

PROCESSUS STOCHASTIQUES À TEMPS DISCRET

Yueyun Hu et Laurent Tournier

AVANT-PROPOS

Les sections Complément dans les chapitres 0, 1, et 2, correspondent à l'étude plus approfondie des sujets concernés.

Ce texte est seulement un support du cours, et il n'est pas destiné à être distribué ou mis en ligne.

Nous ne revendiquons aucune originalité des contenus de ce texte, en particulier les Chapitres 4 et 5 sont entièrement issus du livre de Lamberton et Lapeyre.

Table des matières

0	Introduction	7
0.1	Espace de probabilités	7
0.2	Processus stochastiques	12
0.3	Deux résultats fondamentaux	13
1	Chaînes de Markov	15
1.1	Définition	15
1.2	Probabilités de transition à m pas	18
1.3	Exemples de chaînes de Markov	19
1.3.1	Chaîne de markov à deux états	19
1.3.2	Marche aléatoire sur \mathbb{Z}	19
1.3.3	Marches aléatoires avec barrières	20
1.3.4	Modèle d'Ehrenfest	20
1.3.5	Processus du renouvellement	20
1.3.6	Processus de branchement	21
1.4	Propriété de Markov	22
1.5	Visites à un sous-ensemble	25
1.6	Réurrence et transience	27
1.7	Décomposition en classes de communication	28
1.8	Exemples	30
1.8.1	Chaîne à deux états	30
1.8.2	Chaîne sur \mathbb{N}	31
1.8.3	Marche aléatoire sur \mathbb{Z}	31
1.8.4	Marche aléatoire simple symétrique sur \mathbb{Z}^d	31
1.9	Probabilités d'absorption	31
1.10	Mesures et Probabilités invariantes	33
1.11	Périodicité et existence de la loi limite au sens fort	38

1.12	Complément sur le processus canonique	43
1.12.1	Première construction	43
1.12.2	Autre approche : extension de lois fini-dimensionnelles compatibles	44
2	Espérance conditionnelle	47
2.1	Introduction	47
2.2	Définition	49
2.3	Propriétés élémentaires	51
2.4	Inégalité de Jensen	51
2.5	Lois et densités conditionnelles	52
2.6	Application statistique : estimateur bayésien.	54
2.7	Propriété de Markov forte et ses applications	55
2.8	Exercices : Espérance conditionnelle	58
3	Martingales en temps discret	61
3.1	Introduction et définitions	61
3.2	Martingales, sur- et sous-martingales.	62
3.3	Propriétés de martingales	63
3.4	Transformée de martingale, théorème d'arrêt	66
3.5	Inégalités maximales	72
3.6	Convergence de martingales	73
3.6.1	Cas des martingales de carré intégrable :	73
3.6.2	Convergence presque sûre	74
3.6.3	Preuve de Doob du Théorème 3.34 :	74
3.6.4	Preuve de Jacques Neveu du Théorème 3.34 :	76
3.6.5	Cas des martingales L^1	77
3.7	Critères de récurrence d'une chaîne de Markov	78
3.8	Étude de la file attente $M/G/1$	81
3.9	Exercices : Martingales à temps discrets	83
4	Modèle de Cox-Ross-Rubinstein	87
4.1	Généralités sur les marchés financiers	87
4.2	Formalisme des marchés financiers discrets	89
4.3	Marché viable et complet	90
4.4	Évaluation et couverture dans un marché complet	92
4.4.1	Exemple du call et du put européens	93

4.5	Le modèle de Cox-Ross-Rubinstein	93
4.6	Pricing dans le modèle C.R.R.	95
4.7	Stratégie de couverture	96
4.8	Exercices : Modèles financiers à temps discret	97
5	Options américaines	99
5.1	Arrêt optimal et enveloppe de Snell	99
5.1.1	Enveloppe de Snell du processus Z	99
5.1.2	Arrêt optimal	100
5.1.3	Décomposition de sur-martingales et plus grand temps d'arrêt optimal	102
5.2	Cas des chaînes de Markov	103
5.3	Applications aux options américaines	103
5.3.1	Évaluation d'une option américaine : suite des profits et suite des primes. . .	104
5.3.2	Exercice optimal d'une option américaine	105
5.3.3	Couverture des options américaines	105
5.4	Options américaines et options européennes	106

Chapitre 0

Introduction et Rappels

Les probabilités ont pour but l'étude des phénomènes aléatoires. En particulier, les processus stochastiques permettent de modéliser l'évolution dans le temps d'un phénomène aléatoire. Nous rappelons d'abord rapidement les principales notions de probabilité dont nous aurons besoin.

0.1 Espace de probabilités

Un **espace de probabilité** $(\Omega, \mathcal{F}, \mathbb{P})$ se compose de :

- Un **ensemble** Ω : ensemble des résultats possibles de l'expérience aléatoire considérée.
- Une **tribu** (ou σ -algèbre) de sous-ensembles de Ω . Un élément de \mathcal{F} s'appelle un **événement**.

On rappelle :

Définition 0.1. Une classe \mathcal{F} de sous-ensembles de Ω est appelée une **tribu** (ou σ -algèbre) sur l'ensemble Ω si

- (i) $\Omega \in \mathcal{F}, \emptyset \in \mathcal{F}$.
- (ii) pour tout $A \in \mathcal{F}, A^c \in \mathcal{F}$.
- (iii) pour toute suite $(A_n)_{n \in \mathbb{N}}$ à valeurs dans $\mathcal{F}, \bigcup_{n=0}^{\infty} A_n \in \mathcal{F}$.

Exemple 0.2. Sur Ω , on note $\mathcal{P}(\Omega)$ l'ensemble de tous les sous-ensembles de Ω y compris l'ensemble vide. Alors $\mathcal{P}(\Omega)$ est la plus grande tribu dite tribu discrète et $\{\emptyset, \Omega\}$ est la plus petite tribu dite tribu grossière ou tribu triviale. Un espace de probabilité est dit discret si la tribu \mathcal{F} est la tribu discrète. C'est en général le cas lorsque Ω est dénombrable.

Proposition 0.3. L'intersection d'une famille de tribus est encore une tribu.

Preuve : preuve laissée comme exercice. □

Pour tout ensemble de parties $\mathcal{C} \subset \mathcal{P}(\Omega)$, on note $\sigma(\mathcal{C})$ la plus petite tribu sur Ω contenant \mathcal{C} , appelée **tribu engendrée par \mathcal{C}** (elle existe par la proposition précédente, car c'est l'intersection de

toutes les tribus contenant \mathcal{C}). Sur \mathbb{R}^n , la **tribu des boréliens** notée $\mathcal{B}(\mathbb{R}^n)$ est la tribu engendrée par les ouverts (elle est aussi engendrée par les fermés, et par les pavés fermés).

• **Une probabilité \mathbb{P} .** On rappelle : une **probabilité** \mathbb{P} sur (Ω, \mathcal{F}) est une mesure (positive) de masse totale égale à 1. C'est donc une application

$$\mathbb{P} : \mathcal{F} \rightarrow [0, 1],$$

telle que $\mathbb{P}(\Omega) = 1$ et qui est σ -additive : pour toute suite (A_n) d'événements *disjoints*,

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

Une propriété sur Ω est vraie **presque sûrement** par rapport à la probabilité \mathbb{P} (ce que l'on écrit en abrégé : “ \mathbb{P} -p.s.”) si elle est vérifiée partout sauf sur un événement de probabilité nulle.

Définition 0.4. Soit (E, \mathcal{E}) un espace mesurable. Une **variable aléatoire (v.a.) à valeurs dans E** est une application $X : (\Omega, \mathcal{F}) \rightarrow (E, \mathcal{E})$ mesurable, c'est-à-dire telle que

$$\forall B \in \mathcal{E}, \quad X^{-1}(B) = \{X \in B\} = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}.$$

On parle de **variable aléatoire (v.a.) réelle** pour une v.a. à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, et de **vecteur aléatoire réel** pour une v.a. à valeurs dans $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, avec $n \geq 2$.

De même une application $f : (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)) \rightarrow (\mathbb{R}^q, \mathcal{B}(\mathbb{R}^q))$ est **borélienne** si elle est mesurable pour les tribus des boréliens correspondantes :

$$\forall B \in \mathcal{B}(\mathbb{R}^q), \quad f^{-1}(B) = \{x \in \mathbb{R}^n | f(x) \in B\} \in \mathcal{B}(\mathbb{R}^n).$$

Définition 0.5. La **loi** d'une variable aléatoire $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (E, \mathcal{E})$ est la probabilité \mathbb{P}_X définie sur (E, \mathcal{E}) par :

$$\forall B \in \mathcal{E}, \quad \mathbb{P}_X(B) = \mathbb{P}(X \in B).$$

Au besoin, on peut préciser qu'il s'agit de la **loi de X sous \mathbb{P}** .

Pour une variable aléatoire réelle X , la loi de X est donc une probabilité sur \mathbb{R} .

Si X ne prend qu'un ensemble $I = X(\Omega)$ *dénombrable* de valeurs dans E (par exemple si X est une v.a. réelle dont les valeurs sont entières), on dit que la loi de X est **discrète** sur E . En introduisant, pour tout $x \in E$, la mesure de Dirac δ_x en x , et en posant, pour tout $x \in E$, $p_x = \mathbb{P}(X = x)$ ($= 0$ si $x \notin I$), on voit que la loi de la v.a. discrète X s'écrit $\mathbb{P}_X = \sum_{x \in I} p_x \delta_x$. Pour tout $B \subset I$ (ou $B \subset E$), on a en effet :

$$\mathbb{P}(X \in B) = \sum_{x \in B} p_x.$$

Un autre cas important est celui où \mathbb{P}_X admet une densité f par rapport à la mesure de Lebesgue dx sur \mathbb{R}^n , noté $\mathbb{P}_X = f(x)dx$. On dit alors que la loi de X est **absolument continue** sur \mathbb{R}^n . Dans ce cas pour tout borélien $B \in \mathcal{B}(\mathbb{R}^n)$:

$$\mathbb{P}_X(B) = \int_B f(x)dx.$$

Le lemme suivant s'avèrera souvent utile.

Lemme 0.6 (Lemme d'unicité). *Dans (Ω, \mathcal{F}) considérons une classe \mathcal{C} de parties de \mathcal{F} , stable par intersections finies. Si \mathbf{P} et \mathbf{Q} sont deux probabilités telles que*

$$\mathbf{P}(C) = \mathbf{Q}(C), \quad \forall C \in \mathcal{C},$$

alors

$$\mathbf{P}(A) = \mathbf{Q}(A), \quad \forall A \in \sigma(\mathcal{C}).$$

Par exemple, supposons que X et Y sont des v.a. à valeurs dans \mathbb{R}^d telles que, pour tout pavé fermé C , $\mathbb{P}(X \in C) = \mathbb{P}(Y \in C)$. Alors leurs lois \mathbb{P}_X et \mathbb{P}_Y sont des probabilités sur \mathbb{R}^d qui coïncident sur la famille \mathcal{C} des pavés fermés de \mathbb{R}^d . Or \mathcal{C} est stable par intersections finies et engendre $\mathcal{B}(\mathbb{R}^d)$, donc $\mathbb{P}_X = \mathbb{P}_Y$: X et Y ont même loi.

Pour une v.a. X réelle positive, ou à valeurs dans \mathbb{R}^d et intégrable pour la probabilité \mathbb{P} , on définit son **espérance** comme son intégrale sur Ω par rapport à \mathbb{P} :

$$\mathbb{E}(X) = \int_{\Omega} X(\omega) \mathbb{P}(d\omega).$$

La v.a. X à valeurs dans \mathbb{R}^d est donc intégrable si $\mathbb{E}(|X|) < \infty$.

Pour une v.a. X à valeurs dans \mathbb{R}^d , pour toute fonction borélienne $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$, la variable $Y = \varphi(X)$ est intégrable pour la probabilité \mathbb{P} sur Ω si, et seulement si, φ est intégrable pour \mathbb{P}_X sur \mathbb{R}^d , et

$$\mathbb{E}(\varphi(X)) = \int_{\mathbb{R}^d} \varphi(x) \mathbb{P}_X(dx).$$

C'est le *théorème de transfert*. Si $\varphi \geq 0$, cette formule reste valable sans supposer $\varphi(X)$ intégrable.

Si X est une v.a. discrète à valeurs dans un ensemble dénombrable I , pour toute fonction $\varphi : I \rightarrow \mathbb{R}$ on a donc

$$\mathbb{E}(\varphi(X)) = \sum_{x \in I} \varphi(x) \mathbb{P}(X = x),$$

dès lors que $\varphi \geq 0$ ou $\sum_{x \in I} |\varphi(x)| \mathbb{P}(X = x) < \infty$. Et si X est une v.a. sur \mathbb{R}^d ayant pour densité f_X , pour toute fonction mesurable $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$ on a

$$\mathbb{E}(\varphi(X)) = \int \varphi(x) f_X(x) dx.$$

dès lors que $\varphi \geq 0$ ou $\int |\varphi(x)| f_X(x) dx < \infty$.

Citons pour mémoire les lois classiques comme : Bernoulli, géométrique, binomiale, Poisson, pour les lois discrètes ; exponentielles, de Gauss, pour les lois absolument continues.

Définition 0.7. *Pour une v.a. X sur (Ω, \mathcal{F}) , à valeurs dans (E, \mathcal{E}) , on définit la **tribu engendrée par X** comme la plus petite tribu sur Ω rendant X mesurable. On la note $\sigma(X)$ et on a :*

$$\sigma(X) = \{X^{-1}(B) \mid B \in \mathcal{E}\} \subset \mathcal{F}.$$

C'est la tribu de tous les événements relatifs à X , c'est-à-dire de la forme $\{X \in B\}$.

Connaître la valeur de X revient à savoir si les événements de la tribu $\sigma(X)$ sont réalisés ou non. De façon générale, on utilisera souvent les sous-tribus de \mathcal{F} pour décrire des ensembles d'informations dont on peut disposer sur le résultat de l'expérience aléatoire.

Le résultat suivant montre qu'une variable aléatoire Z est $\sigma(X)$ -mesurable si, et seulement si c'est une fonction (mesurable) de X .

Proposition 0.8. *Considérons une application $X : \Omega \rightarrow E$, où (E, \mathcal{E}) est un espace mesurable. Alors une application $Z : \Omega \rightarrow \mathbb{R}$ est $\sigma(X)$ -mesurable si et seulement si il existe une application mesurable $f : E \rightarrow \mathbb{R}$ telle que $Z = f(X)$.*

Preuve : On écrit

$$Z = \lim_{n \rightarrow +\infty} \sum_{k=-\infty}^{+\infty} \frac{k}{2^n} \mathbf{1}_{[k2^{-n}, (k+1)2^{-n}[}(Z).$$

Puisque Z est $\sigma(X)$ -mesurable, pour tous $n \geq 0$ et $k \in \mathbb{Z}$ il existe $A_{n,k} \in \mathcal{E}$ tel que

$$\{Z \in [k2^{-n}, (k+1)2^{-n}[\} = \{X \in A_{n,k}\},$$

c'est à dire tel que $\mathbf{1}_{[k2^{-n}, (k+1)2^{-n}[}(Z) = \mathbf{1}_{A_{n,k}}(X)$. On a alors $Z = f(X)$ en définissant

$$\text{pour tout } x \in E, \quad f(x) = \liminf_{n \rightarrow +\infty} \sum_{k=-\infty}^{+\infty} \frac{k}{2^n} \mathbf{1}_{A_{n,k}}(x),$$

d'où le lemme. □

La relation d'équivalence entre les variables aléatoires définie par l'égalité presque sûre,

$$X = Y \quad \mathbb{P}\text{-p.s.},$$

définit des classes de v.a. sur $(\Omega, \mathcal{F}, \mathbb{P})$. On note $L^1(\Omega, \mathcal{F}, \mathbb{P})$ et $L^2(\Omega, \mathcal{F}, \mathbb{P})$ les espaces vectoriels des classes de v.a. réelles intégrables et de carré intégrable pour la probabilité \mathbb{P} . Contrairement aux espaces $L^1(dx)$ et $L^2(dx)$ pour la mesure de Lebesgue sur \mathbb{R}^n , on a toujours

$$L^2(\Omega, \mathcal{F}, \mathbb{P}) \subset L^1(\Omega, \mathcal{F}, \mathbb{P}).$$

De plus $L^1(\Omega, \mathcal{F}, \mathbb{P})$ est un espace de Banach pour la norme $\|X\|_1 = \mathbb{E}(|X|)$ et $L^2(\Omega, \mathcal{F}, \mathbb{P})$ est un espace de Hilbert pour le produit scalaire suivant : $\langle X, Y \rangle = \mathbb{E}(XY)$.

Définition 0.9. *Sur $(\Omega, \mathcal{F}, \mathbb{P})$.*

- Deux événements A et A' sont **indépendants** si

$$\mathbb{P}(A \cap A') = \mathbb{P}(A)\mathbb{P}(A').$$

- Des événements A_1, \dots, A_n sont **indépendants** si

$$\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \dots \mathbb{P}(A_{i_k}),$$

pour tous $2 \leq k \leq n$ et $1 \leq i_1 < \dots < i_k \leq n$.

- Des tribus $\mathcal{A}_1, \dots, \mathcal{A}_n$ contenues dans \mathcal{F} sont **indépendantes** si

$$\mathbb{P}(A_1 \cap A_2 \cdots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2) \cdots \mathbb{P}(A_n),$$

pour tous $A_1 \in \mathcal{A}_1, \dots, A_n \in \mathcal{A}_n$.

- Une v.a. X est dite **indépendante de la tribu \mathcal{G}** si $\sigma(X)$ et \mathcal{G} sont indépendantes.
- Des variables aléatoires X_1, \dots, X_n sont **indépendantes** si les tribus $\sigma(X_1), \dots, \sigma(X_n)$ sont indépendantes ; autrement dit, dans le cas de v.a. réelles, si

$$\mathbb{P}\left(\bigcap_{i=1}^n \{X_i \in B_i\}\right) = \prod_{i=1}^n \mathbb{P}(X_i \in B_i),$$

pour tous $B_1 \in \mathcal{B}(\mathbb{R}), \dots, B_n \in \mathcal{B}(\mathbb{R})$.

Pour X_1 à valeurs dans $(E_1, \mathcal{E}_1), \dots, X_n$ à valeurs dans (E_n, \mathcal{E}_n) , l'indépendance de X_1, \dots, X_n équivaut à dire que la loi du vecteur (X_1, \dots, X_n) est la **loi produit** des lois de X_1, \dots, X_n : $\mathbb{P}_{(X_1, \dots, X_n)} = \mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_n}$. En conséquence, par les théorèmes de Fubini, on a :

$$\mathbb{E}[\varphi(X_1, \dots, X_n)] = \int \cdots \int \varphi(x_1, \dots, x_n) d\mathbb{P}_{X_1}(x_1) \cdots d\mathbb{P}_{X_n}(x_n),$$

dès lors que $\varphi \geq 0$ ou que $\varphi(X_1, \dots, X_n)$ est intégrable, où les intégrales peuvent être calculées dans un ordre quelconque. En particulier, pour toutes fonctions $f_1 : E_1 \rightarrow \mathbb{R}, \dots, f_n : E_n \rightarrow \mathbb{R}$ mesurables,

$$\mathbb{E}[f_1(X_1) \cdots f_n(X_n)] = \mathbb{E}[f_1(X_1)] \cdots \mathbb{E}[f_n(X_n)],$$

dès lors que ces fonctions sont positives, ou que $f_1(X_1), \dots, f_n(X_n)$ sont intégrables.

Le lemme d'unicité fournit la proposition suivante qui simplifie souvent la vérification de l'indépendance :

Proposition 0.10. Soient $\mathcal{C}_1 \subset \mathcal{F}$ et $\mathcal{C}_2 \subset \mathcal{F}$ deux classes stables par intersection finie, engendrant deux tribus \mathcal{A}_1 et \mathcal{A}_2 . Si pour tout $A \in \mathcal{C}_1$ et $B \in \mathcal{C}_2$,

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B),$$

alors les tribus \mathcal{A}_1 et \mathcal{A}_2 sont indépendantes.

Preuve : Soit $B \in \mathcal{C}_2$. Si $\mathbb{P}(B) = 0$, l'égalité est évidente car $\mathbb{P}(A \cap B) \leq \mathbb{P}(B) = 0$. Supposons $\mathbb{P}(B) > 0$. Alors l'égalité s'écrit $\mathbb{P}(A|B) = \mathbb{P}(A)$. Il résulte alors du lemme d'unicité (avec $\mathbf{Q} = \mathbb{P}(\cdot|B)$) que, pour tout $A \in \mathcal{A}_1$, $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Soit $A \in \mathcal{A}_1$. De même, si $\mathbb{P}(A) > 0$, il faut montrer que $\mathbb{P}(B) = \mathbb{P}(B|A)$ pour tout $B \in \mathcal{A}_2$; or c'est vrai pour tout $B \in \mathcal{C}_2$ par ce qui précède, et le lemme permet de conclure, comme précédemment. \square

On étend ces définitions à des familles infinies :

Définition 0.11. Une famille infinie d'événements (resp. de variables aléatoires, resp. de tribus) est **indépendante** si toute sous-famille finie de celle-ci est indépendante.

En général, disposer de l'information qu'un événement A est réalisé modifie la probabilité des événements. Il s'agit alors de la probabilité conditionnelle sachant A :

Définition 0.12. *Pour deux événements A et B tels que $\mathbb{P}(A) > 0$, la **probabilité conditionnelle de B sachant A** est*

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

Remarquons que $\mathbb{P}(\cdot | A)$ est une probabilité sur (Ω, \mathcal{F}) . On peut donc considérer l'espérance associée, notée $\mathbb{E}(\cdot | A)$; pour toute variable aléatoire X intégrable, on a

$$\mathbb{E}(X | A) = \frac{\mathbb{E}(X \mathbf{1}_A)}{\mathbb{P}(A)}.$$

On peut aussi considérer la loi d'une variable aléatoire X sous $\mathbb{P}(\cdot | A)$ (on parlera de **loi conditionnelle sachant A**), ou dire que deux variables aléatoires X et Y sont indépendantes sous $\mathbb{P}(\cdot | A)$ (on parlera d'**indépendance conditionnelle sachant A**). Ces notions joueront un rôle important.

Durant ce cours, on donnera un sens à ces définitions dans certains cas où $\mathbb{P}(A) = 0$ et aussi dans des cas de conditionnement par une tribu.

0.2 Processus stochastiques

Comme il a été dit au début, les *processus stochastiques* permettent de représenter l'évolution dans le temps de phénomènes aléatoires. Le mouvement d'une particule dans l'espace, la transmission d'un signal, le passage dans le temps d'un système à différents états, la variation des cours d'actifs sur un marché, etc., en sont des exemples.

Dans la pratique, on se fixe un ensemble $T \subset \mathbb{R}_+$ d'instantanés d'observation du phénomène étudié.

1. Si T est un intervalle $[a, b]$ on dit que l'étude se fait en **temps continu**.
2. Si T est formé d'une suite d'observations $t_0 = 0 < t_1 < t_2 < \dots < t_n < \dots$, on dit que l'étude se fait en **temps discret**.

Le principe est alors le suivant : pour chaque temps t de T , le phénomène étudié est à un état aléatoire représenté par une v.a. X_t prenant ses valeurs parmi tous les états a priori possibles pour le phénomène. Le processus stochastique correspondant est alors défini par la donnée de la famille de v.a. $(X_t)_{t \in T}$. Autrement dit :

Définition 0.13. *Un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ et une partie T de \mathbb{R}_+ étant donnés, un **processus stochastique** sur l'espace des états $E \subset \mathbb{R}^n$ est une famille $X = (X_t)_{t \in T}$ de v.a.*

$$X_t : \Omega \rightarrow E, \quad t \in T.$$

Pour tout $\omega \in \Omega$ fixé, l'application

$$t \mapsto X_t(\omega)$$

s'appelle une **trajectoire** du processus.

Nous n'étudierons ici que des processus à temps discret (cas 2 pour T) ; on se limite alors à prendre $T \subset \mathbb{N}$; on voit que dans ce cas un processus stochastique revient à la donnée d'une suite de v.a..

On rappelle que l'espace $E^{\mathbb{N}}$ des suites à valeurs dans un espace mesurable (E, \mathcal{E}) peut être muni de la **tribu produit**, $\mathcal{E}^{\otimes \mathbb{N}}$, qui est la tribu engendrée par l'ensemble des **cylindres**

$$B_1 \times \cdots \times B_k \times E^{\mathbb{N}} \subset E^{\mathbb{N}}, \quad \text{pour } B_1, \dots, B_k \in \mathcal{E}, \quad k \in \mathbb{N}.$$

Proposition 0.14. *Soit $X = (X_n)_{n \in \mathbb{N}}$ un processus stochastique à temps discret à valeurs dans E .*

- *Alors $X : (\Omega, \mathcal{F}) \rightarrow (E^{\mathbb{N}}, \mathcal{E}^{\otimes \mathbb{N}})$ est une variable aléatoire. On peut donc considérer sa loi.*
- *La loi de X est caractérisée par la donnée, pour tout $n \geq 0$, de la loi de (X_0, \dots, X_n) .*

Le second point vient du lemme d'unicité 0.6 appliqué aux cylindres.

Si E est dénombrable, se donner la loi de X revient donc à connaître, pour tout $n \geq 0$, pour tous $x_0, \dots, x_n \in E$, la probabilité élémentaire

$$\mathbb{P}(X_0 = x_0, \dots, X_n = x_n),$$

et si ces probabilités sont les mêmes pour deux processus $X = (X_n)_n$ et $Y = (Y_n)_n$, alors les probabilités de tous les événements relatifs à ces processus seront donc égales.

0.3 Deux résultats fondamentaux

On rappelle une abréviation extrêmement courante :

Définition 0.15. *Une famille $(X_i)_{i \in I}$ de v.a. est dite **i.i.d.** (pour **indépendante et identiquement distribuée**) si ces variables sont indépendantes et ont toutes la même loi.*

Théorème 0.16 (Loi (forte) des grands nombres). *Soit $(X_n)_{n \geq 1}$ une suite de v.a. i.i.d., intégrables à valeurs dans \mathbb{R}^d , ayant même loi qu'une v.a. X . On a*

$$\lim_{n \rightarrow \infty} \frac{1}{n} (X_1 + \cdots + X_n) = \mathbb{E}(X), \quad p.s.$$

Rappelons qu'une v.a. réelle X est dite **gaussienne** si sa densité est de la forme

$$x \mapsto \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-m)^2/(2\sigma^2)}, \quad x \in \mathbb{R}.$$

Dans ce cas, $\mathbb{E}(X) = m$ et $\text{Var}(X) = \sigma^2$ et on note $X \sim \mathcal{N}(m, \sigma^2)$. La fonction caractéristique déterminant de façon unique la loi, $X \sim \mathcal{N}(m, \sigma^2)$ si et seulement si

$$\mathbb{E} e^{i\lambda X} = e^{i\lambda m - \sigma^2 \lambda^2 / 2}, \quad \forall \lambda \in \mathbb{R}.$$

Un vecteur (X_1, \dots, X_n) est dit **gaussien** si toute combinaison linéaire $\sum_{j=1}^n \lambda_j X_j$, $\lambda_j \in \mathbb{R}$, est une v.a. réelle gaussienne. On peut montrer que la loi d'un tel vecteur est caractérisée par son espérance $m = (\mathbb{E}X_1, \dots, \mathbb{E}X_n)$ et sa matrice de covariance $\Gamma = (\Gamma_{kl})_{1 \leq k, l \leq n}$ avec $\Gamma_{k,l} = \text{cov}(X_k, X_l)$; sa loi est alors notée $\mathcal{N}(m, \Gamma)$.

Théorème 0.17 (Théorème Central Limite). *Soit $(X_n)_n$ une suite de v.a. i.i.d., de carré intégrable, à valeurs dans \mathbb{R}^d , ayant même loi qu'une v.a. X . Alors quand $n \rightarrow \infty$,*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \mathbb{E}(X_i)) \xrightarrow{(d)} \mathcal{N}(0, \Gamma),$$

où $\mathcal{N}(0, \Gamma)$ est une loi gaussienne sur \mathbb{R}^d , centrée et de matrice de variance-covariance Γ donnée par celle de X :

$$\Gamma_{k,l} = \text{Cov}(X^k, X^l), \quad 1 \leq k, l \leq d,$$

en ayant noté (X^1, \dots, X^d) les composantes de X .

Chapitre 1

Chaînes de Markov sur un espace d'états discret

Intuitivement, les processus de Markov sont des processus stochastiques tels qu'à chaque instant leur comportement dans le futur dépend seulement de leur valeur présente et non de toutes leurs valeurs passées. Les chaînes de Markov sont des processus de Markov à temps discret. Ces processus permettent de modéliser de nombreux phénomènes ; ils sont par exemple utilisés pour des problèmes de télécommunication, de théorie du signal, d'imagerie informatique, etc.

Nous nous limiterons ici à des chaînes ayant un espace d'états E dénombrable (souvent, on aura $E \subset \mathbb{Z}$). Cet espace sera muni de la tribu grossière $\mathcal{P}(E)$, et la loi d'une variable aléatoire X à valeurs dans E sera donc donnée par les probabilités élémentaires $\mathbb{P}(X = x)$ pour tout x de E .

Soit $(X_n, n \geq 0)$ un processus stochastique à valeurs dans E . Les éléments de E sont souvent notés par i, j, k ou x, y, z .

1.1 Définition

Définition 1.1. Le processus $(X_n, n \geq 0)$ est appelé une **chaîne de Markov** s'il existe une famille $(P(x, y))_{x, y \in E}$ de réels, appelés **probabilités de transition**, telle que, pour tout $n \geq 0$, et tous $x_0, \dots, x_n, x_{n+1} \in E$,

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

dès que $\mathbb{P}(X_0 = x_0, \dots, X_n = x_n) > 0$.

Cette définition équivaut à vérifier les deux propriétés suivantes simultanément :

(i) Pour tout $n \geq 1$, et tous $x_0, \dots, x_{n+1} \in E$,

$$\mathbb{P}(X_{n+1} = x_{n+1} \mid X_0 = x_0, \dots, X_n = x_n) = \mathbb{P}(X_{n+1} = x_{n+1} \mid X_n = x_n), \quad (1.1)$$

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

(ii) pour tous $x, y \in E$, la probabilité $\mathbb{P}(X_{n+1} = y \mid X_n = x)$ ne dépend pas du temps n , tel que $\mathbb{P}(X_n = x) > 0$. (On peut donc la noter $P(x, y)$)

La propriété (i) est appelée la **propriété de Markov**. Elle exprime le fait que la loi de X_{n+1} ne dépend de X_0, \dots, X_n qu'à travers la valeur de X_n : le “présent” (X_n) donne autant d'information sur le “futur” (X_{n+1}) que si l'on connaissait tout le “passé” (X_0, \dots, X_n).

La propriété (ii) exprime l'**homogénéité** en temps du processus : la probabilité de transition de l'état x à l'état y est constante au cours du temps.

La famille $(P(x, y))_{x, y \in E}$ de la définition est appelée **noyau de transition** ou **matrice de transition** : si E est fini, on peut en effet numéroté les états $E = \{1, \dots, r\}$ de telle sorte que $(P(x, y))_{1 \leq x, y \leq r}$ est une matrice carrée. On notera indifféremment $P(x, y)$ ou $P_{x, y}$.

Dans la suite de ce chapitre, sauf mention contraire, $X = (X_n)_{n \geq 0}$ désignera une chaîne de Markov ayant une matrice de transition notée P .

Soit $x \in E$ pour lequel il existe $n \geq 0$ tel que $\mathbb{P}(X_n = x) > 0$. D'après la définition, la ligne x de la matrice P , c'est-à-dire la famille $(P(x, y))_{y \in E}$, donne la loi de X_{n+1} sachant $X_n = x$, d'où

$$P(x, y) = \mathbb{P}(X_{n+1} = y \mid X_n = x) \geq 0, \quad \forall y \in E,$$

et

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

Définition 1.2. Une matrice $P = (p_{x, y})_{x, y \in E}$ vérifiant

$$p_{x, y} \geq 0, \quad \forall x, y \in E \quad \text{et} \quad \sum_{y \in E} p_{x, y} = 1, \quad \forall x \in E$$

s'appelle une matrice **stochastique**.

De manière intuitive la probabilité de transition indique comment la chaîne “tourne”, il reste à dire comment elle “démarré”. C'est le rôle de la loi initiale.

Définition 1.3. La **loi initiale** de la chaîne de Markov $(X_n)_{n \geq 0}$ est la loi de X_0 .

La loi initiale de $(X_n)_{n \geq 0}$ est donc la famille $\nu = (\nu(x))_{x \in E}$ où $\nu(x) = \mathbb{P}(X_0 = x)$, $x \in E$.

Le lemme suivant montre que le noyau de transition et la loi initiale déterminent complètement la loi du processus.

Lemme 1.4. Un processus $(X_n)_{n \geq 0}$ à valeurs dans E est une chaîne de Markov de matrice de transition $P = (P(x, y))_{x, y \in E}$ et de loi initiale $\nu = (\nu(x), x \in E)$ si, et seulement si pour tous $x_0, \dots, x_n \in E$,

$$\mathbb{P}(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) = \nu(x_0)P(x_0, x_1) \cdots P(x_{n-1}, x_n).$$

Preuve : Le sens “chaîne de Markov” \Rightarrow “la formule” s'obtient par récurrence sur n , laissée comme exercice. La réciproque s'en déduit car ces probabilités caractérisent la loi du processus. \square

En raison de cette propriété, on pourra donc parler sans ambiguïté de la loi de la chaîne de Markov de loi initiale ν et de matrice de transition P , sans nécessairement préciser la façon dont ce processus est construit. Il sera constamment utile dans la suite du cours de conserver la même matrice de transition mais de faire varier la loi initiale de la chaîne de Markov :

Définition 1.5. Supposons qu'une matrice de transition $P = (P(x, y))_{x, y \in E}$ est donnée. Pour toute loi ν sur E , on notera \mathbb{P}_ν une probabilité sur Ω telle que, sous \mathbb{P}_ν (c'est-à-dire sur l'espace de probabilités (Ω, \mathbb{P}_ν)), $(X_n)_{n \geq 0}$ est une chaîne de Markov de matrice de transition P et de loi initiale ν .

Cette définition suppose admise la propriété suivante d'existence (pour des détails, voir la section 1.12 de complément sur le processus canonique) :

Proposition 1.6. Soit P une matrice stochastique sur E . Il existe un espace mesurable (Ω, \mathcal{F}) , et un processus $(X_n)_{n \geq 0}$ à valeurs dans E tel que, pour toute probabilité ν sur E , il existe une probabilité \mathbb{P}_ν sur Ω sous laquelle $(X_n)_{n \geq 0}$ est une chaîne de Markov ayant pour loi initiale ν et pour matrice de transition P .

Pour $x \in E$, on notera \mathbb{P}_x au lieu de \mathbb{P}_{δ_x} la loi de la chaîne de Markov de matrice P et de loi initiale δ_x , c'est-à-dire telle que $X_0 = x$ p.s.. On dira que c'est la chaîne de Markov **issue de** x .

Avec ces notations, on vérifie par exemple que conditionner par $X_0 = x$ revient à prendre \mathbb{P}_x :

Lemme 1.7. Soit ν une loi sur E . Pour tout $x \in E$ tel que $\nu(x) > 0$, pour tout événement A dépendant de X_0, X_1, \dots ,

$$\mathbb{P}_\nu(A \mid X_0 = x) = \mathbb{P}_x(A).$$

Pour tout événement A dépendant de X_0, X_1, \dots ,

$$\mathbb{P}_\nu(A) = \sum_{x \in E} \mathbb{P}_x(A) \nu(x),$$

Preuve : Il suffit de traiter le cas où $A = \{X_0 = x_0, \dots, X_n = x_n\}$, pour tous $x_0, \dots, x_n \in E$ (par le Lemme d'unicité 0.6). On a

$$\begin{aligned} \mathbb{P}_\nu(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n \mid X_0 = x) &= \frac{\delta_x(x_0) \mathbb{P}_\nu(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n)}{\mathbb{P}_\nu(X_0 = x)} \\ &= \frac{\delta_x(x_0) \nu(x) P(x, x_1) \cdots P(x_{n-1}, x_n)}{\nu(x)} \\ &= \delta_x(x_0) P(x, x_1) \cdots P(x_{n-1}, x_n) \\ &= \mathbb{P}_x(X_0 = x_0, \dots, X_n = x_n). \end{aligned}$$

Ceci implique la première égalité. La seconde égalité du lemme se déduit alors de la première et de la formule des probabilités totales. \square

Dans les cas simples, on pourra représenter graphiquement les transitions possibles en 1 pas sous la forme d'un graphe orienté.

Définition 1.8. Le **graphe** de la chaîne de Markov $(X_n)_{n \geq 0}$ d'espace d'états E et de matrice de transition $P = (P(x, y))_{x, y \in E}$ est le graphe orienté $\mathcal{G} = (S, A)$ dont l'ensemble des sommets est $S = E$ et l'ensemble des arêtes est

$$A = \{(x, y) \in E \times E \mid P(x, y) > 0\}.$$

1.2 Probabilités de transition à m pas

On a vu que la matrice P décrit les probabilités de transition “à 1 pas”, c’est-à-dire pour passer de X_0 à X_1 , ou plus généralement de X_n à X_{n+1} : la ligne i de P donne la loi de X_1 sachant que $X_0 = i$:

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

Pour étendre ceci au calcul des transitions de X_0 à X_m , on introduit une notion de produit de noyaux :

Définition 1.9. Soit P, Q deux noyaux de transition sur E . Leur **produit** PQ est le noyau de transition sur E défini par :

$$PQ(x, y) = \sum_{z \in E} P(x, z)Q(z, y).$$

Dans le cas où $E = \{1, \dots, r\}$, P et Q sont des matrices carrées et cela correspond au produit matriciel. On définit de même par récurrence $P^n = P^{n-1}P$, avec $P^0 = \text{Id}_E$.

Proposition 1.10. Soit $m \in \mathbb{N}$. Pour tous $x, y \in E$, pour tout $n \geq 0$,

$$\mathbb{P}_x(X_m = y) = P^m(x, y).$$

Ainsi, la loi de X_m sous \mathbb{P}_x est donnée par la ligne x de P^m .

Preuve : On procède par récurrence sur m . Pour $m = 0$, la propriété est triviale. Soit $m \in \mathbb{N}$. On suppose la propriété vraie pour m . Pour tous $x, y \in E$, pour tout $n \in \mathbb{N}$, on a

$$\begin{aligned} \mathbb{P}_x(X_{m+1} = y) &= \sum_{z \in E} \mathbb{P}_x(X_{m+1} = y | X_m = z) \mathbb{P}_x(X_m = z) \\ &= \sum_{z \in E} P(z, y) P^m(x, z) \end{aligned}$$

d’après la définition de P et l’hypothèse de récurrence. Ceci se réécrit $\mathbb{P}_x(X_{m+1} = y) = P^m P(x, y) = P^{m+1}(x, y)$, ce qui conclut. \square

La notion de produit s’étend aux vecteurs comme dans le cas usuel des matrices :

Définition 1.11. Soit P un noyau de transition sur E , ν une probabilité sur E et $f : E \rightarrow \mathbb{R}^+$ une fonction. La probabilité νP est la probabilité définie par

$$\nu P(x) = \sum_{y \in E} \nu(y) P(y, x).$$

La fonction $Pf : E \rightarrow \mathbb{R}^+$ est définie par

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

Le réel νf est défini par

$$\nu f = \sum_{y \in E} \nu(y) f(y) = \int_E f d\nu.$$

Dans le cas où $E = \{1, \dots, r\}$, νP est le produit du *vecteur-ligne* ν par la matrice P , Pf est le produit de la matrice P par le *vecteur-colonne* f , et de même pour νf .

Corollaire 1.12. *Soit $m \in \mathbb{N}$. Pour toute loi ν sur E , la loi de X_m sous \mathbb{P}_ν est νP^m . Ainsi, pour toute fonction $f : E \rightarrow \mathbb{R}_+$,*

$$\mathbb{E}_\nu(f(X_m)) = \nu P^m f,$$

et en particulier pour tout $x \in E$,

$$\mathbb{E}_x(f(X_m)) = P^m f(x).$$

1.3 Exemples de chaînes de Markov

1.3.1 Chaîne de markov à deux états

Soit E un ensemble à deux éléments, notés par exemple $E = \{0, 1\}$: 0 peut représenter un système à l'arrêt et 1 le même système en fonctionnement. Soit $X = (X_n, n \geq 0)$ une chaîne de Markov sur E (ainsi, X_n modélise par exemple l'état de ce système après n jours de fonctionnement). En notant

$$\alpha = \mathbb{P}(X_{n+1} = 1 \mid X_n = 0), \quad \beta = \mathbb{P}(X_{n+1} = 0 \mid X_n = 1),$$

la matrice de transition de X est alors donnée par

$$P = \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix}.$$

Comme exercice, on vérifie (par récurrence en n) le résultat suivant :

Proposition 1.13. *On a*

$$P^n = \frac{1}{\alpha + \beta} \begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix} + \frac{[1 - \alpha - \beta]^n}{\alpha + \beta} \begin{pmatrix} \alpha & -\alpha \\ -\beta & \beta \end{pmatrix}$$

On peut en déduire (exercice) la limite de P^n quand $n \rightarrow \infty$; on notera alors que le cas $\alpha = \beta = 1$ est particulier.

1.3.2 Marche aléatoire sur \mathbb{Z}

Un exemple important est la notion de marche aléatoire. Soit $(X_n, n \geq 1)$ une suite de v.a. i.i.d. de loi commune q sur \mathbb{Z} :

$$\mathbb{P}(X_n = x) = q(x), \quad n \geq 1, x \in \mathbb{Z},$$

et soit x_0 un élément de \mathbb{Z} . On définit la marche aléatoire partant de x_0 comme le processus $S = (S_n, n \geq 0)$ avec :

$$S_n = x_0 + \sum_{i=1}^n X_i, \quad n \geq 0.$$

C'est une chaîne de Markov sur \mathbb{Z} du fait de l'indépendance de la suite des X_n qui représentent les pas de la marche (exercice : vérifier que c'est une chaîne de Markov). On vérifie facilement que sa probabilité de transition est

$$P(a, b) = q(b - a), \quad a, b \in \mathbb{Z}.$$

Un exemple classique est la “marche de l'ivrogne” : pour p fixé, $0 < p < 1$,

$$\mathbb{P}(X_n = 1) = 1 - \mathbb{P}(X_n = -1) = p.$$

(À chaque instant, on fait un pas à droite avec la probabilité p ou un pas à gauche avec la probabilité $1 - p$). On vérifiera facilement que la probabilité de transition s'écrit :

$$P(z, \cdot) = p\delta_{z+1} + (1 - p)\delta_{z-1}, \quad z \in \mathbb{Z}.$$

S_n décrit par exemple la fortune accumulée après n parties d'un jeu de hasard qui rapporte un euro en cas de gain, et fait perdre un euro sinon, en partant d'une fortune de x_0 euros.

1.3.3 Marches aléatoires avec barrières

1.3.4 Modèle d'Ehrenfest

Ce modèle a été introduit par Ehrenfest pour l'expérience suivante : N molécules sont distribuées dans deux récipients A et B . On choisit une molécule au hasard et on la déplace dans l'autre récipient. On répète ceci indéfiniment.

Soit X_n le nombre de molécules dans A au n -ième essai. Alors $(X_n)_n$ est une chaîne de Markov à valeurs dans $E = \{0, \dots, N\}$ dont le noyau de transition est donné par $p_{0,1} = 1 = p_{N,N-1}$ et, pour $1 \leq i \leq N - 1$ et $0 \leq j \leq N$,

$$p_{i,j} = \begin{cases} i/N, & \text{si } j = i - 1; \\ (N - i)/N, & \text{si } j = i + 1; \\ 0, & \text{sinon.} \end{cases}$$

1.3.5 Processus du renouvellement

On examine le fonctionnement d'une machine à temps discrets $n = 0, 1, \dots$. Si la machine tombe en panne, on la remplace immédiatement par une nouvelle. Supposons que les durées de vie des machines (Y_1, Y_2, \dots) sont i.i.d. et de même loi qu'une v.a. Y à valeurs dans \mathbb{N} . On s'intéresse à X_n , l'âge de la machine courante au temps n . Alors en posant

$$Z_k := \sum_{i=1}^k Y_i$$

et

$$\ell_n := \max\{k : Z_k \leq n\},$$

on a

$$X_n = n - Z_{\ell_n}.$$

Proposition 1.14. *Le processus $(X_n)_{n \geq 0}$ est une chaîne de Markov à valeurs dans \mathbb{N} et de probabilité de transition P de la forme*

$$P = \begin{pmatrix} p_{0,0} & p_{0,1} & 0 & 0 & 0 & 0 & \dots \\ p_{1,0} & 0 & p_{1,2} & 0 & 0 & 0 & \dots \\ p_{2,0} & 0 & 0 & p_{2,3} & 0 & 0 & \dots \\ p_{3,0} & 0 & 0 & 0 & p_{3,4} & 0 & \dots \\ \dots & 0 & 0 & 0 & \dots & \dots & \dots \\ \dots & 0 & 0 & \dots & \dots & \dots & \dots \end{pmatrix},$$

avec

$$p_{i,0} = \mathbb{P}(Y = i + 1 \mid Y > i), \quad p_{i,i+1} = \mathbb{P}(Y > i + 1 \mid Y > i).$$

1.3.6 Processus de branchement

Nous étudions l'évolution d'une population. La population originelle s'appelle la 0-ième génération. Les enfants de la $(n - 1)$ -ième génération constituent la n -ième génération pour tout $n \geq 1$.

On note X_n le nombre d'individus de la n -ième génération.

On suppose que le nombre d'enfants de différents individus sont des variables aléatoires indépendantes et ont tous la même loi qu'une v.a. Z à valeurs dans \mathbb{N} .

Proposition 1.15. *Le processus $(X_n, n \geq 0)$ est une chaîne de Markov à valeurs dans $E = \mathbb{N}$ et de noyau de transition $P = (p_{i,j})_{i,j \geq 0}$ donné par*

$$p_{i,j} = \begin{cases} \mathbb{P}(Z_1 + \dots + Z_i = j), & \text{si } i > 0 \text{ et } j \geq 0; \\ 1, & \text{si } i = j = 0; \\ 0, & \text{si } i = 0 \text{ et } j > 0, \end{cases}$$

où $(Z_j)_{j \geq 1}$ sont i.i.d. et de même loi que Z .

On s'intéresse à la probabilité d'extinction :

$$\mathbb{P}(\mathcal{M}), \quad \text{où } \mathcal{M} = \bigcup_{n=0}^{\infty} \{X_n = 0\}.$$

Il est clair que $\{X_n = 0\} \subset \{X_{n+1} = 0\}$, donc ces événements forment une suite croissante et

$$\mathbb{P}(\mathcal{M}) = \lim_{n \rightarrow \infty} \mathbb{P}(X_n = 0).$$

Théorème 1.16. *Supposons que $X_0 = 1$. Alors la probabilité d'extinction est la plus petit nombre $a \geq 0$ tel que*

$$g(a) = a,$$

où $g(a) = \mathbb{E}(a^Z) = \sum_{j=0}^{\infty} a^j \mathbb{P}(Z = j)$ est la fonction génératrice de Z .

Remarquons que si $\mathbb{P}(Z = 0) = 0$, alors $g(0) = 0$ et clairement la probabilité d'extinction est nulle.

Preuve : On établit d'abord

$$g_n = g_{n-1} \circ g.$$

Pour cela, soit $s \in [0, 1]$. On a

$$g_n(s) = \mathbb{E}s^{X_n} = \sum_{i=0}^{\infty} \mathbb{E}(s^{X_n} \mid X_{n-1} = i) \mathbb{P}(X_{n-1} = i),$$

or d'après l'expression des probabilités de transition,

$$\mathbb{E}(s^{X_n} \mid X_{n-1} = i) = \sum_{j \geq 0} p_{ij} s^j = \sum_{j \geq 0} \mathbb{P}(Z_1 + \dots + Z_i = j) s^j = \mathbb{E}(s^{Z_1 + \dots + Z_i}) = (\mathbb{E}s^Z)^i = (g(s))^i,$$

d'où $g_n(s) = g_{n-1}(g(s))$.

Puisque $g_n = g(\dots(g(s)\dots)$ avec n compositions, on a aussi $g_n = g(g_{n-1})$.

Montrons que la probabilité d'extinction $q := \mathbb{P}(\mathcal{M})$ est un point fixe de g : on a

$$q = \lim_{n \rightarrow \infty} \mathbb{P}(X_n = 0) = \lim_{n \rightarrow \infty} g_n(0) = \lim_{n \rightarrow \infty} g(g_{n-1}(0)).$$

Or g est continue, donc on a bien

$$q = g(q).$$

Montrons que q est le plus petit point fixe. Soit $a \in [0, 1]$ avec $g(a) = a$. Comme g est croissante,

$$g(0) \leq g(a) = a,$$

d'où

$$g_2(0) = g(g(0)) \leq g(a) = a,$$

ainsi par récurrence, pour tout $n \geq 1$, $g_n(0) \leq a$ d'où, en passant à la limite, $q \leq a$. \square

Théorème 1.17. *Supposons que $X_0 = 1$ et $\mathbb{P}(Z = 1) < 1$. Alors la probabilité d'extinction vaut 1 si et seulement si $\mathbb{E}(Z) \leq 1$.*

Preuve : On remarque que $\mathbb{E}(Z) = g'(1)$ (dérivée à gauche, $\in [0, +\infty]$). C'est un résultat d'analyse réelle élémentaire que dans ce cas, $g'(1) \leq 1$ si et seulement si 1 est l'unique point fixe de g . \square

1.4 Propriété de Markov

On souhaite étendre la propriété qui définit les chaînes de Markov à des événements qui concernent tout le “futur” ou tout le “passé” du processus.

On commence par une extension directe de la définition :

Proposition 1.18 (Propriété de Markov simple au temps k). *Soit $k, n \geq 1$. Pour tous $x_0, \dots, x_k, x_{k+1}, \dots, x_{k+n} \in E$,*

$$\mathbb{P}_\nu(X_{k+1} = x_{k+1}, \dots, X_{k+n} = x_{k+n} \mid X_0 = x_0, \dots, X_k = x_k) = \mathbb{P}_{x_k}(X_1 = x_{k+1}, \dots, X_n = x_{k+n}).$$

Preuve : Par le Lemme 1.4, la probabilité de gauche s'écrit :

$$\begin{aligned} \frac{\mathbb{P}_\nu(X_0 = x_0, \dots, X_{k+n} = x_{k+n})}{\mathbb{P}_\nu(X_0 = x_0, \dots, X_k = x_k)} &= \frac{\nu(x_0)P(x_0, x_1) \cdots P(x_{k+n-1}, x_{k+n})}{\nu(x_0)P(x_0, x_1) \cdots P(x_{k-1}, x_k)} \\ &= P(x_k, x_{k+1}) \cdots P(x_{k+n-1}, x_{k+n}) \\ &= \mathbb{P}_{x_k}(X_1 = x_{k+1}, \dots, X_n = x_{k+n}), \end{aligned}$$

comme annoncé. \square

On étend ceci à des événements plus généraux :

Proposition 1.19 (Propriété de Markov simple au temps k). *Soit $k, n \geq 1$. Pour tout état $x \in E$, et tous sous-ensembles $A \subset E^{n+1}$, $B \subset E^{k+1}$,*

$$\mathbb{P}_\nu((X_k, \dots, X_{k+n}) \in A \mid (X_0, \dots, X_k) \in B, X_k = x) = \mathbb{P}_x((X_0, \dots, X_n) \in A),$$

dès lors que l'événement par lequel on conditionne n'a pas une probabilité nulle.

Preuve : La proposition précédente s'écrit aussi, en explicitant la probabilité conditionnelle :

$$\mathbb{P}_\nu(X_0 = x_0, \dots, X_{k+n} = x_{k+n}) = \mathbb{P}_\nu(X_0 = x_0, \dots, X_k = x_k) \mathbb{P}_{x_k}(X_1 = x_{k+1}, \dots, X_n = x_{k+n}),$$

et il suffit alors de prendre $x_k = x$ et de sommer cette identité sur tous les $x_0, \dots, x_{k-1}, x_{k+1}, \dots, x_{k+n}$ tels que $(x_0, \dots, x_{k-1}, x) \in B$ et $(x, x_{k+1}, \dots, x_{k+n}) \in A$. \square

La suite de cette partie contient différents énoncés de la propriété de Markov simple au temps k , qui sont simplement différentes formulations de la proposition précédente.

Proposition 1.20. *Soit $k \geq 1$. Pour tout $x \in E$ et tout $B \subset E^{k+1}$, la loi du processus (X_k, X_{k+1}, \dots) sous $\mathbb{P}_\nu(\cdot \mid (X_0, \dots, X_k) \in B, X_k = x)$ est \mathbb{P}_x : c'est une chaîne de Markov de matrice P issue de x .*

On rappelle que les événements de la forme $\{(X_0, \dots, X_k) \in B\}$ forment la tribu $\sigma(X_0, \dots, X_k)$; on prendra l'habitude d'utiliser la notation suivante et, en général, de penser en terme de tribus :

Définition 1.21. *Pour tout $k \geq 0$, la **tribu du passé avant le temps k** est*

$$\mathcal{F}_k = \sigma(X_0, \dots, X_k).$$

La proposition précédente a pour conséquence directe, en terme d'espérance, les formules suivantes :

Proposition 1.22 (Propriété de Markov au temps k). *Soit $k \geq 1$. Pour tout $x \in E$, pour toute fonction $f : E^\mathbb{N} \rightarrow \mathbb{R}$ mesurable positive ou bornée, pour tout événement $H \in \mathcal{F}_k$,*

$$\mathbb{E}_\nu(f(X_k, X_{k+1}, \dots) \mid H \cap \{X_k = x\}) = \mathbb{E}_x(f(X_0, X_1, \dots)).$$

Plus généralement,

$$\mathbb{E}_\nu(f(X_k, X_{k+1}, \dots) \mathbf{1}_H) = \mathbb{E}_\nu(F(X_k) \mathbf{1}_H)$$

où la fonction $F : E \rightarrow \mathbb{R}$ est définie par

$$F(x) = \mathbb{E}_x(F(X_0, X_1, \dots)), \quad x \in E.$$

Plus généralement, pour toute fonction $g : E^k \rightarrow \mathbb{R}$ positive ou bornée, avec la même notation,

$$\mathbb{E}_\nu(f(X_k, X_{k+1}, \dots)g(X_0, \dots, X_k)) = \mathbb{E}_\nu(F(X_k)g(X_0, \dots, X_k)).$$

Preuve : La première égalité vient directement de la dernière proposition : la loi de (X_k, X_{k+1}, \dots) sous $\mathbb{P}_\nu(\cdot | H \cap \{X_k = x\})$ est \mathbb{P}_x , donc l'espérance de f sous ces deux lois est la même. En explicitant l'espérance conditionnelle (rappelons que $\mathbb{E}(Z | A) = \mathbb{E}(Z \mathbf{1}_A) / \mathbb{P}(A)$), cette égalité devient :

$$\begin{aligned} \mathbb{E}_\nu(f(X_k, X_{k+1}, \dots) \mathbf{1}_H \mathbf{1}_{\{X_k=x\}}) &= \mathbb{E}_x(f(X_0, X_1, \dots)) \mathbb{E}_\nu(\mathbf{1}_H \mathbf{1}_{\{X_k=x\}}) \\ &= F(x) \mathbb{E}_\nu(\mathbf{1}_H \mathbf{1}_{\{X_k=x\}}) = \mathbb{E}_\nu(F(x) \mathbf{1}_H \mathbf{1}_{\{X_k=x\}}) \\ &= \mathbb{E}_\nu(F(X_k) \mathbf{1}_H \mathbf{1}_{\{X_k=x\}}). \end{aligned}$$

En sommant cette égalité sur toutes les valeurs $x \in E$, on obtient la seconde égalité de l'énoncé.

Ceci prouve la dernière égalité lorsque $g(X_1, \dots, X_k) = \mathbf{1}_H$. On en déduit le cas général de façon classique : par linéarité on obtient le cas où g est étagée, puis par convergence monotone ou dominée le cas où g est positive ou bornée. \square

On généralise alors facilement le Lemme 1.4 au cas de transitions à plusieurs pas :

Lemme 1.23. Pour tous états $x_0, \dots, x_r \in E$ et tous temps $0 \leq n_1 \leq \dots \leq n_r$,

$$\mathbb{P}_{x_0}(X_{n_1} = x_1, \dots, X_{n_r} = x_r) = P^{n_1}(x_0, x_1) P^{n_2 - n_1}(x_1, x_2) \dots P^{n_r - n_{r-1}}(x_{r-1}, x_r),$$

et plus généralement, pour toute loi ν sur E ,

$$\mathbb{P}_\nu(X_{n_1} = x_1, \dots, X_{n_r} = x_r) = (\nu P^{n_1})(x_1) P^{n_2 - n_1}(x_1, x_2) \dots P^{n_r - n_{r-1}}(x_{r-1}, x_r).$$

Preuve : On procède par récurrence sur $r \geq 1$. Pour $r = 1$, c'est la Proposition 1.10. Supposons le résultat vrai pour r . Soit $x_0, \dots, x_{r+1} \in E$ et $0 \leq n_1 \leq \dots \leq n_{r+1}$. On a, d'après la propriété de Markov simple au temps n_1 ,

$$\begin{aligned} \mathbb{P}_{x_0}(X_{n_1} = x_1, \dots, X_{n_{r+1}} = x_{r+1}) &= \mathbb{P}_{x_0}(X_{n_1} = x_1) \mathbb{P}_{x_1}(X_{n_2 - n_1} = x_2, \dots, X_{n_{r+1} - n_r} = x_{r+1}) \\ &= P^{n_1}(x_0, x_1) \mathbb{P}_{x_1}(X_{n_2 - n_1} = x_2, \dots, X_{n_{r+1} - n_r} = x_{r+1}) \end{aligned}$$

(la seconde égalité vient de la Proposition 1.10), ce qui conclut, par l'hypothèse de récurrence. Le cas d'une loi initiale ν s'en déduit en décomposant selon la valeur de X_0 (ou via le Lemme 1.7). \square

Quitte à construire la chaîne de Markov sur $\Omega = E^{\mathbb{N}}$ (cf. Appendice), muni de la tribu produit, on peut toujours supposer qu'il existe une application $\theta : \Omega \rightarrow \Omega$ mesurable telle que, pour tout $n \in \mathbb{N}$,

$$X_n \circ \theta = X_{n+1}.$$

Cette application s'appelle le **décalage** (ou "shift" en anglais). On posera $\theta_0 = \text{Id}$, $\theta_1 = \theta$, \dots , $\theta_{n+1} = \theta_n \circ \theta$. Alors $X_n \circ \theta_k = X_{n+k}$, pour tous $n, k \geq 0$. Avec ces notations, la propriété de Markov simple s'écrit comme suit :

Proposition 1.24 (Propriété de Markov simple au temps k). *Pour tout $k \geq 1$ et pour toute v.a. réelle Z qui est $\sigma(X)$ -mesurable, positive ou bornée, et pour toute v.a. réelle W qui est \mathcal{F}_k -mesurable, pour toute loi initiale ν , on a*

$$\mathbb{E}_\nu(Z \circ \theta_k W) = \mathbb{E}_\nu(F(X_k)W),$$

où

$$F(x) := \mathbb{E}_x(Z).$$

Preuve : Toute v.a. Z qui est $\sigma(X)$ -mesurable s'écrit en effet $Z = f(X) = f(X_0, X_1, \dots)$ et toute v.a. W qui est \mathcal{F}_k -mesurable est de la forme $W = g(X_0, \dots, X_k)$. Comme $Z \circ \theta_k = f(X \circ \theta_k) = f(X_k, X_{k+1}, \dots)$, cette proposition ré-exprime la dernière formule de la Proposition 1.22. \square

1.5 Visites à un sous-ensemble

Soit $X = (X_n, n \geq 0)$ une chaîne de Markov de probabilité de transition P sur l'espace d'états E . On introduit de nouvelles variables aléatoires, fonctions de X , relatives aux visites à un sous-ensemble A de E .

Définition 1.25. Soit $A \subset E$. Le **premier temps d'entrée** de la chaîne dans A est la v.a.

$$T_A := \begin{cases} \inf\{n \geq 0 : X_n \in A\}, & \text{si } \{\dots\} \neq \emptyset; \\ \infty, & \text{sinon.} \end{cases}$$

Le **premier temps de retour** de la chaîne dans A est la v.a.

$$\tau_A := \begin{cases} \inf\{n \geq 1 : X_n \in A\}, & \text{si } \{\dots\} \neq \emptyset; \\ \infty, & \text{sinon.} \end{cases}$$

Le **nombre de visites** de la chaîne à A est la v.a.

$$N_A = \sum_{n=0}^{\infty} \mathbf{1}_{(X_n \in A)} \in \mathbb{N} \cup \{\infty\}.$$

Lorsque $A = \{x\}$ est un singleton, on note simplement T_x , τ_x et N_x . Remarquons que si $y \neq x$, $\tau_y = T_y$, \mathbb{P}_x -p.s.

Le théorème de convergence monotone pour les séries donne immédiatement la formule suivante pour tous $x, y \in E$:

$$\mathbb{E}_x(N_y) = \sum_{n=0}^{\infty} P^n(x, y).$$

Donnons une application de la propriété de Markov :

Proposition 1.26. Pour tout couple (x, y) d'éléments de E , on a

$$\mathbb{E}_x(N_y) = \mathbb{P}_x(T_y < \infty) \mathbb{E}_y(N_y).$$

Preuve : Pour tout entier $k \geq 0$,

$$\begin{aligned}
\mathbb{P}_x(N_y = k) &= \sum_{n=0}^{\infty} \mathbb{P}_x\left(T_y = n, \sum_{j=0}^{\infty} \mathbf{1}_{(X_{n+j}=y)} = k\right) \\
&= \sum_{n=0}^{\infty} \mathbb{P}_x\left(X_0 \neq y, \dots, X_{n-1} \neq y, X_n = y, \sum_{j=0}^{\infty} \mathbf{1}_{(X_{n+j}=y)} = k\right) \\
&= \sum_{n=0}^{\infty} \mathbb{P}_x(T_y = n) \mathbb{P}_y\left(\sum_{j=0}^{\infty} \mathbf{1}_{(X_j=y)} = k\right) \\
&= \sum_{n=0}^{\infty} \mathbb{P}_x(T_y = n) \mathbb{P}_y(N_y = k),
\end{aligned}$$

où la troisième égalité est due à la propriété de Markov au temps n . Nous avons donc

$$\mathbb{P}_x(N_y = k) = \mathbb{P}_x(T_y < \infty) \mathbb{P}_y(N_y = k),$$

d'où la proposition en sommant $\sum_{k \geq 0} k \mathbb{P}_x(N_y = k)$, et en remarquant que si $\mathbb{P}_x(N_y = \infty) > 0$ alors $\mathbb{P}_y(N_y = \infty) > 0$. \square

Donnons une autre démonstration, qui utilise la propriété de Markov sous forme d'espérance :

Preuve : On a

$$\mathbb{E}_x(N_y) = \sum_{n=0}^{\infty} \mathbb{E}_x(N_y \mathbf{1}_{\{T_y=n\}}),$$

et on note que si $T_y = n$ alors $N_y = N_y \circ \theta_n$ (si la première visite en y a lieu au temps n , alors le nombre total de visites en y est le même que le nombre de visites en y à partir du temps n). Ainsi, pour tout $n \geq 0$, par la propriété de Markov au temps n , vu que $\mathbf{1}_{\{T_y=n\}}$ est \mathcal{F}_n -mesurable,

$$\mathbb{E}_x(N_y \mathbf{1}_{\{T_y=n\}}) = \mathbb{E}_x(N_y \circ \theta_n \mathbf{1}_{\{T_y=n\}}) = \mathbb{E}_x(F(X_n) \mathbf{1}_{\{T_y=n\}}) = \mathbb{E}_x(F(y) \mathbf{1}_{\{T_y=n\}}),$$

où $F(x) = \mathbb{E}_x(N_y)$. En sommant sur $n \in \mathbb{N}$, on obtient

$$\mathbb{E}_x(N_y \mathbf{1}_{\{T_y < \infty\}}) = \mathbb{E}_y(N_y) \mathbb{P}_x(T_y < \infty),$$

et le premier terme est égal à $\mathbb{E}_x(N_y)$ car sur l'événement $\{T_y = \infty\}$ on a $N_y = 0$. \square

Définition 1.27. La **fonction de Green** de la chaîne de Markov X est la fonction $G : E \times E \rightarrow \mathbb{R} \cup \{\infty\}$ donnée par : pour tous $x, y \in E$,

$$G(x, y) = \mathbb{E}_x(N_y) = \sum_{n=0}^{+\infty} P^n(x, y) = \sum_{n=0}^{+\infty} \mathbb{E}_x(\mathbf{1}_y(X_n)) = \mathbb{E}_x\left(\sum_{n=0}^{+\infty} \mathbf{1}_y(X_n)\right)$$

Avec cette notation, la proposition précédente s'écrit $G(x, y) = \mathbb{P}_x(T_y < \infty) G(y, y)$, d'où en particulier

$$G(x, y) \leq G(y, y).$$

1.6 Récurrence et transience

Nous allons étudier le comportement des trajectoires d'une chaîne de Markov X_n lorsque $n \rightarrow +\infty$. On va ici essentiellement distinguer deux types de comportement suivant que la chaîne "tend vers l'infini" ou non.

Définition 1.28. Un état x est **récurrent** si $\mathbb{P}_x(\tau_x < +\infty) = 1$. Dans le cas contraire, x est dit **transient**.

Théorème 1.29. Si pour la chaîne X l'état x est transient, le nombre de visites N_x suit, sous la probabilité \mathbb{P}_x , la loi géométrique de paramètre

$$a = 1 - \mathbb{P}_x(\tau_x < \infty) = \mathbb{P}_x(\tau_x = \infty),$$

c'est-à-dire que, pour tout $k \geq 1$,

$$\mathbb{P}_x(N_x = k) = a(1 - a)^{k-1}, \quad k \geq 1.$$

En particulier, $\mathbb{P}_x(N_x < \infty) = 1$ et $G(x, x) = \mathbb{E}_x(N_x) = 1/a$.

Preuve : Soit x transient. Sous \mathbb{P}_x , $X_0 = x$ et $\sum_{j=0}^{\tau_x-1} \mathbf{1}_{(X_j=x)} = 1$. Il vient, pour tout $k \geq 1$,

$$\mathbb{P}_x(N_x = k + 1) = \sum_{n=1}^{\infty} \mathbb{P}_x(\tau_x = n, N_x \circ \theta_n = k),$$

où $N_x \circ \theta_n = \sum_{j=0}^{\infty} \mathbf{1}_{(X_{j+n}=x)}$ est le nombre de visite en x par la chaîne après le temps n . D'après la propriété de Markov au temps n (Proposition 1.24), on a $\mathbb{P}_x(\tau_x = n, N_x \circ \theta_n = k) = \mathbb{E}_x(\mathbf{1}_{\{\tau_x=n\}} \mathbf{1}_{\{N_x \circ \theta_n=k\}}) = \mathbb{E}_x(\mathbf{1}_{\{\tau_x=n\}} \mathbb{E}_{X_n}(\mathbf{1}_{\{N_x=k\}})) = \mathbb{P}_x(\tau_x = n) \mathbb{P}_x(N_x = k)$ d'où

$$\mathbb{P}_x(N_x = k + 1) = \sum_{n=1}^{\infty} \mathbb{P}_x(\tau_x = n) \mathbb{P}_x(N_x = k) = \mathbb{P}_x(\tau_x < \infty) \mathbb{P}_x(N_x = k).$$

Par récurrence en k , on obtient

$$\mathbb{P}_x(N_x = k + 1) = \mathbb{P}_x(\tau_x < \infty)^k \mathbb{P}_x(N_x = 1).$$

Ce qui prouve le théorème puisque $\mathbb{P}_x(N_x = 1) = \mathbb{P}_x(\tau_x = \infty) = a$. □

Par la Proposition 1.26, on en déduit que si x est transient alors le nombre de visites N_x est fini presque sûrement, quel que soit le point de départ : $\mathbb{P}_y(N_x < \infty) = 1$ pour tout $y \in E$.

Théorème 1.30. Les assertions suivantes sont équivalentes :

- (1) x est récurrent.
- (2) $N_x = +\infty$, \mathbb{P}_x -p.s.
- (3) $G(x, x) = +\infty$.

Preuve :

- $1 \implies 2$: On a montré dans la preuve du Théorème 1.29 que la propriété de Markov donne, pour tout $k \in \mathbb{N}$, $\mathbb{P}_x(N_x = k+1) = \mathbb{P}_x(\tau_x < \infty) \mathbb{P}_x(N_x = k)$. Si x est récurrent on a donc, pour tout $k \in \mathbb{N}$, $\mathbb{P}_x(N_x = k) = \mathbb{P}_x(N_x = k+1)$. Comme $\sum_{k \in \mathbb{N}} \mathbb{P}_x(N_x = k) = \mathbb{P}_x(N_x < \infty) < \infty$ et que tous les termes de la somme sont égaux entre eux, ils sont nécessairement tous nuls et donc $\mathbb{P}_x(N_x < \infty) = 0$.
- $2 \implies 3$: trivial.
- $3 \implies 1$: théorème précédent.

□

Corollaire 1.31. *Si l'espace d'états E est fini, alors il existe au moins un état récurrent.*

Preuve : Prenons un état initial z quelconque. On a, \mathbb{P}_z -p.s., $\infty = \sum_{x \in E} N_x$ (le temps total passé dans E est infini), mais $N_x < \infty$ \mathbb{P}_z -p.s. pour tout x transient, d'où la conclusion. □

Remarquons que, si x est récurrent, on peut parfois avoir $N_x < \infty$ si le point de départ n'est pas x ; cela dépend des classes de communication de x et du point de départ.

1.7 Décomposition en classes de communication

Notation : $P^n(x, y)$ désigne un élément de la matrice de transition P^n . Attention à ne pas confondre avec $(P(x, y))^n$.

Définition 1.32. *Un état x est dit **absorbant** si $P(x, x) = 1$.*

Bien sûr, un état absorbant est récurrent.

Définition 1.33. *On dit que l'état y est **accessible à partir de l'état** x , noté par $x \rightarrow y$, s'il existe $n > 0$ tel que $P^n(x, y) > 0$. On dit que les deux états x et y **communiquent** si $x \rightarrow y$ et $y \rightarrow x$, noté par $x \leftrightarrow y$.*

Proposition 1.34 (Transitivité). *Si $x \rightarrow y$ et $y \rightarrow z$, alors $x \rightarrow z$.*

Preuve : Par hypothèses, il existe $n, m \geq 1$ tels que $P^n(x, y) > 0$ et $P^m(y, z) > 0$. Or

$$P^{n+m}(x, z) = \sum_{w \in E} P^n(x, w) P^m(w, z) \geq P^n(x, y) P^m(y, z) > 0,$$

d'où le résultat. □

On a donc

Corollaire 1.35. *La relation \leftrightarrow est une relation symétrique et transitive.*

Étant donné un état x , on note $C(x)$ l'ensemble des états qui communiquent avec x . Il est possible que $C(x) = \emptyset$; dans ce cas, $x \not\leftrightarrow x$, et x est appelé un **état de non-retour**. Remarquons que tout état de non-retour est clairement transient. On note E_{nr} l'ensemble des états de non-retour, et $E_r = E \setminus E_{nr}$. Alors sur E_r , la relation \leftrightarrow est une relation d'équivalence. Ses classes d'équivalence

sont appelées les **classes de communication** (ou plus simplement les **classes**) de la chaîne de Markov : ce sont donc les ensembles $\mathcal{C}_1, \dots, \mathcal{C}_r$ (où $r \in \mathbb{N}^* \cup \{\infty\}$) tels que E est partitionné en

$$E = E_{nr} \cup \bigcup_{k=1}^r \mathcal{C}_k,$$

et pour $i = 1, \dots, r$, tous les états de \mathcal{C}_i communiquent entre eux, et ne communiquent avec aucun autre : pour tout $i \in \{1, \dots, r\}$ et $x, y \in \mathcal{C}_i$, on a $x \rightarrow y$, et pour tous $i, j \in \{1, \dots, r\}$ distincts, et tous $x \in \mathcal{C}_i, y \in \mathcal{C}_j$, on a $x \not\rightarrow y$ ou $y \not\rightarrow x$.

Définition 1.36. Une chaîne de Markov est **irréductible** si tous ses états communiquent, autrement dit s'il n'y a pas d'état de non-retour et qu'il n'y a qu'une classe de communication.

Définition 1.37. Un ensemble $\mathcal{C} \subset E$ est dit **fermé** si aucun état $y \notin \mathcal{C}$ n'est accessible à partir d'un état quelconque $x \in \mathcal{C}$, autrement dit si

$$P^n(x, y) = 0, \quad \forall n \geq 1, x \in \mathcal{C}, y \notin \mathcal{C}.$$

Ainsi, si $\mathcal{C} \subset E$ est fermé alors pour tout $x \in \mathcal{C}$, \mathbb{P}_x -p.s., pour tout $n \geq 1$, $X_n \in \mathcal{C}$ et donc, sous \mathbb{P}_x , $(X_n)_{n \geq 0}$ est une chaîne de Markov d'espace d'états \mathcal{C} et de matrice de transition $(P(x, y))_{x, y \in \mathcal{C}}$. On l'appelle la chaîne de Markov **restreinte à \mathcal{C}** .

Naturellement, les transitions en 1 pas suffisent à déterminer si un ensemble d'états est fermé :

Proposition 1.38. Un ensemble $\mathcal{C} \subset E$ est fermé si, et seulement si

$$P(x, y) = 0, \quad \forall x \in \mathcal{C}, y \notin \mathcal{C}.$$

Preuve : La partie “seulement si” est immédiate. Pour la partie “si”, on procède par récurrence en n . □

Proposition 1.39. Toute partie non vide $\mathcal{C} \subset E$ fermée et finie contient au moins un état récurrent.

Preuve : Cela vient du corollaire 1.31 appliqué à la chaîne de Markov restreinte à \mathcal{C} . □

Proposition 1.40. Si x est récurrent et que $x \rightarrow y$, alors $y \rightarrow x$ et y est aussi récurrent.

Preuve : Nous montrons par l'absurde que $y \rightarrow x$. Supposons que $y \not\rightarrow x$, i.e. pour tout $n \geq 1$, $P^n(y, x) = 0$. Comme $x \rightarrow y$, il existe un $n \geq 1$ tel que $P^n(x, y) = \mathbb{P}_x(X_n = y) > 0$. Remarquons que

$$\left\{ \sum_{l=n+1}^{\infty} \mathbf{1}_{(X_l=x)} = 0 \right\} \subset \{N_x \leq n+1\}.$$

D'après la propriété de Markov, on a

$$\begin{aligned}
\mathbb{P}_x(N_x \leq n+1 \mid X_n = y) &\geq \frac{1}{\mathbb{P}_x(X_n = y)} \mathbb{P}_x\left(X_n = y, \sum_{l=n+1}^{\infty} \mathbf{1}_{(X_l=x)} = 0\right) \\
&= \frac{1}{\mathbb{P}_x(X_n = y)} \mathbb{P}_x(X_n = y) \mathbb{P}_y\left(\sum_{l=1}^{\infty} \mathbf{1}_{(X_l=x)} = 0\right) \quad (\text{Markov en } n) \\
&= \mathbb{P}_y\left(\sum_{l=1}^{\infty} \mathbf{1}_{(X_l=x)} = 0\right) = 1,
\end{aligned}$$

la dernière égalité venant du fait que $\sum_{l=1}^{\infty} P^l(y, x) = 0$, i.e. $\mathbb{E}_y(\sum_{l=1}^{\infty} \mathbf{1}_{(X_l=x)}) = 0$.

On obtient alors $\mathbb{P}_x(N_x < \infty) \geq \mathbb{P}_x(N_x \leq n) \geq \mathbb{P}_x(N_x \leq n \mid X_n = y) \mathbb{P}_x(X_n = y) = \mathbb{P}_x(X_n = y) > 0$, ce qui contredit la récurrence de x . Donc $y \rightarrow x$.

Par conséquent, $x \leftrightarrow y$. Soient $j, l \geq 1$ tels que $P^j(x, y) > 0$ et $P^l(y, x) > 0$. On a pour tout $n \geq 0$,

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

En sommant par rapport à n , on obtient

$$\begin{aligned}
G(y, y) &\geq \sum_{m=n+j+l}^{\infty} P^m(y, y) \\
&\geq \sum_{n=0}^{\infty} P^l(y, x) P^n(x, x) P^j(x, y) \\
&= P^l(y, x) P^j(x, y) G(x, x) = \infty,
\end{aligned}$$

d'après la récurrence de x . D'après le théorème 1.30, y est aussi récurrent. □

Corollaire 1.41.

- Deux états qui communiquent ont la même nature (tous récurrents ou tous transients).
 \leadsto On dira qu'une classe de communication est **récurrente** (resp. **transiente**) si tous ses états le sont. Toute classe est donc ou bien récurrente, ou bien transiente.
- Une classe récurrente est fermée. Par conséquent, une classe non fermée est transiente.
- Une classe fermée et finie est récurrente.

Remarque. Ce résultat permet donc de déterminer la nature de tous les états d'une chaîne de Markov sur un espace d'états *fini* à partir de son graphe (ou de sa matrice de transition). En revanche elle ne permet pas toujours de conclure quand l'espace d'états est infini, car savoir qu'une classe de communication infinie est fermée ne suffit pas à connaître sa nature.

1.8 Exemples

1.8.1 Chaîne à deux états

Si $0 < p_0 \leq 1$ et si $0 < p_1 \leq 1$, la chaîne est irréductible récurrente.

Si $p_0 = 0$ et $0 < p_1 \leq 1$, $\mathcal{T} = \{1\}$ et $\mathcal{C} = \{0\}$.

Si $p_0 = p_1 = 0$, $\mathcal{C}_1 = \{0\}$ et $\mathcal{C}_2 = \{1\}$.

1.8.2 Chaîne sur \mathbb{N}

On vérifie, en calculant $\mathbb{P}_0(\tau_0 = n) = (1 - p_{n-1}) \prod_{i=0}^{n-2} p_i$, que

- la chaîne est transiente si $\prod_{n=0}^{\infty} p_n > 0$;
- la chaîne est récurrente si $\prod_{n=0}^{\infty} p_n = 0$.

1.8.3 Marche aléatoire sur \mathbb{Z}

Comme $0 < p < 1$, tous les états communiquent. Il suffit donc de regarder par exemple la nature de l'état 0. On a

$$\mathbb{P}_0(X_n = 0) = \begin{cases} 0, & \text{si } n \text{ est impair;} \\ C_n^{n/2} p^{n/2} (1-p)^{n/2}, & \text{si } n \text{ est pair.} \end{cases}$$

À l'aide de la formule de Stirling (c'est-à-dire $n! \sim (n/e)^n \sqrt{2\pi n}$), on obtient :

- la marche est récurrente si $p = 1/2$;
- la marche est transiente si $p \neq 1/2$.

1.8.4 Marche aléatoire simple symétrique sur \mathbb{Z}^d

Un calcul direct de $\mathbb{P}_0(X_n = 0)$ montre que la marche est récurrente quand $d = 2$ et transiente quand $d \geq 3$: pour n pair, on a en effet

$$\mathbb{P}_0(X_n = 0) = (2d)^{-n} \sum_{n_1 + \dots + n_d = \frac{n}{2}} \frac{n!}{(n_1! \dots n_d!)^2} \sim 2 \left(\frac{d}{2n\pi} \right)^{d/2}.$$

1.9 Probabilités d'absorption

Notons \mathcal{T} l'ensemble des états transients, et $\mathcal{C}_1, \dots, \mathcal{C}_N$ les classes de récurrence (avec $N \in \mathbb{N} \cup \{\infty\}$). On a vu que ce sont des classes fermées : une fois entrée dans l'une d'elle, la chaîne n'en sort plus. Pour $i = 1, \dots, N$, le temps d'atteinte $T_{\mathcal{C}_i}$ est ainsi appelé **temps d'absorption par la classe \mathcal{C}_i** , et pour tout état $x \in E$ on définit la **probabilité d'absorption par la classe \mathcal{C}_i en partant de x** par

$$q_i(x) = \mathbb{P}_x(T_{\mathcal{C}_i} < \infty).$$

Remarquons que $q_i(x) = 0$ si $x \in \mathcal{C}_j$ où $j \neq i$, et $q_i(x) = 1$ si $x \in \mathcal{C}_i$.

Proposition 1.42. Soit $i \in \{1, \dots, N\}$. On a :

$$\text{pour tout } x \in \mathcal{T}, \quad q_i(x) = \sum_{y \in \mathcal{C}_i} P(x, y) + \sum_{z \in \mathcal{T}} P(x, z) q_i(z). \quad (1.2)$$

De plus, si \mathcal{T} est fini, les relations (1.2) caractérisent les probabilités $(q_i(x))_{x \in \mathcal{T}}$.

Preuve : Soit $i \in \{1, \dots, N\}$. Pour simplifier on note $T_i = T_{\mathcal{C}_i} = \inf\{n \geq 1 : X_n \in \mathcal{C}_i\}$ le temps d'entrée dans la classe \mathcal{C}_i . Soit $x \in \mathcal{T}$. On remarque que, sous \mathbb{P}_x , on a $T_i > 0$, et que $T_i = T_i \circ \theta_1$. Par la propriété de Markov au temps 1, on a donc :

$$\begin{aligned} \mathbb{P}_x(T_i < \infty) &= \sum_{z \in E} \mathbb{P}_x(X_1 = z, T_i < \infty) = \sum_{z \in E} \mathbb{P}_x(X_1 = z, T_i \circ \theta_1 < \infty) \\ &= \sum_{z \in E} \mathbb{P}_x(X_1 = z) \mathbb{P}_z(T_i < \infty) = \sum_{z \in E} P(x, z) q_i(z). \end{aligned}$$

En écrivant $E = \mathcal{T} \cup \mathcal{C}_i \cup \bigcup_{j \neq i} \mathcal{C}_j$, et vu que $q_i(z) = 1$ si $z \in \mathcal{C}_i$ et que $q_i(z) = 0$ si $z \in \mathcal{C}_j$ avec $j \neq i$, on obtient (1.2). Remarquons que la première somme de (1.2) est égale à $\mathbb{P}_x(X_1 \in \mathcal{C}_i) = \mathbb{P}_x(T_i = 1)$.

Supposons que \mathcal{T} est fini et qu'une fonction $f : \mathcal{T} \rightarrow [0, 1]$ vérifie (1.2) pour tout $x \in \mathcal{T}$. Alors par itération,

$$\begin{aligned} f(x) &= \mathbb{P}_x(T_i = 1) + \sum_{z \in \mathcal{T}} P(x, z) \left(\mathbb{P}_z(T_i = 1) + \sum_{z' \in \mathcal{T}} P(z, z') f(z') \right) \\ &= \mathbb{P}_x(T_i \leq 2) + \sum_{y \in \mathcal{T}} P^2(x, y) f(y). \end{aligned}$$

Par récurrence, on obtient que pour tout n ,

$$f(x) = \mathbb{P}_x(T_i \leq n) + \sum_{y \in \mathcal{T}} P^n(x, y) f(y).$$

Or pour tout état transient y , d'après le lemme suivant, $\lim_{n \rightarrow \infty} P^n(x, y) = 0$. Ce qui entraîne, comme \mathcal{T} est fini, que $f(x) = \lim_{n \rightarrow \infty} \mathbb{P}_x(T_i \leq n) = \mathbb{P}_x(T_i < \infty) = q_i(x)$. \square

Lemme 1.43. Si $y \in E$ est transient on a, pour tout $x \in E$,

$$\lim_{n \rightarrow \infty} P^n(x, y) = 0.$$

Preuve : Par la proposition 1.26, $G(x, y) \leq G(y, y) < \infty$, or $G(x, y) = \mathbb{E}_x(N_y) = \sum_n P^n(x, y)$, d'où le résultat. \square

La proposition suivante raccourcit souvent le calcul des probabilités d'absorption :

Proposition 1.44. Si \mathcal{T} est fini alors, pour tout $x \in \mathcal{T}$,

$$\sum_{i=1}^N q_i(x) = 1.$$

Preuve : On a en effet, pour tout $n \geq 0$,

$$1 = \mathbb{P}_x(X_n \in E) = \mathbb{P}_x(X_n \in \mathcal{T}) + \sum_{i=1}^N \mathbb{P}_x(X_n \in \mathcal{C}_i) = \sum_{y \in \mathcal{T}} P^n(x, y) + \sum_{i=1}^N \mathbb{P}_x(T_i \leq n)$$

d'où, en passant à la limite quand $n \rightarrow \infty$, à l'aide du lemme précédent, et du fait que \mathcal{T} est fini,

$$1 = \sum_{y \in \mathcal{T}} 0 + \sum_{i=1}^N \mathbb{P}_x(T_i < \infty),$$

ce qui conclut. □

1.10 Mesures et Probabilités invariantes

Dans la théorie des chaînes et des processus de Markov, la notion de probabilité invariante est essentielle. Elle correspond à celle d'état d'équilibre en physique.

Définition 1.45. On dit qu'une mesure (resp. une loi) m sur E est une **mesure invariante** (resp. une **loi invariante**) de la chaîne de Markov de noyau de transition P si $mP = m$, i.e.

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

Définition 1.46. Un processus $(X_n, n \geq 0)$ est dit **stationnaire** si, pour tout $k \in \mathbb{N}$, les deux processus (X_0, X_1, \dots) et (X_k, X_{k+1}, \dots) ont la même loi.

Théorème 1.47. Une chaîne de Markov $(X_n, n \geq 0)$ est stationnaire si et seulement si la loi initiale est invariante.

Preuve : On note m la loi initiale de X et P sa matrice de transition. D'après le Corollaire 1.12, mP est la loi de X_1 . Donc si X est stationnaire, X_1 a la même loi que X_0 et par conséquent, $mP = m$, m est une loi invariante.

D'autre part, si $mP = m$, on a par itération $mP^k = m$ pour tout k . D'après le Corollaire 1.12 et la Proposition 1.20, pour tout k , (X_0, X_1, \dots) et (X_k, X_{k+1}, \dots) ont donc la même loi. □

Théorème 1.48. Soit x un état récurrent fixé de la chaîne de Markov. La formule

$$m(y) = \mathbb{E}_x \left(\sum_{k=0}^{\tau_x-1} \mathbf{1}_y(X_k) \right), \quad y \in E,$$

définit une mesure invariante. De plus, pour tout $y \in E$, $m(y) < \infty$.

Remarques.

- On a $m(x) = 1$.
- Comme une classe récurrente est fermée, la mesure m définie ci-dessus est portée par la classe contenant x .

- La masse totale $m(E)$ de la mesure m est égale à l'espérance du temps de retour en x :

$$m(E) = \sum_{y \in E} m(y) = \mathbb{E}_x(\tau_x) \quad (\in]0, \infty]).$$

- Pour toute fonction $f : E \rightarrow \mathbb{R}_+$, on a

$$\int f \, dm = \mathbb{E}_x \left(\sum_{k=0}^{\tau_x-1} f(X_k) \right).$$

En effet, la définition exprime cette égalité quand $f = \mathbf{1}_y$, et le cas général s'en déduit par linéarité (ou convergence monotone) car $f = \sum_y f(y) \mathbf{1}_y$.

Preuve : Sous \mathbb{P}_x , $X_0 = X_{\tau_x} = x$. On a donc, pour toute fonction $f : E \rightarrow \mathbb{R}^+$

$$\begin{aligned} \int f \, dm &= \mathbb{E}_x \left(\sum_{k=0}^{\tau_x-1} f(X_k) \right) = \mathbb{E}_x \left(\sum_{k=0}^{\tau_x-1} f(X_{k+1}) \right) = \mathbb{E}_x \left(\sum_{k=0}^{\infty} \mathbf{1}_{\{\tau_x > k\}} f(X_{k+1}) \right) \\ &= \sum_{k=0}^{+\infty} \mathbb{E}_x (\mathbf{1}_{\{\tau_x > k\}} f(X_{k+1})) = \sum_{k=0}^{\infty} \mathbb{E}_x (\mathbf{1}_{\{\tau_x > k\}} F(X_k)) \end{aligned}$$

par la propriété de Markov au temps k , avec, pour $y \in E$, $F(y) = \mathbb{E}_y(f(X_1)) = Pf(y)$. On poursuit :

$$\int f \, dm = \mathbb{E}_x \left(\sum_{k=0}^{\infty} \mathbf{1}_{\{\tau_x > k\}} F(X_k) \right) = \mathbb{E}_x \left(\sum_{k=0}^{\tau_x-1} F(X_k) \right) = \int F \, dm.$$

Comme $\int f \, dm = \sum_{x \in E} m(x) f(x) = mf$ (m vecteur-ligne, f vecteur-colonne, cf. Définition 1.11), on obtient, pour toute $f : E \rightarrow \mathbb{R}_+$, $mf = mF = mPf$. Par suite, $m = mP$: m est invariante.

Il reste cependant à s'assurer que $m(y) < \infty$ pour tout $y \in E$. C'est évident pour $y = x$ car $m(x) = 1$, et pour y hors de la classe de x car alors $m(y) = 0$. Enfin, si $x \leftrightarrow y$, il existe n tel que $P^n(x, y) > 0$ et l'équation $m(x) = mP^n(x)$ s'écrit

$$1 = \sum_{z \in E} m(z) P^n(x, z) \geq m(y) P^n(x, y),$$

ce qui implique $m(y) < \infty$. □

Si x est récurrent, le théorème suivant montre que la mesure m relie asymptotiquement le temps passé en chaque état y au temps passé en x :

Théorème 1.49 (Théorème ergodique). *Soit x un état récurrent. Pour tout état y , \mathbb{P}_x -presque sûrement,*

$$\frac{\sum_{k=0}^{n-1} \mathbf{1}_y(X_k)}{\sum_{k=0}^{n-1} \mathbf{1}_x(X_k)} \xrightarrow{n \rightarrow \infty} m(y),$$

où m est la mesure (dépendant de x) définie par le théorème précédent. Plus généralement, pour toutes fonctions $f, g : E \rightarrow \mathbb{R}^+$, \mathbb{P}_x -presque sûrement,

$$\lim_{n \rightarrow +\infty} \frac{\sum_{k=0}^{n-1} f(X_k)}{\sum_{k=0}^{n-1} g(X_k)} = \frac{\int f \, dm}{\int g \, dm}$$

si $\int g \, dm$ est fini et non nul.

La démonstration de ce théorème repose sur l'utilisation de la propriété de Markov “forte” et sera présentée comme complément dans la section 2.7 (Théorème 2.19).

Nous étudions l'unicité des mesures invariantes. Le lemme qui suit nous permet de se ramener à étudier une chaîne irréductible, quitte à restreindre à une classe fermée.

Lemme 1.50. *Soit \mathcal{C} une classe de communication fermée. Soit m une mesure invariante pour la restriction $(P(x, y))_{x, y \in \mathcal{C}}$ de P à \mathcal{C} . Si on étend m sur E par $\tilde{m}(x) := 0$ si $x \notin \mathcal{C}$, alors \tilde{m} est une mesure invariante pour P .*

Preuve : Si $x \in \mathcal{C}$, comme m est invariante on a

$$\tilde{m}P(x) = \sum_{y \in E} \tilde{m}(y)P(y, x) = \sum_{y \in \mathcal{C}} m(y)P(y, x) = m(x) = \tilde{m}(x).$$

Si $x \notin \mathcal{C}$, comme \mathcal{C} est fermée on a $P(y, x) = 0$ si $y \in \mathcal{C}$ et donc

$$\tilde{m}P(x) = \sum_{y \in E} \tilde{m}(y)P(y, x) = \sum_{y \in \mathcal{C}} \tilde{m}(y)P(y, x) = 0 = \tilde{m}(x)$$

□

Lemme 1.51. *Si m est une mesure invariante d'une chaîne irréductible, et $m \not\equiv 0$, alors $m(y) > 0$ pour tout $y \in E$.*

Preuve : Il existe au moins un état x tel que $m(x) \neq 0$. Soit $y \in E$. Il existe n tel que $P^n(x, y) > 0$, et on a alors

$$m(y) = mP^n(y) = \sum_{z \in E} m(z)P^n(z, y) \geq m(x)P^n(x, y) > 0.$$

□

Lemme 1.52. (Unicité) *Si une chaîne de Markov irréductible possède une probabilité invariante π , toute autre mesure invariante est proportionnelle à π . En particulier, π est la seule probabilité invariante.*

Preuve : On suppose que π est une probabilité invariante. Soit m une mesure invariante et x un état fixé de E . Si $m \equiv 0$, le lemme est vrai. Supposons $m \not\equiv 0$, d'où $m(y) > 0$ pour tout $y \in E$ par le lemme précédent. Posons $\lambda = \frac{\pi(x)}{m(x)}$ et $m' = \lambda m$. Remarquons que m' est invariante et

$m'(x) = \pi(x)$. Il faut montrer que $m' = \pi$. Définissons une mesure μ sur E , en posant, pour tout $y \in E$, $\mu(y) = \min(m'(y), \pi(y))$. Alors

$$(\mu P)(y) = \sum_{z \in E} \mu(z) P(z, y) \leq \sum_{z \in E} m'(z) P(z, y) = m'(y)$$

car m' est une mesure invariante. On voit de la même façon que $(\mu P)(y) \leq \pi(y)$. On a donc

$$(\mu P)(y) \leq \min(m'(y), \pi(y)) = \mu(y).$$

Mais on a aussi

$$\sum_{y \in E} \mu P(y) = \sum_{y \in E} \sum_{z \in E} \mu(z) P(z, y) = \sum_{z \in E} \mu(z) \sum_{y \in E} P(z, y) = \sum_{z \in E} \mu(z)$$

et ces sommes sont finies car $\sum_y \mu(y) \leq \sum_y \pi(y) = 1$, ce qui permet de calculer leur différence, et

$$\sum_{y \in E} (\mu(y) - \mu P(y)) = 0.$$

Les termes de la somme étant positifs ou nuls, ils sont donc tous nuls : $\mu P(y) = \mu(y)$ pour tout $y \in E$. Ainsi, μ est invariante. Comme π et m' aussi sont invariantes, $m_1 = \pi - \mu (\geq 0)$ et $m_2 = m' - \mu (\geq 0)$ sont également des mesures invariantes. Or $m'(x) = \pi(x) = \mu(x)$ (grâce au choix de λ), d'où $m_1(x) = m_2(x) = 0$, si bien que le lemme précédent implique que $m_1 = m_2 \equiv 0$, d'où $\pi = \mu = m'$. \square

Théorème 1.53. *Considérons une chaîne de Markov (X_n) irréductible. Les conditions suivantes sont équivalentes :*

- (i) *il existe un état $x \in E$ tel que $\mathbb{E}_x(\tau_x) < +\infty$;*
- (ii) *pour tout état $x \in E$, $\mathbb{E}_x(\tau_x) < +\infty$;*
- (iii) *la chaîne de Markov possède une probabilité invariante π .*

*Sous ces conditions la chaîne est dite **récurrente positive**. π est alors la seule mesure invariante (à une constante multiplicative près) et la seule probabilité invariante. De plus, pour tout $y \in E$,*

$$\pi(y) = \frac{1}{\mathbb{E}_y(\tau_y)}$$

et, pour tous $x, y \in E$,

$$\pi(y) = \frac{1}{\mathbb{E}_x(\tau_x)} \mathbb{E}_x \left(\sum_{n=0}^{\tau_x-1} \mathbf{1}_y(X_n) \right).$$

Preuve : Montrons d'abord que (i) implique (iii). Supposons donc (i). En particulier $\tau_x < \infty$, \mathbb{P}_x -p.s., donc x est récurrent. On note m la mesure invariante associée à l'état récurrent x définie au Théorème 1.48. Se rappelant que $m(E) = \mathbb{E}_x(\tau_x)$, on définit alors une probabilité invariante par

$$\pi = \frac{m}{\mathbb{E}_x(\tau_x)}, \tag{1.3}$$

ce qui prouve (iii).

Montrons que (iii) entraîne (ii). Il résulte du principe du maximum (cf. Proposition 1.26) que, si π est une probabilité invariante, pour tout état x ,

$$\sum_{y \in E} \pi(y)G(y, x) \leq \sum_{y \in E} \pi(y)G(x, x) = G(x, x).$$

Or,

$$\sum_{y \in E} \pi(y)G(y, x) = \sum_{n=0}^{\infty} \sum_{y \in E} \pi(y)P^n(y, x) = \sum_{n=0}^{\infty} \pi(x) = +\infty$$

car on sait que $\pi(x) > 0$ (Lemme 1.51). Donc $G(x, x) = +\infty$ et l'état x est récurrent. Puisque la mesure m associée à x est une mesure invariante, il résulte du lemme 1.52 que m et π sont proportionnelles, donc que $m(E)$ est fini. Or $m(E) = \mathbb{E}_x(\tau_x)$, donc $\mathbb{E}_x(\tau_x)$ est fini, ce qui montre (ii). Puisque (ii) entraîne (i) de façon évidente, ceci prouve les équivalences du théorème.

L'unicité de π résulte du lemme 1.52. La deuxième formule donnant π est la formule (1.3). Cette formule appliquée à x donne

$$\pi(x) = \frac{\mathbb{E}_x(\sum_{n=0}^{\tau_x-1} \mathbf{1}_x(X_n))}{\mathbb{E}_x(\tau_x)} = \frac{1}{\mathbb{E}_x(\tau_x)}.$$

(Par l'unicité, π ne dépend pas du choix de x dans (1.3)) □

Remarque : On a vu que quand la chaîne entre dans une classe récurrente, elle n'en sort plus. On peut donc appliquer le théorème précédent à toute classe récurrente. Les conditions du théorème peuvent alors être satisfaites ou non selon la classe, d'où les définitions suivantes.

Définition 1.54. *Un état récurrent x est **récurrent positif** si $\mathbb{E}_x(\tau_x) < \infty$, et **récurrent nul** sinon. On étend ces définitions à toute une classe de communication.*

Corollaire 1.55. *Lorsque E est fini, toute chaîne irréductible est récurrente positive.*

Preuve : L'existence d'un état récurrent résulte du Corollaire 1.31. On sait qu'à cet état récurrent, on peut associer une mesure invariante. Elle est de masse finie car l'espace est fini. □

Pour les chaînes récurrentes positives, le théorème ergodique prend la forme d'une loi des grands nombres :

Théorème 1.56 (Théorème ergodique des chaînes de Markov récurrentes positives). *Si (X_n) est une chaîne de Markov irréductible récurrente positive de probabilité invariante π , alors pour toute loi initiale ν , pour tout état $y \in E$, \mathbb{P}_ν -presque sûrement,*

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{1}_y(X_k) = \pi(y),$$

et plus généralement pour toute fonction $f : E \rightarrow \mathbb{R}^+$, \mathbb{P}_ν -presque sûrement,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(X_k) = \int f d\pi. \quad (*)$$

et pour tous $x, y \in E$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} P^k(x, y) = \pi(y).$$

Preuve : Pour tout état x , il résulte du Théorème 1.49, appliqué à $g = 1$, que $(*)$ est vraie \mathbb{P}_x -presque sûrement (se souvenir que π ne dépend pas de x). C'est encore vrai \mathbb{P}_ν -p.s. car

$$\mathbb{P}_\nu(\cdot) = \sum_{x \in E} \nu(x) \mathbb{P}_x(\cdot).$$

En prenant $f = \mathbf{1}_y$, on obtient la première limite, et en prenant de plus $\nu = \delta_x$ le théorème de convergence dominée (domination par 1) donne la dernière limite. \square

Ce théorème donne deux moyens pratiques d'approcher π si, comme c'est souvent le cas, on ne sait pas la calculer explicitement. La première façon est la méthode de Monte Carlo, qui consiste à simuler sur ordinateur une longue trajectoire $X_n(\omega)$ de la chaîne, puis de faire la moyenne de f le long de cette trajectoire. D'après $(*)$ on obtient ainsi à peu près $\int f d\pi$. L'autre façon est de calculer itérativement P^n , par exemple dans le cas où E est fini. Puis de faire la moyenne des $P^n(x, y)$ pour approcher $\pi(y)$ (on peut montrer que la vitesse de convergence est exponentielle). C'est très souvent beaucoup plus rapide que la recherche d'un vecteur propre de la transposée de P .

1.11 Périodicité et existence de la loi limite au sens fort

Nous avons vu que pour une matrice de transition irréductible et récurrente positive, la limite

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n P^k(x, y) = \pi(y)$$

existe et définit une loi invariante (et l'unique loi invariante). Parfois, le résultat plus fort suivant est valable :

$$\lim_{n \rightarrow \infty} P^n(x, y) = \pi(y),$$

cependant ce n'est pas toujours le cas. Cela dépend de la notion de périodicité :

Définition 1.57. Si x est un état tel que $x \rightarrow x$, la **période** de x est

$$d(x) = \text{pgcd}(\{n \in \mathbb{N}^* \mid P^n(x, x) > 0\}).$$

Lemme 1.58. Pour tous $x, y \in E$, si $x \leftrightarrow y$ alors $d(x) = d(y)$. Autrement dit, la période est un nombre qui ne dépend que de la classe à laquelle appartient l'état considéré.

Preuve : Soit $x, y \in E$ tels que $x \leftrightarrow y$. Il existe $m, l > 0$ tels que $P^m(x, y) > 0$ et $P^l(y, x) > 0$. On a alors

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

donc $d(x)$ divise $m + l$. Pour tout n tel que $P^n(y, y) > 0$, on a

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

donc $d(x)$ divise $m + n + l$, ce qui implique, avec ce qui précède, que $d(x)$ divise n . Par définition de $d(y)$, on a donc $d(x) \leq d(y)$. Par symétrie on a aussi $d(y) \leq d(x)$, donc $d(x) = d(y)$. \square

Définition 1.59. Une chaîne de Markov irréductible (resp. une classe) est **apériodique** si la période commune à tous les états de E (resp. à toute la classe) est 1, **périodique** sinon.

Proposition 1.60. Soit C une classe récurrente. On note d la période de cette classe. Fixons $x \in C$. Pour tout $y \in C$, il existe un entier $\nu_y \in \{0, \dots, d-1\}$ tel que

- a) $P^n(x, y) > 0 \implies n \equiv \nu_y \pmod{d}$.
- b) il existe n_y tel que si $n \geq n_y$ alors $P^{nd+\nu_y}(x, y) > 0$.

Preuve : a) Soit $y \in C$. Choisissons r et m tels que $P^r(y, x) > 0$ et $P^m(x, y) > 0$. Si $P^n(x, y) > 0$ alors on a

$$P^{m+r}(x, x) > 0 \quad \text{et} \quad P^{n+r}(x, x) > 0,$$

d'où, par la définition de période, $d|m+r$ et $d|n+r$, et donc $d|m-n$ et $n \equiv m \pmod{d}$. L'entier ν_y est le reste de la division euclidienne de m par d .

b) Soit $(n_l)_{l \geq 1}$ la famille des entiers positifs tels que $P^{n_l}(x, x) > 0$. Pour tout $s \geq 1$, posons $d_s := \text{pgcd}(n_1, \dots, n_s)$. La suite d'entiers positifs $(d_s)_{s \geq 1}$ est décroissante donc stationne en une valeur $d_t : d_s = d_t$ pour tout $s \geq t$. Or d_t divise tous les n_l donc divise aussi leur pgcd d : on a $d_t|d$. D'autre part, $d|n_l$ pour $l = 1, \dots, t$ donc $d|d_t$. Ainsi, $d = d_t = \text{pgcd}(n_1, \dots, n_t)$.

Par le lemme qui suit, il existe donc un entier N tel que, pour tout $n \geq N$, on a $nd = \alpha_1 n_1 + \dots + \alpha_t n_t$ pour certains entiers naturels $\alpha_1, \dots, \alpha_t$ et donc

$$P^{nd}(x, x) \geq \prod_{l=1}^t (P^{n_l}(x, x))^{\alpha_l} > 0.$$

Soit $y \in C$. D'après a) et le fait que $x \rightarrow y$, il existe $m \geq 0$ tel que $P^{md+\nu_y}(x, y) > 0$. Posons $n_y = N + m$. Pour tout $n \geq n_y$, on a $n = n' + m$ avec $n' \geq N$, et donc $nd + \nu_y = n'd + md + \nu_y$, par conséquent

$$P^{nd+\nu_y}(x, y) \geq P^{n'd}(x, x) P^{md+\nu_y}(x, y) > 0,$$

d'où b). \square

Lemme 1.61. Soient $n_1, \dots, n_t \in \mathbb{N}$. On pose $d = \text{pgcd}(n_1, \dots, n_t)$. Alors il existe un entier N tel que pour tout $n \geq N$, on a

$$nd = \alpha_1 n_1 + \dots + \alpha_t n_t$$

pour certains entiers $\alpha_1, \dots, \alpha_t \in \mathbb{N}$.

Le problème de Frobenius en théorie de nombres consiste à déterminer le plus petit nombre $N = N(n_1, \dots, n_d)$ vérifiant une telle propriété.

Preuve : En divisant tous les n_l par d , on réduit le problème au cas $d = 1$. D'après le lemme de Bézout, il existe une combinaison linéaire $\beta_1 n_1 + \dots + \beta_t n_t = 1$, avec des coefficients $\beta_l \in \mathbb{Z}$.

En rassemblant séparément les termes positifs et négatifs, on obtient alors $p - q = 1$ où p et q sont des sommes de termes positifs. Posons $N = q^2 - 1$. Alors si $n \geq N$, la division de n par q donne

$$n = \alpha q + \beta,$$

où $0 \leq \beta < q$, d'où $\alpha \geq q - 1 \geq \beta$, et l'écriture suivante prouve le lemme :

$$n = \alpha q + \beta(p - q) = (\alpha - \beta)q + \beta p.$$

□

Supposons $x \in C$ fixé. Pour $\nu \in \{0, \dots, d-1\}$, on définit

$$C^{(\nu)} \equiv C^{(\nu)}(x) := \left\{ j \in C \mid \nu_j \text{ de la proposition 1.60 a) vaut } \nu \right\}.$$

$C^{(0)}, \dots, C^{(d-1)}$ sont appelées les **sous-classes cycliques** de C . Nous étendons la définition de $C^{(\nu)}$ à tout entier $\nu \geq 0$ en posant $C^{(s)} := C^{(\nu)}$ si $s \equiv \nu \pmod{d}$.

Proposition 1.62. *Soit C une classe récurrente de période d .*

- a) $C^{(0)}, \dots, C^{(d-1)}$ sont deux à deux disjointes et leur réunion donne C .
- b) si $j \in C^{(\nu)}$ et $P_{jk}^{(n)} > 0$ ($n > 0$), alors $k \in C^{(\nu+n)}$. Donc

$$\sum_{k \in C^{(\nu+n)}} P^n(j, k) = 1.$$

Preuve : a) est une conséquence de la proposition 1.60. Quant à b), considérons un $m > 0$ tel que $P^m(i, j) > 0$, on a

$$P^{m+n}(i, k) \geq P^m(i, j)P^n(j, k) > 0.$$

Ceci en vue de la proposition précédente (b) implique que

$$m + n \equiv \nu_k \pmod{d},$$

donc $k \in C^{(m+n)} = C^{(\nu+n)}$ puisque $m = \nu \pmod{d}$. □

Cette proposition justifie l'attribut "cyclique", car elle met en évidence l'évolution cyclique de la chaîne partant d'un état de C : si la chaîne part d'un état j d'une sous-classe, elle atteint à nouveau les états de cette sous-classe aux temps $d, 2d, 3d, \dots$

D'un état j on passe en un pas à un état de la sous-classe suivante.

Il est intéressant d'observer que si la chaîne part de $i \in C^{(0)}$, alors $(X_{nd})_{n \geq 0}$ est une chaîne de Markov ayant $C^{(0)}$ pour ensemble d'états et $(P^d(i, j))_{ij \in C^{(0)}}$ comme matrice de transition. (*exercice*). Remarquons que $C^{(0)}$ est une classe fermée pour P^d et elle est irréductible et de période 1. En effet, $P^{nd}(i, i) > 0$ pour tout $n \geq N$.

Nous allons montrer que les sous-classes $C^{(0)}(i), \dots, C^{(d-1)}(i)$ ne vont pas beaucoup dépendre de i . Pour cela, soit $j \in C^{(\nu)}(i)$, on considère les sous-classes associées $C^{(0)}(j), \dots, C^{(d-1)}(j)$. Soit $k \in C^{(s)}(j)$. Soient m et n tels que

$$P^m(i, j) > 0, \quad P^n(j, k) > 0.$$

On a

$$m = \nu \bmod d, \quad n = s \bmod d.$$

D'autre part,

$$P^{m+n}(i, k) \geq P^m(i, j)P^n(j, k) > 0,$$

donc $k \in C^{(m+n)}(i) = C^{(\nu+s)}(i)$, car $m + n = \nu + s \bmod d$.

Il s'ensuit que $C^{(s)}(j) \subset C^{(\nu+s)}(i)$. Mais les sous-classes font une partition de C , donc

$$C^{(s)}(j) = C^{(\nu+s)}(i), \quad \forall s = 0, \dots, d-1.$$

Ainsi les sous-classes sont les mêmes à un décalage d'indices près.

Exemple : Soit $E = \{0, 1, 2, 3, 4\}$ et

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

Supposons que la chaîne part de $0 \in C^{(0)}$. En un seul pas, elle peut aller à 2 ou 3 qui vont appartenir forcément à $C^{(1)}$. ensuite on a à 1 donc $1 \in C^{(2)}$. ensuite de 1 on va 0, 4 $\in C^{(3)}$, ceci implique $C^{(3)} = C^{(0)}$. La période vaut donc 3 et $C^{(0)} = \{0, 4\}$, $C^{(1)} = \{2, 3\}$ et $C^{(2)} = \{1\}$.

Remarque : Si la chaîne est apériodique, on a $P^n(i, j) > 0$ pour tout n suffisamment grand.

Théorème 1.63. *Considérons une chaîne de Markov X de matrice de transition P irréductible et récurrente positive. On note π la loi invariante.*

(1) *(Convergence en loi pour la chaîne de Markov) Si la matrice est apériodique alors pour toute loi initiale ν ,*

$$\lim_{n \rightarrow \infty} \mathbb{P}_\nu(X_n = j) = \pi_j, \quad \forall j \in E.$$

En particulier, on a

$$\lim_{n \rightarrow \infty} P^n(i, j) = \pi_j, \quad \forall i, j.$$

(2) *Si la matrice est périodique et de période d , alors pour tout couple i, j , il existe un entier $a \in [0, d-1]$ tel que $P^m(i, j) = 0$ si $m \not\equiv a \bmod d$, et*

$$\lim_{n \rightarrow \infty} P^{nd+a}(i, j) = \frac{\pi_j}{\pi(C^{(a)})} = d\pi_j,$$

où $\pi(C^{(a)}) = \sum_{j \in C^{(a)}} \pi_j = \frac{1}{d}$.

On rappelle que si la chaîne irréductible est apériodique, $\forall i, j \in E$, on a $P^n(i, j) > 0$ pour tout n suffisamment grand.

Preuve : (1) On utilise l'argument de couplage : Soit Y une autre chaîne de Markov de noyau de transition P , de loi initiale π , indépendante de X . Une telle chaîne Y existe : pour construire le couple (X, Y) , il suffit de se placer sur l'espace produit $E^{\mathbb{N}} \times E^{\mathbb{N}}$ muni de la probabilité produit $\mathbb{P}_\nu \times \pi_\pi$. Soit $W_n = (X_n, Y_n)$ la chaîne couplée. W est une chaîne de Markov de loi initiale $\nu \times \pi$ et de

matrice de transition $\tilde{P}((i, k), (j, l)) = P(i, j)P(k, l)$. La chaîne W est irréductible et apériodique, car pour tout n , $\tilde{P}^n((i, k), (j, l)) = P^n(i, j)P^n(k, l)$ qui est strictement positif pour tout n assez grand (c'est ici on a utilisé l'apériodicité de P , sinon on ne peut pas garantir que $P^n(i, j)P^n(k, l) > 0$ pour tout couple (i, k) et (j, l) , donc W ne serait même pas irréductible).

Il est immédiat de vérifier que la loi $\tilde{\pi}(i, l) := \pi_i \pi_l$ définie sur $E \times E$ est une loi invariante pour \tilde{P} .

Alors W est une chaîne récurrente positive irréductible et apériodique. Le temps d'atteinte T de la diagonale de $E \times E$ par W :

$$T := \inf\{n \geq 1 : X_n = Y_n\},$$

est $\tilde{\mathbb{P}}_{\nu \times \pi}$ -p.s. fini car T est plus petit que le premier temps d'atteinte de n'importe quel (i, i) par W , qui est presque sûrement fini. On définit maintenant une nouvelle chaîne (Z_n) par

$$Z_n(\omega) := \begin{cases} X_n(\omega), & \text{sur } \{T(\omega) > n\}; \\ Y_n(\omega), & \text{sur } \{T(\omega) \leq n\}. \end{cases}$$

On remarque que $Z_T = X_T = Y_T$ et que $Z_0 = X_0$ a pour loi initiale ν . On vérifie maintenant que sous $\tilde{\mathbb{P}}_{\nu \times \pi}$, Z est une chaîne de Markov de noyau de transition P ; En fait,

$$\tilde{\mathbb{P}}_{\nu \times \pi}(Z_{n+1} = j | Z_n = i_n, \dots, Z_0 = i_0) = \frac{\tilde{\mathbb{P}}_{\nu \times \pi}(Z_{n+1} = j, Z_n = i_n, \dots, Z_0 = i_0)}{\tilde{\mathbb{P}}_{\nu \times \pi}(Z_n = i_n, \dots, Z_0 = i_0)}.$$

Pour simplifier les notations, soit $A_k := \{Z_0 = i_0, \dots, Z_k = i_k\}$ pour tout $1 \leq k \leq n$. Alors

$$\begin{aligned} \tilde{\mathbb{P}}_{\nu \times \pi}(Z_{n+1} = j, A_n) &= \tilde{\mathbb{P}}_{\nu \times \pi}(Z_{n+1} = j, T > n, A_n) + \tilde{\mathbb{P}}_{\nu \times \pi}(Z_{n+1} = j, T = n, A_n) \\ &\quad + \sum_{k=0}^{n-1} \tilde{\mathbb{P}}_{\nu \times \pi}(Z_{n+1} = j, T = k, A_n). \end{aligned} \quad (1.4)$$

On discute ces trois cas séparément. Sur $\{T > n\}$, $Z_{n+1} = X_{n+1}$ et $A_n = \{X_0 = i_0, \dots, X_n = i_n\}$. D'après la propriété de Markov de X en n (X et Y sont indépendants),

$$\tilde{\mathbb{P}}_{\nu \times \pi}(Z_{n+1} = j, T > n, A_n) = P(i, j) \tilde{\mathbb{P}}_{\nu \times \pi}(T > n, X_n = i_n, \dots, X_0 = i_0) = P(i, j) \tilde{\mathbb{P}}_{\nu \times \pi}(T > n, A_n).$$

Sur $\{T = n\}$, $Z_{n+1} = Y_{n+1}$, $Z_n = Y_n$ et $A_n = \{Y_n = i_n, X_{n-1} = i_{n-1}, \dots, X_0 = i_0\}$. Donc la propriété de Markov de Y en n entraîne que

$$\begin{aligned} \tilde{\mathbb{P}}_{\nu \times \pi}(Z_{n+1} = j, T = n, A_n) &= P(i, j) \tilde{\mathbb{P}}_{\nu \times \pi}(T = n, Y_n = i_n, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) \\ &= P(i, j) \tilde{\mathbb{P}}_{\nu \times \pi}(T = n, A_n). \end{aligned}$$

Finalement sur $\{T = k\}$ avec $k \leq n-1$, $\{Z_{n+1} = j, T = k, A_n\} = \{Y_{n+1} = j, Y_n = i_n, \dots, Y_k = i_k, T = k, X_{k-1} = i_{k-1}, \dots, X_0 = i_0\}$. On applique la propriété de Markov en n et obtient

$$\begin{aligned} \tilde{\mathbb{P}}_{\nu \times \pi}(Z_{n+1} = j, T = k, A_n) &= P(i, j) \tilde{\mathbb{P}}_{\nu \times \pi}(Y_n = i_n, \dots, Y_k = i_k, T = k, X_{k-1} = i_{k-1}, \dots, X_0 = i_0) \\ &= P(i, j) \tilde{\mathbb{P}}_{\nu \times \pi}(T = k, A_n), \end{aligned}$$

donc $\tilde{\mathbb{P}}_{\nu \times \pi}(Z_{n+1} = j | Z_n = i_n, \dots, Z_0 = i_0) = P(i, j)$ en sommant ces trois cas dans (1.4).

On voit que X et Z ont la même loi initiale et même noyau de transition donc ont la même loi. Donc

$$\mathbb{P}_\nu(X_n = j) - \pi_j = \tilde{\mathbb{P}}_{\nu \times \pi}(X_n = j) - \tilde{\mathbb{P}}_{\nu \times \pi}(Y_n = j) = \tilde{\mathbb{P}}_{\nu \times \pi}(Z_n = j) - \tilde{\mathbb{P}}_{\nu \times \pi}(Y_n = j).$$

Remarquer que sur $\{T \leq n\}$, $Z_n = Y_n$. Donc $\tilde{\mathbb{P}}_{\nu \times \pi}(Z_n = j) - \tilde{\mathbb{P}}_{\nu \times \pi}(Y_n = j) = \tilde{\mathbb{P}}_{\nu \times \pi}(Z_n = j, n < T) - \tilde{\mathbb{P}}_{\nu \times \pi}(Y_n = j, n < T)$ et

$$|\mathbb{P}_\nu(X_n = j) - \pi_j| \leq |\tilde{\mathbb{P}}_{\nu \times \pi}(Z_n = j, n < T) - \tilde{\mathbb{P}}_{\nu \times \pi}(Y_n = j, n < T)| \leq \tilde{\mathbb{P}}_{\nu \times \pi}(n < T) \rightarrow 0,$$

car T est presque sûrement fini.

(2) On vérifie d'abord que $\pi(C^{(a)}) = \frac{1}{d}$. Comme $\pi = \pi P$ et que pour $j \in C^{(0)}$, $\pi_j = \sum_{i \in E} \pi_i P_{i,j} = \sum_{i \in C^{(d-1)}} \pi_i P_{i,j}$, donc

$$\pi(C^{(0)}) = \sum_{j \in C^{(0)}} \sum_{i \in C^{(d-1)}} \pi_i P_{i,j} = \sum_{i \in C^{(d-1)}} \pi_i \sum_{j \in C^{(0)}} P_{i,j} = \sum_{i \in C^{(d-1)}} \pi_i$$

car $\sum_{j \in C^{(0)}} P_{i,j} = 1$ pour tout $i \in C^{(d-1)}$; donc $\pi(C^{(0)}) = \pi(C^{(d-1)})$. En considérant $\pi = \pi P^2 = \dots = \pi P^{d-1}$, on obtient que $\pi(C^{(0)}) = \pi(C^{(1)}) = \dots = \pi(C^{(d-1)}) = \frac{1}{d}$ puisque les sommes de $\pi(C^{(1)}), \dots, \pi(C^{(d)})$ vaut 1.

On traite par exemple le cas $a = 0$. La matrice P^d apériodique et pour tout $i \in C^{(0)}$, $P_{i,j}^d = 0$ si $j \notin C^{(0)}$. Alors la mesure $\hat{\pi}_j := \frac{\pi_j}{\pi(C^{(0)})}$, $j \in C^{(0)}$ est une loi sur $C^{(0)}$, de plus, elle est invariante pour P^d . Alors appliquer (1) on obtient la limite annoncée. \square

$$\frac{1}{n} \sum_{k=1}^n \mathbf{1}_{(\mu(Y_{k+1})Q(Y_{k+1}, X_k) > \mu(X_k)Q(X_k, Y_{k+1}))}.$$

1.12 Complément sur le processus canonique

Soit P une matrice stochastique, sur un espace d'états E , et μ une loi sur E . On a jusque-là admis (cf. Proposition 1.6) l'existence d'une chaîne de Markov de matrice P et de loi initiale μ . Justifions-la ici.

1.12.1 Première construction

Quitte à numéroté les états, E étant dénombrable on peut supposer $E \subset \mathbb{N}$, et même $E = \mathbb{N}$ quitte à ajouter des états où la mesure μ est nulle. Remarquons que, si U est une variable aléatoire de loi uniforme sur $[0, 1]$, alors la variable aléatoire $X = f_\mu(U)$ suit la loi μ , en définissant la fonction

$$f_\mu : u \mapsto \min\{n \in \mathbb{N} \mid \mu(0) + \mu(1) + \dots + \mu(n) \geq U\}.$$

En effet, pour tout $n \in \mathbb{N}$, $f_\mu(U) = n$ si, et seulement si $U \in]\mu(0) + \dots + \mu(n-1), \mu(0) + \dots + \mu(n)]$, et cet intervalle a pour largeur $\mu(n)$.

Supposons donnée une suite $(U_n)_{n \geq 0}$ de variables aléatoires indépendantes et de loi uniforme sur $[0, 1]$. On discutera de l'existence (non évidente) d'une telle suite plus bas.

Alors on peut définir par récurrence la suite $(X_n)_{n \geq 0}$ de variables aléatoires par : $X_0 = f_\mu(U_0)$ et, pour tout $n \geq 0$,

$$X_{n+1} = f_{P(X_n, \cdot)}(U_{n+1}),$$

où $P(X_n, \cdot)$ est la mesure de probabilité donnée par la ligne X_n de la matrice P . Par cette construction, X_0 suit la loi μ et, pour tout n , la loi de X_{n+1} sachant $\{X_0 = x_0, \dots, X_n = x_n\}$ est la loi de $f_{P(x_n, \cdot)}(U_{n+1})$ (car U_{n+1} est indépendante de X_0, \dots, X_n), c'est-à-dire $P(x_n, \cdot)$. Le processus $(X_n)_{n \geq 0}$ est donc une chaîne de Markov de loi initiale μ et de matrice de transition P . Notons \mathbb{P}_μ sa loi : c'est une mesure de probabilité sur l'espace $E^\mathbb{N}$ des suites à valeurs dans E , muni de sa tribu produit (voir rappels). L'existence de \mathbb{P}_μ étant maintenant acquise, on va remplacer X par une version plus "standard".

Considérons l'espace **canonique** $\Omega = E^\mathbb{N}$ des suites à valeurs dans E , muni de sa tribu produit \mathcal{F} . Le **processus canonique** sur E est défini par $X = \text{Id}_\Omega$: c'est la fonction identité $X = (X_n)_{n \geq 0} : \Omega \rightarrow E^\mathbb{N}$. Pour toute probabilité μ sur E , sous \mathbb{P}_μ , ce processus X est une chaîne de Markov de loi initiale μ et de matrice P .

Avec l'espace canonique $\Omega = E^\mathbb{N}$, on travaille donc toujours avec le même processus $X : E^\mathbb{N} \rightarrow E^\mathbb{N}$ (l'identité), mais sous différentes lois \mathbb{P}_μ . De plus, on peut définir l'**opérateur de décalage** (ou **shift**) $\theta : \Omega \rightarrow \Omega$ par

$$\theta((x_n)_{n \geq 0}) = (x_{n+1})_{n \geq 0}, \quad \text{pour tout } (x_n)_{n \geq 0} \in \Omega,$$

qui vérifie $X_n \circ \theta = X_{n+1}$.

Justifions enfin l'existence de suites $(U_n)_{n \geq 0}$ de variables aléatoires indépendantes et de loi uniforme sur $[0, 1]$. Prenons $\Omega = [0, 1]$, muni de la loi uniforme, et $U : [0, 1] \rightarrow [0, 1]$ définie par $U(\omega) = \omega$, de telle sorte que U suit la loi uniforme sur $[0, 1]$. On rappelle que, si $U = \overline{0, X_0 X_1 \dots}$ est le développement de U en base 2 (avec $X_0, X_1, \dots \in \{0, 1\}$), alors les variables $(X_n)_{n \geq 0}$ sont indépendantes et de même loi $\mathcal{B}(1/2)$ (et vice-versa, si $(X_n)_n$ est ainsi, U suit la loi uniforme sur $[0, 1]$). Considérons une bijection $\varphi : \mathbb{N}^2 \rightarrow \mathbb{N}$ (on pourrait définir φ explicitement). Alors, en définissant, pour tout $n \geq 0$, la variable aléatoire U_n par

$$U_n = \overline{0, X_{\varphi(n,0)} X_{\varphi(n,1)} \dots} = \sum_{k=0}^{\infty} \frac{X_{\varphi(n,k)}}{2^{k+1}},$$

les variables $(U_n)_{n \geq 0}$ sont indépendantes, du fait de la propriété d'indépendance par paquets, et suivent toutes la loi uniforme sur $[0, 1]$, comme rappelé plus haut.

1.12.2 Autre approche : extension de lois fini-dimensionnelles compatibles

On décrit ici, sans démonstration, comment cette existence s'inscrit dans un cadre plus général d'existence de processus.

Soit X un processus à valeurs dans un espace (E, \mathcal{E}) (sans hypothèse de dénombrabilité). On note μ_n la loi de (X_0, X_1, \dots, X_n) . C'est une probabilité sur $(E^{n+1}, \mathcal{E}^{\otimes(n+1)})$. Si on note $\pi_{n+1,n}$ la projection canonique de E^{n+1} sur E^n i.e. l'application $(x_0, \dots, x_{n-1}, x_n) \mapsto (x_0, \dots, x_{n-1})$, on a

$$\pi_{n+1,n}(\mu_n) = \mu_{n-1}.$$

Ceci est équivalent à, pour tout $A_k \in \mathcal{E}$,

$$\mu_{n-1}(A_0 \times A_1 \times \dots \times A_{n-1}) = \mu_n(A_0 \times A_1 \times \dots \times A_{n-1} \times E). \quad (1.5)$$

Les probabilités $(\mu_n)_{n \geq 0}$ s'appellent les répartitions finies du processus X .

Réciproquement, si on se donne des probabilités μ_n sur $(E^{n+1}, \mathcal{E}^{\otimes(n+1)})$ vérifiant la consistance (1.5), se pose la question de savoir s'il existe un processus ayant pour répartitions finies les μ_n . On introduit l'espace canonique $\Omega = E^{\mathbb{N}}$ pour $\omega = (\omega_n)_{n \geq 0}$, $X_n(\omega) = \omega_n$, $\mathcal{F}_n = \sigma(X_k, k \leq n)$, $\mathcal{F} = \sigma(X_k, k \geq 0)$.

Soit $A \in \mathcal{F}_n$, A est de la forme $A = B \times E \times \dots \times E \times \dots$ avec $B \in \mathcal{E}^{\otimes(n+1)}$. On définit alors une probabilité \mathbb{P}_n sur (Ω, \mathcal{F}_n) en posant,

$$\mathbb{P}_n(A) = \mu_n(B) \text{ puis une fonction d'ensembles sur } \cup_n \mathcal{F}_n \text{ par}$$

$$\mathbb{P}(A) = \mathbb{P}_n(A), \quad A \in \mathcal{F}_n. \quad (1.6)$$

Il s'agit de prolonger \mathbb{P} en une probabilité sur $\sigma(\cup_n \mathcal{F}_n)$. Remarquons que $\cup_n \mathcal{F}_n$ étant stable par intersection finie, ce prolongement sera unique. L'existence de ce prolongement a été montrée par Kolmogorov et on a :

Théorème 1.64. *Soit $(\mu_n)_{n \geq 0}$ une famille de probabilités sur $(E^{n+1}, \mathcal{E}^{\otimes(n+1)})$ vérifiant (1.5). Il existe une unique probabilité \mathbb{P} sur l'espace canonique (Ω, \mathcal{F}) défini par (1.6) telle que $(\Omega, \mathcal{F}, (X_n)_{n \geq 0}, \mathbb{P})$ soit un processus de répartitions finies $(\mu_n)_{n \geq 0}$.*

Exemple 1. Soient $\nu_0, \dots, \nu_n, \dots$ une suite de probabilités sur (E, \mathcal{E}) . On veut construire un modèle pour une suite de v.a. indépendantes de lois $\nu_0, \dots, \nu_n, \dots$. On définit μ_n sur $(E^{n+1}, \mathcal{E}^{\otimes(n+1)})$ par $\mu_n = \nu_0 \otimes \dots \otimes \nu_n$ et on applique le Théorème 1.64. On obtient une probabilité \mathbb{P} sur (Ω, \mathcal{F}) telle que $(X_n)_n$ soit une suite de v.a. indépendantes de loi $\nu_0, \dots, \nu_n, \dots$.

Exemple 2. Cet exemple fournit la construction des chaînes de Markov. On considère un ensemble E dénombrable muni d'une probabilité μ et d'une matrice de transition $Q(x, y)$, $x, y \in E$, c'est à dire une matrice à termes positifs telle que pour tous $x, y \in E$,

$$Q(x, y) \geq 0, \quad \sum_{y \in E} Q(x, y) = 1.$$

On définit μ_n sur E^{n+1} par $\mu_n(x_0, x_1, \dots, x_n) = \mu(x_0)Q(x_0, x_1) \dots Q(x_{n-1}, x_n)$ et on applique le Théorème 1.64. On obtient une probabilité \mathbb{P}_μ sur (Ω, \mathcal{F}) telle que les vecteurs (X_0, X_1, \dots, X_n) aient pour loi μ_n .

