par indépendance la loi jointe de (X, Y) est la mesure produit $\mathbb{P}_X \otimes \mathbb{P}_Y$. Soit $A = \{Y \in B\}$ un élément de $\sigma(Y)$. En appliquant Fubini on a

$$\begin{aligned}
\mathbb{E}[f(X, Y)\mathbb{1}_A] &= \int_{\mathbb{R}^2} f(x, y)\mathbb{1}_B(y)\, \mathbb{P}_X(dx)\mathbb{P}_Y(dy) \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y)\, \mathbb{P}_X(dx) \right) \mathbb{1}_B(y)\, \mathbb{P}_Y(dy) = \mathbb{E}[F(Y)\mathbb{1}_A],
\end{aligned}$$

ce qu'il fallait démontrer. □

Deuxième partie

Martingales

Dans ce qui suit on appelle *processus stochastique* une suite (X_n) de variables aléatoires définies sur un même espace de probabilités $(\Omega, \mathcal{F}, \mathbb{P})$.

5 Filtrations, martingales

Definition II.1. Une *filtration* est une famille $(\mathcal{F}_n)_{n\geq 0}$ de sous-tribus de $\mathcal{F}$ telles que $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$ pour tout $n \geq 0$.

Exemple. Si (X_n) est un processus stochastique, on appelle filtration naturelle associée à (X_n) la filtration $(\mathcal{F}_n)$ donnée par

$$\mathcal{F}_n = \sigma(X_0, \dots, X_n), \quad n \geq 0.$$

Désormais on interprète l'indice n comme une variable de temps. La tribu $\mathcal{F}_n$ s'interprète alors comme l'information dont on dispose au temps n.

Definition II.2. Soit (X_n) un processus stochastique et $(\mathcal{F}_n)$ une filtration. On dit que (X_n) est *adapté* à $(\mathcal{F}_n)$ si X_n est $\mathcal{F}_n$-mesurable pour tout $n \geq 0$. On dit que $(X_n)_{n\geq 0}$ est *prévisible* si X_0 est constante et si X_n est $\mathcal{F}_{n-1}$-mesurable pour tout $n \geq 1$.

Remarque II.3. La filtration naturelle de (X_n) est la plus petite filtration qui rende (X_n) adapté.

Un processus adapté (X_n) est un processus que l'on découvre progressivement : à l'instant n on connaît alors les valeurs de X_k pour $k \leqslant n$, à savoir les valeurs passées de X. En revanche on ne peut pas a priori déterminer les valeurs futures de X. Un processus prévisible est un processus (X_n) pour lequel on connaît au temps n toutes les valeurs passées, ainsi que la valeur suivante X_{n+1}.

Exemple (marche aléatoire)**.** Soit (X_n) une suite de variables i.i.d. et soit $(\mathcal{F}_n)$ sa filtration naturelle. Le processus (S_n) défini par

$$S_n = X_0 + \dots + X_n$$

est adapté à $(\mathcal{F}_n)$.

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité, muni d'une filtration $(\mathcal{F}_n)$.

Definition II.4. Soit (X_n) un processus stochastique. On dit que (X_n) est une *martingale* si

- Le processus (X_n) est adapté ;
- La variable X_n est intégrable, pour tout $n \geq 0$;
- On a l'égalité $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = X_n$ pour tout $n \geq 0$.

Si la dernière condition est remplacée par $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] \geq X_n$ on dit que (X_n) est une *sous-martingale*. À l'inverse si $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] \leq X_n$ on dit que (X_n) est une *sur-martingale*.

Remarque II.5. La notion de martingale est donc relative à une filtration donnée. Lorsque la filtration n'est pas précisée, on dit que (X_n) est une martingale si c'est une martingale par rapport à sa filtration naturelle, ce qui revient à dire que

$$\mathbb{E}[X_{n+1} \mid X_0, \dots, X_n] = X_n,$$

pour tout n.

Dans la suite, pour $n \geq 1$, on notera $\Delta X_n = X_n - X_{n-1}$ l'accroissement du processus (X_n) au temps n. Si (X_n) est adapté et intégrable on a

$$\mathbb{E}[\Delta X_n \mid \mathcal{F}_{n-1}] = \mathbb{E}[X_n - X_{n-1} \mid \mathcal{F}_{n-1}] = \mathbb{E}[X_n \mid \mathcal{F}_{n-1}] - X_{n-1}.$$

On en déduit que le processus (X_n) est une martingale si et seulement s'il est adapté, intégrable et vérifie

$$\mathbb{E}[\Delta X_n \mid \mathcal{F}_{n-1}] = 0, \quad \forall n \geq 1.$$

Si on pense X_n comme la fortune au temps n d'un joueur, alors (X_n) est une martingale si le jeu est équitable, une sur-martingale si le jeu est défavorable et une sous-martingale si le jeu est favorable.

Exemple. On reprend l'exemple de la marche aléatoire, en supposant maintenant que les (X_i) sont intégrables. On a vu que le processus (S_n) était adapté. De plus S_n est intégrable comme somme de variables intégrables. Par ailleurs, le caractère i.i.d. des (X_i) montre que X_n est indépendant de $\mathcal{F}_{n-1}$. Donc

$$\mathbb{E}[\Delta S_n \mid \mathcal{F}_{n-1}] = \mathbb{E}[X_n \mid \mathcal{F}_{n-1}] = \mathbb{E}[X_n] = \mathbb{E}[X_1].$$

Par conséquent (S_n) est une martingale si et seulement si $\mathbb{E}[X_1] = 0$.

Exemple (Martingale fermée)**.** Soit Z une variable aléatoire intégrable. Alors le processus (X_n) défini par $X_n = \mathbb{E}[Z|\mathcal{F}_n]$ pour tout n est une martingale. En effet (X_n) est alors intégrable et adapté par définition et d'après la propriété d'emboitement de l'espérance conditionnelle

$$\mathbb{E}[X_n \mid \mathcal{F}_{n-1}] = \mathbb{E}\left[\mathbb{E}[Z \mid \mathcal{F}_n] \mid \mathcal{F}_{n-1}\right] = \mathbb{E}[Z \mid \mathcal{F}_{n-1}] = X_{n-1}.$$

Une martingale de cette forme est dite fermée.

Lemme II.6. *Si (X_n) est une martingale alors $\mathbb{E}[X_n] = \mathbb{E}[X_0]$ pour tout n. Autrement dit l'espérance d'une martingale est constante au cours du temps. De même l'espérance d'une sous-martingale est croissante et celle d'une sur-martingale est décroissante.*

Démonstration. En prenant l'espérance de l'égalité $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = X_n$ on obtient $\mathbb{E}[X_{n+1}] = \mathbb{E}[X_n]$, ce qui implique le résultat. □

6 Le théorème d'arrêt

Considérons un jeu de pile ou face où l'on mise 1 euro à chaque fois. Appelons X_n notre gain à la n-ième partie (X_n vaut donc 1 ou -1 avec probabilité $1/2$) et notons

$$S_n = X_1 + \cdots + X_n$$

la somme des gains accumulés au temps n. On a vu précédemment que (S_n) est une martingale (le jeu est équitable). En particulier $\mathbb{E}[S_n] = 0$ pour tout n, ce qui montre que le gain moyen est nul. On voudrait maintenant se permettre de quitter le jeu à un instant dépendant de l'issue même du jeu, donc à un temps T aléatoire. La question qui se pose est la suivante : est-il possible de choisir T de sorte que notre espérance de gain en quittant le jeu au temps T soit strictement positive ? La question n'est intéressante que si T vérifie certaines hypothèses raisonnables. Considérons quelques exemples :

— On quitte le jeu dès qu'on perd pour la première fois

$$T = \inf\{n \geq 1 : X_n = -1\};$$

— On quitte le jeu dès qu'on atteint les 10 euros de gain

$$T = \inf\{n \geq 1 : S_n = 10\};$$

— On quitte le jeu juste avant de perdre pour la première fois :

$$T = \inf\{n \geq 0 : X_{n+1} = -1\}.$$

Les deux premières stratégies semblent parfaitement légitimes. En revanche la troisième ne l'est pas, elle suppose de connaître au temps n le résultat du $n+1$-ième lancer de pièce. De manière plus formelle la bonne notion de stratégie à considérer est donnée par la définition suivante. Le cadre est toujours le même : on considère un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ muni d'une filtration $(\mathcal{F}_n)$.

Definition II.7. Soit $T\colon \Omega \to \mathbb{N} \cup \{+\infty\}$ une variable aléatoire. On dit que T est un *temps d'arrêt* si on a

$$\{T = n\} \in \mathcal{F}_n, \quad \forall n \in \mathbb{N}.$$

Noter qu'un temps d'arrêt peut prendre la valeur $+\infty$.

Remarque II.8. On ne change rien à la définition en remplaçant l'ensemble $\{T = n\}$ par l'ensemble $\{T \leq n\}$. En effet si $\{T = n\} \in \mathcal{F}_n$ pour tout n alors $\{T = k\} \in \mathcal{F}_k \subset \mathcal{F}_n$ pour tout $k \leq n$. Et donc

$$\{T \leq n\} = \{T = 0\} \cup \cdots \cup \{T = n\} \in \mathcal{F}_n.$$

La réciproque se montre de manière similaire.

Exemple. On reprend les trois exemples précédents la définition, la filtration étant bien sûr la filtration naturelle des (X_n). Le premier temps est bien un temps d'arrêt puisque

$$\{T \leq n\} = \{X_1 = -1\} \cup \cdots \cup \{X_n = -1\} \in \mathcal{F}_n.$$

De même, le deuxième est aussi un temps d'arrêt. Le troisième n'est pas un temps d'arrêt, puisque $\{T = n\}$ dépend de X_{n+1}, qui n'est pas $\mathcal{F}_n$-mesurable. Plus précisément, on a par exemple

$$\{T = 1\} = \{X_1 = 1\} \cap \{X_2 = -1\} \notin \mathcal{F}_1.$$

Lemme II.9. *Soit (X_n) un processus adapté et B un Borélien. Alors le temps T défini par*

$$T = \inf\{n \geq 0 : \, X_n \in B\}$$

est un temps d'arrêt, appelé temps d'atteinte de B.

Démonstration. Comme (X_n) est adapté $\{X_k \in B\} \in \mathcal{F}_k \subset \mathcal{F}_n$ pour tout $k \leq n$. Donc

$$\{T \leq n\} = \{X_0 \in B\} \cup \cdots \cup \{X_n \in B\} \in \mathcal{F}_n,$$

ce qu'il fallait démontrer. □

Remarque II.10. Dans la suite on adoptera la convention $\inf(\emptyset) = +\infty$. Donc si $X_n \notin B$ pour tout n on dit que le temps d'atteinte de B est infini.

Étant donné un processus (X_n) et un temps aléatoire T on considère le processus arrêté en T

$$X_{n\wedge T} = \begin{cases} X_n, & \text{si } n \leq T \\ X_T, & \text{sinon.} \end{cases}$$

Il faut bien comprendre que X_T est la variable aléatoire $\omega \mapsto X_{T(\omega)}(\omega)$. On en arrive maintenant au théorème d'arrêt de Doob.

Théorème II.11. *Si T est un temps d'arrêt et (X_n) est une martingale, alors $(X_{n\wedge T})$ est une martingale. On a le même résultat pour les sous-martingales et les sur-martingales.*

Remarque II.12. Le théorème est vrai sans aucune hypothèse d'intégrabilité sur le temps T qui peut même prendre la valeur $+\infty$ avec probabilité positive.

On va en fait montrer un résultat un peu plus général. Rappelons que pour tout $n \geq 1$ on note $\Delta X_n = X_n - X_{n-1}$ l'accroissement du processus X au temps n.

Definition II.13 (Transformée de martingale)**.** Soit (X_n) un processus adapté et $(C_n)_{n\geq 1}$ un processus prévisible. On définit un nouveau processus, noté $C \cdot X$ en posant $(C \cdot X)_0 = 0$ et $\Delta(C \cdot X)_n = C_n \Delta X_n$ pour tout $n \geq 1$. Autrement dit

$$(C \cdot X)_n = \sum_{i=1}^{n} C_i (X_i - X_{i-1}).$$

Remarque II.14. La variable $(C \cdot X)_n$ est $\mathcal{F}_n$-mesurable comme somme de variables $\mathcal{F}_n$-mesurables. Le processus $C \cdot X$ est donc adapté.

Lemme II.15. *Soit $(C_n)_{n\geq 1}$ un processus prévisible, tel que $C_n \in L^\infty$ pour tout entier n. Alors pour toute martingale (X_n) le processus $C \cdot X$ est une martingale. Si de plus le processus (C_n) est positif, alors on a le même résultat pour les sous-martingales et les sur-martingales.*

Démonstration. On a déjà vu que $C \cdot X$ est adapté. Comme X_n et X_{n-1} sont intégrables ΔX_n aussi. Et comme C_n est bornée $\Delta(C \cdot X)_n = C_n \Delta X_n$ est aussi intégrable. Donc $(C \cdot X)_n$ est intégrable comme somme de variables intégrables. En utilisant la prévisibilité de (C_n) on obtient

$$\mathbb{E}[\Delta(C \cdot X)_n | \mathcal{F}_{n-1}] = \mathbb{E}[C_n \Delta X_n | \mathcal{F}_{n-1}] = C_n \mathbb{E}[\Delta X_n | \mathcal{F}_{n-1}]$$

pour tout $n \geq 1$. Par conséquent, si $\mathbb{E}[\Delta X_n \mid \mathcal{F}_{n-1}] = 0$ alors

$$\mathbb{E}[\Delta(C \cdot X)_n \mid \mathcal{F}_{n-1}] = 0,$$

ce qui montre que $C \cdot X$ est une martingale. De même si $\mathbb{E}[\Delta X_n \mid \mathcal{F}_{n-1}] \geq 0$ et $C_n \geq 0$ alors

$$\mathbb{E}[\Delta(C \cdot X)_n \mid \mathcal{F}_{n-1}] \geq 0,$$

donc $C \cdot X$ est une sous-martingale. □

Démonstration du Théorème II.11. Remarquons qu'un processus arrêté $(X_{n \wedge T})$ vérifie

$$X_{n \wedge T} - X_{n-1 \wedge T} = \begin{cases} X_n - X_{n-1} & \text{si } n \leq T \\ 0 & \text{sinon.} \end{cases}$$

Autrement dit

$$\Delta X_{n \wedge T} = \mathbb{1}_{\{n \leq T\}} \Delta X_n$$

Par conséquent

$$X_{n \wedge T} = X_0 + (C \cdot X)_n, \quad \forall n \geq 0,$$

où (C_n) est le processus donné par $C_n = \mathbb{1}_{\{n \leq T\}}$. Remarquons que (C_n) est positif et borné par 1. De plus si T est un temps d'arrêt alors

$$\{n \leq T\} = \{T \leq n-1\}^c \in \mathcal{F}_{n-1}$$

ce qui montre que (C_n) est prévisible. En appliquant le lemme précédent on obtient le résultat cherché. □

On a donc montré que si (X_n) est une martingale et T un temps d'arrêt alors $(X_{n \wedge T})$ est une martingale. En particulier, l'espérance du processus arrêté est constante au cours du temps et on obtient

$$\mathbb{E}[X_0] = \mathbb{E}[X_{0 \wedge T}] = \mathbb{E}[X_{n \wedge T}]$$

pour tout entier n. On se pose maintenant la question de savoir si $\mathbb{E}[X_T] = \mathbb{E}[X_0]$. Remarquons d'abord que si T prend la valeur $+\infty$ avec probabilité positive alors la variable X_T n'est pas forcément bien définie : on ne sait pas si la suite (X_n) a une limite presque sure (ce sera l'objet du chapitre suivant) et on ne peut pas parler a priori de X_∞. L'exemple suivant montre que même si T est fini presque surement l'égalité $\mathbb{E}[X_0] = \mathbb{E}[X_T]$ peut s'avérer fausse.

Exemple. Soit (S_n) la marche aléatoire simple sur $\mathbb{Z}$ partant de 0. Le processus (S_n) est donc donné par

$$S_0 = 0, \quad S_n = \sum_{i=1}^{n} X_i, \quad \forall n \geq 1,$$

où (X_i) est une suite de variables i.i.d. prenant les valeurs 1 et -1 avec probabilité 1/2. On sait que (S_n) est une martingale par rapport à la filtration naturelle des (X_i). Soit

$$T = \inf\{n \geq 0 \colon S_n = 1\}$$

le temps d'atteinte de 1. Alors T est un temps d'arrêt et donc

$$\mathbb{E}[S_{n \wedge T}] = \mathbb{E}[S_0] = 0$$

pour tout n. De plus, on peut montrer que T est fini presque surement, et par définition de T, on a $S_T = 1$ p.s. Donc $\mathbb{E}[S_T] = 1 \neq \mathbb{E}[S_0]$.

Théorème II.16 (Théorème d'arrêt optionnel de Doob)**.** *Soit $(X_n)_{n\geq 0}$ une martingale et soit T un temps d'arrêt. La variable X_T est intégrable et vérifie $\mathbb{E}[X_T] = \mathbb{E}[X_0]$ dans chacun des cas suivants :*

(i) T est borné ;

(ii) $(X_{n\wedge T})$ est bornée dans L^∞ et $T < +\infty$ p.s. ;

(iii) T est intégrable et la suite (ΔX_n) est bornée dans L^∞.

On a un résultat analogue pour les sous-martingales, on obtient dans ce cas $\mathbb{E}[X_0] \leq \mathbb{E}[X_T]$.

Démonstration. On sait d'après le Théorème II.11 que $(X_{n\wedge T})$ est une martingale et donc que pour tout $n \geq 1$
$$\mathbb{E}[X_0] = \mathbb{E}[X_{0\wedge T}] = \mathbb{E}[X_{n\wedge T}]. \tag{1}$$
Si T est borné, il existe un entier N tel que $T \leq N$ p.s. Donc $N \wedge T = T$ p.s. et en appliquant l'égalité précédente à $n = N$ on obtient le résultat cherché.
Si T est fini presque surement, alors la variable X_T est bien définie et avec probabilité 1, on a $X_{n\wedge T} = X_T$ pour n assez grand. En particulier
$$X_{n\wedge T} \to X_T, \quad \text{p.s.}$$
Si de plus $(X_{n\wedge T})$ est bornée alors par convergence dominée on obtient
$$\mathbb{E}[X_{n\wedge T}] \to \mathbb{E}[X_T].$$
On obtient donc le résultat cherché en passant à la limite dans (1).
Dans le troisième cas, on sait qu'il existe une constante K telle que $|\Delta X_n| \leq K$ p.s. pour tout n et on écrit
$$|X_{n\wedge T}| = \left| X_0 + \sum_{k=1}^{n\wedge T} \Delta X_k \right| \leq |X_0| + K(n\wedge T) \leq |X_0| + KT.$$
Comme T est intégrable, la suite $(X_{n\wedge T})$ est dominée par une variable intégrable, et à nouveau le théorème de convergence dominée s'applique. □

Remarque II.17. Les trois cas décrits dans le théorème sont loin d'être exhaustifs. Comme le montre la démonstration, on aura $\mathbb{E}[X_T] = \mathbb{E}[X_0]$ dès qu'on peut appliquer le théorème de convergence dominée à la suite $(X_{n\wedge T})$.

Exemple. On a vu que dans le cas de la marche aléatoire simple (S_n) et du temps T d'atteinte de 1, la conclusion du théorème était fausse. Par ailleurs on a dans ce cas $\Delta S_n = 1$ ou -1. En particulier (ΔS_n) est bornée. Le point (iii) du théorème montre donc $\mathbb{E}[T] = \infty$. La marche aléatoire atteint la valeur 1 presque surement, mais le temps qu'elle met à le faire est en moyenne infini.

Exemple (Ruine du joueur)**.** Soient x et N deux entiers tels que $x \leq N$. Soit (X_i) une suite i.i.d. de variables aléatoires valant 1 et -1 avec probabilité $1/2$. On pose
$$S_n = x + \sum_{i=1}^{n} X_i$$
pour tout n, on pose
$$T = \inf\{n \geq 1 \colon S_n = 0 \text{ ou } N\},$$
et on s'intéresse au processus $(S_{n\wedge T})$. C'est la fortune au temps n d'un joueur qui possède initialement x euros qui mise 1 euro sur un jeu de pile ou face à chaque instant et qui décide de s'arrêter dès qu'il est ruiné ou qu'il a atteint la somme de N euros. On veut évaluer la probabilité de ruine $\mathbb{P}(S_T = 0)$. On sait que (S_n) est une martingale (le jeu est équitable) et que T est un temps d'arrêt (c'est le temps d'atteinte de l'ensemble $\{0, N\}$). De plus, par définition de T on a $0 \leq S_{n\wedge T} \leq N$ pour tout n p.s. La suite $(S_{n\wedge T})$ est donc bornée. Remarquons maintenant que T est majoré par le temps S d'attente d'une suite de $N+1$ piles consécutifs et il est clair que S est fini p.s. (détails laissés en exercice). D'après le point (ii) du théorème d'arrêt on en déduit
$$x = \mathbb{E}[S_0] = \mathbb{E}[S_T].$$

Mais par définition de T, la variable S_T ne prend que deux valeurs 0 et N. Par conséquent

$$\mathbb{E}[S_T] = N\,\mathbb{P}(S_T = N)$$

et on en déduit que la probabilité de ruine vérifie

$$\mathbb{P}(S_T = 0) = 1 - \frac{x}{N}.$$

En conclusion les théorèmes d'arrêt répondent par la négative à la question initialement posée : dans un jeu équitable il n'y a pas de stratégie permettant de rendre l'espérance de gain strictement positive.

7 Convergence des martingales

On s'intéresse maintenant à la convergence des martingales. Commençons par remarquer que toutes les martingales ne convergent pas. Par exemple la marche aléatoire simple (S_n) est une martingale vérifiant presque surement

$$|S_n - S_{n-1}| = 1, \quad \forall n \in \mathbb{N}.$$

En particulier, avec probabilité 1 la suite (S_n) n'est pas de Cauchy.

On dit qu'un processus est borné dans L^1 si la suite $(\|X_n\|_1)$ est bornée, autrement dit s'il existe une constante $C > 0$ telle que

$$\mathbb{E}[|X_n|] \le C, \quad \forall n \in \mathbb{N}.$$

Le théorème fondamental suivant est encore dû à Doob.

Théorème II.18. *Une sous-martingale $(X_n)_{n\ge1}$ bornée dans L^1 converge presque surement. De plus la limite X_∞ appartient à L^1. En particulier $|X_\infty| < +\infty$ p.s.*

Remarque II.19. Évidemment le résultat est aussi vrai pour les sur-martingales (remplacer X_n par $-X_n$).

En revanche la convergence de (X_n) vers X_∞ n'a pas forcément lieu dans L_1, comme le montre l'exemple suivant.

Exemple (Quitte ou double)**.** Soit (U_n) une suite de variables i.i.d. valant 0 et 2 avec probabilité 1/2. Soit $X_0 = 1$ et

$$X_n = \prod_{i=1}^{n} U_i, \quad \forall n \ge 1.$$

À chaque temps on double nos gains avec probabilité 1/2 ou on perd tout avec probabilité 1/2. Il est facile de voir que (X_n) est une martingale pour la filtration naturelle des (U_n). De plus $X_n \ge 0$ donc

$$\|X_n\|_1 = \mathbb{E}[X_n] = 1,$$

ce qui montre que (X_n) est bornée dans L^1. D'après le théorème précédent (X_n) converge presque surement. Il est même facile d'identifier la limite. En effet avec probabilité 1, il existe k tel que $U_k = 0$, et donc tel que $X_n = 0$ pour tout $n \ge k$. Par conséquent $X_n \to 0$ p.s. Comme par ailleurs $\mathbb{E}[X_n] = 1$ pour tout n, la suite (X_n) ne tend pas vers 0 dans L^1.

Si (X_n) est une sur-martingale positive alors

$$\mathbb{E}[|X_n|] = \mathbb{E}[X_n] \le \mathbb{E}[X_0].$$

Donc (X_n) est bornée dans L_1. Plus généralement une sur-martingale minorée est bornée dans L_1. De même pour une sous-martingale majorée. Le Théorème II.18 admet donc le corollaire suivant.

Corollaire II.20. *Une sous-martingale majorée converge presque surement. De plus la limite appartient à L_1. De même pour une sur-martingale minorée.*

Pour retenir ce résultat il suffit de se rappeler qu'une sous-martingale est un jeu favorable, donc croissant en moyenne. Le corollaire précédent est donc une version martingale du théorème d'analyse selon lequel une suite croissante et majorée converge.

On va décomposer la démonstration du Théorème II.18 en plusieurs étapes. La première étape est juste un résultat d'analyse. Soit (x_i) une suite réelle (déterministe) et soient $a < b$. On appelle nombre de montées de a à b au temps n, noté $U_n(a, b)$ le nombre de fois que la suite (x_i) est passée en dessous de a puis au dessus de b jusqu'au temps n. Plus précisément $U_n(a, b)$ est l'entier k maximal tel qu'il existe une suite croissante $(i_1, \ldots, i_{2k})$ d'entiers compris entre 0 et n tels que pour tout $l \in \{1, \ldots, k\}$ on ait

$$x_{i_{2l-1}} < a \quad \text{et} \quad x_{i_{2l}} > b.$$

Enfin on appelle $U(a, b) = \lim_n U_n(a, b)$ le nombre total de montées de a à b pour la suite (x_i).

Lemme II.21. *Si pour tous rationnels $p < q$ on a $U(p, q) < +\infty$ alors $\lim_n x_n$ existe dans $\mathbb{R} \cup \{+\infty, -\infty\}$.*

Démonstration. Posons $a = \liminf x_n$ et $b = \limsup x_n$, il s'agit de montrer que $a = b$. Raisonnons par contraposée. Si $a < b$ on peut par densité de $\mathbb{Q}$ dans $\mathbb{R}$ trouver deux rationnels p, q tels que

$$a < p < q < b.$$

Alors par définition de a et b la suite (x_n) passe une infinité de fois en dessous de p et une infinité de fois au dessus de q. On en déduit facilement que $U(p, q) = +\infty$, ce qu'il fallait démontrer. ☐

Le lemme clé dans la démonstration du théorème est le suivant.

Lemme II.22. *Soit (X_n) une sur-martingale et soient $a < b$ des réels. Alors pour tout entier n, le nombre de montées $U_n(a, b)$ au temps n du processus (X_i) vérifie*

$$\mathbb{E}[U_n(a, b)] \le \frac{\mathbb{E}[(X_n - a)_-]}{b - a}$$

où $(X_n - a)_-$ désigne la partie négative de $X_n - a$.

Démonstration. On interprète (X_n) comme un jeu défavorable, et on considère la stratégie suivante : on commence à jouer quand (X_n) passe en dessous de a pour la première fois, on accumule à ce moment les gains ΔX_n jusqu'à ce que (X_n) dépasse la valeur b. On s'arrête alors de jouer, jusqu'à ce que (X_n) repasse en dessous de de a, moment auquel on se remet à jouer, et ainsi de suite. Appelons (W_n) le processus des gains accumulés au temps n avec cette stratégie. On a alors $W_0 = 0$ et

$$\Delta W_n = C_n \Delta X_n$$

avec $C_n = 0$ ou 1 selon que l'on soit en train de jouer ou pas. Autrement dit $W_n = (C \cdot X)_n$. La description de la stratégie nous donne $C_1 = \mathbb{1}_{\{X_0 < a\}}$ et pour $n \ge 2$

$$C_n = \mathbb{1}_{\{C_{n-1}=0, X_{n-1}<a\}} + \mathbb{1}_{\{C_{n-1}=1, X_{n-1}\le b\}}.$$

On en déduit facilement que (C_n) est prévisible, et donc, en utilisant le théorème d'arrêt, que (W_n) est une sur-martingale. En particulier $\mathbb{E}[W_n] \le 0$ pour tout n. Par ailleurs, on a l'inégalité

$$W_n \ge (b - a)U_n(a, b) - (X_n - a)_-.$$

En effet chaque montée complète de a à b nous a fait gagner au moins $b - a$, et la montée qu'on est éventuellement en train de faire au temps n nous a fait perdre $(X_n - a)_-$ au pire. En prenant l'espérance et en utilisant l'inégalité $\mathbb{E}[W_n] \le 0$ on obtient le résultat. ☐

Démonstration du Théorème II.18. Soit (X_n) une sur-martingale bornée dans L^1. Il existe C tel que $\mathbb{E}[|X_n|] \le C$ pour tout n. Soient $a < b$ des réels, d'après le Lemme II.22, le nombre de montées au temps n du processus (X_n) vérifie

$$\mathbb{E}[U_n(a, b)] \le \frac{\mathbb{E}[(X_n - a)_-]}{b - a} \le \frac{\mathbb{E}[|X_n|] + |a|}{b - a} \le \frac{C + |a|}{b - a},$$

Soit $U(a,b) = \lim_n U_n(a,b)$ le nombre total de montées. Par convergence monotone on peut passer à la limite dans l'inégalité précédente et on obtient

$$\mathbb{E}[U(a,b)] \leq \frac{C + |a|}{b - a} < +\infty.$$

Par conséquent $U(a,b)$ est intégrable, donc fini presque surement. Et ce pour tous $a < b$. Comme $\mathbb{Q}$ est dénombrable, on en déduit qu'avec probabilité 1, $U(p,q) < +\infty$ pour tous rationnels $p < q$. D'après le Lemme II.21, ceci implique qu'avec probabilité 1 la suite (X_n) admet une limite X_∞ (pouvant éventuellement prendre les valeurs $\pm\infty$). Il reste à montrer qu'en fait X_∞ est intégrable. D'une part $|X_n| \to |X|$ p.s. et d'autre part $\mathbb{E}[|X_n|] \leq C$ pour tout n. Donc par Fatou

$$\mathbb{E}[|X_\infty|] = \mathbb{E}\left[\liminf |X_n|\right] \leq \liminf \mathbb{E}[|X_n|] \leq C,$$

ce qui montre que X_∞ est intégrable. □

On a vu qu'une martingale bornée dans L^1 convergeait presque surement mais qu'il n'y avait pas nécessairement convergence dans L^1. On va voir maintenant que ce phénomène ne se produit pas dans L^2.

Théorème II.23. *Soit $(M_n)_{n\geq 0}$ une martingale bornée dans L^2. Alors (M_n) converge presque surement et dans L^2.*

Évidemment les limites p.s. et L^2 coïncident (par exemple parce que les deux modes de convergence impliquent la convergence en probabilité). Si (M_n) est bornée dans L^2 elle l'est aussi dans L^1. Le fait que (M_n) converge p.s. est donc une conséquence du Théorème II.18. Par ailleurs la convergence L^2 implique la convergence L^1. Le résultat implique donc qu'une martingale bornée dans L^2 converge dans L^1. Ceci n'est pas en contradiction avec l'exemple du quitte ou double qui n'est pas borné dans L^2 puisqu'il vérifie $\mathbb{E}[M_n^2] = 2^n$ pour tout n.

Démonstration. D'après ce qui précède il suffit de montrer que (M_n) converge dans L^2. Remarquons que les incréments de (M_n) sont orthogonaux : pour $i \neq j$, on a

$$\mathbb{E}[\Delta M_i \Delta M_j] = 0.$$

En effet supposons par exemple que $i < j$, comme ΔM_i est $\mathcal{F}_i$-mesurable et $\Delta M_i \Delta M_j$ est intégrable, on a

$$\mathbb{E}[\Delta M_i \Delta M_j] = \mathbb{E}\left[\Delta M_i \mathbb{E}[\Delta M_j \mid \mathcal{F}_i]\right].$$

Par ailleurs la propriété de martingale donne $\mathbb{E}[\Delta M_j \mid \mathcal{F}_i] = 0$, d'où le résultat. Par Pythagore on obtient donc

$$\sum_{i=1}^n \|\Delta M_i\|_2^2 = \|M_n - M_0\|_2^2.$$

De plus le membre de droite est borné par hypothèse. On en déduit que la série $\sum \|\Delta M_i\|_2^2$ est convergente. Par conséquent, pour $m < n$ entiers,

$$\|M_n - M_m\|_2^2 = \sum_{i=m+1}^n \|\Delta M_i\|_2^2$$

tend vers 0 quand $m,n \to +\infty$. Autrement dit la suite (M_n) est de Cauchy dans L^2, donc elle converge. □

8 Application : séries de variables indépendantes

Dans cette section on se donne une suite de variables aléatoires indépendantes (X_i) et on s'intéresse à la convergence de la série $\sum X_i$. La convergence est ici à prendre au sens de la convergence simple : on dit que la série $\sum X_i$ converge si la suite des sommes partielles admet une limite finie.

Théorème II.24. *Si les variables X_i sont de carré intégrable et d'espérance nulle on a l'implication suivante*

$$\sum \mathrm{var}(X_i) < +\infty \quad \Rightarrow \quad \sum X_i \text{ converge p.s.}$$

Si de plus la suite (X_i) est bornée dans L^∞ la réciproque est vraie.

Démonstration. On pose $S_n = \sum_{i=1}^n X_i$ pour tout n. En utilisant l'indépendance des (X_i) et leur caractère centré, on montre facilement que (S_n) est une martingale par rapport à la filtration naturelle des (X_i), notée $(\mathcal{F}_i)$ par la suite. De plus comme S_n est de moyenne nulle on a par indépendance

$$\mathbb{E}[S_n^2] = \mathrm{var}(S_n) = \sum_{i=1}^n \mathrm{var}(X_i) = \sum_{i=1}^n \mathrm{var}(X_i).$$

Si $\sum \mathrm{var}(X_i) < +\infty$ on en déduit que la martingale (S_n) est bornée dans L^2, donc qu'elle converge p.s. et dans L^2. En particulier la limite est finie p.s. ce qui démontre la première partie du théorème. Montrons maintenant la réciproque, en supposant de plus que $|X_i| \leq C$ p.s. pour tout i et pour une certaine constante $C > 0$. On pose $A_n = \sum_{i=1}^n \mathrm{var}(X_i)$. Comme les (X_n) sont indépendantes et centrées on a

$$\mathbb{E}[S_n^2 \mid \mathcal{F}_{n-1}] = \mathbb{E}[S_{n-1}^2 + 2X_nS_{n-1} + X_n^2 \mid \mathcal{F}_{n-1}] = S_{n-1}^2 + \mathrm{var}(X_n),$$

ce qui montre que le processus $(S_n^2 - A_n)$ est une martingale. On fixe $K > 0$ et on pose

$$T = \inf\{n \geq 1 \colon |S_n| > K\}.$$

Comme T est un temps d'arrêt on a d'après le théorème d'arrêt

$$\mathbb{E}[S_{n\wedge T}^2 - A_{n\wedge T}] = \mathbb{E}[S_1 - A_1] = 0.$$

De plus par définition de T, on a $|S_{n\wedge T}| = |S_n| \leq K$ si $n < T$ et

$$|S_{n\wedge T}| = |S_T| \leq |S_{T-1}| + |X_T| \leq K + C, \quad \text{si } T \leq n.$$

Dans tous les cas $|S_{n\wedge T}| \leq K + C$ p.s. Par conséquent $\mathbb{E}[A_{n\wedge T}] \leq (K+C)^2$ pour tout entier n. D'un autre côté, comme (S_n) converge p.s. $\sup_n |S_n| < +\infty$ p.s. Par conséquent si K est suffisamment grand on a $\sup_n |S_n| \leq K$ et donc $T = +\infty$ avec probabilité positive. Et comme $A_{n\wedge T} \geq A_n \mathbb{1}_{\{T=\infty\}}$ on en déduit

$$\sum_{i=1}^n \mathrm{var}(X_i) \leq \frac{(K+C)^2}{\mathbb{P}(T=+\infty)}, \quad \forall n \geq 1.$$

En faisant tendre n vers $+\infty$ on obtient la convergence de la série $\sum \mathrm{var}(X_i)$. □

C'est un théorème très puissant. Combiné au lemme suivant, il permet notamment de retrouver la loi des grands nombres.

Lemme II.25 (Lemme de Kronecker)**.** *Soit (x_n) une suite de réels. Si la série $\sum n^{-1}x_n$ converge alors*

$$\frac{1}{n}\sum_{k=1}^n x_k \to 0.$$

Démonstration. On pose $u_n = \sum_{k\leq n} k^{-1}x_k$ pour tout $n \geq 1$ et on appelle l la limite de (u_n). Remarquons que

$$\sum_{k=1}^n u_k = (n+1)u_n - \sum_{k=1}^n x_k.$$

Or par Cesàro $n^{-1}\sum_{k\leq n} u_k \to l$. Donc en divisant l'égalité précédente par n et en faisant tendre n vers l'infini on obtient le résultat. □

On est maintenant en position de démontrer la loi forte des grands nombres.

Théorème II.26 (Loi forte des grands nombres)**.** *Soit (X_n) une suite i.i.d. de variables aléatoires intégrables. On a*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to \mathbb{E}[X_1], \quad p.s.$$

Démonstration. Les X_i ne sont pas supposées de carré intégrable. En vue d'utiliser le Théorème II.24 on commence par les tronquer : on pose $Y_i = X_i \mathbb{1}_{\{|X_i| \le i\}}$ pour tout entier i. Notons que puisque les X_i sont identiquement distribuées et intégrables

$$\sum_{i=1}^{\infty} \mathbb{P}(X_i \neq Y_i) = \sum_{i=1}^{\infty} \mathbb{P}(|X_1| > i) \le \mathbb{E}[|X_1|] < +\infty.$$

On a donc $X_i = Y_i$ à partir d'un certain rang, d'après Borel–Cantelli. Par conséquent si $n^{-1}\sum_{i\le n} Y_i$ converge alors $n^{-1}\sum_{i\le n} X_i$ converge aussi et les limites coïncident. Il suffit donc de montrer que $n^{-1}\sum_{i\le n} Y_i \to \mathbb{E}[X_1]$ p.s. Posons $Z_i = Y_i - \mathbb{E}[Y_i]$. Les Z_i sont indépendantes et centrées. De plus

$$\sum_{i=1}^{\infty} \mathrm{var}(Z_i/i) \le \sum_{i=1}^{\infty} i^{-2}\mathbb{E}[Y_i^2] = \sum_{i=1}^{\infty} i^{-2}\mathbb{E}[X_1^2 \mathbb{1}_{\{|X_1|\le i\}}] \le C\,\mathbb{E}[|X_1|],$$

pour une certaine constante C. Comme X_1 est intégrable on voit que la série $\sum \mathrm{var}(Z_i/i)$ converge. D'après le Théorème II.24 on en déduit que $\sum Z_i/i$ converge p.s. Donc, d'après le lemme de Kronecker

$$\frac{1}{n}\sum_{i=1}^{n} Z_i \to 0 \quad \text{p.s.}$$

Par ailleurs, on a $\mathbb{E}[Y_i] = \mathbb{E}[X_1 \mathbb{1}_{\{|X_1|\le i\}}] \to \mathbb{E}[X_1]$ par convergence dominée. Donc $n^{-1}\sum_{i\le n}\mathbb{E}[Y_i] \to \mathbb{E}[X_1]$ par Cesàro. On obtient donc

$$\frac{1}{n}\sum_{i\le n} Y_i \to \mathbb{E}[X_1] \quad \text{p.s.}$$

ce qu'il fallait démontrer. □

Revenons aux séries de variables indépendantes. Soit (X_i) une suite de variables indépendantes. On suppose que les variables X_i sont de carré intégrable, mais on ne suppose plus qu'elles sont centrées.

Théorème II.27 (Théorème des deux séries)**.** *Si les deux séries $\sum \mathbb{E}[X_i]$ et $\sum \mathrm{var}(X_i)$ sont convergentes alors la série $\sum X_i$ converge p.s. Encore une fois la réciproque est vraie à condition de supposer que la suite (X_i) est bornée dans L^∞.*

Démonstration. Le sens direct est immédiat : en appliquant le Théorème II.24 à la suite (Y_i) donnée par $Y_i = X_i - \mathbb{E}[X_i]$, on obtient la convergence p.s. de $\sum Y_i$. Mais comme $\sum \mathbb{E}[X_i]$ converge ceci implique que $\sum X_i$ converge p.s. La réciproque est plus astucieuse. On pose $Z_i = X_i - X_i'$, où la suite (X_i') est une copie indépendante de la suite (X_i). On peut alors appliquer la partie « réciproque » du Théorème II.24 aux variables Z_i. On en déduit que $\sum \mathrm{var}(Z_i) < +\infty$. Mais comme $\mathrm{var}(Z_i) = 2\mathrm{var}(X_i)$, on obtient $\sum \mathrm{var}(X_i) < +\infty$. En appliquant maintenant le sens direct du théorème on obtient la convergence p.s. de la série $\sum Y_i$. Mais si les séries $\sum X_i$ et $\sum Y_i$ convergent toutes les deux, alors $\sum \mathbb{E}[X_i]$ converge aussi. □

On en vient au théorème principal de cette section.

Théorème II.28 (Théorème des trois séries)**.** *Soit (X_i) une suite de variables indépendantes. Les X_i sont supposées finies p.s. mais pas forcément intégrables. La série $\sum X_i$ converge p.s. si et seulement s'il existe une constante $C > 0$ telle que les trois séries suivantes convergent :*

(i) $\sum \mathbb{P}(|X_i| > C)$,

(ii) $\sum \mathbb{E}\left[X_i \mathbb{1}_{\{|X_i|>C\}}\right]$,

(iii) $\sum \mathrm{var}\left(X_i \mathbb{1}_{\{|X_i|>C\}}\right)$.

De plus les trois séries convergent alors pour toute constante $C > 0$.

Démonstration. Supposons qu'il existe C tel que les trois séries convergent et posons $Y_i = X_i \mathbb{1}_{\{|X_i| \leq C\}}$. On sait que $\sum \mathbb{E}[Y_i]$ et $\sum \text{var}(Y_i)$ convergent. Donc $\sum Y_i$ converge p.s. d'après le théorème des deux séries. De plus par Borel-Cantelli, la convergence de la série (i) implique qu'avec probabilité 1 on a $Y_i = X_i$ à partir d'un certain rang. Donc $\sum X_i$ converge p.s. Réciproquement, on suppose que $\sum X_i$ converge p.s. et on se donne $C > 0$. Alors $X_i \to 0$ et donc $X_i \leq C$ à partir d'un certain rang, avec probabilité 1. Comme les X_i sont indépendantes la réciproque de Borel-Cantelli assure alors la convergence de la série (i). À nouveau on pose $Y_i = X_i \mathbb{1}_{\{|X_i| \leq C\}}$. Comme $Y_i = X_i$ à partir d'un certain rang, on a convergence p.s. de $\sum Y_i$, et comme les (Y_i) sont toutes bornées par C, la réciproque du théorème des deux séries assure alors la convergence des séries (ii) et (iii). □

Troisième partie

Chaines de Markov

9 Définitions et premières propriétés

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité. On va s'intéresser dans cette partie à des processus (X_n) à valeurs dans un espace M discret, c'est-à-dire que l'ensemble M sera toujours supposé fini ou dénombrable.

Definition III.1. Soit (X_n) un processus à valeurs dans M. On dit que (X_n) est une *chaine de Markov* si pour tout entier n et pour toute suite $x_0, \ldots, x_{n+1}$ d'éléments de M vérifiant

$$\mathbb{P}(X_0 = x_0, \ldots, X_n = x_n) > 0,$$

on a

$$\mathbb{P}(X_{n+1} = x_{n+1} \mid X_0 = x_0, \ldots, X_n = x_n) = \mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n)$$

Si de plus cette dernière quantité ne dépend pas de n mais seulement de x_n et x_{n+1} on dit que la chaine est *homogène*. L'ensemble M est appelé *espace d'états*.

Autrement dit, un processus (X_n) est une chaine de Markov si pour tout temps n, la loi conditionnelle du futur X_{n+1} sachant le passé $X_0, \ldots, X_n$ ne dépend que du présent X_n et éventuellement de l'instant n. Si de plus cette loi ne dépend pas du temps n la chaine est homogène. Par la suite, on ne va s'intéresser qu'à des chaines homogènes, par conséquent l'adjectif homogène sera dorénavant sous-entendu.

Definition III.2. Une *matrice de transition* sur M est une application $P \colon M \times M \to \mathbb{R}_+$ telle que pour tout $x \in M$

$$\sum_{y \in M} P(x, y) = 1.$$

Si (X_n) est une chaine de Markov à valeurs dans M la matrice de transition P associée à (X_n) est la matrice P donnée par

$$P(x, y) = \mathbb{P}(X_1 = y \mid X_0 = x).$$

Remarque III.3. Si M est de cardinal n, P est une matrice de taille $n \times n$. Si M est infini P est une *matrice* avec une infinité de lignes et de colonnes.

Exemple (Marche aléatoire sur $\mathbb{Z}$)**.** Un joueur lance successivement une pièce de monnaie truquée, tombant sur pile avec probabilité p. Chaque fois qu'il obtient un pile, il reçoit un euro, chaque fois qu'il obtient un face, il perd un euro. Alors la fortune S_n du joueur à l'étape n vérifie

$$S_n = \begin{cases} S_{n-1} + 1 & \text{avec probabilité } p \\ S_{n-1} - 1 & \text{avec probabilité } 1 - p \end{cases}$$

indépendamment du passé. Ceci montre que (S_n) est une chaine de Markov sur $\mathbb{Z}$ de matrice de transition P donnée par

$$P(k, k+1) = p, \quad P(k, k-1) = 1 - p$$

pour tout $k \in \mathbb{Z}$ et $P(k, l) = 0$ dans tous les autres cas.

Exemple (Marche aléatoire sur un graphe)**.** Soit $G = (V, E)$ un graphe non-orienté : V est un ensemble non vide et E est un ensemble d'arêtes sur V c'est-à-dire un ensemble de paires $\{x, y\}$ d'éléments de V. Si l'arête $\{x, y\}$ est présente dans E on dit que x et y sont voisins et on note cette relation $x \sim y$. Pour $x \in V$, on note $d(x)$ le *degré* de x, c'est-à-dire le nombre de voisins de x. La marche aléatoire sur G est le processus (X_n) à valeurs dans V qui évolue ainsi : pour tout temps n, conditionnellement au passé $X_0, \ldots, X_n$, la variable X_{n+1} est uniforme sur l'ensemble des voisins de X_n. Autrement dit, (X_n) est une chaine de Markov sur V de matrice de transition P donnée par

$$P(x, y) = \begin{cases} \frac{1}{d(x)} & \text{si } x \sim y \\ 0 & \text{sinon.} \end{cases}$$

Exemple (Urne d'Ehrenfest)**.** N molécules de gaz sont réparties dans un récipient divisé en deux enceintes séparées par une paroi poreuse. À chaque instant une particule choisie uniformément au hasard change d'enceinte. On note X_n le nombre de particules présente dans la première enceinte au temps n. Sachant le passé $X_1, \dots, X_{n-1}$ on a donc

$$X_n = \begin{cases} X_{n-1} + 1 & \text{avec probabilité } 1 - \frac{X_{n-1}}{N}; \\ X_{n-1} - 1 & \text{avec probabilité } \frac{X_{n-1}}{N}. \end{cases}$$

Autrement (X_n) est une chaine de Markov, d'espace d'état $M = \{0, \dots, N\}$ et de matrice de transition P donnée par $P(0,1) = 1$, $P(N, N-1) = 1$,

$$P(x, x+1) = 1 - \frac{x}{N}, \quad P(x, x-1) = \frac{x}{N}, \quad x \in \{1, \dots, N-1\}$$

et $P(x,y) = 0$ dans tous les autres cas.

Par définition (X_n) est une chaine de Markov de matrice de transition P si et seulement si pour tout n et pour toute suite $x_0, \dots, x_n$ d'éléments de M on a

$$\mathbb{P}(X_0 = x_0, \dots, X_n = x_n) = \mathbb{P}(X_0 = x_0, \dots, X_{n-1} = x_{n-1}) P(x_{n-1}, x_n).$$

En appliquant cette égalité de manière répétée on obtient l'égalité suivante.

Proposition III.4 (Équation de Chapman-Kolmogorov)**.** *Pour tout entier n et pour toute suite $x_0, \dots, x_n$ d'éléments de M on a*

$$\mathbb{P}(X_0 = x_0, \dots, X_n = x_n) = \mathbb{P}(X_0 = x_0) P(x_0, x_1) \cdots P(x_{n-1}, x_n),$$

En particulier la loi de X_0 et la matrice P déterminent la loi du n-uplet $(X_0, \dots, X_n)$ pour tout n.

Si P est une matrice de transition sur M on note P^2 la matrice donnée par

$$P^2(x,y) = \sum_{z \in M} P(x,z) P(z,y), \quad \forall x, y \in M.$$

Il est facile de voir que c'est encore une matrice de transition. Plus généralement, pour tout entier n la matrice P^n définie par

$$P^n(x,y) = \sum_{x_1, \dots, x_{n-1}} P(x, x_1) P(x_1, x_2) \cdots P(x_{n-1}, y), \quad \forall x, y \in M$$

est une matrice de transition.

Lemme III.5. *On a*

$$P^n(x,y) = \mathbb{P}(X_n = y \mid X_0 = x),$$

pour tout n et pour tous $x, y \in M$.

Démonstration. D'après Chapmann-Kolmogorov

$$\begin{aligned} \mathbb{P}(X_0 = x, X_n = y) &= \sum_{x_1, \dots, x_{n-1}} \mathbb{P}(X_0 = x, X_1 = x_1, \dots, X_{n-1} = x_{n-1}, X_n = y) \\ &= \sum_{x_1, \dots, x_{n-1}} \mathbb{P}(X_0 = x) P(x, x_1) \cdots P(x_{n-1}, y) = \mathbb{P}(X_0 = x) P^n(x,y). \end{aligned}$$

D'où le résultat. □

Dans la suite on identifiera une mesure μ sur M avec le vecteur ligne $(\mu(x))_{x \in M}$ et une fonction $f \colon M \to \mathbb{R}$ avec le vecteur colonne $(f(x))_{x \in M}$. Si P est une matrice de transition sur M, alors μP est la mesure donnée par

$$\mu P(x) = \sum_{y \in M} \mu(y) P(y,x), \quad x \in M,$$

et Pf est la fonction

$$Pf(x) = \sum_{y \in M} P(x,y)f(y), \quad x \in M.$$

En utilisant l'égalité $P^n(x,y) = \mathbb{P}(X_n = y \mid X_0 = x)$ on voit tout de suite que μP^n est la loi de X_n lorsque X_0 a pour loi μ. De manière analogue $P^n f(x)$ est la valeur moyenne de $f(X_n)$ sachant $X_0 = x$:

$$P^n f(x) = \mathbb{E}[f(X_n) \mid X_0 = x].$$

Definition III.6 (Récurrence aléatoire)**.** Soient $(E, \mathcal{E})$ un espace muni d'une tribu, soit $\xi_1, \xi_2, \ldots$ une suite de variables aléatoires i.i.d. à valeurs dans E et soit $f \colon E \times M \to M$ une application mesurable. Soit X_0 une variable à valeurs dans M, indépendante de la suite (ξ_i), on définit récursivement

$$X_{n+1} = f(\xi_{n+1}, X_n), \quad \forall n \geq 0.$$

La suite (X_n) est appelé récurrence aléatoire.

Lemme III.7. *Soit*

$$X_{n+1} = f(\xi_{n+1}, X_n)$$

une récurrence aléatoire, alors (X_n) *est une chaine de Markov de matrice de transition* P *donnée par*

$$P(x,y) = \mathbb{P}(f(\xi_1, x) = y), \quad \forall x, y \in M.$$

Démonstration. La définition de (X_n) montre que pour tout n la variable X_n est mesurable par rapport à la tribu

$$\sigma(X_0, \xi_1, \ldots, \xi_n).$$

Comme par hypothèse ξ_{n+1} est indépendante de cette tribu, on en déduit que ξ_{n+1} est indépendante de X_n, et plus généralement que ξ_{n+1} est indépendante de

$$\sigma(X_1, \ldots, X_n).$$

Par conséquent

$$\begin{aligned}
\mathbb{P}(X_{n+1} = x_{n+1} \mid X_0 = x_0, \ldots, X_n = x_n)& \\
&= \mathbb{P}(f(\xi_{n+1}, X_n) = x_{n+1} \mid X_0 = x_0, \ldots, X_n = x_n) \\
&= \mathbb{P}(f(\xi_{n+1}, x_n) = x_{n+1} \mid X_0 = x_0, \ldots, X_n = x_n) \\
&= \mathbb{P}(f(\xi_{n+1}, x_n) = x_{n+1}) \\
&= \mathbb{P}(f(\xi_1, x_n) = x_{n+1}),
\end{aligned}$$

ce qu'il fallait démontrer. □

Ce lemme fournit un critère concret pour montrer qu'un processus donné est une chaine de Markov.

Exemple. Reprenons l'exemple de la marche aléatoire et posons $\xi_n = 1$ si le n-ième lancer donne pile et -1 sinon. Alors (ξ_n) est une suite i.i.d. et on a

$$S_{n+1} = S_n + \xi_{n+1}$$

ce qui montre que (S_n) peut s'écrire comme une récurrence aléatoire.

Remarque III.8. On peut montrer que la réciproque du lemme précédent est vraie et que toute chaine de Markov peut s'écrire comme une récurrence aléatoire, mais nous n'utiliserons pas ce fait.

10 Espace canonique et propriété de Markov abstraite

Rappelons que l'espace M est supposé dénombrable et qu'on prendra toujours l'ensemble $\mathcal{P}(M)$ comme tribu sur M. Comme on l'a vu, on appelle processus à valeurs dans M une suite (X_n) de variables aléatoires à valeurs dans M définies sur un même espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. On voudrait parler de la loi d'un processus, qui est donc une mesure de probabilité sur $M^{\mathbb{N}}$. Pour ce faire il faut d'abord préciser la tribu considérée sur $M^{\mathbb{N}}$.

Definition III.9. On dit qu'un sous-ensemble A de M est *cylindrique* s'il ne dépend que d'un nombre fini de coordonnées. Autrement dit s'il existe un entier n et un sous-ensemble B de M^n tel que

$$A = \{\omega \in M^n \colon (\omega_1, \ldots, \omega_n) \in B\}.$$

La classe des ensembles cylindriques sera notée $\mathcal{C}$.

Definition III.10. On pose $\mathcal{G} = \sigma(\mathcal{C})$. Cette tribu est appelée *tribu produit* sur $M^{\mathbb{N}}$.

Le théorème suivant permet de construire des mesures sur $M^{\mathbb{N}}$ muni de la tribu produit. Pour tout entier n on note P_n la projection qui ne garde que les $n+1$ premiers termes d'une suite :

$$P_n(\omega) = (\omega_0, \ldots, \omega_n).$$

Rappelons la définition de la mesure image. Si μ est une mesure sur $M^{\mathbb{N}}$, la projection de μ sur M^{n+1} notée $P_n\#\mu$ est la mesure sur M^{n+1} définie par

$$P_n\#\mu = \mu(P_n^{-1}(A)), \quad \forall A \subset M^{n+1}.$$

Théorème III.11 (Théorème d'extension de Kolmogorov)**.** *Soit (μ_n) une suite de mesures telle que μ_n soit une mesure de probabilité sur M^{n+1} pour tout entier n. On suppose que les μ_n sont compatibles entre elles, au sens où pour tout entier n la projection de μ_{n+1} sur M^{n+1} coïncide avec μ_n. De manière plus explicite on a*

$$\mu_{n+1}(A \times M) = \mu_n(A), \quad \forall A \subset M^{n+1},\ \forall n \in \mathbb{N}. \tag{2}$$

Alors il existe une unique mesure de probabilité μ sur $(M^{\mathbb{N}}, \mathcal{G})$ dont la projection sur M^{n+1} soit égale à μ_n pour tout n.

Démonstration (idée). L'unicité est une conséquence du lemme de classe monotone. En effet supposons que μ et ν sont deux mesures de probabilité sur $M^{\mathbb{N}}$ vérifiant $P_n\#\mu = P_n\#\nu$ pour tout n. Ceci revient à dire que μ et ν coïncident sur la classe $\mathcal{C}$ des ensembles cylindriques. Mais comme $\mathcal{C}$ est une classe stable par intersections finies qui engendre la tribu $\mathcal{G}$, le lemme de classe monotone assure que μ et ν sont égales. Pour l'existence, on commence par définir une application $\widetilde{\mu}$ sur $\mathcal{C}$ de la manière suivante. Soit $B \in \mathcal{C}$. Il existe un entier n et un sous-ensemble A de M^{n+1} tel que $B = (P_n)^{-1}(A)$. On pose alors $\widetilde{\mu}(B) = \mu_n(A)$. Évidemment n n'est pas unique, on pourrait par exemple écrire aussi $B = (P_{n+1})^{-1}(A \times M)$. La condition de compatibilité (2) assure justement que la définition de $\widetilde{\mu}(B)$ ne dépend pas du n choisi. On montre ensuite que cette application définit une *pré-mesure* sur la classe $\mathcal{C}$ des ensembles cylindriques. Le théorème d'extension de Caratheodory assure ensuite l'existence d'une mesure μ sur $\mathcal{G} = \sigma(\mathcal{C})$ dont la restriction à $\mathcal{C}$ coïncide avec $\widetilde{\mu}$. □

Exemple (Mesure produit)**.** Soit ν une probabilité sur M. Il existe une unique mesure de probabilité sur $(M^{\mathbb{N}}, \mathcal{G})$ dont la projection sur M^n soit égale à la mesure produit $\nu^{\otimes n}$, pour tout entier n. Cette mesure est notée $\nu^{\otimes \mathbb{N}}$.

Exemple (Chaine de Markov)**.** Soit ν une mesure de probabilité sur M et soit P une matrice de transition. Pour $x = (x_0, \ldots, x_n) \in M^{n+1}$ on pose

$$\mu_n(\{x\}) = \nu(x_0) P(x_0, x_1) \cdots P(x_{n-1}, x_n).$$

Ceci définit une mesure de probabilité sur M^{n+1} et il est facile de voir que la suite (μ_n) vérifie la condition de compatibilité (2). Le théorème d'extension de Kolmogorov assure qu'il existe une unique mesure de probabilité μ sur $M^{\mathbb{N}}$ telle que

$$\mu(\{\omega \in M^{\mathbb{N}} \colon \omega_0 = x_0, \ldots, \omega_n = x_n\}) = \nu(x_0) P(x_0, x_1) \cdots P(x_{n-1}, x_n),$$

pour tout n et pour tout $x \in M^{n+1}$.

Pour tout entier n on note $C_n\colon \omega \in M^{\mathbb{N}} \mapsto \omega_n \in M$ l'application coordonnée correspondante. On peut reformuler l'équation précédente de la manière suivante : pour tout n et pour toute suite $x_0, \dots, x_n$ d'éléments de M on a

$$\mu(C_0 = x_0, \dots, C_n = x_n) = \nu(x_0)P(x_0, x_1) \cdots P(x_{n-1}, x_n).$$

Autrement dit, la suite (C_n) vue comme un processus défini sur l'espace de probabilité $(M^{\mathbb{N}}, \mathcal{G}, \mu)$ est une chaine de Markov, ayant loi ν au temps 0 et matrice de transition P. Cet espace est appelé *espace canonique*. Le processus de coordonnées (C_n) est appelé *processus canonique*. La mesure μ est la loi de la chaine de Markov sur M partant de ν et de matrice de transition P.

Venons en à la propriété de Markov. On introduit pour tout entier n l'application $\theta_n\colon M^{\mathbb{N}} \to M^{\mathbb{N}}$ qui décale de n rangs vers la gauche les coordonnées d'une suite :

$$\theta_n(\omega_0, \omega_1, \dots) = (\omega_n, \omega_{n+1}, \dots),$$

pour tout $\omega \in M^{\mathbb{N}}$. Remarquons que l'image réciproque d'un ensemble cylindrique est encore un ensemble cylindrique. Par conséquent l'application θ_n est mesurable pour la tribu produit $\mathcal{G}$. Comme on l'a vu plus haut, si $(X_n)_{n\geq 0}$ est une chaine de Markov sur M, sa loi est complètement déterminée par la loi de X_0 et la matrice de transition. Dans la suite on précisera parfois la loi de X_0 en indice des signes $\mathbb{P}$ et $\mathbb{E}$. De plus si $\mu = \delta_x$ on écrira $\mathbb{P}_x$ plutôt que $\mathbb{P}_{\delta_x}$.

Théorème III.12 (Propriété de Markov)**.** *Soit (X_n) une chaine de Markov à valeurs dans M et soit $(\mathcal{F}_n)$ sa filtration naturelle. Alors pour tout entier n, pour toute fonction $F\colon M^{\mathbb{N}} \to \mathbb{R}$ qui soit mesurable pour la tribu $\mathcal{G}$ et positive ou bornée on a*

$$\mathbb{E}[F \circ \theta_n(X) \mid \mathcal{F}_n] = u(X_n),$$

en posant $u(x) = \mathbb{E}_x[F(X)]$.

Démonstration. Il suffit de montrer le résultat pour $F = \mathbb{1}_A$ avec $A \in \mathcal{G}$. Par un argument de classe monotone on peut même supposer que A est cylindrique. Enfin comme M est dénombrable on peut se ramener au cas où A est l'image réciproque d'un singleton par P_k pour un certain k. Il suffit donc de montrer que pour tous entiers n et k et pour toute suite $x_0, \dots, x_k$ d'éléments de A on a

$$\mathbb{P}(X_n = x_0, \dots, X_{n+k} = x_k \mid \mathcal{F}_n) = \mathbb{P}_{x_0}(X_0 = x_0, \dots, X_k = x_k).$$

Or cette égalité est une conséquence directe de l'équation de Chapman-Kolmogorov. □

En appliquant ceci à une fonction F ne dépendant que d'une coordonnée on obtient le résultat suivant.

Corollaire III.13 (Propriété de Markov simple)**.** *Soit $f\colon M \to \mathbb{R}$. On a $\mathbb{E}[f(X_{n+k}) \mid \mathcal{F}_n] = P^k f(X_n)$ pour tous entiers k, n.*

Donnons maintenant un exemple d'application où la fonction F dépend de toute la trajectoire. Soit (X_n) une chaine de Markov à valeurs dans M, soit $y \in M$ et soit

$$T_y = \inf\{n \geq 0\colon X_n = y\}$$

le temps d'atteinte de y. Pour $x \in M$ on note $u(x) = \mathbb{E}_x[T_y]$ le temps moyen mis pour atteindre y partant de x.

Lemme III.14. *La fonction u vérifie le système suivant*

$$u(y) = 0, \quad u(x) = 1 + Pu(x),\ x \neq y.$$

Démonstration. Il est évident que $u(y) = 0$. Pour $\omega \in M^{\mathbb{N}}$ posons $F(\omega) = \inf\{n \geq 0\colon \omega_n = y\}$, de sorte que $T_y = F(X)$. Notons que F est bien $\mathcal{G}$-mesurable. En effet, pour tout entier n

$$F^{-1}(\{n\}) = \{\omega \in M^{\mathbb{N}}\colon \omega_0 \neq x, \dots, \omega_{n-1} \neq x, \omega_n = x\}$$

est un ensemble cylindrique. Par ailleurs, si $\omega_0 \neq y$ alors

$$F(\omega) = \inf\{n \geq 1\colon \omega_n = y\} = 1 + \inf\{n \geq 0\colon \omega_{n+1} = y\} = 1 + F \circ \theta_1(\omega).$$

Soit $x \neq y$. En utilisant l'égalité précédente et la propriété de Markov on obtient (en notant $\mathcal{F}_1$ la tribu engendrée par X_0 et X_1)

$$\mathbb{E}_x[T_y \mid \mathcal{F}_1] = \mathbb{E}_x[F(X) \mid \mathcal{F}_1] = 1 + \mathbb{E}_x[F \circ \theta_1(X) \mid \mathcal{F}_1] = 1 + u(X_1).$$

En prenant l'espérance de cette égalité on trouve

$$u(x) = \mathbb{E}_x[T_y] = 1 + \mathbb{E}_x[u(X_1)] = 1 + Pu(x)$$

ce qui est le résultat annoncé. □

On peut étendre la propriété de Markov au cas où le décalage n est aléatoire, à condition que ce soit un temps d'arrêt.

Definition III.15. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité muni d'une filtration $(\mathcal{F}_n)$ et soit $T\colon \Omega \to \mathbb{N} \cup \{+\infty\}$ un temps d'arrêt. On appelle tribu arrêtée au temps T l'ensemble

$$\mathcal{F}_T = \{A \in \mathcal{F} \colon A \cap \{T = n\} \in \mathcal{F}_n, \, \forall n \geq 0\}.$$

Proposition III.16. *On a les propriétés suivantes :*
(i) $\mathcal{F}_T$ est une tribu,
(ii) T est $\mathcal{F}_T$-mesurable,
(iii) si S est un autre temps d'arrêt vérifiant $S \leq T$ p.s. alors $\mathcal{F}_S \subset \mathcal{F}_T$.

La démonstration est laissée en exercice. Intuitivement, la tribu $\mathcal{F}_T$ s'interprète comme l'information dont on dispose au temps aléatoire T.

Théorème III.17 (Propriété de Markov forte)**.** *Soit (X_n) une chaine de Markov, soit $(\mathcal{F}_n)$ la filtration naturelle du processus (X_n) et soit T un temps d'arrêt. Pour toute fonction $F\colon M^{\mathbb{N}} \to \mathbb{R}$ mesurable pour la tribu produit et positive ou bornée on a en posant $u(x) = \mathbb{E}_x[F(X)]$*

$$\mathbb{E}[F \circ \theta_T(X) \mid \mathcal{F}_T] = u(X_T) \quad \text{sur l'événement} \quad \{T < +\infty\}.$$

Démonstration. Soit $A \in \mathcal{F}_T$. Par définition de la tribu $\mathcal{F}_T$ on a $A \cap \{T = n\} \in \mathcal{F}_n$ pour tout n. D'après la propriété de Markov on a aussi

$$E[F \circ \theta_n(X) \mid \mathcal{F}_n] = u(X_n).$$

pour tout entier n. En décomposant suivant les valeurs prises par T on obtient donc

$$\begin{aligned}
\mathbb{E}\left[F \circ \theta_T(X)\, \mathbb{1}_A\, \mathbb{1}_{\{T<+\infty\}}\right] &= \sum_{n=0}^{\infty} E\left[F \circ \theta_n(X)\, \mathbb{1}_{A \cap \{T=n\}}\right] \\
&= \sum_{n=0}^{\infty} E\left[u(X_n)\, \mathbb{1}_{A \cap \{T=n\}}\right] = \mathbb{E}\left[u(X_T) \mathbb{1}_A\, \mathbb{1}_{\{T<+\infty\}}\right],
\end{aligned}$$

ce qu'il fallait démontrer. □

11 Récurrence, transience

Soit (X_n) une chaine de Markov sur M de matrice de transition P

Definition III.18. On dit que x *communique* avec y s'il existe un entier n tel que $P^n(x, y) > 0$.

Remarque III.19. De manière équivalente, x communique avec y s'il existe un entier n et une suite finie $x_0, x_1, \ldots, x_n$ d'éléments de M telle que $x_0 = x$, $x_n = y$ et $P(x_i, x_i + 1) > 0$ pour tout $i \in \{0, \ldots, n\}$. En effet comme

$$P^n(x, y) = \sum_{x_1, \ldots, x_n} P(x, x_1) P(x_1, x_2) \cdots P(x_{n-1}, y)$$

et les termes de la somme sont positifs ou nuls, la somme est strictement positive si et seulement si l'un des termes au moins est strictement positif.

Lemme III.20. *La relation communique est transitive : si x communique avec y et y communique avec z alors x communique avec z.*

Démonstration. Il existe k et n tels que $P^k(x,y) > 0$ et $P^n(y,z) > 0$. Alors $P^{n+k}(x,z) \geq P^k(x,y)P^n(y,z) > 0$, ce qui montre que x communique avec z. □

Si x communique avec y et y communique avec x ou si $x = y$ alors on dit que x et y *communiquent* entre eux. On note $x \leftrightarrow y$ cette relation. Le lemme précédent montre que c'est une relation d'équivalence, les classes d'équivalence (qui forment donc une partition de M) sont appelées *classes de communication.*

Definition III.21. Si M est formé d'une seule classe de communication, autrement dit si tous les états communiquent entre eux, on dit que la chaine (X_n) est *irréductible.* On dit aussi que la matrice P est irréductible.

Exemple. La marche aléatoire sur $\mathbb{Z}$ est irréductible. En effet, soient $x \neq y \in \mathbb{Z}$, et posons $k = |x-y|$. Alors

$$P^k(x,y) = 2^{-k} > 0.$$

Donc tous les éléments de $\mathbb{Z}$ communiquent entre eux.

Exemple. Soit (X_n) la chaine de Markov sur $\{1,2,3\}$ de matrice de transition

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

Remarquons que l'état 3 vérifie $P(3,3) = 1$. Un tel état est dit *absorbant.* En effet ceci implique immédiatement $P^n(3,3) = 1$ pour tout n et donc $P^n(3,x) = 0$ pour tout $x \neq 3$. Autrement dit, une fois que la chaine atteint l'état 3 elle ne peut plus en repartir. En particulier, la chaine n'est pas irréductible. Plus précisément les classes de communication sont $\{1,2\}$ et $\{3\}$.

Pour $x \in M$ on note

$$N_x = \sum_{n=0}^{+\infty} \mathbb{1}_{\{X_n = 0\}},$$

le nombre de passages en x de la chaine (X_n). On appelle

$$T_x = \inf\{n \geq 1 : \ X_n = x\}$$

le temps d'atteinte de x (ou temps de retour en x si $X_0 = x$). Le résultat suivant est fondamental.

Théorème III.22. *Sous $\mathbb{P}_x$, autrement dit sachant $X_0 = x$, on a l'alternative suivante :*
- *Soit $N_x = +\infty$ presque surement.*
- *Soit N_x est fini presque surement. De plus, dans ce cas $\mathbb{P}_x(T_x = +\infty) > 0$ et N_x suit une loi géométrique de paramètre $\mathbb{P}_x(T_x = +\infty)$. En particulier N_x est intégrable et*

$$\mathbb{E}_x[N_x] = \frac{1}{\mathbb{P}_x(T_x = +\infty)}.$$

Definition III.23. Dans le premier cas x est dit *récurrent.* Dans le deuxième cas x est dit *transient.*

On va décomposer la démonstration du Théorème III.22 en plusieurs étapes. On commence par introduire les différents temps de retour en x. On pose $S_1 = T_x$ et pour $n \geq 1$ on définit récursivement

$$S_{n+1} = \inf\{k \geq S_n + 1 : \ X_k = x\}$$

Autrement dit S_n est le temps du n-ième retour en x. Il est facile de voir que les S_n sont des temps d'arrêt pour la filtration naturelle du processus (X_k). Rappelons que $\mathcal{F}_{S_n}$ désigne la tribu arrêtée au temps S_n (voir la Définition III.15).

Lemme III.24 (Régénération)**.** *Conditionnellement à l'événement* $\{S_n < +\infty\}$ *la variable* $S_{n+1} - S_n$ *est indépendante de* $\mathcal{F}_{S_n}$ *et sa loi coïncide avec la loi de* $S_1 = T_x$ *sous* $\mathbb{P}_x$*. En particulier si* x *est récurrent, la suite* $(S_{n+1} - S_n)_{n\geq 1}$ *est i.i.d.*

Démonstration. Pour $\omega \in M^{\mathbb{N}}$ on définit récursivement une suite (F_n) de fonctionnelles $M^{\mathbb{N}} \to \mathbb{N} \cup \{+\infty\}$ en posant

$$\begin{aligned} &F_1(\omega) = \inf\{k \geq 1 \colon \omega_k = x\} \\ &F_{n+1}(\omega) = \inf\{k \geq F_n(\omega) + 1 \colon \omega_k = x\}, \quad \forall n \geq 1, \end{aligned}$$

de sorte que $S_n = F_n(X)$ pour tout n. Notons que la deuxième équation peut se réécrire $F_{n+1} = F_n + F_1 \circ \theta_{F_n}$, du moins sur $\{F_n < +\infty\}$. En utilisant la propriété de Markov forte on a donc

$$\mathbb{P}(S_{n+1} - S_n \in A \mid \mathcal{F}_{S_n}) = \mathbb{P}(F_1 \circ \theta_{S_n}(X) \in A \mid \mathcal{F}_{S_n}) = u(X_{S_n}) \quad \text{sur} \quad \{S_n < \infty\},$$

en posant $u(y) = \mathbb{P}_y(F_1(X) \in A)$. Mais par définition de S_n, si $S_n < +\infty$ alors $X_{S_n} = x$. L'équation précédente peut donc se réécrire

$$\mathbb{P}(S_{n+1} - S_n \in A \mid \mathcal{F}_{S_n}) = \mathbb{P}_x(T_x \in A) \quad \text{sur} \quad \{S_n < \infty\},$$

ce qu'il fallait démontrer. □

Corollaire III.25. *On a pour tout entier* n

$$\mathbb{P}_x(N_x > n) = \mathbb{P}_x(T_x < +\infty)^n.$$

Démonstration. Sous $\mathbb{P}_x$ on a $X_0 = x$ et donc $N_x \geq 1$ p.s. L'égalité est donc trivialement vérifiée en $n = 0$. Soit $n \geq 1$, remarquons que puisque S_n est le temps du n-ième retour en x

$$N_x > n \quad \Leftrightarrow \quad S_n < +\infty,$$

Donc en utilisant le lemme de régénération

$$\begin{aligned} \mathbb{P}_x(N_x > n+1) &= \mathbb{P}_x(S_{n+1} < +\infty) \\ &= \mathbb{P}_x(S_n < +\infty, S_{n+1} - S_n < +\infty) \\ &= \mathbb{P}_x(S_n < +\infty)\mathbb{P}_x(T_x < +\infty) \\ &= \mathbb{P}_x(N_x > n)\mathbb{P}_x(T_x < +\infty). \end{aligned}$$

On conclut par récurrence. □

Démonstration du Théorème III.22. On deux cas possibles : si $\mathbb{P}_x(T_x < +\infty) = 1$, alors d'après le corollaire précédent $\mathbb{P}_x(N_x > n) = 1$ pour tout n. En faisant tendre n vers $+\infty$ on obtient $N_x = +\infty$ p.s. Autrement dit x est récurrent. Si au contraire $\mathbb{P}_x(T_x < +\infty) < 1$ alors $\mathbb{P}_x(T_x = +\infty) > 0$ et le corollaire précédent montre que N_x suit une loi géométrique de paramètre $\mathbb{P}_x(T_x = +\infty)$, ce qui termine la preuve. □

Corollaire III.26. *Un état* x *est récurrent si et seulement si*

$$\sum_{n=0}^{+\infty} P^n(x,x) = +\infty.$$

Démonstration. On se place sous $\mathbb{P}_x$. On a vu que soit x est récurrent et dans ce cas $N_x = +\infty$ p.s. soit x est transient et dans ce cas N_x suit une loi géométrique, en particulier N_x est intégrable. Donc x est récurrent si et seulement si N_x n'est pas intégrable. Mais d'après Fubini

$$\mathbb{E}_x[N_x] = \mathbb{E}_x\left[\sum_{n=0}^{+\infty} \mathbb{1}_{\{X_n = x\}}\right] = \sum_{n=0}^{+\infty} \mathbb{P}_x(X_n = x) = \sum_{n=0}^{+\infty} P^n(x,x).$$

D'où le résultat. □

Corollaire III.27. *Si x et y communiquent entre eux alors ils sont tous les deux du même type, à savoir tous deux récurrents ou tous deux transients.*

Démonstration. Il suffit de montrer que si x est transient alors y aussi. Si x, y communiquent, alors il existe k, l tels que $P^k(x,y) > 0$ et $P^l(y,x) > 0$. On pose $\alpha = P^k(x,y)P^l(y,x)$, on sait que α est strictement positif. Remarquons que pour tout entier n

$$P^{k+n+l}(x,x) \geq P^k(x,y)P^n(y,y)P^l(y,x) = \alpha\, P^n(y,y)$$

Donc

$$\sum_{n\geq 0} P^n(y,y) \leq \frac{1}{\alpha}\sum_{n\geq 0} P^{k+n+l}(x,x) \leq \frac{1}{\alpha}\sum_{n\geq 0} P^n(x,x).$$

Si x est transient alors le membre de droite est fini, et donc y est aussi transient. □

Remarque III.28. Dans le même ordre idée, si x est transient on a $\sum_{n\geq 0} P^n(y,x) < +\infty$ pour tout y de M. En effet en utilisant la propriété de Markov forte, on montre que si N_x est le nombre de passages en x et T_x le temps d'atteinte de x alors

$$\mathbb{E}_y[N_x] = \mathbb{E}_x[N_x]\,\mathbb{P}_y(T_x < +\infty),$$

pour tout $y \neq x$. En particulier $\mathbb{E}_y[N_x] \leq \mathbb{E}_x[N_x]$ pour tout y.

Par conséquent si la chaine est irréductible, les états sont tous transients ou tous récurrents. On dit donc que la chaine elle-même est transiente ou récurrente.

Exemple. Soit (S_n) la marche aléatoire simple symétrique sur $\mathbb{Z}$, c'est-à-dire la chaine de Markov sur $\mathbb{Z}$ donnée par

$$P(k,k+1) = P(k,k-1) = \frac{1}{2}, \quad \forall k \in \mathbb{Z}.$$

Alors (S_n) est récurrente. En effet, pour tout entier n la quantité $\mathbb{P}(S_{2n} = 0)$ est la probabilité de faire exactement n piles sur $2n$ lancers de pièce, on a donc

$$\mathbb{P}_0(S_{2n} = 0) = \binom{2n}{n} 2^{-2n}.$$

Par ailleurs en utilisant la formule de Stirling on vérifie facilement que

$$\binom{2n}{n} \sim \frac{2^{2n}}{\sqrt{\pi n}}.$$

Par conséquent $\mathbb{P}(S_{2n} = 0) \sim (\pi n)^{-1/2}$. Comme la série $\sum n^{-1/2}$ diverge on en déduit

$$\sum_{n\in\mathbb{N}} P^n(0,0) = \sum_{n\in\mathbb{N}} P^{2n}(0,0) = +\infty,$$

ce qui montre que 0 est récurrent, donc que la chaine elle-même est récurrente puisqu'elle est irréductible. Supposons maintenant que (S_n) est la marche biaisée, de matrice de transition

$$P(k,k+1) = p,\ P(k,k-1) = 1-p, \quad \forall k \in \mathbb{Z},$$

p étant un paramètre fixé entre 0 et 1, strictement. On a dans ce cas

$$\mathbb{P}_0(S_{2n} = 0) = \binom{2n}{n} p^n(1-p)^n,$$

et donc

$$\mathbb{P}_0(S_{2n} = 0) \sim \frac{(4p(1-p))^n}{\sqrt{\pi n}},$$

Si la marche possède un biais, c'est-à-dire si $p \neq 1/2$, alors $4p(1-p) < 1$, et donc la série

$$\sum_{n \in \mathbb{N}} \mathbb{P}_0(S_{2n} = 0)$$

converge. Par conséquent la marche est dans ce cas transiente. On peut retrouver ce fait en utilisant la loi des grands nombres. En effet, la loi des grands nombres montre que S_n/n converge vers $2p-1$ p.s. Si par exemple $p > 1/2$ alors $2p-1 > 0$, et on obtient $S_n \to +\infty$ p.s. En particulier, avec probabilité 1, S_n ne passe qu'un nombre fini de fois en 0.

Lemme III.29. *Une chaine de Markov irréductible sur un espace d'état fini est nécessairement récurrente.*

Démonstration. Soit $x \in M$. On a

$$\sum_{y \in M} \sum_{n \geq 0} P^n(x,y) = \sum_{n \geq 0} \mathbb{P}_x(X_n \in M) = +\infty.$$

Comme M est fini, on en déduit qu'il existe $y \in M$ tel que

$$\sum_{n \geq 0} P^n(x,y) = +\infty.$$

Ceci montre que y est récurrent, et donc que tous les états sont récurrents puisque la chaine est irréductible. $\square$

12 Marche aléatoire sur $\mathbb{Z}^d$

Soit (S_n) la marche aléatoire simple sur $\mathbb{Z}^d$. C'est-à-dire la chaine de Markov sur $\mathbb{Z}^d$ de matrice P donnée par

$$P(x,y) = \begin{cases} \frac{1}{2d} & \text{si } \|x-y\|_1 = 1 \\ 0 & \text{sinon} \end{cases}$$

pour tous $x, y \in \mathbb{Z}^d$. Le but cette section est de montrer le théorème suivant.

Théorème III.30 (Théorème de Pólya)**.** *La marche simple sur $\mathbb{Z}^d$ est récurrente en dimension 1 et 2, elle est transiente pour $d \geq 3$.*

Démonstration. Il est clair que la marche est irréductible en toute dimension. Comme on l'a vu à la section précédente il suffit d'étudier la convergence de la série $\sum P^n(0,0)$. En dimension 1 on a vu que cette série divergeait et que donc la marche était récurrente. Plaçons nous maintenant en dimension 2. De la même manière qu'en dimension 1, on a $P^n(0,0) = 0$ si n est impair. En effet, pour revenir en à son point de départ au bout de n étapes il faut avoir fait autant de pas vers la droite que vers la gauche et autant de pas vers le haut que vers le bas, ce qui implique que n est pair. Soit $k \leq n$, un chemin allant de 0 à 0 en $2n$ étapes et faisant k pas vers la droite doit aussi faire k pas vers la gauche, $n-k$ pas vers le bas et $n-k$ pas vers le haut. Il y a donc

$$\binom{2n}{k}\binom{2n-k}{k}\binom{2n-2k}{n-k} = \binom{2n}{n}\binom{n}{k}^2$$

tels chemins. La probabilité individuelle de chacun de ces chemins étant $(1/4)^{2n}$ on en déduit

$$P^{2n}(0,0) = \binom{2n}{n} \sum_{k=0}^{n} \binom{n}{k}^2 4^{-2n} = \binom{2n}{n}^2 4^{-2n} \sim \frac{1}{\pi n},$$

où on a utilisé la formule $\sum_k \binom{n}{k}^2 = \binom{2n}{n}$ (exercice) ainsi que l'équivalent $\binom{2n}{n} \sim 4^n/\sqrt{\pi n}$, lequel se déduit de la formule de Stirling. Comme la série des $1/n$ diverge on en déduit $\sum P^n(0,0) = +\infty$ ce qui montre que la marche est récurrente. En dimension 3, un argument similaire montre que $P^n(0,0)$ est nul si n est impair et que pour tout entier n

$$P^{2n}(0,0) = 6^{-2n}\binom{2n}{n} \sum_{i+j+k=n} \binom{n}{i,j,k}^2,$$

où $\binom{n}{i,j,k} = \frac{n!}{i!j!k!}$ est le nombre de façon de placer n boules numérotées dans 3 urnes numérotées en plaçant i boules dans la première urne, j dans la deuxième et k dans la troisième. Comme d'une part $\sum \binom{n}{i,j,k} = 3^n$ et d'autre part $\binom{n}{i,j,k} \leq C\,3^n/n$ pour une certaine constante C (voir ci-dessous) on a

$$\sum_{i+j+k=n} \binom{n}{i,j,k}^2 = O\left(\frac{3^{2n}}{n}\right),$$

et donc $P^{2n}(0,0) = O(n^{-3/2})$. Comme la série des $n^{-3/2}$ converge on en déduit $\sum P^n(0,0) < +\infty$ ce qui montre que cette fois la marche est transiente. Pour les dimensions supérieures on peut écrire un argument similaire ou se ramener à la dimension 3 en considérant les trois premières coordonnées de la marche. Les détails sont laissés en exercice. □

Lemme III.31. *Il existe une constante C telle que pour tous entiers n, i, j, k tels que $n = i + j + k$ on ait*

$$\binom{n}{i,j,k} \leq C\,\frac{3^n}{n}.$$

Démonstration. On étend la fonction factorielle à $\mathbb{R}_+$ en posant $x! = \int_0^\infty t^x e^{-t}\,dt$ pour tout réel positif x. En utilisant Hölder on voit que la fonction $\log x!$ est convexe. Comme $i + j + k = n$ on en déduit

$$\log(n/3)! \leq \frac{1}{3}(\log i! + \log j! + \log k!).$$

Par conséquent $\binom{n}{i,j,k} \leq \binom{n}{n/3,n/3,n/3}$. En utilisant Stirling il est facile de voir que $\binom{n}{n/3,n/3,n/3} = O(3^n/n)$. □

13 Mesures stationnaires

Soit (X_n) une chaine de Markov. Rappelons que l'espace d'état M est supposé discret, et notons P la matrice de transition.

Definition III.32. Soit μ une mesure sur M, on dit que μ est *stationnaire* (ou *invariante*) si $\mu P = \mu$, c'est-à-dire si

$$\sum_{x \in M} \mu(x) P(x,y) = \mu(y), \quad \forall y \in M.$$

Remarque III.33. En raisonnant par récurrence on obtient alors $\mu P^n = \mu$ pour tout entier n.

Si μ est une mesure de probabilité on a vu que μP^n est la loi de X_n sous $\mathbb{P}_\mu$. Donc si μ est une probabilité stationnaire alors la loi de (X_n) est constante au cours du temps sous $\mathbb{P}_\mu$. On dit que la chaine est à l'équilibre.

Remarque III.34. Il peut y avoir plusieurs probabilités stationnaires. Par exemple considérons la chaine sur $\{1, 2, 3\}$ de matrice de transition

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{3}$$

Il est facile de voir que δ_3 est stationnaire, et que $(\delta_1 + \delta_2)/2$ est aussi stationnaire.

Évidemment la mesure nulle est toujours stationnaire, de même que la mesure μ vérifiant $\mu(x) = +\infty$ pour tout x. S'il y a plusieurs classes de communication, il est aussi facile de voir qu'une mesure μ vérifiant $\mu(x) = +\infty$ sur certaines classes de communication et $\mu(x) = 0$ sur les autres est stationnaire. De telles mesures stationnaires ne sont pas intéressantes. Dans la suite on dira qu'une mesure stationnaire est non triviale s'il existe $x \in M$ tel que $0 < \mu(x) < +\infty$.

Lemme III.35. *Si (X_n) est une chaine irréductible et si μ est une mesure stationnaire non triviale alors $0 < \mu(y) < +\infty$ pour tout $y \in M$.*

Démonstration. Comme μ est supposée non triviale, il existe $x \in M$ tel que $0 < \mu(x) < +\infty$. Soit $y \in M$, comme P est irréductible, il existe un entier k tel que $P^k(x,y) > 0$ et un entier l tel que $P^l(y,x) > 0$. En utilisant la stationnarité on obtient

$$\mu(y) = \mu P^k(y) = \sum_{z \in M} \mu(z) P^k(z,y) \geq \mu(x) P^k(x,y),$$

ce qui montre que $\mu(y) > 0$. De même

$$\mu(x) = \mu P^l(x) \geq \mu(y) P^l(y,x),$$

ce qui montre que $\mu(y) < +\infty$. □

Théorème III.36. *Si l'espace d'état M est fini, alors il existe une probabilité invariante. Si de plus la chaine est irréductible, la probabilité invariante est unique.*

L'existence se déduit du lemme d'analyse suivant.

Lemme III.37. *Soit $A\colon \mathbb{R}^n \to \mathbb{R}^n$ une application affine et soit K un sous-ensemble convexe compact de $\mathbb{R}^n$ vérifiant $AK \subset K$. Alors A admet un point fixe dans K, c'est-à-dire qu'il existe $x \in K$ tel que $Ax = x$.*

Démonstration. On fixe $x_1 \in K$ et on pose pour $n \geq 2$

$$x_n = \frac{1}{n} \sum_{k=0}^{n-1} A^k x_1,$$

(où A^k désigne l'application A composée k fois). Comme $AK \subset K$ on a $A^k x_1 \in K$ pour tout k et donc $x_n \in K$ pour tout n par convexité de K. Comme K est compact on peut extraire une sous-suite convergente de (x_n). Il existe une suite croissante (n_k) d'entiers telle que (x_{n_k}) converge vers un élément x de K. Alors comme A est continue

$$Ax_{n_k} \longrightarrow Ax.$$

Par ailleurs, comme A est affine

$$Ax_n = \frac{1}{n} \sum_{k=0}^{n-1} A^{k+1}(x_1) = x_n + \frac{1}{n}(A^n x_1 - x_1).$$

Et comme la suite $(A^n x_1 - x_1)$ est bornée on en déduit que

$$Ax_n - x_n \longrightarrow 0.$$

Par conséquent $Ax_{n_k} - x_{n_k} \to 0$, ce qui montre que $Ax = x$. □

Démonstration du Théorème III.36. On peut supposer que l'espace d'état est $M = \{1, \ldots, n\}$. On identifie une mesure de probabilité μ sur M avec le vecteur ligne

$$(\mu(1), \ldots, \mu(n))$$

L'ensemble des mesures de probabilité sur M est donc décrit par le simplexe

$$\Delta_{n-1} = \left\{ x \in \mathbb{R}^n;\ x_i \geq 0,\ \forall i \quad \text{et} \quad \sum_{i=1}^n x_i = 1 \right\}$$

qui est un sous-ensemble convexe compact de $\mathbb{R}^n$. L'application

$$\mu \mapsto \mu P$$

est linéaire et préserve Δ_n. D'après le lemme précédent elle admet un point fixe dans Δ_n ce qui prouve l'existence d'une probabilité invariante.

On suppose maintenant que la chaine est irréductible. Soit π, μ deux probabilités stationnaires, on veut montrer que $\pi = \mu$. D'après le Lemme III.35 on a $\pi(j) > 0$ pour tout j. Soit $i \leq n$ un indice tel que le rapport $\mu(i)/\pi(i)$ soit minimal. On a donc

$$\mu(j) \geq \frac{\mu(i)}{\pi(i)}\pi(j), \quad \forall j \leq n.$$

En utilisant la stationnarité de π et μ on obtient pour tout entier k

$$\begin{aligned}\mu(i) = \mu P^k(i) &= \sum_{j \leq n} \mu(j) P^k(j,i) \\ &\geq \sum_{j \leq n} \frac{\mu(i)}{\pi(i)}\pi(j) P^k(j,i) = \frac{\mu(i)}{\pi(i)}\pi P^k(i) = \mu(i).\end{aligned}$$

L'inégalité du milieu est donc une égalité, ce qui implique

$$\mu(j) P^k(j,i) = \frac{\mu(i)}{\pi(i)}\pi(j) P^k(j,i)$$

pour tout $j \leq n$. De plus comme la chaine est irréductible, étant donné $j \leq n$ on peut trouver un entier k tel que $P^k(j,i) > 0$, auquel cas l'égalité précédente devient

$$\mu(j) = \frac{\mu(i)}{\pi(i)}\pi(j).$$

Ceci étant valable pour tout $j \leq n$, la mesure μ est donc un multiple de la mesure π. Enfin comme μ et π sont des mesures de probabilité on obtient bien $\mu = \pi$. □

Exemple (Urne d'Ehrenfest)**.** Dans le modèle de l'urne d'Ehrenfest, le nombre X_n de particules dans la première enceinte est une chaine de Markov sur $\{0, \ldots, N\}$ de matrice de transition P donnée par $P(0,1) = 1$, $P(N, N-1) = 1$ et

$$P(x, x+1) = 1 - \frac{x}{N}, \quad P(x, x-1) = \frac{x}{N}$$

pour tout $x \in \{1, \ldots, N-1\}$. Il est facile de voir que cette chaine est irréductible. Comme l'espace d'état est fini, le théorème précédent montre qu'il existe une unique probabilité stationnaire. Vérifions le théorème dans ce cas. L'équation de stationnarité devient

$$\left\{\begin{array}{rll} \mu(1)\frac{1}{N} & = \mu(0) & \\ \mu(x-1)\frac{N-x+1}{N} + \mu(x+1)\frac{x+1}{N} & = \mu(x), & \quad x = 1, \ldots, N-1 \\ \mu(N-1)\frac{1}{N} & = \mu(N). & \end{array}\right.$$

On obtient alors $\mu(1) = N\mu(0)$ puis $\mu(2) = \frac{N(N-1)}{2}\mu(0)$ et on montre facilement par récurrence que

$$\mu(x) = \binom{N}{x}\mu(0), \quad \forall x \in \{0, \ldots, N\}.$$

Les mesures stationnaires sont donc toutes de la forme $\mu(x) = \binom{N}{x}c$ pour une certaine constante c. En particulier, l'unique probabilité stationnaire est donnée par

$$\pi(x) = \binom{N}{x}2^{-N}, \quad x \in \{0, \ldots, N\}.$$

Lorsque l'espace d'état est infini, il n'existe pas forcément de probabilité stationnaire.

Exemple. Considérons la marche aléatoire simple symétrique sur $\mathbb{Z}$. Soit π une mesure stationnaire, on a alors

$$\pi(x) = \frac{1}{2}\pi(x-1) + \frac{1}{2}\pi(x+1)$$

pour tout $x \in \mathbb{Z}$. Posons $u(x) = \pi(x) - \pi(x-1)$. L'équation précédente devient $u(x+1) = u(x)$ pour tout x. Donc $u(x)$ est constante, donc il existe deux constantes a, b telles que $\pi(x) = ax + b$. Comme par ailleurs π est une mesure on doit avoir $\pi(x) \geq 0$ pour tout $x \in \mathbb{Z}$ donc on a nécessairement $a = 0$ et $b \geq 0$. Autrement dit π est un multiple positif de la mesure de comptage sur $\mathbb{Z}$. Donc, à un facteur multiplicatif près, la mesure de comptage sur $\mathbb{Z}$ est l'unique mesure invariante pour la marche aléatoire. Ceci montre en particulier que la marche aléatoire n'admet pas de probabilité invariante.

Si l'espace d'état est infini, le Théorème III.36 reste essentiellement vrai à condition d'ajouter une hypothèse de récurrence. Plus précisément, on a le résultat suivant, qu'on démontrera dans la prochaine section.

Théorème III.38. *Soit (X_n) une chaine de Markov admettant au moins un état récurrent. Alors il existe une mesure stationnaire non triviale. Si de plus la chaine est irréductible, alors les mesures stationnaires sont toutes multiples les unes des autres.*

Exemple. Soit $p \in]0,1[$ et soit (S_n) la marche aléatoire biaisée, de matrice de transition P donnée par

$$P(x, x+1) = p, \ P(x, x-1) = 1 - p, \quad \forall x \in \mathbb{Z}.$$

Remarquons d'abord que cette chaine de Markov est toujours irréductible, quelle que soit la valeur de $p \in]0,1[$. Si $p = 1/2$ on a vu précédemment que la chaine était récurrente. Le théorème précédent s'applique donc, et comme on l'a vu plus haut les mesures invariantes sont en effet toutes multiples de la mesure de comptage sur $\mathbb{Z}$. Si $p \neq 1/2$, la mesure de comptage sur $\mathbb{Z}$ est toujours invariante, mais la chaine admet maintenant une deuxième mesure invariante linéairement indépendante de la première : la mesure μ donnée par

$$\mu(x) = \left(\frac{p}{1-p} \right)^x, \quad x \in \mathbb{Z}.$$

Ceci montre donc que le théorème est faux sans l'hypothèse de récurrence. Au passage on obtient ainsi une troisième démonstration du fait que la marche biaisée est transiente.

14 Excursions

Soit (X_n) une chaine de Markov, soit M son espace d'état et soit P sa matrice de transition. On note $(\mathcal{F}_n)$ la filtration naturelle du processus (X_n) : pour tout entier n

$$\mathcal{F}_n = \sigma(X_0, \ldots, X_n).$$

Le résultat principal de cette section est le suivant.

Théorème III.39. *Soit ν une mesure de probabilité sur M et soit T une variable aléatoire vérifiant les deux conditions suivantes :*
(i) T temps d'arrêt pour la filtration $(\mathcal{F}_n)$;
(ii) T est fini p.s. et loi de X_T sous $\mathbb{P}_\nu$ est égale à ν. De manière plus explicite, on a

$$\mathbb{P}_\nu(X_T = x) = \nu(x), \quad \forall x \in M.$$

Alors la mesure μ définie par

$$\mu(x) = \mathbb{E}_\nu \left[\sum_{n=0}^{T-1} \mathbb{1}_{\{X_n = x\}} \right], \quad \forall x \in M,$$

est stationnaire.

Démonstration. Soit $f \colon M \to \mathbb{R}$ une fonction positive. Comme $X_0 = X_T$ en loi sous $\mathbb{P}_\nu$ on a

$$\int_M f \, d\mu = \mathbb{E}_\nu \left[\sum_{n=0}^{T-1} f(X_n) \right] = \mathbb{E}_\nu \left[\sum_{n=1}^{T} f(X_n) \right]$$

De plus T étant un temps d'arrêt, on a $\{n \leq T\} = \{T \leq n-1\}^c \in \mathcal{F}_{n-1}$. Donc en utilisant la propriété de Markov

$$\begin{aligned}
\int_M f\, d\mu = \mathbb{E}_\nu \left[\sum_{n=1}^{T} f(X_n)\right] &= \sum_{n=1}^{\infty} \mathbb{E}_\nu \left[f(X_n)\mathbb{1}_{\{n\leq T\}}\right] \\
&= \sum_{n=1}^{\infty} \mathbb{E}_\nu \left[\mathbb{E}_\nu \left[f(X_n) \mid \mathcal{F}_{n-1}\right] \mathbb{1}_{\{n\leq T\}}\right] \\
&= \sum_{n=1}^{\infty} \mathbb{E}_\nu \left[Pf(X_{n-1})\mathbb{1}_{\{n\leq T\}}\right] \\
&= \mathbb{E}_\nu \left[\sum_{n=0}^{T-1} Pf(X_n)\right] = \int_M Pf\, d\mu.
\end{aligned}$$

On a donc $\int_M f\, d\mu = \int_M Pf\, d\mu$ pour toute $f \geq 0$, ce qui revient à dire que μ est stationnaire (exercice). □

Lorsque la chaine admet un état récurrent x, on peut appliquer le théorème à la mesure $\nu = \delta_x$ et au temps

$$T = T_x = \inf\{n \geq 1 \colon X_n = x\}.$$

En effet le temps de retour en x est bien un temps d'arrêt, il est fini presque surement puisque x est supposé récurrent, et on a bien sûr $X_{T_x} = x$ p.s. par définition de T_x. La condition (ii) est donc vérifiée. On obtient donc le corollaire suivant.

Corollaire III.40. *Soit (X_n) une chaine de Markov admettant un état récurrent x. Alors la mesure μ_x définie par*

$$\mu_x(y) = \mathbb{E}_x \left[\sum_{k=0}^{T_x-1} \mathbb{1}_{\{X_n=y\}}\right], \quad \forall y \in M.$$

est stationnaire.

Remarque III.41. Par définition de T_x on a $\mu_x(x) = 1$, ce qui montre que la mesure μ_x est non triviale. Le résultat précédent donne donc la première partie du Théorème III.38 : s'il existe un état récurrent, alors il existe une mesure stationnaire non triviale.

Montrons maintenant l'unicité de la mesure stationnaire dans le cas irréductible récurrent.

Proposition III.42. *Soit (X_n) une chaine de Markov irréductible et récurrente, les mesures stationnaires sont toutes multiples les unes des autres.*

Démonstration. Soit x un élément arbitraire de M. On va montrer que les mesures invariantes sont toutes proportionnelles à la mesure μ_x définie plus haut. Soit donc ν une mesure stationnaire, supposons qu'elle est normalisée de sorte que $\nu(x) = 1$ et montrons que $\nu = \mu_x$. Soit $y \neq x$, comme ν est stationnaire et $\nu(x) = 1$ on a

$$\nu(y) = \nu P(y) = P(x,y) + \sum_{y_1 \neq x} \nu(y_1)P(y_1,y)$$

On ré-applique cette égalité aux y_1 à l'intérieur de la somme et ainsi de suite de manière inductive

$$\nu(y) = P(x,y) + \sum_{y_1\neq x} P(x,y_1)P(y_1,y) + \sum_{y_1\neq x, y_2\neq x} \nu(y_2)P(y_2,y_1)P(y_1,y) = \cdots$$

Ceci montre en particulier que

$$\begin{aligned}
\nu(y) &\geq P(x,y) + \sum_{y_1\neq x} P(x,y_1)P(y_1,y) + \sum_{y_1\neq x, y_2\neq x} P(x,y_1)P(y_1,y_2)P(y_2,y) + \cdots \\
&= \mathbb{P}_x(X_1 = y) + \mathbb{P}_x(X_1 \neq x, X_2 = y) + \mathbb{P}_x(X_1 \neq x, X_2 \neq x, X_3 = y) + \cdots \\
&= \sum_{n=1}^{\infty} \mathbb{P}_x(X_n = y; n < T_x) = \mathbb{E}_x \left[\sum_{n=1}^{T_x-1} \mathbb{1}_{\{X_n=y\}}\right] = \mu_x(y).
\end{aligned}$$

Les mesures ν et μ_x sont toutes les deux stationnaires et vérifient donc

$$\begin{aligned} \nu(x) &= \mu_x(x) = 1, \\ \nu(y) &\geq \mu_x(y), \quad \forall y \neq x \end{aligned}$$

Autrement dit le rapport $\nu(y)/\mu_x(y)$ atteint son maximum 1 en $y = x$. En raisonnant de la même manière que dans la preuve du Théorème III.36 on en déduit facilement que $\nu(y)/\mu_x(y) = 1$ pour tout y, c'est-à-dire que $\nu = \mu_x$. □

On se pose maintenant la question de savoir s'il existe une mesure de probabilité stationnaire. Remarquons que la masse totale de la mesure μ_x définie plus haut est

$$\mu_x(M) = \mathbb{E}_x\left[\sum_{k=0}^{T_x-1} \mathbb{1}_{\{X_n\in M\}}\right] = \mathbb{E}_x[T_x].$$

Definition III.43. Soit x un état récurrent. Si $\mathbb{E}_x[T_x] < +\infty$ on dit que x est *récurrent positif.* Si au contraire $\mathbb{E}_x[T_x] = +\infty$ on dit que x est *récurrent nul.*

Proposition III.44. *Soit (X_n) une chaine irréductible. Les propositions suivantes sont équivalentes.*
(i) Il existe un état récurrent positif.
(ii) Tous les états sont récurrents positifs.
(iii) La chaine admet une probabilité stationnaire.
De plus la probabilité stationnaire π est alors unique et donnée par la formule

$$\pi(x) = \frac{1}{\mathbb{E}_x[T_x]}, \quad \forall x \in M. \tag{4}$$

Démonstration. Montrons d'abord $(i) \Rightarrow (iii)$. Soit $x \in M$ récurrent positif. D'après le Corollaire III.40 la mesure μ_x est stationnaire, et on a

$$\mu_x(M) = \mathbb{E}_x[T_x] < +\infty.$$

Donc la mesure μ_x est finie. Il suffit donc de la normaliser pour obtenir une probabilité stationnaire.
Montrons $(iii) \Rightarrow (ii)$. Supposons qu'il existe une probabilité stationnaire π et soit $x \in M$. Commençons par montrer que x est récurrent. Soit $N_x = \sum_{n\geq 0} \mathbb{1}_{\{X_n=x\}}$ le nombre de passages en x. Comme π est stationnaire et la chaine irréductible on a $\pi(x) > 0$ d'après le Lemme III.35 et donc

$$\mathbb{E}_\pi[N_x] = \sum_{n\geq 0} \pi P^n(x) = \sum_{n\geq 0} \pi(x) = +\infty$$

ce qui montre que x est récurrent (remarque III.28). Donc la mesure μ_x donnée par l'excursion partant de x est stationnaire. Comme la chaine est supposée irréductible toutes les mesures stationnaires sont multiples les unes des autres. En particulier μ_x est proportionnelle à π. Comme $\mu_x(x) = 1$ et $\mu_x(M) = \mathbb{E}_x[T_x]$, on obtient

$$\mathbb{E}_x[T_x] = \frac{\mu_x(M)}{\pi(M)} = \frac{\mu_x(x)}{\pi(x)} = \frac{1}{\pi(x)} < +\infty.$$

Ceci montre que x est récurrent positif, ainsi que l'égalité (4). Ceci termine la preuve puisque l'implication $(ii) \Rightarrow (i)$ est triviale. □

Lorsque cette situation se produit on dit que la chaine elle-même est récurrente positive.

Exemple. La marche aléatoire simple sur $\mathbb{Z}$ est récurrente nulle.

Une conséquence immédiate du résultat précédent est qu'une chaine irréductible sur un espace d'état fini est nécessairement récurrente positive. En effet on sait dans ce cas qu'elle possède une probabilité invariante. De plus l'égalité (4) permet de calculer explicitement les temps de retour moyens.

Exemple. Soit $G = (V, E)$ un graphe fini et soit (X_n) la marche aléatoire sur G. La chaine (X_n) est irréductible si et seulement si pour tout $x, y \in V$, il existe un chemin $x_0, \dots, x_n$ tels que $x_0 = x$, $x_n = y$ et $x_i \sim x_{i+1}$ pour tout i. On dit que le graphe est *connexe*. Supposons que cette hypothèse est vérifiée. Pour tout sommet x on note $d(x)$ le degré de x. Il est facile de voir que la mesure

$$\mu(x) = d(x), \quad \forall x \in V$$

est stationnaire. En effet pour tout $x \in V$ on a

$$d(x) = \sum_{y \in V} \mathbb{1}_{x \sim y} = \sum_{y \in V} d(y) P(y, x).$$

Par conséquent la mesure π donnée par

$$\pi(x) = \frac{\mu(x)}{\mu(V)} = \frac{d(x)}{2|E|},$$

pour tout $x \in V$ est l'unique probabilité stationnaire de la chaine. Pour tout sommet x le temps moyen de retour en x est donc donné par la formule

$$\mathbb{E}_x[T_x] = \frac{2|E|}{d(x)}.$$

Exemple (Urne d'Ehrenfest)**.** Le modèle de l'urne d'Erhenfest est irréductible, donc récurrent puisque l'espace d'état est fini. En particulier la chaine revient presque surement en N. On peut alors s'interroger sur la crédibilité de ce modèle, puisqu'on imagine mal dans la réalité la totalité des molécules de gaz se concentrer dans la première enceinte. Pour sauver le modèle, calculons l'espérance du temps de retour en N. On a vu que l'unique probabilité stationnaire est donnée par

$$\nu(x) = 2^{-N} \binom{N}{x}, \quad x \in \{0, \dots, N\}.$$

D'après le théorème précédent le temps moyen de retour en N est donc

$$\mathbb{E}_N[T_N] = \frac{1}{\nu(N)} = 2^N.$$

Si N est de l'ordre du nombre d'Avogadro, disons 10^{23}, alors 2^N est un nombre monstrueux. Même si l'unité de temps est minuscule (disons 10^{-23} seconde) le temps moyen de retour en N dépasse de plusieurs ordres de grandeur l'âge de l'univers. Notons par ailleurs que le temps de retour en $N/2$ est lui beaucoup plus raisonnable puisque

$$\mathbb{E}_{N/2}[T_{N/2}] = \frac{1}{\nu(N/2)} = \frac{2^N}{\binom{N}{N/2}} \sim \sqrt{\pi N/2}.$$

15 Mesures réversibles

Soit (X_n) une chaine de Markov, d'espace d'états M et matrice de transition P.

Definition III.45. Soit μ une mesure sur M, on dit que μ est réversible si

$$\mu(x)P(x, y) = \mu(y)P(y, x), \quad \forall x, y \in M.$$

Exemple. Si P est une matrice symétrique, c'est-à-dire si $P(x, y) = P(y, x)$ pour tous x, y, alors la mesure de comptage sur M est réversible.

Lemme III.46. *Si μ est réversible alors μ est stationnaire.*

Démonstration. En utilisant la réversibilité, on obtient pour tout $y \in M$

$$\sum_{x \in M} \mu(x) P(x, y) = \sum_{x \in M} \mu(y) P(y, x) = \mu(y),$$

ce qui montre que μ est stationnaire. □

Exemple. Soit $p \in]0,1[$ et soit (S_n) la marche aléatoire biaisée sur $\mathbb{Z}$, de matrice de transition P donnée par

$$P(x,x+1) = p,\ P(x,x-1) = 1-p, \quad \forall x \in \mathbb{Z}.$$

Si $p = 1/2$ la mesure de comptage sur $\mathbb{Z}$ est réversible, donc stationnaire. Si $p \neq 1/2$ la mesure de comptage est toujours stationnaire, mais elle n'est plus réversible.

De même que pour la stationnarité, lorsque μ est une mesure de probabilité, la réversibilité s'interprète de manière probabiliste.

Lemme III.47. *Soit π une mesure de probabilité. Alors π est réversible si et seulement si sous $\mathbb{P}_\pi$ on a pour tout entier n*

$$(X_0, X_1, \ldots, X_n) = (X_n, X_{n-1}, \ldots, X_0), \quad \text{en loi.}$$

Démonstration. En utilisant l'équation de Chapman-Kolmogorov et l'égalité $\pi(x)P(x,y) = \pi(y)P(y,x)$ de manière répétée on obtient

$$\begin{aligned}
\mathbb{P}_\pi(X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n) &= \pi(x_0)P(x_0,x_1)P(x_1,x_2)\cdots P(x_{n-1},x_n) \\
&= \pi(x_1)P(x_1,x_0)P(x_1,x_2)\cdots P(x_{n-1},x_n) \\
&= \cdots \\
&= \pi(x_n)P(x_1,x_0)P(x_2,x_1)\cdots P(x_n,x_{n-1}) \\
&= \mathbb{P}_\pi(X_0 = x_n, X_1 = x_{n-1}, \ldots, X_n = x_0),
\end{aligned}$$

ce qui montre que $(X_0, \ldots, X_n) = (X_n, \ldots, X_0)$ en loi. La réciproque est laissée en exercice. □

Autrement dit, une mesure de probabilité π est réversible si partant de π la loi de la chaine est invariante par retournement du temps.

Exemple (L'urne d'Erhenfest)**.** Le modèle de l'urne d'Erhenfest à l'équilibre devrait être invariant par retournement du temps. Vérifions cette intuition par le calcul. L'équation $\pi(x)P(x,y) = \pi(y)P(y,x)$ se ramène dans ce cas à

$$\pi(x-1)\left(1 - \frac{x-1}{N}\right) = \pi(x)\frac{x}{N}, \quad x = 1, \ldots, N.$$

Ceci revient à $\pi(x) = \frac{N-x+1}{x}\pi(x-1)$ pour tout x, ce qui se résout en

$$\pi(x) = \frac{(N-x+1)\cdots(N-1)N}{x(x-1)\cdots 1}\pi(0) = \binom{N}{x}\pi(0).$$

Autrement dit la loi binomiale de paramètre 1/2 est réversible. Comme on le voit sur cet exemple, un des intérêt des mesures réversibles est qu'elles sont en général plus faciles à déterminer que les mesures stationnaires.

La proposition suivante fournit un critère d'existence de mesure réversible.

Proposition III.48 (Condition de cycle de Kolmogorov)**.** *Soit (X_n) une chaine irréductible. Les propriétés suivantes sont équivalentes :*

(i) (X_n) admet une mesure réversible non triviale ;

(ii) Pour tout entier n et pour toute suite $x_0, \ldots, x_n$ d'éléments de M vérifiant $x_0 = x_n$ on a

$$\prod_{i=1}^{n-1} P(x_i, x_{i+1}) = \prod_{i=1}^{n-1} P(x_{i+1}, x_i).$$

Autrement dit, il existe une mesure réversible si pour tout cycle $x_1, \ldots, x_n = x_1$ la probabilité de parcourir le cycle dans un sens ou dans l'autre est la même.

Démonstration. Soit μ une mesure réversible. En raisonnant comme dans le lemme précédent on obtient

$$\mu(x_0) \prod_{i=1}^{n-1} P(x_i, x_{i+1}) = \mu(x_n) \prod_{i=1}^{n-1} P(x_{i+1}, x_i),$$

pour toute suite $x_0, \ldots, x_n$. De plus comme la chaine est irréductible la mesure μ charge tous les points, donc $\mu(x_0) > 0$. Par conséquent si $x_0 = x_n$ on obtient bien l'égalité (ii). Réciproquement supposons que (ii) est vérifiée. Remarquons qu'on a alors

$$P(x, y) > 0 \ \Rightarrow \ P(y, x) > 0, \quad \forall x, y \in M.$$

En effet soient $x, y \in M$. Si $x = y$ l'implication est trivialement vérifiée. Si $x \neq y$, comme la chaine est irréductible il existe une suite $x_1, \ldots, x_n$ telle que $x_1 = y$, $x_n = x$ et $P(x_i, x_{i+1}) > 0$ pour tout i. On pose alors $x_0 = x$ et on utilise la condition (ii) pour le cycle $x_0, x_1, \ldots, x_n$:

$$\prod_{i=0}^{n-1} P(x_i, x_{i+1}) = \prod_{i=0}^{n-1} P(x_{i+1}, x_i).$$

Si $P(x, y) > 0$ le membre de gauche est alors strictement positif par construction, donc le membre de droite aussi. Ceci implique en particulier

$$P(y, x) = P(x_1, x_0) > 0,$$

ce qu'il fallait démontrer.
Montrons maintenant qu'il existe une mesure réversible. On fixe $x_0 \in M$ arbitrairement et on pose $\mu(x_0) = 1$. Soit $x \neq x_0$. Comme la chaine est irréductible il existe une suite $x_1, \ldots, x_n$ telle que $x_n = x$ et $P(x_i, x_{i+1}) > 0$ pour tout i. Alors d'après ce qui précède $P(x_{i+1}, x_i) > 0$ pour tout i et on peut poser

$$\mu(x) = \prod_{i=0}^{n-1} \frac{P(x_i, x_{i+1})}{P(x_{i+1}, x_i)}.$$

Montrons que ceci ne dépend pas de la suite (x_i) choisie. Soit $y_1, \ldots, y_m$, une suite vérifiant $y_m = x$ et $P(y_i, y_i + 1) > 0$ pour tout i (en posant $y_0 = x_0$), il s'agit de montrer que

$$\prod_{i=0}^{m-1} \frac{P(y_i, y_{i+1})}{P(y_{i+1}, y_i)} = \prod_{i=0}^{n-1} \frac{P(x_i, x_{i+1})}{P(x_{i+1}, x_i)}.$$

Mais en réarrangeant les facteurs on voit tout de suite que cette égalité se ramène à la condition de cycle (ii) pour la suite

$$(y_0, y_1, \ldots, y_{m-1}, x, x_{n-1}, \ldots, x_1, x_0).$$

La mesure μ est donc bien définie. Il est ensuite très facile de vérifier qu'elle est réversible. □

16 Le théorème ergodique

On considère une chaine irréductible récurrente positive (X_n). On a vu qu'elle admet une unique probabilité stationnaire π. On considère la mesure empirique

$$L_n = \frac{1}{n} \sum_{k=1}^{n} \delta_{X_k}.$$

C'est une mesure de probabilité aléatoire sur M. On a pour tout $y \in M$

$$L_n(y) = \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{\{X_k = y\}}.$$

Autrement dit $L_n(y)$ est la fréquence au temps n des passages en y de la chaine (X_n).

Théorème III.49. *Avec probabilité* 1*, la mesure aléatoire* L_n *converge en loi vers la probabilité stationnaire* π*. De manière plus explicite, pour toute fonction bornée* f *sur* M *on a*

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) \longrightarrow \int_M f\, d\pi, \quad \textit{p.s.}$$

Remarque III.50. Le théorème est valable quelle que soit la loi de X_0.

Démonstration. Dans le cas particulier où $f = \mathbb{1}_{\{y\}}$ le théorème devient

$$\frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{\{X_k=y\}} \longrightarrow \pi(y), \quad \text{p.s.}$$

On va se contenter de faire la démonstration dans ce cas. On fixe y, on note $S_1 = T_y = \inf\{n \geq 1 \colon X_n = y\}$ le temps de retour en y et pour $k \geq 2$ on note S_k le temps du k-ième retour en y. Comme la chaine est récurrente on sait qu'avec probabilité 1, elle passe une infinité de fois en y et donc que les S_k sont finis. Pour $k \geq 1$ on note $\tau_k = S_{k+1} - S_k$ le temps qui s'écoule entre les k-ième et $k+1$-ième retours en y. Le lemme de régénération III.24 assure que les τ_k sont i.i.d. et que leur loi commune est donnée par

$$\mathbb{P}(\tau_k = i) = \mathbb{P}_y(T_y = i), \quad \forall i \geq 1.$$

En particulier, comme la chaine est récurrente positive

$$\mathbb{E}[\tau_k] = \mathbb{E}_y[T_y] = \frac{1}{\pi(y)}.$$

D'après la loi des grands nombres

$$\frac{1}{k}\sum_{i=1}^{n} \tau_i \longrightarrow \frac{1}{\pi(y)}, \quad \text{p.s.}$$

Comme S_1 est fini p.s. et comme $S_k = S_1 + \sum_{i=1}^{k-1} \tau_i$, on en déduit que

$$\frac{S_k}{k} \longrightarrow \frac{1}{\pi(y)}, \quad \text{p.s.}$$

Ce qu'on peut réécrire $S_k \sim k/\pi(y)$. Comme $k \sim k+1$ ceci implique en particulier $S_k \sim S_{k+1}$. Soit

$$N_n = nL_n(y) = \sum_{j=1}^{n} \mathbb{1}_{\{X_j=y\}}$$

le nombre de passages en y au temps n. Encore une fois comme la chaine est récurrente $N_n \to +\infty$ p.s. Donc par composition des limites $T_{N_n} \sim N_n/\pi(y)$ et $T_{N_n} \sim T_{N_n+1}$. Or par définition, si $N_n = k$, alors on a fait exactement k passages en y au temps n, et donc $S_k \leq n < S_{k+1}$. Autrement dit

$$S_{N_n} \leq n < S_{N_n+1}.$$

Comme $T_{N_n+1} \sim T_{N_n}$ on en déduit $T_{N_n} \sim n$ et donc $N_n \sim n\pi(y)$, le tout avec probabilité 1, ce qu'il fallait démontrer. $\square$

17 Convergence à l'équilibre

Le théorème ergodique affirme que la suite de variables aléatoires

$$\frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{\{X_k=y\}}$$

converge presque surement vers la constante $\pi(y)$. Comme cette suite est de plus bornée (comprise entre 0 et 1) on a aussi convergence des espérances, d'après le théorème de convergence dominée. On obtient le corollaire suivant.

Corollaire III.51. *Soit (X_n) une chaine de Markov récurrente positive et soit π sa probabilité invariante, alors pour tout $y \in M$*

$$\frac{1}{n}\sum_{k=1}^{n}\mathbb{P}(X_k = y) \longrightarrow \pi(y).$$

Remarque III.52. Encore une fois on ne précise pas en indice la loi de X_0 parce que le résultat est valable pour toute loi initiale.

On peut se poser la question de savoir si la moyenne de Cesàro est nécessaire dans ce résultat : a-t-on $\mathbb{P}(X_n = y) \to \pi(y)$? Autrement dit est-ce que (X_n) converge vers π en loi ? La réponse est non en général, comme le montre l'exemple très simple suivant.

Exemple. On considère la chaine (X_n) sur $\{1, 2\}$ de matrice de transition

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

Autrement dit si la chaine est en 1 au temps n alors elle est en 2 au temps $n + 1$ et réciproquement. Il est clair que cette chaine est irréductible et récurrente positive de probabilité invariante $\pi = (\delta_1 + \delta_2)/2$. Supposons par exemple que $X_0 = 1$ on aura alors $X_n = 1$ pour tout n pair et $X_n = 2$ pour tout n impair, et ce presque surement. En particulier $\mathbb{P}_1(X_n = 1)$ vaut 0 ou 1 selon que n soit pair ou impair, ce qui montre que $\mathbb{P}_1(X_n = 1)$ ne converge pas vers $\pi(1) = 1/2$.

L'objectif de cette section est de donner une condition sous laquelle $\mu P^n \to \pi$ pour toute mesure de probabilité μ sur M. On commence par introduire une distance entre mesures de probabilité appelée distance en variation totale.

Definition III.53. Soit μ et ν deux mesures de probabilités sur un espace muni d'une tribu $(M, \mathcal{A})$. On appelle distance en variation totale entre μ et ν, notée $\mathrm{TV}(\mu, \nu)$, la quantité

$$\mathrm{TV}(\mu, \nu) = \sup_{A \in \mathcal{A}} \{\mu(A) - \nu(A)\}.$$

Remarque III.54. Remarquons aussi que comme la tribu $\mathcal{A}$ est stable par passage au complémentaire on a en fait

$$\mathrm{TV}(\mu, \nu) = \sup_{A \in \mathcal{A}} \{|\mu(A) - \nu(A)|\}.$$

Par définition $\mathrm{TV}(\mu, \nu) \in [0, 1]$ et $\mathrm{TV}(\mu, \nu) = 0$ si et seulement si $\mu = \nu$.Il est aussi facile de voir que la variation totale vérifie l'inégalité triangulaire :

$$\mathrm{TV}(\mu, \nu) \leq \mathrm{TV}(\mu, \pi) + \mathrm{TV}(\pi, \nu)$$

pour toutes mesures de probabilité μ, ν, π. On en déduit donc que la variation totale défini une distance sur l'espace des mesures de probabilités.

Definition III.55. Pour $x \in M$, on appelle *période* de x le plus grand commun diviseur de l'ensemble $\{n \in \mathbb{N};\ P^n(x, x) > 0\}$.

Lemme III.56. *Si x et y communiquent entre eux alors ils ont la même période.*

Démonstration. Il existe k, l tels que $P^k(x, y) > 0$ et $P^l(y, x) > 0$. Soit n vérifiant $P^n(y, y) > 0$. On a

$$P^{k+l}(x, x) \geq P^k(x, y)P^l(y, x) > 0$$
$$P^{k+l+n}(x, x) \geq P^k(x, y)P^n(y, y)P^l(y, x) > 0.$$

Alors par définition, la période p de x divise à la fois $k+l$ et $k+l+n$, donc p divise n. Comme ceci est valable pour tous les entiers n vérifiant $P^n(y, y) > 0$, on en déduit que p divise leur plus grand commun diviseur, à savoir la période q de y. En échangeant les rôles de x et y on obtient q divise p, et donc $p = q$. □

Par conséquent si (X_n) est irréductible tous les états ont la même période, qu'on appelle donc période de (X_n). Lorsqu'une chaine est irréductible de période 1, on dit qu'elle est *apériodique.*

Exemples. La chaine ayant $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ comme matrice de transition est de période 2. La marche aléatoire simple sur $\mathbb{Z}$ est également de période 2. Si P est irréductible et s'il existe x tel que $P(x,x) > 0$ alors x est de période 1, donc tout le monde est de période 1 donc P est apériodique. En particulier si P est n'importe quelle matrice irréductible, la matrice $(I+P)/2$ est irréductible et apériodique. Cette matrice est parfois appelée version *paresseuse* de P. En effet, si une chaine admet $(I+P)/2$ comme matrice de transition alors à chaque instant, la chaine se déplace selon la transition donnée par P avec probabilité $1/2$ ou reste en place avec probabilité $1/2$. Il également facile de voir qu'une chaine et sa version paresseuse ont les mêmes mesures invariantes.

Lemme III.57. *Un élément x est de période 1 si et seulement si $P^n(x,x) > 0$ pour tout n suffisamment grand. La matrice P est irréductible et apériodique si et seulement si pour tous x, y on a $P^n(x,y) > 0$ pour n assez grand.*

Démonstration. C'est juste une question d'arithmétique. L'ensemble $A = \{n \in \mathbb{N};\ P^n(x,x) > 0\}$ est toujours stable par addition. Si x est de période 1, alors A est de pgcd 1. Montrons qu'un tel ensemble contient tous les entiers à partir d'un certain rang. On commence par montrer qu'il contient deux entiers consécutifs. Comme A est de pgcd 1, d'après le théorème de Bezout il existe des entiers relatifs $k_1, \dots, k_r$ et des éléments $n_1, \dots, n_r$ de A tels que $\sum k_i n_i = 1$, ce qu'on réécrit

$$\sum_{i=1}^{r} (k_i)_+ n_i = \sum_{i=1}^{r} (k_i)_- n_i + 1.$$

Posons $n_0 = \sum (k_i)_- n_i$. Comme A est stable par addition $\sum (k_i)_- n_i$ et $\sum (k_i)_+ n_i$ appartiennent à A. On a donc montré que n_0 et $n_0 + 1$ appartiennent à A. Montrons maintenant que A contient tous les entiers supérieurs à n_0^2. Soit n un entier et soit $n = n_0 q + r$ la division Euclidienne de n par n_0. Si $n \geq n_0^2$ alors $q \geq n_0 > r$. Donc $q - r$ et r sont des entiers naturels ce qui montre que $n = (q-r)n_0 + r(n_0+1)$ appartient à A, comme annoncé.
Si de plus P est irréductible alors pour tous x, y il existe k tel que $P^k(x,y) > 0$. Comme d'après ce qui précède il existe l tel que $P^n(x,x) > 0$ pour tout $n \geq l$, on obtient

$$P^n(x,y) \geq P^{n-k}(x,x) P^k(x,y) > 0$$

pour tout $n \geq k + l$. Les réciproques sont évidentes. □

Théorème III.58. *Soit (X_n) une chaine de Markov récurrente positive et apériodique. On appelle π la probabilité stationnaire. Alors pour tout $x \in M$ on a*

$$\mathrm{TV}(\delta_x P^n, \pi) \to 0.$$

En particulier pour tous $x, y \in M$

$$\mathbb{P}_x(X_n = y) \to \pi(y).$$

La démonstration utilise une méthode dite de couplage.

Lemme III.59. *Soit (X_n, Y_n) une chaine de Markov sur $M \times M$. On suppose que les processus (X_n) et (Y_n) sont des chaines de Markov sur M ayant la même matrice de transition P et on pose*

$$T = \inf\{n \geq 0 \colon X_n = Y_n\}.$$

Alors en appelant μ et ν les lois respectives de X_0 et Y_0, on a l'inégalité

$$\mathrm{TV}(\mu P^n, \nu P^n) \leq \mathbb{P}(T > n), \quad \forall n \in \mathbb{N}.$$

Remarque III.60. En général le fait que (X_n, Y_n) soit une chaine de Markov n'implique pas que les marginales (X_n) et (Y_n) le soient aussi, c'est une hypothèse qu'on fait. En appelant Q et P les matrices de transition respectives de (X_n, Y_n) et (X_n) cette hypothèse se traduit ainsi : pour tous x, x', y on a

$$\sum_{y' \in M} Q((x,y),(x',y')) = P(x,x').$$

Notons en particulier que $\sum_{y' \in M} Q((x,y),(x',y'))$ ne dépend pas de y. Plus généralement on aura

$$\sum_{y' \in M} Q^n((x,y),(x',y')) = P^n(x,y),$$

pour tout entier n.

Démonstration. Montrons d'abord que X_n et Y_n ont la même loi conditionnelle sachant $T \leq n$. Soit $(\mathcal{F}_n)$ la filtration naturelle du processus (X_n, Y_n). Comme T est un temps d'arrêt on a pour tout $k \leq n$ et $x \in M$

$$\mathbb{P}(X_n = x; T = k) = \mathbb{E}\left[\mathbb{P}(X_n = x \mid \mathcal{F}_k)\mathbb{1}_{\{T=k\}}\right].$$

En utilisant la propriété de Markov pour le processus (X_n, Y_n) et la remarque faite plus haut on obtient

$$\begin{aligned}\mathbb{P}(X_n = x \mid \mathcal{F}_k) &= \sum_{y \in M} \mathbb{P}((X_n, Y_n) = (x,y) \mid \mathcal{F}_k) \\ &= \sum_{y \in M} Q^{n-k}((X_k, Y_k),(x,y)) = P^{n-k}(X_k, x).\end{aligned}$$

Puis, en utilisant $X_T = Y_T$ p.s.,

$$\begin{aligned}\mathbb{P}(X_n = x; T = k) &= \mathbb{E}\left[P^{n-k}(X_k, x)\mathbb{1}_{\{T=k\}}\right] \\ &= \mathbb{E}\left[P^{n-k}(Y_k, x)\mathbb{1}_{\{T=k\}}\right] = \mathbb{P}(Y_n = x; T = k).\end{aligned}$$

En sommant cette égalité sur les $k \leq n$ on obtient

$$\mathbb{P}(X_n = x; T \leq n) = \mathbb{P}(Y_n = x; T \leq n)$$

pour tout x, ce qu'il fallait démontrer. La fin de la démonstration est facile. Pour $A \subset M$ on écrit

$$\begin{aligned}\mu P^n(A) - \nu P^n(A) &= \mathbb{P}(X_n \in A) - \mathbb{P}(Y_n \in A) \\ &\leq \mathbb{P}(T > n) + \mathbb{P}(X_n \in A; T \leq n) - \mathbb{P}(Y_n \in A; T \leq n) = \mathbb{P}(T > n),\end{aligned}$$

d'après ce qui précède. En prenant le supremum en A on obtient l'inégalité cherchée. $\square$

Démonstration du Théorème III.58. Soient (X_n) et (Y_n) deux chaines de Markov de matrice de transition P qui évoluent indépendamment l'une de l'autre. Autrement dit (X_n, Y_n) est une chaine de Markov de matrice de transition Q donnée par

$$Q((x,y),(x',y')) = P(x,x')P(y,y'), \quad \forall x, x', y, y' \in M,$$

On pose $T = \inf\{n \geq 0;\ X_n = Y_n\}$. D'après le lemme précédent on a (en appelant μ et ν les lois respectives de X_0 et Y_0)

$$\mathrm{TV}(\mu P^n, \nu P^n) \leq \mathbb{P}(T > n).$$

Montrons maintenant que T est fini p.s. C'est ici qu'on va utiliser l'apériodicité. D'après le Lemme III.57, pour tous x, y, x', y' on a $P^n(x,y) > 0$ et $P^n(x',y') > 0$ pour n suffisamment grand. Et donc $Q^n((x,y),(x',y')) > 0$ à partir d'un certain rang. En particulier la matrice Q est irréductible. Par ailleurs il est très facile de voir que la mesure produit $\pi \otimes \pi$ est stationnaire pour Q. Donc Q est irréductible et admet une probabilité invariante. Ceci implique que Q est récurrente positive. La chaine (X_n, Y_n) visite donc tous les points de $M \times M$, en particulier la diagonale $\{(x,x),\ x \in M\}$, ce qui montre que T est fini p.s. Par conséquent $\mathbb{P}(T > n) \to 0$. On obtient donc

$$\mathrm{TV}(\mu P^n, \nu P^n) \to 0.$$

Ceci est valable pour toutes mesures de probabilité μ, ν. En appliquant à $\nu = \pi$ on obtient le résultat. $\square$

Pour les applications il est important de savoir à quelle vitesse a lieu la convergence. On a donc besoin d'estimer $\mathbb{P}(T > n)$. Le couplage indépendant utilisé dans la preuve précédente permet de voir que $\delta_x P^n \to \pi$ mais il donne rarement le bon ordre de grandeur pour la vitesse de convergence. Étant donnée une chaine concrète, le jeu consiste à trouver un bon couplage auquel appliquer le Lemme III.59.

Exemple. Soit $\mathbb{Z}_L$ l'ensemble des entiers modulo L. On considère la marche aléatoire paresseuse sur $\mathbb{Z}_L$, de matrice de transition P donnée par

$$P(x,x) = \frac{1}{2},\; P(x,x+1) = \frac{1}{4} = P(x,x-1), \quad x \in \mathbb{Z}_L$$

Alors P est irréductible et apériodique et la mesure stationnaire est bien sûr uniforme sur $\mathbb{Z}_L$. Soient $x, y \in \mathbb{Z}_L$, et soit (X_n) une marche paresseuse partant de x. Soit (Y_n) une marche partant de y, couplée à (X_n) de la manière suivante. Lorsque X_n saute à droite ou à gauche, Y_n ne bouge pas, et si X_n reste en place, alors Y_n saute à droite ou à gauche de manière équiprobable. De manière plus formelle (X_n, Y_n) est une chaine de Markov sur $\mathbb{Z}_L \times \mathbb{Z}_L$ de matrice de transition Q donnée par

$$Q((x,y),(x+1,y)) = Q((x,y),(x-1,y)) = Q((x,y) = (x,y+1)) = Q((x,y),(x,y-1)) = \frac{1}{4}, \quad \forall x,y \in \mathbb{Z}_L.$$

En appelant T le temps de couplage, on obtient d'après le Lemme III.59

$$\mathrm{TV}(\delta_x P^n, \delta_y P^n) \leq \mathbb{P}(T > n) \leq \frac{\mathbb{E}[T]}{n}.$$

Déterminons maintenant $\mathbb{E}[T]$. Posons $Z_n = X_n - Y_n$. On remarque alors que le processus (Z_n) est également Markovien et que sa matrice de transition R est donnée par

$$R(x,x+1) = R(x,x-1) = \frac{1}{2}, \quad \forall x \in \mathbb{Z}_L.$$

Autrement dit (Z_n) est une marche non paresseuse sur $\mathbb{Z}_L$ partant de $x - y$. On a vu dans le problème sur la ruine du joueur que le temps T quelle met à toucher 0 vérifie

$$\mathbb{E}[T] = |x-y|\,(L - |x-y|) \leq \frac{L^2}{4}.$$

On obtient donc

$$\mathrm{TV}(\delta_x P^n, \delta_y P^n) \leq \frac{L^2}{4n}, \quad \forall n \in \mathbb{N},$$

pour tous $x, y \in M$. Il est facile de voir que ceci implique

$$\mathrm{TV}(\delta_x P^n, \pi) \leq \frac{L^2}{4n}, \quad \forall n \in \mathbb{N},$$

pour tout $x \in M$. Par conséquent, au bout d'un temps n de l'ordre de L^2, la loi de X_n est proche en variation totale de la loi uniforme sur $\mathbb{Z}_L$.