Université Pierre et Marie Curie 2004-2005

Master de Mathématiques

Spécialité: Probabilités et Applications

Processus de Markov, application à la dynamique des populations

Jean Jacod

Chapitre 1

Chaînes de Markov

1.1 Introduction

L'idée des chaînes de Markov est très simple: il s'agit d'une suite $(X_n)_{n\in\mathbb{N}}$ de variables aléatoires à valeurs dans un espace mesurable $(E,\mathcal{E})$, définies sur un espace de probabilité $(\Omega,\mathcal{F},\mathbb{P})$, et telle que pour tout n on ait la propriété suivante:

$$\left.\begin{array}{l}\text{Conditionnellement à la valeur de } X_n\text{, les variables } (X_0,\ldots,X_{n-1})\\ \text{d'une part, } (X_{n+1},X_{n+2},\ldots) \text{ d'autre part, sont indépendantes.}\end{array}\right\} \tag{1.1.1}$$

Ce modèle permet de rendre compte d'un très grand nombre de situations concrètes.

Mathématiquement, la propriété précédente s'exprime ainsi: soit $\mathcal{F}_n=\sigma(X_0,\ldots,X_n)$ ($\sigma(.)$ désigne la tribu "engendrée par..."). On doit alors avoir pour tout $n\geq 0$:

$$\mathbb{E}(g(X_{n+1},X_{n+2},\ldots)|\mathcal{F}_n)=\mathbb{E}(g(X_{n+1},X_{n+2},\ldots)|\sigma(X_n)) \tag{1.1.2}$$

pour toute fonction g mesurable bornéee (ou positive) sur E^∞. Ci–dessus, $\mathbb{E}(\cdot\ |\mathcal{F}_n)$ désigne l'espérance conditionnelle, et il faut évidemment définir la mesurabilité d'une fonction définie sur E^∞. On verra que cela se ramène à la propriété apparemment plus faible suivante: pour tout $n\geq 0$ on a

$$\mathbb{E}(f(X_{n+1})|\mathcal{F}_n)=\mathbb{E}(f(X_{n+1})|\sigma(X_n)) \tag{1.1.3}$$

pour toute fonction f mesurable bornée (ou positive) sur E. Ainsi, en itérant (1.1.3), on voit que la construction d'une chaîne de Markov se fait en deux étapes:

- on se donne la loi de la "variable initiale" X_0,
- pour chaque $n\geq 0$ on se donne un mécanisme décrivant X_{n+1}, ou plutôt sa loi, en fonction de X_n.

On voit donc que l'étude mathématique des chaînes de Markov nécessite un certain nombre de préliminaires que nous introduisons dans le paragraphe suivant.

1.2 Préliminaires

1.2.1 Espérances conditionnelles et classes monotones

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, et $\mathcal{G}$ une sous–tribu de $\mathcal{F}$. Rappelons d'abord la

Définition 1.2.1. On appelle *espérance conditionnelle* de la variable aléatoire réelle Z (supposée positive, resp. intégrable) si $\mathcal{G}$ (ou, "par rapport à $\mathcal{G}$"), et on note $\mathbb{E}(Z|\mathcal{G})$, toute variable aléatoire U telle que:

1. U est $\mathcal{G}$–mesurable.

2. $\mathbb{E}(ZV) = \mathbb{E}(UV)$ pour toute variable aléatoire V positive (resp. bornée) et $\mathcal{G}$–mesurable.

L'espérance conditionnelle existe et est positive (resp. intégrable) si Z est positive (resp. intégrable). Elle est "unique à un ensemble de mesure nulle près": si U est une version de l'espérance conditionelle, la variable $\mathcal{G}$–mesurable U' en est une autre si et seulement si $U = U'$ $\mathbb{P}$–p.s.

Il y a essentiellement deux façons d'introduire l'espérance conditionnelle:

1. Le théorème de Radon-Nikodym: si Z est une variable aléatoire réelle positive, on définit la mesure (positive) $\mu(A) = \mathbb{E}(Z1_A)$. La restriction $\mu_{\mathcal{G}}$ de la mesure μ à la sous-tribu $\mathcal{G}$ est absolument continue par rapport à $\mathbb{P}_{\mathcal{G}}$ et l'espérance conditionnelle est définie comme la densité de $\mu_{\mathcal{G}}$ par rapport à $\mathbb{P}_{\mathcal{G}}$. Pour Z de signe quelconque, on procède par différence.

2. Le théorème de projection dans les espaces de Hilbert: si $F = L^2(\Omega, \mathcal{F}, \mathbb{P})$ et $G = L^2(\Omega, \mathcal{G}, \mathbb{P})$ alors l'espérance conditionnelle de $Z \in F$ est définie comme sa projection orthogonale sur G. Cette définition est ensuite étendue aux variables positives ou intégrables.

Un résultat très utile pour la caractérisation des espérances conditionnelles est fourni par le théorème des classes monotones:

Théorème 1.2.2. *Soit $\mathcal{H}$ un espace vectoriel de fonctions réelles bornées définies sur Ω et $\mathcal{C}$ un ensemble de parties de Ω stable par intersection finie. On suppose que*

(i) $1 \in \mathcal{H}$,

(ii) si $f_n \in \mathcal{H}$ et si $0 \leq f_n \uparrow f$ et f est bornée, alors $f \in \mathcal{H}$.

(iii) pour tout $A \in \mathcal{C}$, $1_A \in \mathcal{H}$.

Alors $\mathcal{H}$ contient toutes les fonctions $\sigma(\mathcal{C})$–mesurables bornées.

A titre d'application, on en déduit facilement le résultat suivant: soit X une variable définie sur $(\Omega, \mathcal{F}, \mathbb{P})$ et à valeurs dans un espace mesurable $(E, \mathcal{E})$. On suppose que la sous–tribu $\mathcal{G}$ est engendrée par une famille dénombrable (Y_n) de variables à valeurs dans un espace mesurable $(Y, \mathcal{Y})$ et soit f une fonction réelle $\mathcal{E}$–mesurable, positive (resp. bornée). Pour tester qu'une variable U, $\mathcal{G}$–mesurable positive (resp. bornée) est bien une version de l'espérance conditionnelle de $f(X)$ si $\mathcal{G}$ il suffit de vérifier que:

$$\mathbb{E}(f(X) \prod_{k=1}^{n} f_k(Y_k)) = \mathbb{E}(U \prod_{k=1}^{n} f_k(Y_k)) \tag{1.2.1}$$

pour tout $n \geq 1$ et tout système $f_1, f_2, \ldots f_n$ de fonctions $\mathcal{Y}$–mesurables, positives (resp. bornées).

1.2.2 Espérances conditionnelles et lois conditionnelles

Soit X une variable définie sur $(\Omega, \mathcal{F}, \mathbb{P})$, à valeurs dans un espace mesurable $(E, \mathcal{E})$, et $\mathcal{G}$ une sous–tribu de $\mathcal{F}$. On définit la "loi conditionnelle de X si $\mathcal{G}$" par la formule:

$$\mathbb{P}_X(B|\mathcal{G}) = \mathbb{E}(1_B \circ X|\mathcal{G}), \qquad B \in \mathcal{E}.$$

Bien sûr, pour B fixé cette formule ne définit $\mathbb{P}_X(B)$ que comme une classe d'équivalence de variables aléatoires, donc le théorème de convergence monotone ne fournit qu'une σ–additivité $\mathbb{P}$–presque sûre. Il est important de savoir s'il est possible de trouver pour chaque $B \in \mathcal{E}$ une version de cette classe d'équivalence, telle que pour chaque ω la fonction $B \mapsto \mathbb{P}_X(B|\mathcal{G})(\omega)$ soit une "vraie" probabilité sur $(E, \mathcal{E})$.

De manière plus précise introduisons la notion suivante, qui sera fondamentale pour les chaînes de Markov:

Définition 1.2.3. Si $(E, \mathcal{E})$ et $(F, \mathcal{F})$ sont deux espaces mesurables, on appelle *mesure de transition* de $(E, \mathcal{E})$ dans $(F, \mathcal{F})$ toute famille $Q = (Q(x, B) : x \in E, B \in \mathcal{F})$ de nombres dans $[0, \infty]$, qui vérifie:

1. $x \mapsto Q(x, A)$ est $\mathcal{E}$–mesurable pour tout $A \in \mathcal{F}$;

2. $A \mapsto Q(x, A)$ est une mesure sur $(F, \mathcal{F})$, pour tout $x \in E$.

Si de plus $Q(x, F) = 1$ pour tout x, on dit que Q est une *probabilité de transition.* On écrit aussi Q ainsi: $Q(x, dy)$.

On peut montrer (c'est le théorème de Jirina) que, dès que la tribu $\mathcal{E}$ est séparable (= engendrée par une *algèbre* dénombrable, comme le sont les tribus boréliennes de $\mathbb{R}$ ou de $\mathbb{R}^n$), il existe une probabilité de transition Q de $(\Omega, \mathcal{G})$ dans $(E, \mathcal{E})$, telle que pour tout $B \in \mathcal{E}$, la variable aléatoire $\omega \mapsto Q(\omega, B)$ soit une version de l'espérance conditionnelle $\mathbb{P}_X(B|\mathcal{G})(\omega)$. On écrit souvent $\mathbb{P}_{X/\mathcal{G}}$ au lieu de Q, et on a donc pour toute fonction $\mathcal{E}$–mesurable f, positive ou bornée:

$$\mathbb{E}(f(X)|\mathcal{G}) = \int_E f(x)\ \mathbb{P}_{X/\mathcal{G}}(\cdot, dx) \tag{1.2.2}$$

(cette égalité est vraie si $f = 1_B$ pour $B \in \mathcal{E}$ par définition même; par linéarité elle est vraie si f est mesurable étagée positive; par limite monotone elle est vraie si f est mesurable positive; par différence elle est vraie si f est mesurable bornée).

La probabilité de transition $\mathbb{P}_{X/\mathcal{G}}$ s'appelle *une version régulière de la loi conditionnelle.* Il est important de noter qu'elle n'est pas définie de manière unique: si par exemple A est un ensemble $\mathcal{G}$–mesurable négligeable, si on pose $\mathbb{P}'_{X/\mathcal{G}}(\omega, dx) = \mathbb{P}_{X/\mathcal{G}}(\omega, dx)$ pour $\omega \notin A$ et $\mathbb{P}'_{X/\mathcal{G}}(\omega, dx) = \mu(dx)$ pour $\omega \in A$ (où μ est une probabilité arbitraire sur E), on obtient une autre version régulière de la loi conditionnelle.

1.2.3 Opérations sur les mesures de transition

Dans (1.2.2) on a vu apparaitre de manière détournée une première opération sur les mesures de transition. Plus généralement, si $Q(x, dy)$ est une mesure de transition de $(E, \mathcal{E})$ dans $(F, \mathcal{F})$, on pose

$$Qf(x) = \int_F Q(x, dy)f(y), \qquad \forall x \in E \tag{1.2.3}$$

pour toute fonction $\mathcal{F}$–mesurable f sur F telle que cette formule ait un sens pour tout $x \in E$ (par exemple si $f \geq 0$, ou si f est bornée quand Q est une probabilité de transition). Noter que la fonction Qf est $\mathcal{E}$–mesurable: c'est évident si $f = 1_B$, par linéarité c'est vrai si f est positive mesurable étagée, par limite monotone c'est vrai si f est mesurable positive, et par différence c'est vraie si f est mesurable et si l'intégrale (1.2.3) existe pour tout x.

On a une opération "duale": si μ est une mesure sur $(E, \mathcal{E})$, on pose

$$\mu Q(B) = \int_E \mu(dx)Q(x, B), \qquad \forall B \in \mathcal{F}, \tag{1.2.4}$$

ce qui définit (d'après la linéarité de l'intégrale et le théorème de limite monotone) une nouvelle mesure μQ sur $(F, \mathcal{F})$. Si ε_x désigne la masse de Dirac en x, on a $\varepsilon_x Q = Q(x, \cdot)$.

Plus généralement, si Q est comme ci–dessus et si R est une mesure de transition de $(F, \mathcal{F})$ dans un autre espace $(G, \mathcal{G})$, on pose

$$(QR)(x, C) = \int_F Q(x, dy)R(y, C), \qquad \forall x \in E, \forall C \in \mathcal{G}. \tag{1.2.5}$$

Noter que $(QR)(x, C) = Qf(x)$ si $f(y) = R(y, C)$, tandis que $(QR)(x, C) = \mu R(C)$ si $\mu(\cdot) = Q(x, \cdot)$: par suite QR est une mesure de transition de $(E, \mathcal{E})$ dans $(G, \mathcal{G})$.

Par itération, si $(E_i, \mathcal{E}_i)$ est une suite d'espaces mesurables, si Q_i est une mesure de transition de $(E_{i-1}, \mathcal{E}_{i-1})$ dans $(E_i, \mathcal{E}_i)$, si μ_0 est une mesure sur $(E_0, \mathcal{E}_0)$, et si f_i est une fonction mesurable sur $(E_i, \mathcal{E}_i)$, on a ainsi:

- $Q_1 Q_2 \dots Q_i$: une mesure de transition de $(E_0, \mathcal{E}_0)$ dans $(E_i, \mathcal{E}_i)$;
- $\mu_0 Q_1 Q_2 \dots Q_i$: une mesure sur $(E_i, \mathcal{E}_i)$;
- $Q_1 Q_2 \dots Q_i f_i$: une fonction mesurable sur $(E_0, \mathcal{E}_0)$;
- $\mu_0 Q_1 Q_2 \dots Q_i f_i$: un nombre !

Si on suppose de plus que μ_0 est une probabilité et que les Q_i sont des probabilités de transition, notre objectif maintenant est de leur associer une probabilité sur l'espace produit $\prod E_i$.

1.2.4 Probabilités sur un produit: le cas fini

Commençons par le cas d'un produit de deux espaces $(E, \mathcal{E})$ et $(F, \mathcal{F})$. On pose $G = E \times F$ et $\mathcal{G} = \mathcal{E} \otimes \mathcal{F}$ (rappelons que $\mathcal{G}$ est la plus petite tribu sur G contenant les pavés mesurables $A \times B$, i.e. tels que $A \in \mathcal{E}$ et $B \in \mathcal{F}$).

Rappelons l'essentiel du théorème de Fubini, dans le cas des probabilités: si μ et ν sont des probabilités sur les deux espaces ci–dessus, $\eta = \mu \otimes \nu$ est l'unique mesure sur $(G,\mathcal{G})$ telle que $\eta(A \times B) = \mu(A)\nu(B)$ pour tous $A \in \mathcal{E}$ et $B \in \mathcal{F}$, et on a pour toute fonction $\mathcal{G}$–mesurable f, bornée ou positive:

$$\int f d\eta = \int_E \mu(dx) \left(\int_F \nu(dy) f(x,y) \right).$$

Enfin η est la loi d'un couple de variables (X,Y) lorsque X et Y sont de lois respectives μ et ν, et sont *indépendantes.*

Ce qui suit est une extension du théorème de Fubini, avec essentiellement la même preuve, lorsqu'on veut construire la loi d'un couple de variables (X,Y) avec X de loi μ, et la loi conditionelle de Y sachant que $X = x$ égale à $Q(x,\cdot)$.

Théorème 1.2.4. *Soit μ une probabilité sur $(E,\mathcal{E})$ et Q une probabilité de transition de $(E,\mathcal{E})$ dans $(F,\mathcal{F})$. Il existe une probabilité η et une seule sur le produit $(G,\mathcal{G})$, telle que*

$$\eta(A \times B) = \int \mu(dx)\ 1_A(x)\ Q(x,B), \qquad \forall A \in \mathcal{E}, \forall B \in \mathcal{F}. \tag{1.2.6}$$

De plus, pour toute fonction f mesurable sur $(G,\mathcal{G})$, positive ou bornée, la fonction $x \mapsto \int_F Q(x,dy) f(x,y)$ est $\mathcal{E}$–mesurable, et on a

$$\int f d\eta = \int_E \mu(dx) \left(\int_F Q(x,dy) f(x,y) \right). \tag{1.2.7}$$

Si enfin X et Y sont deux variables aléatoires à valeurs dans E et F respectivement, définies sur un espace de probabilité $(\Omega,\mathcal{A},I\!\!P)$ arbitraire, et telles que la loi du couple (X,Y) soit η, alors on a:

$$\left.\begin{array}{l} \bullet\ \text{la loi de } X \text{ est } \mu, \\ \bullet\ Q(X(\omega),\cdot) \text{ est une version régulière de } I\!\!P_{Y/\sigma(X)}. \end{array}\right\} \tag{1.2.8}$$

Noter à l'inverse qu'il existe toujours un couple (X,Y) vérifiant les dernières propriétés ci–dessus: il suffit de prendre $\Omega = G$, $\mathcal{A} = \mathcal{G}$, $I\!\!P = \eta$, et pour (X,Y) l'application identité de G dans lui-même.

Preuve. 1) Soit $\mathcal{H}$ l'espace de toutes les fonctions f mesurables bornées sur $(G,\mathcal{G})$ pour lesquelles $x \mapsto \int_F Q(x,dy) f(x,y)$ soit $\mathcal{E}$–mesurable, et $\mathcal{C}$ l'ensemble des pavés mesurables $A \times B$, qui est stable par intersection finie. Il est immédiat que $\mathcal{H}$ satisfait les hypothèses du théorème 1.2.2, donc égale la classe de toutes les fonctions $\mathcal{G}$–mesurables bornées. Par limite croissante, on en déduit aussi que $x \mapsto \int_F Q(x,dy) f(x,y)$ est $\mathcal{E}$–mesurable pour toute f mesurable positive.

2) D'après ce qui précède, pour tout $C \in \mathcal{G}$ on peut poser

$$\eta(C) = \int_E \mu(dx) \left(\int_F Q(x,dy) 1_C(x,y) \right).$$

Par linéarité et limite monotone, on voit que cela définit une probabilité η sur $(G,\mathcal{G})$, qui vérifie évidemment (1.2.6). On a (1.2.7) lorsque $f = 1_C$. Par linéarité, limite monotone, puis différence, on en déduit (1.2.7) pour toute fonction f mesurable, positive ou bornée.

3) L'unicité de η vérifiant (1.2.6) découle du théorème de Fubini (via une nouvelle application du théorème des classes monotones).

4) Il reste à prouver (1.2.8). Le fait que $\mathcal{L}(X) = \mu$ (où $\mathcal{L}(X)$ désigne la loi de X) découle de (1.2.6) appliqué avec $B = F$, donc $Q(x,B) = 1$ pour tout x. En appliquant (1.2.7) à $f(x,y) = g(x)1_B(y)$, on obtient

$$\mathbb{E}(g(X)1_B(Y)) = \int_E \mu(dx)g(x)Q(x,B)) = \mathbb{E}(g(X)Q(X,B)).$$

La seconde partie de (1.2.8) en découle immédiatement. □

Cette procédure se généralise aisément à un produit fini d'espaces. Nous nous contentons ci–dessous d'une version adaptée aux chaînes de Markov, mais il est possible de faire plus général.

La situation est la suivante: on a des espaces mesurables $(E_i,\mathcal{E}_i)$, des probabilités de transition Q_i de $(E_{i-1},\mathcal{E}_{i-1})$ dans $(E_i,\mathcal{E}_i)$, et une probabilité μ_0 sur $(E_0,\mathcal{E}_0)$. On pose $F_N = E_0 \times \ldots \times E_N$ et $\mathcal{F}_N = \mathcal{E}_0 \otimes \ldots \otimes \mathcal{E}_N$. On a alors:

Théorème 1.2.5. *(1) Pour toute fonction f mesurable sur $(F_N,\mathcal{F}_N)$, positive ou bornée, les formules $f_N = f$ et*

$$f_n(x_0,\ldots,x_n) = \int Q_{n+1}(x_n,dx_{n+1})f_{n+1}(x_0,\ldots,x_{n+1}) \tag{1.2.9}$$

pour $n = N-1, N-2, \ldots, 0$, définissent (par récurrence descendante) pour chaque n une fonction f_n qui est mesurable sur $(F_n,\mathcal{F}_n)$.

(2) Il existe une probabilité η_N et une seule sur $(F_N,\mathcal{F}_N)$, notée aussi (un peu abusivement) *$\mu_0 \otimes Q_1 \otimes \cdots \otimes Q_N$, telle que pour toute fonction f mesurable sur $(F_N,\mathcal{F}_N)$, positive ou bornée, on ait (avec la notation (1.2.9)):*

$$\int f d\eta_N = \int f_0 d\mu_0. \tag{1.2.10}$$

(3) Si enfin les X_i sont des variables aléatoires à valeurs dans E_i, définies sur un espace de probabilité $(\Omega,\mathcal{A},\mathbb{P})$ arbitraire, et telles que la loi du $N+1$–uplet $(X_0,\ldots,X_N)$ soit η, alors on a (pour $n = 1,\ldots,N$):

$$\left.\begin{array}{l}
\bullet\ \text{la loi de } X_0 \text{ est } \mu_0,\\
\bullet\ Q_n(X_{n-1}(\omega),\cdot) \text{ est une version régulière de } \mathbb{P}_{X_n/\sigma(X_0,X_1,\ldots,X_{n-1})}\\
\bullet\ \Big(Q_n(X_{n-1}(\omega),\cdot\,) \otimes Q_{n+1} \otimes \cdots \otimes Q_N\Big)(\cdot\,) \text{ est une version régulière de}\\
\qquad\qquad \mathbb{P}_{(X_n,X_{n+1},\ldots,X_N)/\sigma(X_0,X_1,\ldots,X_{n-1})}.
\end{array}\right\} \tag{1.2.11}$$

On combine en général (1.2.9) et (1.2.10), pour écrire:

$$\int f d\eta_N = \int \mu_0(dx_0) \int Q_1(x_0, dx_1) \dots \int Q(x_{N-1}, dx_N) f(x_0, \dots, x_N). \tag{1.2.12}$$

Ci–dessus, la convention est – comme d'habitude – que les intégrales sont effectuées de la droite vers la gauche.

Preuve. En posant $Q_n((x_0, \dots, x_{n-1}), B) = Q_n(x_{n-1}, B)$, on voit qu'on peut considérer Q_n comme une probabilité de transition de $(F_{n-1}, \mathcal{F}_{n-1})$ dans $(E_n, \mathcal{E}_n)$. Donc la mesurabilité de f_n découle immédiatement de la mesurabilité de f_{n+1} et du théorème 1.2.4 appliqué à $E = F_n$ et $F = E_{n+1}$, d'où le (1) par récurrence descendante.

Le (2) se montre par récurrence sur N: lorsque $N = 0$ il n'y a rien à montrer. Supposons que ce soit vrai pour $N - 1$. En remarquant que les fonctions f_n associées à f pour $n \leq N$ sont aussi les fonctions associées à f_{N-1} lorsqu'on démarre la récurrence descendante à $N - 1$, on voit que $\int f_0 d\mu_0 = \int f_{N-1} d\eta_{N-1}$. En d'autres termes, la formule (1.2.10) s'écrit aussi

$$\int f d\eta_N = \int \eta_{N-1}(du) \int Q_N(u, dx_N) f(u, x_N),$$

en notant $u = (x_0, \dots, x_{N-1})$. Pour obtenir l'existence et l'unicité de η_N il suffit alors d'appliquer le théorème 1.2.4 à $E = F_{N-1}$ et $F = E_N$.

Enfin pour (3), là encore la première partie de (1.2.11) est évidente. En appliquant (1.2.12) à $f(x_0, \dots, x_N) = 1_B(x_n, \dots, x_N) g(x_0, \dots, x_{n-1})$, on voit que

$$\mathbb{E}(g(X_0, \dots, X_{n-1}) 1_B(X_n, \dots, X_N)) = \mathbb{E}(g(X_0, \dots, X_{n-1}) R_n(X_{n-1}, B)),$$

où nous avons posé $R_n(x, \cdot) = Q_n(x, \cdot) \otimes Q_{n+1} \otimes \dots \otimes Q_N$. La troisième partie de (1.2.11) en découle, et la seconde partie en est un cas particulier. $\square$

On peut déjà observer que (1.2.11) nous donne la "propriété de Markov" (1.1.3), et même (1.1.2), pour la suite finie $(X_n)_{0 \leq n \leq N}$. Il faut maintenant voir ce qui se passe si on veut construire une suite infinie de variables: c'est un problème plus difficile, qui fait l'objet du paragraphe suivant.

1.2.5 Probabilités sur un produit: le cas infini

La situation est encore la suivante: pour $i \in \mathbb{N}$ on a des espaces mesurables $(E_i, \mathcal{E}_i)$, des probabilités de transition Q_i de $(E_{i-1}, \mathcal{E}_{i-1})$ dans $(E_i, \mathcal{E}_i)$ si $i \geq 1$, et une probabilité μ_0 sur $(E_0, \mathcal{E}_0)$. Mais, outre les espaces $(F_N, \mathcal{F}_N)$ définis ci–dessus, nous avons aussi l'espace $(F_\infty, \mathcal{F}_\infty)$, construit ainsi: d'une part $F_\infty = \prod_{i \in \mathbb{N}} E_i$; d'autre part $\mathcal{F}_\infty$ est la plus petite tribu contenant les pavés mesurables $\prod_{i \in \mathbb{N}} A_i$ (avec $A_i \in \mathcal{E}_i$), ou de manière équivalente contenant les pavés mesurables $A_0 \times \dots \times A_n \times \prod_{i \geq n+1} E_i$ avec n fini arbitraire.

Toute fonction f sur un F_n sera aussi considérée ci–dessous comme une fonction sur F_∞, ne dépendant que des "coordonnées" $x_0, \dots, x_n$.

Pour bien comprendre la seconde partie de l'énoncé ci–après, il convient de rappeler que si les X_i sont des variables aléatoires à valeurs dans les espaces $(E_i, \mathcal{E}_i)$, la suite infinie $(X_i)_{i\in\mathbb{N}}$ peut être considérée comme une variable à valeurs dans $(F_\infty, \mathcal{F}_\infty)$: la tribu $\mathcal{F}_\infty$ est construite précisément pour qu'on ait cette propriété. La "loi" de la suite infinie est donc une probabilité sur $(F_\infty, \mathcal{F}_\infty)$. A l'inverse, n'importe quelle probabilité η sur cet espace est la loi d'une suite infinie $(X_i)_{i\in\mathbb{N}}$ de variables sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$: il suffit de prendre $\Omega = F_\infty$, $\mathcal{A} = \mathcal{F}_\infty$, $\mathbb{P} = \eta$, et $X_i(x_0, x_1, \ldots) = x_i$ (les "applications coordonnées").

Théorème 1.2.6. (Ionescu–Tulcea) *(1) Il existe une probabilité η et une seule sur $(F_\infty, \mathcal{F}_\infty)$, notée aussi $\mu_0 \otimes Q_1 \otimes \cdots \otimes Q_n \otimes \cdots$, telle que pour tout $N \in \mathbb{N}$ et toute fonction f mesurable sur $(F_N, \mathcal{F}_N)$, positive ou bornée, on ait (avec la mesure η_N sur $(F_N, \mathcal{F}_N)$ du théorème 1.2.5)):*

$$\int f d\eta = \int f d\eta_N. \tag{1.2.13}$$

(2) Si les X_i sont des variables aléatoires à valeurs dans E_i, définies sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ arbitraire, et telles que la loi de la suite infinie $(X_n)_{n\in\mathbb{N}}$ soit η, alors on a pour tout n:

$$\left.\begin{array}{l}
\bullet \ \text{la loi de } X_0 \text{ est } \mu_0, \\
\bullet \ Q_n(X_{n-1}(\omega), \cdot) \text{ est une version régulière de } \mathbb{P}_{X_n/\sigma(X_0,X_1,\ldots,X_{n-1})} \\
\bullet \ \Big(Q_n(X_{n-1}(\omega), \cdot\,) \otimes Q_{n+1} \otimes \cdots\Big)(\cdot\,) \text{ est une version régulière de} \\
\qquad\qquad\qquad \mathbb{P}_{(X_n,X_{n+1},\ldots)/\sigma(X_0,X_1,\ldots,X_{n-1})}.
\end{array}\right\} \tag{1.2.14}$$

Preuve. a) De même qu'une fonction sur F_n peut être considérée comme une fonction sur F_∞, une partie A de F_n est identifiée à la partie $A \times \prod_{i\geq n+1} E_i$ de F_∞. Ainsi, la tribu $\mathcal{F}_n$ sur F_n peut être considérée comme une tribu sur F_∞. Avec cette convention, on pose $\mathcal{B} = \cup_n \mathcal{F}_n$: la classe $\mathcal{B}$ est une algèbre sur F_∞, qui engendre la tribu $\mathcal{F}_\infty$.

Si $A \in \mathcal{B}$, on a $A \in \mathcal{F}_N$ pour un N fini, et on pose $\eta(A) = \eta_N(A)$, avec η_N définie dans le théorème 1.2.5, qui implique aussi que $\eta_{N=p}(A) = \eta_N(A)$ pour tout $p \geq 1$. Cela définit donc une fonction additive d'ensembles sur $\mathcal{B}$, et d'après le théorème de prolongement des mesures on sait qu'elle s'étend en une probabilité sur $\mathcal{F}_\infty$, notée encore η et nécessairement unique, si et seulement si on a

$$\eta(A_n) \to 0 \qquad \text{pour toute suite } (A_n) \text{ d'éléments de } \mathcal{B} \text{ décroissant vers } \emptyset. \tag{1.2.15}$$

Si c'est le cas, la probabilité η coïncide avec η_N sur $\mathcal{F}_N$, d'où (1.2.13).

b) Pour obtenir (1) il nous reste donc à montrer (1.2.15). Si $n \geq 1$ et $A \in \mathcal{F}_N$ pour un $N \geq n$, on pose

$$R_n(x_0, \ldots, x_{n-1}; A) = \int Q_n(x_{n-1}, dx_n) \ldots \int Q_N(x_{N-1}, dx_N) 1_A(x_0, \ldots, x_N).$$

D'après le (1) du théorème 1.2.5 cette quantité est $\mathcal{F}_{n-1}$–mesurable en $(x_0, \ldots, x_{n-1})$. On notera que cette quantité ne dépend pas de N, pourvu que $N \geq n$ et que $A \in \mathcal{F}_N$, de sorte que $R_n(x_0,\ldots,x_{n-1};A)$ est en fait définie pour tout $A\in\mathcal{B}$, et de plus si $A\in\mathcal{F}_N$ et $N\geq n$ on a avec $u=(x_0,\ldots,x_{n-1})$:

$$\eta(A)=\eta_N(A)=\int\eta_n(du)R_n(u,A). \tag{1.2.16}$$

Soit maintenant une suite A_j dans $\mathcal{B}$, décroissant vers $\emptyset$. La suite $\eta(A_j)$ décroit vers une limite $a\geq 0$. On va supposer que $a>0$ et en déduire une contradiction, ce qui prouvera (1.2.15). D'abord, (1.2.16) appliqué à $n=1$ donne

$$\eta(A_j)=\int\mu_0(dx_0)R_1(x_0;A_j)\downarrow a>0,$$

et comme $R_1(x_0;A_j)$ décroit et est compris entre 0 et 1, on déduit du théorème de Lebesgue qu'il existe au moins un $x_0\in E_0$ tel que $R_1(x_0;A_j)\downarrow a_1>0$.

Supposons alors qu'on ait trouvé pour un $n\geq 1$ des points $x_0,\ldots,x_{n-1}$ tels que

$$R_n(x_0,\ldots,x_{n-1};A_j)\downarrow a_n>0, \tag{1.2.17}$$

lorsque $j\to\infty$. En remarquant que

$$R_n(x_0,\ldots,x_{n-1};A_j)=\int Q_n(x_{n-1},dx_n)R_{n+1}(x_0,\ldots,x_n;A_j)\geq a_n,$$

le même argument que ci-dessus entraine qu'il existe $x_n\in E_n$ et $a_{n+1}>0$ tels que $R_{n+1}(x_0,\ldots,x_{n-1},x_n;A_j)\downarrow a_{n+1}$. On construit ainsi par récurrence une suite $(x_n)_{n\in\mathbb{N}}$ telle que $x_n\in E_n$ et qu'on ait (1.2.17) pour tout n.

Mais chaque A_j est élément de $\mathcal{F}_{N_j}$ pour un certain entier N_j. En revenant à la définition de R_n, on voit que si $n>N_j$ le nombre $R_n(x_0,\ldots,x_{n-1};A_j)$ vaut 1 si $(x_0,\ldots)$ appartient à A_j et vaut 0 autrement. En comparant ceci à (1.2.17), on voit que nécessairement $(x_0,\ldots)\in A_j$. Comme ceci est vrai pour tout j, le point $(x_0,\ldots)$ appartient à $\cap_j A_j$, qui est vide, d'où une contradiction.

c) Plaçons–nous dans la situation de (2). La première partie de (1.2.14) est encore une fois évidente, et la second est un cas particulier de la troisième. Pour celle–ci, posons $R_n(x,\cdot)=Q_n(x,\cdot)\otimes Q_{n+1}\otimes Q_{n+2}\otimes\ldots$ (en vertu de (1), c'est une probabilité sur l'espace $(H_n,\mathcal{H}_n)=(\prod_{i\geq n}E_i,\otimes_{i\geq n}\mathcal{E}_i))$. D'après (1.2.13) et (3) du théorème 1.2.5, $\int f dR_n(X_{n-1},\cdot)$ est une version de l'espérance conditionnelle de $\mathbb{E}(f(X_n,\ldots)|\sigma(X_0,\ldots,X_{n-1}))$ pour toute fonction f bornée mesurable sur $(H_n,\mathcal{H}_{n,m})$ pour un $m\geq n$, où $\mathcal{H}_{n,m}=\otimes_{n\leq i\leq m}\mathcal{E}_i$. Un argument de classe monotone montre qu'il en est de même si f est bornée et $\mathcal{H}_n$–mesurable, ce qui donne le résultat. $\square$

Remarque: Ce théorème – plutôt difficile – n'est en aucune manière plus simple lorsque les espaces E_i sont égaux à $\mathbb{R}$, ou sont dénombrables, ou même finis ! (sauf bien–sûr s'ils sont réduits à un seul point). De même il n'est pas plus simple lorsque les probabilités de transition sont de la forme $Q_i(x,dy)=\mu_i(dy)$, où μ_i est une probabilité sur $(E_i,\mathcal{E}_i)$. Dans ce dernier cas, en vertu de la partie (2) du théorème précédent la mesure η est la loi d'une suite de variables (X_n) qui sont *indépendantes*, chaque X_n étant à valeurs dans E_n et de loi μ_n: comme corollaire, on obtient ainsi l'existence d'une suite de variables indépendantes de lois données.

1.3 Chaînes de Markov: définitions

Venons–en maintenant aux chaînes de Markov. On se donne un "espace d'état" mesurable $(E,\mathcal{E})$, qui est *a priori* quelconque. Il existe plusieurs définitions possibles d'une chaîne de Markov.

• La notion la plus faible consiste à dire qu'on a un espace probabilisé $(\Omega,\mathcal{F},\mathbb{P})$ et une suite (X_n) de variables à valeurs dans $(E,\mathcal{E})$, telle qu'on ait (1.1.2) (où g "mesurable sur E^∞" signifie mesurable par rapport à la tribu produit tensoriel $\mathcal{E}^{\otimes\infty}$). Par conditionnements successifs, cette propriété se ramène à (1.1.3). Cette dernière propriété donne moralement la loi conditionnelle de X_n sachant $(X_0,\ldots,X_{n-1})$, mais on n'est pas sûr de l'existence de versions régulières de ces lois conditionnelles.

• Une seconde définition possible consiste à dire qu'on a (1.1.3) *avec de plus* des probabilités conditionnelles régulières: cela revient à dire qu'on a des probabilités de transition P_n de $(E,\mathcal{E})$ dans lui-même, telles que pour tout n et toute fonction mesurable bornée ou positive f sur E,

$$\mathbb{E}(f(X_n)|\sigma(X_0,\ldots,X_{n-1})) = P_nf(X_{n-1}). \tag{1.3.1}$$

(Comme l'espérance conditionnelle n'est définie qu'à un ensemble de probabilité nulle près, nous omettons systématiquement d'écrire le "p.s." dans des égalités faisant intervenir les espérances ou probabilités conditionnelles). De ce point de vue, la "loi initiale" (i.e., la loi de X_0) est considérée comme donnée.

• Une situation un peu plus générale consiste à supposer que l'espace $(\Omega,\mathcal{F})$ est muni d'une *filtration*: une suite $(\mathcal{F}_n)_{n\in\mathbb{N}}$ de sous–tribus de $\mathcal{F}$, avec $\mathcal{F}_n \subset \mathcal{F}_{n-1}$ et X_n est $\mathcal{F}_n$–mesurable pour tout n (on dit que (X_n) est *adapté* à la filtration); on remplace alors (1.3.1) par

$$\mathbb{E}(f(X_n)|\mathcal{F}_{n-1}) = P_nf(X_{n-1}). \tag{1.3.2}$$

Lorsque $\mathcal{F}_n = \sigma(X_0,\ldots,X_n)$, (1.3.2) se ramène à (1.3.1).

• Une quatrième définition possible consiste à se donner les transitions P_n, considérées comme décrivant le "mécanisme d'évolution" de la chaîne, mais à laisser la loi initiale arbitraire: pour toute probabilité μ sur $(E,\mathcal{E})$ on a donc une probabilité $\mathbb{P}_\mu$ sur $(\Omega,\mathcal{F})$ vérifiant (1.3.2), et telle que $\mathcal{L}(X_0) = \mu$.

Les situations 2 et 3 ci–dessus nous conduisent à poser:

Définition 1.3.1. 1) Soit $(\Omega,\mathcal{F},(\mathcal{F}_n)_{n\in\mathbb{N}},\mathbb{P})$ un espace probabilisé filtré. Une suite $(X_n)_{n\in\mathbb{N}}$ de variables aléatoires sur cet espace, à valeurs dans $(E,\mathcal{E})$, est dite vérifier la $(\mathcal{F}_n)$*–propriété de Markov* si pour tout n la variable X_n est $\mathcal{F}_n$–mesurable et si on a (1.3.2) pour une suite de probabilités de transition (P_n).

2) Si on a $\mathcal{F}_n = \sigma(X_0,\ldots,X_n)$ (donc (1.3.2)=(1.3.1)), on dit simplement qu'on a la *propriété de Markov*.

3) Si $P_n = P$ ne dépend pas de n, la propriété de Markov est dite *homogène*.

La situation 4 ci–dessus n'est réellement intéressante que dans le cas homogène. Dans ce cadre, nous allons donner la "vraie" définition des chaînes de Markov, celle que nous utiliserons dans la suite (on distinguera “chaîne de Markov”, définie ci–après, et “propriété de Markov”, définie ci–dessus). Cette définition peut sembler compliquée, mais en fait elle permet de rendre la suite beaucoup plus simple.

On part donc d'un ensemble mesurable $(\Omega, \mathcal{F})$, muni de variables aléatoires $(X_n)_{n\in\mathbb{N}}$ à valeurs dans $(E, \mathcal{E})$; on pose $\mathcal{F}_n = \sigma(X_0, \ldots, X_n)$, et on suppose que $\mathcal{F} = \bigvee_n \mathcal{F}_n$. Cet espace est aussi muni d'une “translation” θ, qui est une application de Ω dans lui-même telle que

$$X_{n+1} = X_n \circ \theta, \qquad \forall n. \tag{1.3.3}$$

On se donne enfin une famille $(\mathbb{P}_x)_{x\in E}$ de probabilités sur $(\Omega, \mathcal{F})$, telle que $x \mapsto \mathbb{P}_x(A)$ est $\mathcal{E}$–mesurable pour tout $A \in \mathcal{F}$ (donc, $\mathbb{P}_x(d\omega)$ est une probabilité de transition de $(E, \mathcal{E})$ dans $(\Omega, \mathcal{F})$. On suppose que

$$\mathbb{P}_x(X_0 = x) = 1 \qquad \forall x \in E. \tag{1.3.4}$$

On notera $\mathbb{E}_x$ l'espérance, relativement à la probabilité $\mathbb{P}_x$.

Définition 1.3.2. Le terme $\mathbf{X} = (\Omega, \mathcal{F}, (X_n), \theta, (\mathbb{P}_x))$ défini ci–dessus s'appelle une *chaîne de Markov* (sous–entendu: “homogène”) s'il existe une probabilité de transition P de $(E, \mathcal{E})$ dans lui–même telle que pour tout n, tout $x \in \mathbb{E}$, et toute fonction mesurable f sur E, bornée ou positive, on a

$$\mathbb{E}_x(f(X_n)|\mathcal{F}_{n-1}) = Pf(X_{n-1}). \tag{1.3.5}$$

Pour toute probabilité μ sur $(E, \mathcal{E})$, on pose

$$\mathbb{P}_\mu(A) = \int \mu(dx)\mathbb{P}_x(A) \qquad \forall A \in \mathcal{F}, \tag{1.3.6}$$

ce qui en vertu de (1.2.4) définit une nouvelle probabilité sur $(\Omega, \mathcal{F})$. L'espérance relative à $\mathbb{P}_\mu$ sera notée $\mathbb{E}_\mu$.

Proposition 1.3.3. *Si $\mathbf{X}$ est une chaîne de Markov, pour toute probabilité μ sur $(E, \mathcal{E})$, la loi de X_0 sous $\mathbb{P}_\mu$ (appelée “loi initiale”) est μ. De plus pour tout n et toute variable aléatoire réelle Y sur $(\Omega, \mathcal{F})$, positive ou bornée, en notant θ^n la n*–ième *itérée de θ, la variable $Y \circ \theta^n$ est $\mathcal{F}$–mesurable et on a:*

$$\mathbb{E}_\mu(Y \circ \theta^n|\mathcal{F}_n) = \mathbb{E}_{X_n}(Y). \tag{1.3.7}$$

Preuve. D'abord, si $A \in \mathcal{E}$, on a par (1.3.3) et (1.3.4):

$$\mathbb{P}_\mu(X_0 \in A) = \int \mu(dx)\mathbb{P}_x(X_0 \in A) = \int \mu(dx)1_A(x) = \mu(A),$$

d'où la première assertion.

D'après le théorème des classes monotones 1.2.2, et puisque $\mathcal{F} = \sigma(X_0, X_1, \ldots)$, pour la seconde assertion il suffit de montrer le résultat quand Y est de la forme $Y = \prod_{i=0}^m f_i(X_i)$, pour des fonctions bornées mesurables f_i. Dans ce cas, on a $Y \circ \theta^n = \prod_{i=0}^m f_i(X_{n+i})$, donc on voit que $Y \circ \theta^n$ est $\mathcal{F}$–mesurable. Si Z est une variable bornée $\mathcal{F}_n$–mesurable, on a

$$\mathbb{E}_\mu(Z\ Y \circ \theta^n) = \int \mu(dx)\mathbb{E}_x(Z\ Y \circ \theta^n).$$

Il nous reste donc à montrer que

$$\mathbb{E}_x(Z\ Y \circ \theta^n) = \mathbb{E}_x(Z\ \mathbb{E}_{X_n}(Y)). \tag{1.3.8}$$

Cela se montre par récurrence sur m. Lorsque $m = 0$ on a $Y = f_0(X_0)$, donc $\mathbb{E}_x(Y) = f_0(x)$, et (1.3.8) est évident. Supposons donc que (1.3.8) soit vrai pour tout n et pour toute variable de la forme $Y = \prod_{i=0}^{m-1} g_i(X_i)$; nous allons alors montrer (1.3.8) pour $Y = \prod_{i=0}^{m} f_i(X_i)$. Posons $U = \prod_{i=0}^{m-1} f_i(X_i)$ et $V = U\ Pf_m(X_{m-1})$. En vertu de (1.3.5) on a

$$\begin{aligned} \mathbb{E}_x(Z\ Y \circ \theta^n) &= \mathbb{E}_x(Z\ U \circ \theta^n\ f_m(X_{n+m})) \\ &= \mathbb{E}_x\left(\mathbb{E}_x(Z\ U \circ \theta^n\ f_m(X_{n+m}) | \mathcal{F}_{n+m-1})\right) \\ &= \mathbb{E}_x\left(Z\ U \circ \theta^n\ Pf_m(X_{n+m-1})\right) \\ &= \mathbb{E}_x\left(Z\ V \circ \theta^n\right) \\ &= \mathbb{E}_x\left(Z\ \mathbb{E}_{X_n}(V)\right), \end{aligned}$$

où la dernière égalité provient de l'hypothèse de récurrence. En appliquant ce qui précède à $n = 0$ et $Z = 1$, on obtient aussi $\mathbb{E}_y(Y) = \mathbb{E}_y(V)$ pour tout $y \in E$, d'où (1.3.8). $\square$

Ce résultat nous dit que, sous chaque $\mathbb{P}_\mu$, le chaîne (X_n) admet la propriété de Markov homogène; de plus la probabilité de transition est P: il suffit d'appliquer (1.3.7) à $Y = f(X_1)$ et de remarquer que $\mathbb{E}_x(f(X_1)) = Pf(x)$ par (1.3.5). Noter qu'on a aussi démontré (1.1.2) pour $\mathbb{P}_\mu$.

Construction d'une chaîne de Markov: Etant donnée une probabilité de transition P arbitraire de $(E, \mathcal{E})$ dans lui-même, nous sommes en mesure de construire une chaîne de Markov $\mathbf{X}$ de transition P. Cela peut se faire sur beaucoup d'espaces Ω différents, bien–sûr, mais une manière canonique de faire est la suivante:

On prend $\Omega = E^{\mathbb{N}}$, donc un point de Ω est une suite infinie $\omega = (x_0, x_1, \ldots)$. On pose

$$X_n(x_0, \ldots) = x_n, \qquad \theta(x_0, x_1, \ldots) = (x_1, \ldots).$$

Enfin, on prend pour $\mathbb{P}_x$ la probabilité construite dans le théorème de Ionescu–Tulcea 1.2.6, correspondant aux $(E_i, \mathcal{E}_i) = (E, \mathcal{E})$ et $Q_i = P$ et à $\mu_0 = \varepsilon_x$ (masse de Dirac en x). La mesurabilité de $x \mapsto \mathbb{P}_x(A)$ s'obtient par application de (1) du théorème 1.2.5 avec $n = 0$ et (1.2.13) et encore une fois un argument de classe monotone. Noter que $\mathbb{P}_\mu$ est obtenue de la même manière, en prenant $\mu_0 = \mu$.

La chaîne de Markov $\mathbf{X}$ ainsi obtenue s'appelle *la chaîne canonique* associée à la transition P.

1.4 Premières propriétés

Ci–dessous on considère une chaîne de Markov $\mathbf{X} = (\Omega, \mathcal{F}, (X_n), \theta, (\mathbb{P}_x))$ de transition P sur $(E, \mathcal{E})$. On désigne par $P^n = P \ldots P$ la n–ième itérée de P au sens de (1.2.5) (par convention P^0 est la transition "identité", c'est-à-dire $P^0(x, dy) = \varepsilon_x(dy)$).

Proposition 1.4.1. *Pour tous $n, m \geq 0$ et toute fonction mesurable f bornée ou positive, on a*

$$\mathbb{E}_\mu(f(X_{n+m})|\mathcal{F}_n) = P^m f(X_n). \tag{1.4.1}$$

De plus, la loi de X_m sous $\mathbb{P}_\mu$ est μP^m.

Preuve. Pour $m = 0$ (1.4.1) est évident, et pour $m = 1$ c'est (1.3.7) (appliqué à $Y = f(X_1)$). Supposons (1.4.1) vraie pour $m-1$, et toute fonction f et tout n. On a

$$\mathbb{E}_\mu(f(X_{n+m})|\mathcal{F}_n) = \mathbb{E}_\mu\left(\mathbb{E}_\mu(f(X_{n+m})|\mathcal{F}_{n+m-1}|\mathcal{F}_n)\right) = \mathbb{E}_\mu\left(Pf(X_{n+m-1})|\mathcal{F}_n\right),$$

qui vaut $P^{m-1}Pf(X_n) = P^m f(X_n)$ d'après l'hypothèse de récurrence. On a donc (1.4.1) pour tout $n, m \geq 0$.

Enfin, la loi μ_m de X_m sous $\mathbb{P}_\mu$ est donnée par $\mu_m(A) = \mathbb{P}_\mu(X_m \in A)$. Mais

$$\mathbb{P}_\mu(X_m \in A) = \mathbb{E}_\mu(\mathbb{P}_\mu(X_m \in A|\mathcal{F}_0)) = \mathbb{E}_\mu(P^m(X_0, A)) = \int \mu(dx) P^m(x, A) = \mu P^m(A),$$

où on a utilisé (1.4.1) pour la seconde égalité: on a donc la seconde assertion. $\square$

La seconde propriété, fondamentale pour la suite, est la "propriété forte de Markov". Rappelons qu'un temps d'arrêt est une application de Ω dans $\overline{\mathbb{N}} = \mathbb{N} \cup \{\infty\}$ qui vérifie $\{T \leq n\} \in \mathcal{F}_n$ pour tout $n \in \mathbb{N}$, ou de manière équivalente $\{T = n\} \in \mathcal{F}_n$ pour tout n. On associe à un temps d'arrêt T sa "tribu antérieure", qui est

$$\mathcal{F}_T = \{A : \ A \in \mathcal{F}, A \cap \{T \leq n\} \in \mathcal{F}_n \quad \forall n \in \mathbb{N}\},$$

et ci–dessus on peut remplacer $A \cap \{T \leq n\}$ par $A \cap \{T = n\}$ partout si on veut: on vérifie que $\mathcal{F}_T$ est une tribu, et si $T(\omega) = m$ pour tout ω alors T est un temps d'arrêt et $\mathcal{F}_T = \mathcal{F}_m$.

De même qu'on a défini les itérées θ^n de la translation θ, pour tout temps d'arrêt T on peut définir les applications θ^T et X_T de $\{T < \infty\}$ dans Ω et E respectivement, par

$$\left.\begin{aligned} \theta^T(\omega) &= \theta^{T(\omega)}(\omega) = \theta^n(\omega) \\ X_T(\omega) &= X_{T(\omega)}(\omega) = X_n(\omega) \end{aligned}\right\} \quad \text{sur l'ensemble} \ \ \{T = n\}. \tag{1.4.2}$$

D'une part on a $\{X_T \in A\} \cap \{T = n\} = \{X_n \in A\} \cap \{T = n\}$, qui est $\mathcal{F}_n$–mesurable si $A \in \mathcal{E}$: donc X_T est $\mathcal{F}_T$–mesurable en restriction à l'ensemble $\{T < \infty\}$. D'autre part pour toute variable aléatoire Y on a $\{Y \circ \theta^T \in A\} \cap \{T = n\} = \{Y \circ \theta^n \in A\} \cap \{T = n\}$, qui est $\mathcal{F}$–mesurable pour tout borélien A: donc $Y \circ \theta^T$ est $\mathcal{F}$–mesurable en restriction à l'ensemble $\{T < \infty\}$. Ces propriétés de mesurabilité montrent que l'assertion ci-dessous a un sens.

Proposition 1.4.2. (Propriété forte de Markov) *Pour tout temps d'arrêt T et toute variable aléatoire Y positive ou bornée on a*

$$\mathbb{E}_\mu(Y \circ \theta^T|\mathcal{F}_T) = \mathbb{E}_{X_T}(Y) \qquad \textit{sur l'ensemble} \ \ \{T < \infty\}. \tag{1.4.3}$$

Pour bien comprendre (1.4.3) il convient de remarquer que l'ensemble $\{T < \infty\}$ est $\mathcal{F}_T$–mesurable, donc ci-dessus on peut aussi bien ajouter son indicatrice en facteur dans chaque membre, et à gauche faire passer cette indicatrice à l'intérieur de l'espérance conditionnelle: les deux membres ont donc *a priori* un sens.

En appliquant (1.4.3) à $Y = f(X_m)$, on obtient (en utilisant aussi la proposition 1.4.1):

$$\mathbb{E}_\mu(f(X_{T+m})|\mathcal{F}_T) = P^m f(X_T) \qquad \text{sur l'ensemble } \{T < \infty\}. \tag{1.4.4}$$

Preuve. Les variables $U = Y \circ \theta^T$ et $V = \mathbb{E}_{X_T}(Y)$ sont bien définies sur l'ensemble $\{T < \infty\}$, et la seconde est $\mathcal{F}_T$–mesurable. Ainsi il suffit de montrer que si $A \in \mathcal{F}_T$ on a $\mathbb{E}_\mu(U 1_A 1_{\{T<\infty\}}) = \mathbb{E}_\mu(V 1_A 1_{\{T<\infty\}})$, Il suffit même de montrer que pour chaque $n \in \mathbb{N}$ on a

$$\mathbb{E}_\mu(U 1_A 1_{\{T=n\}}) = \mathbb{E}_\mu(V 1_A 1_{\{T=n\}}).$$

Mais le membre de gauche est $E_\mu(1_{A\cap\{T=n\}}\, Y \circ \theta^n)$, tandis que le membre de droite est $E_\mu(1_{A\cap\{T=n\}}\, \mathbb{E}_{X_n}(Y))$, et $A \cap \{T = n\} \in \mathcal{F}_n$: il suffit alors d'appliquer (1.3.7). □

Les notions suivantes seront souvent utilisées; elles sont relatives à la transition P:

Définition 1.4.3. 1) Une mesure (positive) η sur $(E, \mathcal{E})$ est dite

- *invariante* si $\eta P = \eta$,
- *sous–invariante* si $\eta P \leq \eta$.

2) Une fonction mesurable positive sur $(E, \mathcal{E})$ est dite

- *harmonique* si $Pf = f$,
- *surharmonique* si $Pf \leq f$.

Proposition 1.4.4. *(1) Si f est une fonction harmonique on a $P^n f = f$ pour tout n, et si de plus $\mu(f) < \infty$ le processus $f(X_n)$ est une martingale sous la probabilité $\mathbb{P}_\mu$.*

(2) Si f est une fonction surharmonique on a $P^n f \leq f$ pour tout n, et le processus $f(X_n)$ est une surmartingale sous toute probabilité $\mathbb{P}_\mu$.

Preuve. Comme P, en tant qu'opérateur sur les fonctions, est positif, les premières assertions de (1) et (2) sont évidentes. De plus $f(X_n)$ est $\mathcal{F}_n$–mesurable, et on a pour toute probabilité initiale μ:

$$\mathbb{E}_\mu(f(X_{n+m})|\mathcal{F}_n) = P^m f(X_n).$$

Si f est surharmonique on a donc $\mathbb{E}_\mu(f(X_{n+m})|\mathcal{F}_n) \leq f(X_n)$, ce qui est l'inégalité caractéristique des surmartingales. Quand f est harmonique, on a $\mathbb{E}_\mu(f(X_{n+m})|\mathcal{F}_n) = f(X_n)$, ce qui est l'égalité caractéristique des martingales: pour obtenir que $f(X_n)$ est une $\mathbb{P}_\mu$–martingale il faut de plus que chaque $f(X_n)$, ou de manière équivalente $f(X_0)$, soit μ–intégrable, ce qui revient à dire que $\mu(f) < \infty$. □

Définition 1.4.5. Si f est une fonction mesurable positive, son *potentiel* est la fonction

$$Uf = \sum_{n\geq 0} P^n f \qquad (\text{avec } P^0 f = f). \tag{1.4.5}$$

Par le théorème de limite monotone, il est facile de voir qu'on définit ainsi une mesure de transition U, appelée aussi le potentiel de P, de sorte que Uf définie ci–dessus est l'action de U sur f. Pour toute mesure positive μ on a de même la mesure μU. On écrit bien-sûr $U = \sum_{n\geq 0} P^n$.

Comme $Uf = f + \sum_{n\geq 1} P^n f$, on obtient l'équation (pour $f \geq 0$):

$$U = I + PU = I + UP \qquad \text{i.e.,} \quad Uf = f + PUf = f + UPf. \tag{1.4.6}$$

En particulier, Uf est une fonction surharmonique. De même pour toute mesure (positive) μ on a $\mu U = \mu + \mu PU = \mu + \mu UP$, donc μU est une mesure sous–invariante.

On notera aussi que, par le théorème de limite monotone, pour toute fonction mesurable positive f et toute probabilité μ,

$$\mu Uf = I\!\!E_\mu \left(\sum_{n\geq 0} f(X_n) \right). \tag{1.4.7}$$

Pour terminer ce paragraphe, examinons la stationnarité de la chaîne (X_n). On rappelle qu'un processus $(X_n)_{n\in I\!\!N}$ est *stationnaire* si pour tout n la loi du $n+1$–uplet $(X_m, X_{m+1}, \ldots, X_{m+n})$ ne dépend pas de l'entier m.

Proposition 1.4.6. *La chaîne (X_n) est stationnaire sous la probabilité $I\!\!P_\mu$ si et seulement si la probabilité μ est invariante.*

Preuve. Si la chaîne est stationnaire sous $I\!\!P_\mu$, la loi μP de X_1 (cf. proposition 1.4.1) égale la loi μ de X_0, donc μ est invariante.

Inversement supposons μ invariante. Soit $n \geq 0$. La loi $\zeta_{n,m}$ de $(X_m, X_{m+1}, \ldots, X_{m+n})$ est donnée par $\zeta_{n,m}(A) = I\!\!P_\mu((X_m, X_{m+1}, \ldots, X_{m+n}) \in A)$ pour $A \in \mathcal{E}^{\otimes(n+1)}$. Si $Y = 1_A(X_0, \ldots, X_n)$ et $f(x) = I\!\!E_x(Y)$, on a

$$\eta_{n,m}(A) = I\!\!E_\mu(Y \circ \theta^m) = I\!\!E_\mu\left(I\!\!E_\mu(Y \circ \theta^m)|\mathcal{F}_m)\right) = I\!\!E_\mu\left(f(X_m)\right) = \mu P^m f$$

par (1.3.7) et la proposition 1.4.1. L'invariance de μ entraine $\mu P^m f = \mu(f)$, donc $\eta_{n,m}$ ne dépend pas de m. □

Le fait d'avoir une chaîne stationnaire est crucial pour les applications: par exemple (on le verra en exercice) un grand nombre de "files d'attente" sont des chaînes de Markov, et le fait qu'elles soient stationnaires signifie qu'il n'y a pas engorgement de la file. Une bonne partie de ce chapitre sera en fait consacré à l'étude de l'existence d'une probabilité invariante, rendant la chaîne stationnaire.

1.5 Chaînes discrètes: classifications

On dit qu'une chaîne de Markov $\mathbf{X}$ est *discrète* si son espace d'états E est *fini ou dénombrable.* Dans ce cas, $\mathcal{E}$ est toujours la tribu de toutes les parties de E, et toute fonction sur E est mesurable. Les points de E seront – selon l'habitude – notés i, j, k.

1.5.1 Notations

Toute mesure μ sur E est caractérisée par les mesures des singletons $\mu_i = \mu(\{i\})$, via la formule $\mu(A) = \sum_{i\in A} \mu_i$, et toute famille $(\mu_i)_{i\in E}$ de nombres positifs ou nuls correspond à une mesure. Cette mesure est une probabilité si de plus $\sum_i \mu_i = 1$.

Une probabilité de transition P est aussi caractérisée par les nombres $p_{ij} = P(i, \{j\})$, et inversement une double suite (p_{ij}) correspond à une probabilité de transition si et seulement si on a

$$p_{ij} \geq 0, \qquad \sum_j p_{ij} = 1.$$

Ainsi on peut identifier $P = (p_{ij})$ avec une “matrice” infinie, à termes positifs, la somme de chaque ligne valant 1.

Si $Q = (q_{ij})$ est une autre probabilité de transition, alors le produit $R = PQ$ est

$$R = (r_{ij}), \qquad \text{avec} \qquad r_{ij} = \sum_k p_{ik} q_{kj}.$$

On a de même, pour une mesure μ et une fonction f:

$$(\mu P)_i = \sum_j \mu_j p_{ji}, \qquad (Pf)(i) = \sum_j p_{ij} f(j),$$

de sorte qu'en identifiant une mesure à une *matrice ligne*, et une fonction à une *matrice colonne*, les différents produits (1.2.3), (1.2.4) et (1.2.5) sont les produits usuels des matrices (infinies, si E est dénombrable).

Revenons à notre chaîne de Markov de transition P. On définit les temps d'arrêt suivants:

$$T_i = \inf(n \geq 1 : X_n = i), \qquad T_i^1 = T_i, \quad T_i^{n+1} = \inf(m > T_i^n : X_m = i),$$

avec la convention habituelle que l'inf de l'ensemble vide vaut $+\infty$ (Exercice: montrer que ce sont effectivement des temps d'arrêt). Ensuite, on pose

$$f_{ij}^{(n)} = \mathbb{P}_i(T_j^n < \infty), \qquad f_{ij} = f_{ij}^{(1)} = \mathbb{P}_i(T_i < \infty), \qquad N_i = \sum_{n=0}^{\infty} 1_{\{i\}}(X_n).$$

Ainsi, N_i est le nombre (aléatoire) de fois où la chaîne visite l'état i. Les éléments de la transition itérée P^n seront notés $p_{ij}^{(n)}$ (donc $p_{ij}^{(0)} = \delta_{ij}$, le symbole de Kronecker). Le potentiel $U = (u_{ij})$ est donné par

$$u_{ij} = \sum_{n\geq 0} p_{ij}^{(n)} = \mathbb{E}_i(N_j). \tag{1.5.1}$$

1.5.2 Première classification

Définition 1.5.1. Si $i, j \in E$, on dit que *i mène à j*, et on écrit $i \mapsto j$, si on a soit $i = j$, soit $\mathbb{P}_i(T_j < \infty) > 0$. On écrit $i \sim j$ si $i \mapsto j$ *et* $j \mapsto i$.

Proposition 1.5.2. *On a $i \mapsto j$ si et seulement si $u_{ij} > 0$, et la relation $i \sim j$ est une relation d'équivalence.*

Preuve. Si $j = i$ on sait que $u_{ij} \geq 1$ (car $p_{ij}^{(0)} = \delta_{ij}$), et que $i \mapsto j$.

Supposons $j \neq i$. Si $u_{ij} > 0$ il existe $n > 0$ avec $p_{ij}^{(n)} > 0$, et bien-sûr $I\!P_i(T_j < \infty) \geq I\!P_i(X_n = j) = p_{ij}^{(n)} > 0$, donc $i \mapsto j$. Si inversement $i \mapsto j$ il existe $n \geq 1$ avec $0 < I\!P_i(T_j = n) \leq I\!P_i(X_n = j) = p_{ij}^{(n)}$, donc $u_{ij} > 0$.

La relation $\sim$ est évidemment réflexive et symétrique. Pour montrer la transitivité il suffit de vérifier que si $i \mapsto j \mapsto k$, alors $i \mapsto k$. Mais, si $i \mapsto j \mapsto k$, d'après ce qui précède il existe $n \geq 0$ avec $p_{ij}^{(n)} > 0$ et $m \geq 0$ avec $p_{jk}^{(m)} > 0$. Comme $P^{n+m} = P^n P^m$, il vient

$$p_{ik}^{(n+m)} = \sum_{l \in E} p_{il}^{(n)} p_{lk}^{(m)} \geq p_{ij}^{(n)} p_{jk}^{(m)} > 0,$$

d'où le résultat. $\square$

Comme $\sim$ est une relation d'équivalence, on peut considérer les classes d'équivalence associées, qui forment une partition de E et qui s'appellent *les classes* de la chaîne de Markov. La signification intuitive d'une classe est donnée dans la:

Proposition 1.5.3. *Soit C une classe et $T = \inf(n \geq 0 : X_n \notin C)$. Pour tout $i \in C$ on a*

$$I\!P_i(T < \infty, \text{ il existe } n > T \text{ avec } X_n \in C) = 0.$$

En d'autres termes, si la chaîne quitte une classe elle ne peut pas y revenir.

Preuve. Le résultat de l'énoncé équivaut à dire (puisque $I\!N$ et C sont dénombrables) que

$$I\!P_i(T < \infty, X_T = k, X_{T+n} = j) = 0 \qquad \forall n \in I\!N,\ \forall j \in C,\ \forall k \notin C \tag{1.5.2}$$

(puisque $X_T \notin C$ si $T < \infty$ par définition de T). Mais T est un temps d'arrêt, donc le membre de gauche ci–dessus est d'après la propriété de Markov forte:

$$I\!E_i\left(1_{\{T<\infty, X_T = k\}}\ I\!P_i(X_i \circ \theta^T = j | \mathcal{F}_T)\right) = I\!P_i(T < \infty, X_T = k)\ p_{kj}^{(n)} \leq I\!P_i(T_k < \infty)\ p_{kj}^{(n)}.$$

Si l'expression ci–dessus est > 0 on a $p_{kj}^{(n)} > 0$ et $I\!P_i(T_k < \infty) > 0$, donc $i \mapsto k \mapsto j$ et comme aussi $j \mapsto i$ (car $j \in C$) on a $k \in C$: par suite on a (1.5.2). $\square$

Définition 1.5.4. On appelle *période de l'état* i et on note d_i le PGCD de l'ensemble M_i des $n \geq 1$ tels que $p_{ii}^{(n)} > 0$, avec la convention $d_i = \infty$ si cet ensemble est vide.

Proposition 1.5.5. *Si $i \sim j$ on a $d_i = d_j$.*

Etant donnée une classe C de la chaîne, on appellera donc *période de la classe* la valeur commune des périodes de ses éléments.

Preuve. Supposons bien–sûr $i \neq j$. On a vu ci-dessus qu'il existe $n, m \geq 1$ tels que $a = p_{ij}^{(n)} > 0$ et $b = p_{ji}^{(m)} > 0$. En utilisant $P^{n+l+m} = P^n P^l P^m$, le même argument que dans la proposition 1.5.2 montre que $p_{ii}^{(n+l+m)} \geq abp_{jj}^{(l)}$. Donc $l \in M_j$ implique $n+m+l \in N_i$, et aussi évidemment $n+m \in M_i$: donc d_i divise $n+m$ et $n+m+l$, donc l, pour tout $l \in M_j$, donc finalement d_i divise d_j. Symétriquement, d_i divise d_j, d'où le résultat. □

Proposition 1.5.6. *Soit C une classe de période d finie. On peut diviser C en d sous–classes non–vides $C_0, \dots, C_{d-1}$ telles que, si $i \in C_r$, alors $I\!P_i$–presque sûrement la chaîne ne visite C_l qu'en des instants n tels que $n+r = l$ modulo d.*

Preuve. Soit i_0 fixé arbitrairement dans C. Si $j \in C$, il existe $m > 0$ avec $p_{ji_0}^{(m)} > 0$, et on note r_j l'unique entier dans $\{0, 1, \dots, d-1\}$ tel que $m = r_j$ modulo d. Pour tout $n > 0$ avec $p_{i_0 j}^{(n)} > 0$, on a $p_{i_0 i_0}^{(n+m)} > 0$, donc d divise $n+m$ et donc $n = r_j$ modulo d. En d'autres termes, sous $I\!P_{i_0}$ la chaîne ne visite j qu'en des instants de la forme $r_j + nd$ pour $n \in I\!N$, presque sûrement.

On note alors C_r l'ensemble des $j \in C$ tels que $r_j = r$. Les C_r sont deux–à–deux disjoints, de réunion égale C. De plus, si la chaîne part de i_0, à l'instant 1 elle est dans C_1, puis à l'instant 2 dans C_2, puis dans C_3, ... , puis dans C_{d-1}, puis C_0, puis C_1, etc..., jusqu'à ce qu'elle sorte de C, auquel cas elle n'y revient plus d'après la proposition 1.5.3. Cela montre à l'évidence la dernière partie de l'énoncé, et aussi que chaque sous–classe C_r est non vide: en effet C_0 contient au moins i_0, et si C_r était vide pour un $r = 1, \dots, d-1$ la chaîne aurait quitté C $I\!P_{i_0}$–p.s. avant l'instant r, donc ne pourrait pas revenir en i_0, ce qui contredirait le fait que $d = d_{i_0}$ est fini. □

1.5.3 Seconde classification

Définition 1.5.7. Un état i est dit *récurrent* si $f_{ii} = I\!P_i(T_i < \infty) = 1$, et *transient* sinon.

Lemme 1.5.8. *On a $f_{ij}^{(n+1)} = f_{ij} f_{jj}^{(n)}$ et $u_{ij} = \delta_{ij} + \sum_{n\geq 1} f_{ij}^{(n)}$.*

Preuve. On a $T_j^{n+1} = T_j + T_j^n \circ \theta^{T_j}$ sur l'ensemble $\{T_j < \infty\}$, et $T_j^{n+1} = \infty$ sur $\{T_j = \infty\}$. On déduit alors, de la propriété de Markov forte et du fait que $X_{T_j} = j$ si $T_j < \infty$, que

$$
\begin{aligned}
f_{ij}^{(n+1)} &= I\!P_i(T_j < \infty, T_j^n \circ \theta^{T_j} < \infty) = I\!E_i(1_{\{T_j<\infty\}}\, I\!P_i(T_j^n \circ \theta^{T_j} < \infty | \mathcal{F}_{T_j})) \\
&= I\!E_i(1_{\{T_j<\infty\}}\, I\!P_{X_{T_j}}(T_j^n < \infty)) = I\!E_i(1_{\{T_j<\infty\}})\, f_{jj}^{(n)} = f_{ij} f_{jj}^{(n)}.
\end{aligned}
$$

Par ailleurs, N_j étant une variable à valeurs entières, on a $I\!E_i(N_j) = \sum_{n\geq 1} I\!P_i(N_j \geq n)$. Comme $u_{ij} = I\!E_i(N_j)$ et comme $\{N_j \geq n\} = \{T_j^n < \infty\}$ si $X_0 \neq j$ et $\{N_j \geq n\} = \{T_j^{n-1} < \infty\}$ si $X_0 = j$, on obtient immédiatement la seconde propriété énoncée. □

Théorème 1.5.9. *(1) On a les équivalences*

$$
i \text{ récurrent} \iff u_{ii} = +\infty \iff I\!P_i(N_i = \infty) = 1. \tag{1.5.3}
$$

Dans ce cas, si $j \sim i$ alors j est aussi récurrent, et on a $f_{ij} = 1$ et $u_{ij} = \infty$ et $I\!P_i(N_j = \infty) = 1$.

(2) On a les équivalences

$$i \text{ transient} \iff u_{ii} < +\infty \iff I\!P_i(N_i = \infty) = 0. \tag{1.5.4}$$

Dans ce cas, si $i \sim j$ alors j est transient et $u_{ij} < \infty$ et $I\!P_i(N_j = \infty) = 0$.

Ainsi, la récurrence et la transience sont des *propriétés de classe*: on dira que la classe C est récurrente (resp. transiente) si ses points sont récurrents (resp. transients).

Preuve. D'après le lemme, $f_{ii}^{(n)} = (f_{ii})^n$ et $u_{ii} = \sum_{n\geq 0}(f_{ii})^n$, tandis que $I\!P_i(N_i = \infty) = \lim_n f_{ii}^{(n)}$: on a donc les équivalences (1.5.3) et (1.5.4).

Supposons i récurrent et $j \sim i$ avec $j \neq i$. Comme dans la proposition 1.5.5 il existe n et m avec $a = p_{ij}^{(n)} > 0$ et $b = p_{ji}^{(m)} > 0$, donc $p_{jj}^{(n+m+l)} \geq abp_{ii}^{(l)}$ pour tout $l \geq 0$. Par suite $u_{jj} \geq abu_{ii} = \infty$, et donc j est aussi récurrent. On a de même $u_{ij} \geq au_{jj} = \infty$. Enfin d'après la propriété de Markov forte, on a

$$I\!P_j(T_i < \infty, T_j \circ \theta^{T_i} = \infty) = I\!E_j(1_{\{T_i<\infty\}}\ I\!P_j(T_j \circ \theta^{T_j} = \infty | \mathcal{F}_{T_i})) = f_{ji}(1 - f_{ij}).$$

Mais (1.5.3) pour j donne $\{N_j = \infty, T_i < \infty\} = \{T_i < \infty\}$ $I\!P_j$–p.p., donc a fortiori $I\!P_j(T_i < \infty, T_j \circ \theta^{T_i} = \infty) = 0$. Donc $f_{ji}(1 - f_{ij}) = 0$, et comme $f_{ji} > 0$ on a $f_{ij} = 1$.

Enfin, comme $I\!E_i(N_j) = u_{ij} = f_{ij}u_{jj}$ si $j \neq i$, la dernière assertion de (2) découle de ce qui précède. □

Corollaire 1.5.10. *Soit C une classe récurrente, et μ un probabilité sur E ne chargeant que C. On a alors $I\!P_\mu(X_n \in C, \forall n) = 1$ et $I\!P_\mu(N_i = \infty) = 1$ pour tout $i \in C$.*

Preuve. Comme $I\!P_\mu(A) = \sum_{j\in C} \mu_j I\!P_j(A)$ pour tout $A \in \mathcal{F}$, il suffit de montrer les assertions si $\mu = \varepsilon_j$ pour un $j \in C$: la seconde assertion vient de la proposition précédente, et la première de la combinaison de la seconde assertion et de la proposition 1.5.3. □

1.6 Chaînes discrètes: propriétés ergodiques

Ce qu'on appelle propriétés ergodiques pour une chaîne de Markov concerne le comportement à l'infini, soit de la chaîne elle-même, soit de ses probabilités de transition P^n.

On part encore d'une chaîne de Markov $\mathbf{X}$ avec un espace d'état fini ou dénombrable, et on utilise toutes les notations du paragraphe précédent. On commense par étudier les chaînes dites *irréductibles*, i.e. qui ne possèdent qu'une seule classe.

Proposition 1.6.1. *Si $\mathbf{X}$ est une chaîne irréductible récurrente* (i.e., tous ses états sont récurrents), *toute fonction surharmonique finie est constante.*

Preuve. Soit f surharmonique finie. D'après la proposition 1.4.4, le processus $f(X_n)$ est une surmartingale positive sous $I\!P_i$, donc pour tout temps d'arrêt $I\!P_i$–p.s. fini T on a $I\!E_i(f(X_T)) \leq I\!E_i(f(X_0)) = f(i)$. En particulier, comme $I\!P_i(T_j < \infty) = 1$ pour tout j et $f(X_{T_j}) = f(j)$ si $T_j < \infty$, on a $f(j) \leq f(i)$. □

Théorème 1.6.2. *Si* $\mathbf{X}$ *est une chaîne irréductible récurrente, la mesure* $\mu = (\mu_i)$ *définie par*

$$\mu_i = \mathbb{E}_{i_0}\left(\sum_{0\le n<T_{i_0}} 1_{\{i\}}(X_n)\right), \tag{1.6.1}$$

où i_0 *est un point arbitraire de* E*, est invariante et vérifie* $0 < \mu_i < \infty$*. De plus, une mesure est sous–invariante si et seulement si elle est le produit de* μ *par une constante arbitraire dans* $[0,\infty]$.

Cela montre en particulier que si μ' est la mesure associée par (1.6.1) à un autre état i_0', alors $\mu_i' = c\mu_i$ pour tout i, pour une constante $c \in]0,\infty[$. Cela montre aussi que toute mesure sous–invariante est invariante.

Preuve. a) Soit η une mesure sous–invariante. Pour tout n on a $\eta P^n \le \eta$, donc $\eta_i \ge \sum_j \eta_j p_{ji}^{(n)}$. Pour tout couple (i,j) il existe n avec $p_{ji}^{(n)} > 0$, puisque la chaîne est irréductible: donc $\eta_j > 0 \ \Rightarrow\ \eta_i > 0$, et aussi $\eta_i < \infty \ \Rightarrow\ \eta_j < \infty$. Par suite, si η n'est ni identiquement nulle ni identiquement infinie (i.e. $\eta_i = \infty$ pour tout i), on a $0 < \eta_i < \infty$ pour tout i.

b) Montrons maintenant l'unicité de la mesure sous–invariante (ni identiquement nulle ni identiquement infinie), à une constante multiplicative près. Posons

$$\widehat{p}_{ij} = \frac{\eta_j p_{ji}}{\eta_i}, \qquad \widehat{p}_{ij}^{(n)} = \frac{\eta_j p_{ji}^{(n)}}{\eta_i}.$$

La matrice $\widehat{P} = (\widehat{p}_{ij})$ est à éléments positifs et la somme de chaque ligne est ≤ 1 puisque η est sous–invariante. Considérons un point Δ “extérieur” à E, et posons $E_\Delta = E \cup \{\Delta\}$. Ensuite, introduisons la matrice $\overline{P} = (\overline{p}_{ij})_{i,j\in E_\Delta}$ suivante:

$$\overline{p}_{ij} = \begin{cases} \widehat{p}_{ij} & \text{si } \ i,j \in E \\ 1 - \sum_{j\in E} \widehat{p}_{ij} & \text{si } \ i \in E, j = \Delta \\ 0 & \text{si } \ i = \Delta, j \in E \\ 1 & \text{si } \ i = j = \Delta \end{cases} \tag{1.6.2}$$

Cette matrice est associée à une chaîne de Markov $\overline{\mathbf{X}}$ à valeurs dans E_Δ, et sa puissance $n$ième est donnée par (1.6.2) également, à condition de remplacer $\widehat{p}_{ij}$ par $\widehat{p}_{ij}^{(n)}$. Le potentiel associé $\overline{u}_{ij}$ vérifie $\overline{u}_{ij} = \eta_j u_{ji}/\eta_i = \infty$ (car la chaîne initiale $\mathbf{X}$ est irréductible récurrente) pour tous $i,j \in E$, ainsi que $\overline{u}_{\Delta\Delta} = \infty$ et $\overline{u}_{\Delta i} = 0$ si $i \in E$: en comparant au théorème 1.5.9, on en déduit que $\overline{\mathbf{X}}$ admet deux classes récurrentes, à savoir E et $\{\Delta\}$: étant donné le corollaire 1.5.10, si on part d'un point de E, on ne sort jamais de E, ce qui signifie qu'en fait la somme des lignes de la matrice $\widehat{P}$ vaut 1 (et la construction de la chaîne $\overline{\mathbf{X}}$ sur un espace étendu E_Δ était inutile...)

Supposons maintenant que ν soit une autre mesure sous–invariante pour P, encore non identiquement nulle, ni identiquement infinie. Si $f(i) = \nu_i/\eta_i$ on a

$$\widehat{P}f(i) = \sum_j \frac{\eta_j p_{ji}}{\eta_i}\,\frac{\nu_j}{\eta_j} = \frac{(\nu P)_i}{\eta_i} \le f(i)$$

et f est surharmonique pour la chaîne récurrente irréductible associée à la transition $\widehat{P}$. Donc f est constante par la proposition précédente, et ν est un multiple de η.

c) Il nous reste à montrer que (1.6.1) définit une mesure invariante, automatiquement ni identiquement nulle, ni identiquement infinie, puisque par construction $\mu_{i_0} = 1$. On a

$$\begin{aligned}
(\mu P)_i &= \mathbb{E}_{i_0}\left(\sum_{0\leq n<T_{i_0}} p_{X_n,i}\right) = \sum_{n\geq 0} \mathbb{E}_{i_0}\left(p_{X_n,i}\, 1_{\{n<T_{i_0}\}}\right) \\
&= \sum_{n\geq 0} \mathbb{E}_{i_0}\left(1_{\{i\}}(X_{n+1})\, 1_{\{n<T_{i_0}\}}\right) \qquad \text{(propriété de Markov)} \\
&= \mathbb{E}_{i_0}\left(\sum_{0\leq n<T_{i_0}} 1_{\{i\}}(X_{n+1})\right) \\
&= \mathbb{E}_{i_0}\left(\sum_{0\leq n<T_{i_0}} 1_{\{i\}}(X_n)\right) = \mu_i,
\end{aligned}$$

où l'avant–dernière égalité provient de ce que $\sum_{0\leq n<T_{i_0}} 1_{\{i\}}(X_{n+1}) = \sum_{0\leq n<T_{i_0}} 1_{\{i\}}(X_n)$ $\mathbb{P}_{i_0}$–p.s. (distinguer les cas $i = i_0$ et $i \neq i_0$), et la dernière est une nouvelle application de (1.6.1). □

Corollaire 1.6.3. *Si* $\mathbf{X}$ *est une chaîne irréductible récurrente et si* μ *est une mesure invariante ni identiquement nulle ni identiquement infinie, on a deux cas possibles:*

a) $\mu(E) < \infty$ *et* $\mathbb{E}_i(T_i) < \infty$ *pour tout* $i \in E$,

b) $\mu(E) = \infty$ *et* $\mathbb{E}_i(T_i) = \infty$ *pour tout* $i \in E$.

Preuve. La mesure μ de (1.6.1) vérifie clairement $\mu(E) = \mathbb{E}_{i_0}(T_{i_0})$. A cause de "l'unicité", toute les mesures invariantes ni identiquement nulles ni identiquement infinies sont simultanément de masse totale finie (resp. infinie), et le résultat est évident. □

Ce corollaire montre que le temps moyen de retour dans l'état i, soit $m_i = \mathbb{E}_i(T_i)$, joue un rôle important. Cela nous amène à une définition:

Définition 1.6.4. On dit que l'état i est *positif* si $m_i < \infty$, et est *nul* si $m_i = \infty$.

Cette terminologie curieuse vient du fait que c'est l'inverse $\frac{1}{m_i}$ qui va jouer un rôle dans la suite: il est positif (resp. nul) si i est positif (resp. nul).

Proposition 1.6.5. *(1) Tout état transient est nul.*

(2) Les états d'une même classe sont, soit tout positifs, soit tous nuls. On dit alors que la classe est positive*, resp.* nulle.

Preuve. Si i est transient on a $\mathbb{P}_i(T_i = \infty) > 0$, donc *a fortiori* $m_i = \infty$: cela donne (1), et pour (2), il suffit donc de considérer le cas d'une classe récurrente C. On a vu que dans ce cas, si on part d'un point de C, on ne sort jamais de C: si on se contente de considérer des mesures initiales portées par C, on peut donc restreindre l'espace d'état à C lui-même: on obtient ainsi à l'évidence une chaîne irréductible récurrente, et le résultat découle alors immédiatement du corollaire 1.6.3. □

Nous pouvons maintenant passer à un premier théorème ergodique.

Théorème 1.6.6. *Si i est un état de période d_i et si $m_i = \mathbb{E}_i(T_i)$ on a*

$$p_{ii}^{(nd)} \quad \to_{n\to\infty} \quad \begin{cases} \frac{d_i}{m_i} & \text{si } m_i < \infty \text{ (donc } d_i < \infty \text{ aussi)} \\ 0 & \text{si } m_i = \infty. \end{cases} \tag{1.6.3}$$

Comme $p_{ii}^{(n)} = 0$ quand n n'est pas un mutiple de d_i, on obtient ainsi le comportement asymptotique complet de la suite $p_{ii}^{(n)}$: elle tend vers 0 si $m_i = \infty$, et elle alterne entre 0 et une suite convergeant vers d_i/m_i si $m_i < \infty$.

Preuve. a) Si i est transient, on a $u_{ii} < \infty$, donc la suite $p_{ii}^{(n)}$ tend vers 0 quand $n \to \infty$, tandis que $m_i = \infty$ d'après ce qui précède: le résultat est donc démontré. Dans la suite on suppose donc i récurrent.

b) On pose $d = d_i$, $m = m_i$, et

$$h_r = \mathbb{P}_i(T_i = r), \qquad A = \{r > 0 : h_r > 0\}, \qquad B = \{r > 0 : p_{ii}^{(r)} > 0\}.$$

On sait que d est le PGCD de B, qui est non vide (car i est récurrent) et on va montrer que d est aussi le PGCD de A (qui est aussi non vide).

Pour cela, notons d_N (resp. d'_N) le PGCD de $\{r : 1 \leq r \leq N, p_{ii}^{(r)} > 0\}$ (resp. $\{r : 1 \leq r \leq N, h_r > 0\}$). La suite d_N décroit vers d, la suite d'_N décroit vers le PGCD de A, donc il suffit de montrer que $d_N = d'_N$ pour tout N. Comme $h_1 = p_{ii}$, c'est évident pour $N = 1$. Supposons alors que $d_N = d'_N$ pour un $N \geq 1$ donné. D'après la propriété de Markov forte, on a

$$p_{ii}^{(r)} = \mathbb{P}_i(T_i \leq r, X_r = i) = \mathbb{E}_i\left(1_{\{T_i \leq r\}}\, p_{ii}^{(r-T_i)}\right) = \sum_{s=1}^{r} h_s p_{ii}^{(r-s)}. \tag{1.6.4}$$

En particulier, on a

$$p_{ii}^{(N+1)} = h_{N+1} + \sum_{s=1}^{N} h_s p_{ii}^{(N+1-s)}.$$

Si alors $h_{N+1} > 0$, alors d_{N+1} et d'_{N+1} sont tous les deux le PGCD de $\{d_N, N+1\}$. Si ensuite $h_{N+1} = p_{ii}^{(N+1)} = 0$, on a $d'_{N+1} = d_{N+1} = d_N$. Si enfin $h_{N+1} = 0 < p_{ii}^{(N+1)}$, alors $d'_{N+1} = d_N$ tandis que d_{N+1} est le PGCD de $\{d_N, N+1\}$; mais d'après (1.6.4) il existe $s \leq N$ avec $h_s > 0$ et $p_{ii}^{(N+1-s)} > 0$, donc d_N divise s et aussi $N+1-s$, donc *a fortiori* $N+1$, et on a donc $d_{N+1} = d_N$: on a ainsi montré le résultat voulu.

c) D'après (b), on a $h_r = p_{ii}^{(r)} = 0$ si r n'est pas un multiple de d. Donc si $k_n = h_{nd}$ et $p_n = p_{ii}^{(nd)}$, la relation (1.6.4) s'écrit aussi

$$p_n = \sum_{r=1}^{n} k_r p_{n-r}. \tag{1.6.5}$$

Posons $\lambda = \limsup_n p_n$. Il existe une suite n_N croissant vers l'infini, telle que $p_{n_N} \to \lambda$. Donc (1.6.5) donne pour tout q fixé:

$$\begin{aligned}
\lambda &= \lim_N \left(k_q p_{n_N - q} + \sum_{1 \le n \le n_N, n \neq q} k_n p_{n_N - n} \right) \\
&= k_q \liminf_N \; p_{n_N - q} + \limsup_N \sum_{1 \le n \le n_N, n \neq q} k_n p_{n_N - n} \\
&\le k_q \liminf_N \; p_{n_N - q} + \lambda(1 - k_q)
\end{aligned}$$

en appliquant le lemme de Fatou aux "fonctions" $n \mapsto p_{n_N - n} 1_{\{n \neq q\}}$, positives et bornées par 1, et à la "mesure" $n \mapsto k_q$, qui est une probabilité car $f_{ii} = \sum_{r \ge 1} k_r = 1$ (car i est récurrent). Cela implique que

$$p_{n_N} \to \lambda \;\text{ et }\; k_q > 0 \quad \Longrightarrow \quad p_{n_N - q} \to \lambda.$$

En itérant cette propriété, on voit facilement que

$$q = \alpha_1 q_1 + \ldots + \alpha_l q_l, \quad \alpha_i \in \mathbb{N}, \quad k_{q_i} > 0 \quad \Longrightarrow \quad p_{n_N - q} \to \lambda. \tag{1.6.6}$$

d) D'après (b), d est le PGCD de A, donc le PGCD de $\{r : k_r > 0\}$ égale 1. Il existe donc un nombre fini $s_1 < \ldots < s_r$ d'entiers tels que $k_{s_i} > 0$, et que le PGCD de $\{s_1, \ldots, s_r\}$ soit 1. D'après l'identité de Bezout, cela entraine l'existence d'entiers relatifs $\beta_i \in \mathbb{Z}$ tels que $\beta_1 s_1 + \ldots + \beta_r s_r = 1$. Si $M = \sup_i |\beta_i|$ et $s = s_1 + \ldots + s_r$ et $v_0 = Ms^2$, tout $v \ge v_0$ s'écrit $v = sq + v'$ avec $q, v' \in \mathbb{N}$ et $v' \le s - 1$, et donc $q \ge Ms$. En utilisant $\beta_1 s_1 + \ldots + \beta_r s_r = 1$ on obtient alors $v = \sum_{i=1}^{r} (q + v'\beta_i) s_i$, et $q + v'\beta_i \in \mathbb{N}$, et (1.6.6) implique

$$v \ge v_0 \qquad \Longrightarrow \qquad p_{n_N - v} \to \lambda. \tag{1.6.7}$$

e) Posons $l_i = \sum_{n=i+1}^{\infty} k_n$. Comme i est récurrent, on a $l_0 = 1$, et comme $\mathbb{P}_i$–p.s. la variable T_i/d ne prend que des valeurs entières on a

$$\frac{m}{d} = \mathbb{E}_i\left(\frac{T_i}{d}\right) = \sum_{n \ge 1} \mathbb{P}_i(T_i \ge nd) = \sum_{n=0}^{\infty} l_n. \tag{1.6.8}$$

Par ailleurs (1.6.5) donne $p_n = \sum_{r=1}^{n} (l_{r-1} - l_r) p_{n-r}$, d'où l'on déduit immédiatement

$$\sum_{r=0}^{n} l_r p_{n-r} = l_0 p_0 = 1 \qquad \forall n \ge 0. \tag{1.6.9}$$

Nous distinguons maintenant deux cas:

(A) $m < \infty$: Par (1.6.8) on a $\sum_r l_r < \infty$, et $p_n \leq 1$, donc d'après le théorème de Lebesgue, en utilisant (1.6.7) et en écrivant (1.6.9) pour $n = n_N - v_0$, on obtient que $\sum_{r \geq 0} l_r \lambda = 1$, donc $\lambda = \frac{d}{m}$ par (1.6.8).

(B) $m = \infty$: D'après le Lemme de Fatou, en utilisant (1.6.7) et en écrivant (1.6.9) pour $n = n_N - v_0$, on obtient que $\sum_{r \geq 0} l_r \lambda \leq 1$, donc $\lambda = 0$ par (1.6.8).

Dans le cas (B), en se rappelant que $\lambda = \limsup_n p_n$, on voit que $p_n \to 0$ et le résultat est démontré. Dans le cas (A), on pose $\lambda' = \liminf_n p_n$. On peut répéter la preuve précédente en remplaçant partout λ par λ' (il existe en effet une suite n'_N telle que $p_{n'_N} \to \lambda'$) et en changeant les sens des inégalités. On obtient finalement que $\lambda' = \frac{d}{m}$ également, donc $\lambda = \lambda'$, et la suite p_n converge vers $\frac{d}{m}$. $\square$

Voici maintenant la description complète du comportement asymptotique des éléments de la matrice P^n:

Corollaire 1.6.7. *On a les propriétés suivantes, avec $m_i = \mathbb{E}_i(T_i)$ et d_i la période de i:*

(a) Si j est un état nul (i.e. $m_j = \infty$), alors $p_{ij}^{(n)} \to 0$ pour tout $i \in E$.

(b) Si j est un état positif et si $i \sim j$, il existe $r_{ij} \in \{0, \dots, d_j - 1\}$ tel que $p_{ij}^{(nd_j + r_{ij})} \to \frac{d_j}{m_j}$, tandis que $p_{ij}^{(nd_j + r)} = 0$ pour tout $r \in \{0, \dots, d_j - 1\}$ avec $r \neq r_{ij}$.

(c) Si j est un état positif et si i n'est pas dans la même classe que j, pour tout $r \in \{0, \dots, d_j - 1\}$ on a $p_{ij}^{(nd_j + r)} \to f_{ij}(r) \frac{d_j}{m_j}$, où $f_{ij}(r) = \sum_{n \geq 0} \mathbb{P}_i(T_j = nd_j + r)$.

Preuve. Si on pose $h_{ij}^{(n)} = \mathbb{P}_i(T_j = n)$, on a d'après la propriété de Markov forte

$$p_{ij}^{(n)} = \mathbb{P}_i(T_j \leq n, X_n = j) = \mathbb{E}_i\left(1_{\{T_j \leq n\}}\, p_{jj}^{(n - T_i)}\right) = \sum_{s=1}^{n} h_{ij}^{(s)} p_{jj}^{(n-s)}. \tag{1.6.10}$$

Par ailleurs $\sum_{s \geq 1} h_{ij}^{(s)} = f_{ij} \leq 1$.

Si j est nul on a $p_{jj}^{(n)} \to 0$, donc le théorème de Lebesgue appliqué à (1.6.10) donne immédiatement le résultat (a). Dans le cas (b), d'après la proposition 1.5.6 il existe $r_{ij} \in \{0, \dots, d_j - 1\}$ tel que $p_{ij}^{(nd_j + r)} = 0$ et $h_{ij}^{(nd_j + r)} = 0$ pour tout $r' \in \{0, \dots, d_j - 1\}$ avec $r' \neq r_{ij}$. On a aussi $p_{jj}^{(n)} = 0$ si n n'est pas un multiple de d_j, de sorte qu'on peut réécrire (1.6.10) ainsi:

$$p^{(nd_j + r_{ij})} = \sum_{s=1}^{n} h_{ij}^{(sd_j + r_{ij})} p_{jj}^{((n-s)d_j)},$$

et comme $p_{jj}^{((n-s)d_j)} \to \frac{d_j}{m_j}$ un nouvelle application du théorème de Lebesgue donne le résultat.

Enfin dans le cas (c), on peut réécrire (1.6.10) ainsi:

$$p^{(nd_j + r)} = \sum_{s=1}^{n} h_{ij}^{(sd_j + r)} p_{jj}^{((n-s)d_j)},$$

pour tout $r \in \{0, \dots, d_j - 1\}$. Il suffit encore une fois d'appliquer le théorème de Lebesgue pour obtenir le résultat. □

Le comportement décrit ci–dessus est plutôt compliqué, à cause de la période. On obtient un résultat moins fort, mais bien plus simple, si on considère les moyennes de Césaro:

$$\Pi_N = \frac{1}{N} \sum_{n=1}^{N} P^n, \quad \text{de composantes} \quad \pi_{ij}^{(N)} = \frac{1}{N} \sum_{n=1}^{N} p_{ij}^{(n)}. \tag{1.6.11}$$

Corollaire 1.6.8. (Théorème ergodique en moyenne): *Lorsque $N \to \infty$, la matrice Π_N converge (élément par élément) vers la matrice Π de composantes $\pi_{ij} = \frac{f_{ij}}{m_j}$ (avec $\pi_{ij} = 0$ si $m_j = \infty$).*

Preuve. Rappelons que si une suite (a_n) converge vers a, alors $\frac{1}{N}\sum_{n=1}^{N} a_n$ converge *a fortiori* vers a. La convergence $\pi_{ij}^{(n)} \to \pi_{ij}$ découle alors immédiatement du corollaire précédent (noter que dans le cas (c) de ce corollaire, on a $f_{ij} = \sum_{r=0}^{d_j-1} f_{ij}(r)$). □

Pour terminer ce paragraphe, on va déduire des résultats précédents l'existence (ou non...) d'une ou de plusieurs probabilités invariantes pour la chaîne.

Théorème 1.6.9. *Si* **X** *est une chaîne irréductible, il existe une probabilité invariante si et seulement si la chaîne est positive. Dans ce cas la probabilité invariante est unique et donnée par $\mu_i = \frac{1}{m_i}$.*

Preuve. Si μ est une probabilité invariante, on a $\mu P^n = \mu$ pour tout n, donc aussi $\mu\Pi_N = \mu$. Par application du théorème de Lebesgue on obtient alors $\mu\Pi = \mu$: comme $\pi_{ij} = \frac{1}{m_j}$, on en déduit $\mu_j = \frac{1}{m_j}$. Cela implique en particulier qu'on n'a pas $m_j = \infty$ pour tout j, donc la chaîne est positive (et $m_j < \infty$ pour tout j, cf. proposition 1.6.5).

Inversement si la chaîne est positive, l'existence d'une probabilité invariante découle du corollaire 1.6.3. □

Le cas non irréductible est évidemment un peu plus compliqué. On note C_α les *classes positives*, indicées par un ensemble A. La démonstration du résultat suivant, facile, est laissée au lecteur.

Théorème 1.6.10. *Il existe au moins une probabilité invariante si et seulement si l'ensemble A est non vide. Dans ce cas, si on note $\mu(\alpha) = (\mu(\alpha)_i)$, avec $\mu(\alpha)_i = \frac{1}{m_i}$ pour $i \in C_\alpha$ et $\mu(\alpha)_i = 0$ sinon, une mesure μ est une probabilité invariante si et seulement si elle s'écrit*

$$\mu = \sum_{\alpha \in A} c_\alpha \mu(\alpha),$$

où les c_α vérifient $c_\alpha \geq 0$ et $\sum_{\alpha\in A} c_\alpha = 1$.

Ainsi, les $\mu(\alpha)$ sont les probabilités invariantes *extrémales* dans l'ensemble – convexe – de toutes les probabilités invariantes. La probabilité invariante est unique si et seulement s'il existe une classe positive et une seule. S'il n'y a pas de classe positive, il n'y a pas de probabilité invariante.

Enfin, on peut noter le résultat élémentaire suivant:

Proposition 1.6.11. *Si l'espace E est fini, il existe au moins une classe positive (donc récurrente) et donc une probabilité invariante.*

Preuve. On a $\sum_{j\in E} \pi_{ij}^{(N)} = 1$, et comme E est fini on peut passer à la limite en N (cf. corollaire 1.6.8) et on a $\sum_j \pi_{ij} = 1$: il existe donc au moins un point j tel que $\pi_{ij} > 0$, donc $m_j < \infty$. $\square$

Si E est infini, on peut n'avoir que des états transients: par exemple si $E = \mathbb{N}$ et $p_{ij} = 1$ si $j = i+1$ et $p_{ij} = 0$ sinon. On peut aussi n'avoir que des états récurrents nuls: par exemple si $E = \mathbb{Z}$ et $p_{ij} = \frac{1}{2}$ si $j = i-1$ ou $j = i+1$, et $p_{ij} = 0$ sinon; dans ce cas la chaîne $\mathbf{X}$ est une “marche aléatoire de Bernoulli”, i.e. $X_n = X_{n-1} + Y_n$ où les Y_n sont i.i.d. avec $\mathbb{P}(Y_n = 1) = \mathbb{P}(Y_n = -1) = \frac{1}{2}$; il est clair que cette chaîne est irréductible, de période 2, et comme $p_{00}^{(2n)} = C_{2n}^n \ \frac{1}{2^n}$ on vérifie que $u_{00} = \infty$, de sorte que la chaîne est récurrente; enfin la mesure μ définie par $\mu_i = 1$ pour tout i est clairement invariante, et comme elle est de masse totale infinie le corollaire 1.6.3 montre que la chaîne est nulle.

Chapitre 2

Processus de population à temps discret

2.1 Chaînes de Markov de vie et de mort

Les chaînes de Markov (et, à dire vrai, les processus de Markov à voir ci–après encore plus) sont bien adaptés à la description de l'évolution temporelle des populations, en raison du caractère essentiellement aléatoire des naissances et des morts dans une population, et du fait que bien souvent la survenue d'une naissance ou d'une mort dépend de l'effectif total de la population à l'instant considéré, mais pas de ce qui s'est passé avant. On remarquera que le terme "population" est commode, mais peut aussi servir à décrire le nombre de clients dans une file d'attente (une "mort" est un départ, une "naissance" est une arrivée), et bien d'autres choses encore.

La variable X_n est l'effectif de la population à la génération n, et est donc à valeurs dans $E = \mathbb{N}$. Si $X_n = m$, chaque individu vivant à l'instant n meurt (puisqu'on passe de la génération n à la génération $n+1$), en donnant naissance à un nombre aléatoire de descendants, indépendamment des autres: ces variables sont toutes de même loi (dépendant éventuellement de l'effectif m), qu'on note $\nu(m) = (\nu(m)_i)_{i\in\mathbb{N}}$. En outre il y a possibilité "d'immigration" d'un nombre aléatoire d'individus, indépendant de tout le reste, mais dont la loi peut aussi dépendre de $m$et est notée $\eta(m) = (\eta(m)_i)_{i\in\mathbb{N}}$.

Une manière commode de construire mathématiquement cette chaîne est la suivante: sur un espace Ω (par exemple l'espace canonique adéquat) on se donne une famille $(X_0, (Y_{n,m,q} : n,m,q \geq 0), (Z_{n,m} : n,m \geq 0))$ de fonctions à valeurs dans $\mathbb{N}$, et $\mathcal{F}$ est la tribu rendant mesurables toutes ces fonctions. Pour chaque probabilité μ sur $\mathbb{N}$, on note $\mathbb{P}_\mu$ (et $\mathbb{P}_i$, lorsque μ est la masse de Dirac en i) l'unique probabilité sur $(\Omega, \mathcal{F})$ sous laquelle toutes ces variables sont indépendantes, X_0 est de loi μ, les $Y_{n,m,q}$ sont de lois $\nu(m)$, et les $Z_{n,m}$ de lois $\eta(m)$. Partant de X_0, on définit les X_n par récurrence ainsi:

$$X_{n+1} = \sum_{q=1}^{m} Y_{n,m,q} + Z_{n,m} \quad \text{si} \quad X_n = m : \ m = 0, 1, \ldots, \tag{2.1.1}$$

avec la convention qu'une somme "vide" vaut 0.

On remarque que X_{n+1} est une fonction de X_n et des variables $(Y_{n,m,q}, Z_{n,m} : m, q \in \mathbb{N})$. Donc, à cause des propriétés d'indépendance, il est clair qu'on obtient ainsi une chaîne de Markov relativement à la filtration engendrée par le processus (X_n), et aussi par rapport à la filtration plus grosse, définie par $\mathcal{F}_n = \sigma(X_0, Y_{l,m,q}, Z_{l,m} : q, m \in \mathbb{N}, l \leq n-1)$. La probabilité de transition de cette chaîne est donnée par

$$p_{ij} = \sum_{q_1,q_2,\ldots,q_i,r \in \mathbb{N}: q_1+\ldots+q_i+r=j} \nu(i)_{q_1}\nu(i)_{q_2}\ldots\nu(i)_{q_i}\eta(i)_r, \tag{2.1.2}$$

(où un produit "vide" vaut 1). Dit autrement, la loi $(p_{ij} : j \in \mathbb{N})$ est la convolution de $\eta(i)$ et de i fois $\nu(i)$.

Lorsque les probabilités $\nu(i)$ et $\eta(i)$ sont indépendantes de i, on dit que la population est *densité–indépendante*, et on limitera essentiellement à cette situation dans la suite. On dit qu'il n'y a *pas d'immigration* lorsque $\eta(i)_0 = 1$ (i.e., $\eta(i)$ est la masse de Dirac en 0) pour tout i.

Les chaînes précédentes décrivent l'évolution d'une population unique, et isolée lorsqu'il n'y a pas d'immigration. On peut bien–sûr compliquer le modèle de façon à décrire l'évolution de plusieurs populations en interaction, ou en compétition. Donnons deux exemples:

1) Deux populations en interaction, isolées: Il n'y a pas d'immigration, et la génération n comprend $X_n^{(1)}$ individus de type 1, et $X_n^{(2)}$ individus de type 2. La loi de reproduction d'un individu de type $k = 1, 2$ si la population comprend m_1 individus 1 et m_2 individus 2 est $\nu(m_1, m_2)^k = (\nu(m_1, m_2)_i^{(k)})_{i \in \mathbb{N}}$. Si alors $(Y_{n,m_1,m_2,q}^{(k)} : n, m_1, m_2, q \in \mathbb{N}, k = 1, 2)$ sont des variables indépendantes entre elles et aussi des effectifs initiaux $X_0^{(k)}$, chaque $Y_{n,m_1,m_2,q}^{(k)}$ étant de loi $\nu(m_1, m_2)^{(k)}$, on peut poser

$$X_n^{(k)} = \sum_{q=1}^{m_k} Y_{n,m_1,m_2,q}^{(k)} \quad \text{si} \quad X_n^{(1)} = m_1,\ X_n^{(2)} = m_2 :\ m_1, m_2 \in \mathbb{N},\ k = 1, 2. \tag{2.1.3}$$

Cela définit une chaîne de Markov $(X_n^{(1)}, X_n^{(2)})$ à valeurs dans $\mathbb{N}^2$, de transition

$$p_{i_1 i_2, j_1 j_2} = \sum_{q_l, r_m \in \mathbb{N}:\ q_1+\ldots+q_{i_1}=j_1, r_1+\ldots+r_{i_2}=j_2} \prod_{l=1}^{i_1} \nu(i_1, i_2)_{q_l}^{(1)} \prod_{m=1}^{i_2} \nu(i_1, i_2)_{r_m}^{(2)}. \tag{2.1.4}$$

2) Population isolée avec mutation: La situation est la même que ci–dessus, sauf que chaque descendant d'un individu de type 1 (resp. 2) peut "muter" et donc être de type 2 (resp. 1), avec la probabilité α_1 (resp. α_2), les mutations étant indépendantes de tout le reste. La chaîne $(X_n^{(1)}, X_n^{(2)})$est encore markovienne, avec la transition

$$\begin{aligned} p_{i_1 i_2, j_1 j_2} \quad = \quad & \sum_{A(i_1,i_2,j_1,j_2)} \prod_{l=1}^{i_1} \nu(i_1, i_2)_{q_l}^{(1)} C_{q_l} \alpha_1^{q'_l} (1-\alpha_1)^{q_l - q'_l} \\ & \prod_{m=1}^{i_2} \nu(i_1, i_2)_{r_m}^{(2)} C_{r_m} \alpha_2^{r'_m o} (1-\alpha_2)^{r_m - r'_m}, \end{aligned} \tag{2.1.5}$$

où $A(i_1, i_2, j_1, j_2)$ est l'ensemble des familles d'entiers $((q_l, q'_l, r_m, r'_m) : l = 1, \ldots, i_1, m = 1, \ldots, i_2)$ telles que $0 \leq q'_l \leq q_l$ et $0 \leq r'_m \leq r_m$, et avec $j_1 = q'_1 + \ldots + q'_{i_1} + (r_1 - r'_1) + \ldots + (r_{i_2} - r'_{i_2})$ et $j_2 = (q_1 - q'_1) + \ldots + (q_{i_1} - q'_{i_1}) + r'_1 + \ldots + r'_{i_2}$.

On peut encore compliquer (ou rendre plus réalistes) les modèles en supposant une reproduction sexuée, avec des invidus mâles et femelles, ou éventuellement plus de 2 "sexes" (évolution du nombre de germes catalytiques faisant intervenir 3 ou 4 espèces chimiques par exemple).

Les problèmes qui se posent naturellement sont alors le comportement à l'infini: par exemple, y a-t'il extinction, ou un régime stationnaire, ou "explosion" pour une population simple, ou extinction d'un des deux types d'individus pour les exemples 1 et 2 ci–dessus ? D'autres problèmes plus complexes peuvent se poser, par exemple que se passe-t'il lorsque $n \to \infty$ lorsqu'on se place conditionnellement au fait que $X_n > 0$ (la population n'est pas éteinte à l'instant n): c'est l'objet de la "quasi–stationnarité" qu'on étudiera plus loin. On peut aussi étudier la limite "à grands effectifs": si X_n tend vers l'infini, existe-t'il une normalisation (une suite $a_n \to 0$ de nombres positifs) telle que $a_n X_n$ converge, en un sens convenable, vers une limite non triviale ?, ou étudier ce qui se passe si l'effectif initial $X_0 = p$ tend vers l'infini ? Ces problèmes, complexes, ne seront pas abordés ici.

2.2 La chaîne de Bienaymé–Galton–Watson

La chaîne de vie et de mort la plus simple est la chaîne de Bienaymé–Galton–Watson (nous abrégeons en BGW). C'est une chaîne sans immigration, densité–indépendante. On a $\nu(m) = \nu$ pour tout $m \geq 0$, et $\eta(m)$ est la masse de Dirac en 0. Par suite (2.1.1) s'écrit

$$X_{n+1} = \sum_{q=1}^{m} Y_{n,q} \quad \text{si} \quad X_n = m : \; m = 0, 1, \ldots, \tag{2.2.1}$$

où les $Y_{n,q}$ sont indépendantes entre elles et de X_0, et toutes de loi ν.

2.2.1 Résultats élémentaires

Une propriété, immédiate mais fondamentale, est que 0 *est un état absorbant*: si $X_n = 0$, alors $X_{n+1} = 0$. De manière équivalente, le temps d'atteinte de 0, soit $T = \inf(n \geq 0 : X_n = 0)$ est *le temps d'extinction*, au sens où $X_n = 0$ si et seulement si $n \geq T$. On s'intéresse alors de manière prioritaire à la loi du temps d'extinction, ou au moins aux *probabilités d'extinction*, qui sont les nombres $p_i = \mathbb{P}_i(T < \infty)$ pour $i \geq 1$ (bien–sûr $p_0 = 1$).

Une autre propriété importante de cette chaîne est la *propriété de branchement*. Une chaîne de Markov (X_n) à valeurs entières a cette propriété si sa loi (en tant que "processus") sous $\mathbb{P}_i$ est la même que la loi de $(Z_n^{(1)} + \ldots + Z_n^{(i)})_{n \in \mathbb{N}}$, où les $((Z_n^{(j)})_{n \in \mathbb{N}} : j = 1, 2, \ldots)$ sont des chaînes indépendantes entre elles, et qui ont toutes même loi que la chaîne (X_n) sous $\mathbb{P}_1$: en vertu de (2.2.1), c'est bien ce qui se passe ici (si $X_0 = i$ on peut écrire en fait $X_n = \sum_{j=1}^{i} Z_n^{(j)}$, où $Z_n^{(j)}$ représente le nombre d'individus vivant à l'instant n, et descendant du $j$ième individu initial).

Remarquons d'ailleurs que la propriété de branchement pour une chaîne de Markov est *caractéristique* des chaînes BGW; en effet la propriété de branchement implique en particulier que la loi $(p_{ij} : j \in I\!N)$ est la puissance $i$ième de convolution de la loi $(p_{1j} : j \in I\!N)$; en d'autres termes, on a (2.1.2) avec $\nu(m)_i = p_{1i}$ et $\eta(m)_0 = 1$, et la chaîne est BWG. Plus généralement, la propriété de branchement permet de ramener l'étude de $I\!P_i$ à celle de $I\!P_1$ ($I\!P_i$ est une sorte de puissance $i$ième de convolution de $I\!P_1$). Par exemple,

$$\text{la loi } (p_{ij}^{(n)})_{j\in I\!N} \text{ est la puissance } i\text{ième de convolution de la loi } (p_{1j}^{(n)})_{j\in I\!N}. \tag{2.2.2}$$

$$I\!P_i(T \leq n) = I\!P_1(T \leq n)^i \tag{2.2.3}$$

(pour ce denier fait, on remarque que le temps d'extinction de $Z_N^{(1)} + \ldots + Z_n^{(i)}$ est le maximum des temps d'extinction des $Z_n^{(j)}$ pour $1 \leq j \leq i$).

Du point de vue des propriétés de la chaîne, et en particulier du temps d'extinction, un certain nombre de cas sont triviaux:

a) Si $\nu_1 = 1$: on a $X_n = X_0$, l'effectif de la population ne varie pas.

b) Si $\nu_1 < 1$ et $\nu_0 + \nu_1 = 1$: on a $X_{n+1} \leq X_n$, l'effectif de la population décroît. On a $T = 1$ $I\!P_1$–p.s. si $\nu_0 = 1$, et $I\!P_1(T = k) = \nu_0 \nu_1^{k-1}$ si ν_0 (T sui une loi géométrique).

c) Si $\nu_1 < 1$ et $\nu_0 = 0$: on a $X_{n+1} \geq X_n$, l'effectif de la population croît, et tend p.s. vers $+\infty$, et on a $T = \infty$ $I\!P_1$–p.s.

Dans la suite on élimine ces trois cas, en supposant que

$$\nu_0 > 0, \qquad \nu_0 + \nu_1 < 1. \tag{2.2.4}$$

Proposition 2.2.1. *L'état* 0 *est récurrent positif et, sous (2.2.4), les autres états sont transients. Si de plus* d *est le PGCD des* $i \geq 1$ *tels que* $\nu_i > 0$ *(donc* $d \in I\!N^\star$ *sous (2.2.4)), les classes sont* $C = dI\!N^\star = \{nd : n = 1, 2, \ldots\}$*, plus tous les singletons* $\{i\}$ *avec* $i \notin C$.

Preuve. Tout état absorbant est récurrent positif, d'où la première assertion. D'après (2.1.2) ou (2.2.1), si $i \geq 1$ on a $p_{i0} \geq \nu_0^i > 0$, donc i mène à 0, alors que 0 ne mène pas pas à $i \geq 1$ puisque 0 est absorbant: donc l'état i ne peut pas être récurrent. Les $Y_{n,q}$ prennent p.s. leurs valeurs dans $\{i : \nu_i > 0\}$, qui est contenu dans $C \cup \{0\}$, de sorte que $X_n \in C \cup \{0\}$ p.s. également pour tout $n \geq 1$: par suite un état i ne peut mener qu'à 0 et aux points de C, et tout singleton non contenu dans C est une classe. Enfin comme tout point de C s'écrit $n_1 + \ldots + n_l$, avec $n_k \geq 1$ et $\nu_{n_k} > 0$, il est facile de voir que $i \geq 1$ mène en fait à tout point de C, qui constitue ainsi une classe. □

Du point de vue des propriétés ergodiques ou de la stationnarité éventuelle de la chaîne (X_n), tout est dit dans la proposition précédente: il y a une seule classe récurrente $\{0\}$, et une seule probabilté invariante, la masse de Dirac en 0: la chaîne est stationnaire si et seulement si elle part de 0, et elle reste alors toujours en 0...

Du point de vue des applications, en revanche, beaucoup reste à faire: le calcul des p_i, la loi de T, la "vitesse d'explosion" si X_n tend vers $+\infty$, etc...

2.2.2 Le comportement à l'infini

Les propriétés ergodiques nous donnent, plus ou moins, le comportement "en loi" de X_n pour n grand. Il se trouve que pour les chaînes BGW on a bien mieux, à savoir une convergence presque sûre.

Nous avons besoin pour cela d'un outil commode, à savoir la fonction génératrice de la loi ν: c'est la fonction sur $[0,1]$ donnée par

$$\varphi(s) = \sum_{i=0}^{\infty} \nu_i s^i = I\!E(s^{Y_{n,q}}). \tag{2.2.5}$$

On introduit aussi les deux premiers moments de cette loi:

$$m = \sum_{i=1}^{\infty} i\nu_i \quad (\leq +\infty), \qquad m_2 = \sum_{i=1}^{\infty} i^2 \nu_i. \tag{2.2.6}$$

On sait que φ est C^∞ sur $[0,1[$, et est une fois (resp. deux fois) dérivable à gauche en $s=1$ si et seulement si $m<\infty$ (resp. $m_2<\infty$), et dans ce cas $m=\varphi'(1)$ (resp. $m_2 - m = \varphi''(1)$). On note $\varphi^{(n)}$ la $n$ième itérée de φ: on a $\varphi^{(0)}(s) = s$ et $\varphi^{(n+1)} = \varphi \circ \varphi^{(n)}$; remarquer que $\varphi^{(1)} = \varphi$). Enfin on pose

$$s_0 := \inf(s \in [0,1] : \varphi(s) = s). \tag{2.2.7}$$

Commençons par un lemme.

Lemme 2.2.2. *On suppose (2.2.4).*

a) Si $m \leq 1$ on a $s_0 = 1$, et $\varphi^{(n)}(s) \to 1$ pour tout $s \in [0,1]$.

b) Si $m > 1$ (y–compris quand $m = \infty$) on a $0 < s_0 < 1$ et les seules solutions de l'équation $\varphi(s) = s$ dans $[0,1]$ sont $s = s_0$ et $s = 1$, et $\varphi^{(n)}(s) \to s_0$ pour tout $s \in [0,1[$ (et bien-sûr $\varphi^{(n)}(1) = 1$).

c) Si $m_2 < \infty$ et $m = 1$, pour tout $s \in [0,1[$ la suite $1 - \varphi^{(n)}(s)$ est équivalente à $\frac{2}{n(m_2-1)}$ quand $n \to \infty$.

d) On a $1 - \varphi^{(n)}(s) \leq m^n(1-s)$.

Preuve. D'après (2.2.4) la fonction φ est strictement convexe, avec $\varphi(0) = \nu_0 > 0$ et $\varphi(1) = 1$, et la pente de la tangente à sa courbe représentative au point d'abscisse $s = 1$ est m (que m soit fini ou infini). Les premières assertions de (a) et (b) sont alors évidentes. De plus $\varphi^{(n)}(s)$ croît strictement avec n si $s < s_0$, et décroît strictement si $s > s_0$, et $\varphi^{(n)}(s_0) = s_0$. Par suite $\varphi^{(n)}(s)$ tend vers une limite s_1, vérifiant nécessairement $\varphi(s_1) = s_1$ et aussi $s_1 < 1$ si $s > s_0$: donc $s_1 = s_0$ et on a la fin de (a) et (b).

Passons à (c). On pose $a = (m_2 - 1)/2$. On a $a > 0$, et un développement au voisinage de $s = 1$ montre que pour $s \in [0,1[$,

$$1 - \varphi(s) = (1-s)(1-(1-s)(a+\gamma(s))), \qquad \text{où} \quad \lim_{s\to 1} \gamma(s) = 0,$$

puisque $\varphi'(1) = 1$ et $\varphi''(1) = 2a$. Par suite

$$\frac{1}{1-\varphi(s)} = \frac{1}{1-s} + a + \Gamma(s), \qquad \text{où } \lim_{s\to 1} \Gamma(s) = 0.$$

Il vient alors

$$\frac{1}{1-\varphi^{(j+1)}(s)} = \frac{1}{1-\varphi^{(j)}(s)} + a + \Gamma(\varphi^{(j)}(s)),$$

et en sommant ces égalités pour $j = 0, 1, \ldots, n-1$, on arrive à

$$\frac{1}{1-\varphi^{(n)}(s)} = \frac{1}{1-s} + na + \Delta_n(s), \qquad \text{où } \Delta_n(s) = \sum_{j=0}^{n-1} \Gamma(\varphi^{(j)}(s)).$$

Pour $s \in [0,1[$ fixé, on sait par (a) que $\varphi^{(j)}(s) \to 1$, donc $\Gamma(\varphi^{(j)}(s)) \to 0$, quand $j \to \infty$. On en déduit que $\Delta_n(s)/n \to 0$, et donc

$$1 - \varphi^{(n)}(s) = \frac{1}{na} \, \frac{1}{1 + \frac{1}{na(1-s)} + \frac{\Delta_n(s)}{na}} \sim \frac{1}{na} = \frac{2}{n(m_2 - 1)}.$$

On a donc (c). Enfin l'inégalité $1 - \varphi(s) \leq m(1-s)$ est évidente, et en itérant on obtient immédiatement (d). □

Traditionnellement, on distingue trois cas:

- Le cas **sous–critique**, quand $m < 1$,
- Le cas **critique**, quand $m = 1$,
- Le cas **sur–critique**, quand $m > 1$ (y–compris $m = +\infty$).

Théorème 2.2.3. *a) Dans les cas sous–critique et critique ($m \leq 1$) on a $\mathbb{P}_i(T < \infty) = 1$: on a* extinction presque sûre, *et en particulier X_n tend p.s. vers* 0.

b) Dans le cas sous–critique on a $\mathbb{E}_i(T) < \infty$, tandis que $\mathbb{E}_i(T) = \infty$ si $i \geq 1$ dans le cas critique, dès que $m_2 < \infty$.

c) Dans le cas sur–critique ($m > 1$), sous $\mathbb{P}_i$, la suite X_n converge presque surement vers une limite X_∞ qui ne prend que les 2 valeurs 0 *et* $+\infty$*, avec les probabilités:*

$$\mathbb{P}_i(X_\infty = 0) = \mathbb{P}_i(T < \infty) = p_i = s_0^i, \qquad \mathbb{P}_i(X_\infty = +\infty) = 1 - s_0^i. \tag{2.2.8}$$

Preuve. Commençons par un résultat auxiliaire. La fonction génératrice $\psi_n(s) = \mathbb{E}_1(s^{X_n})$ de X_n sous $\mathbb{P}_1$ vérifie, en vertu de (2.2.1):

$$\begin{aligned} \psi_{n+1}(s) &= \sum_{q=0}^{\infty} \mathbb{E}_1\left(s^{Y_{n,1}+\ldots+Y_{n,q}} 1_{\{X_n = q\}}\right) \\ &= \sum_{q=0}^{\infty} \varphi(s)^q \; \mathbb{P}_1(X_n = q) \;=\; \psi_n(\varphi(s)), \end{aligned}$$

où la seconde égalité provient du fait que les $Y_{n,j}$ sont indépendantes entre elles et de X_n, et de fonction génératrice φ. Donc $\psi_n = \varphi^{(n)}$, et (2.2.3) et le lemme 2.2.2 impliquent, puisque $\{T \leq n\} = \{X_n = 0\}$, que

$$I\!\!P_i(T \leq n) = I\!\!P_1(T \leq n)^i = \psi_n(0)^i = (\varphi^{(n)}(0))^i \to s_0^i \tag{2.2.9}$$

quand $n \to \infty$: cela implique $I\!\!P_i(T < \infty) = 1$ si $m \leq 1$, d'où (a), et $I\!\!P_i(T < \infty) = s_0^i$ quand $m > 1$.

On a aussi

$$I\!\!E_i(T) = \sum_{n\geq 0} I\!\!P_i(T > n) = \sum_{n\geq 0}(1 - \varphi^{(n)}(0)^i) = \sum_{n\geq 0}\left(1 - (1 - (1 - \varphi^{(n)}(0)))^i\right).$$

Donc si $m < 1$ le lemme 2.2.2(d) entraine $I\!\!E_i(T) < \infty$, tandis que si $m = 1$ et $m_2 < \infty$ le lemme 2.2.2(c) entraine $I\!\!E_i(T) = \infty$ si $i \geq 1$ (et bien-sûr $I\!\!E_0(T) = 0$): on a donc (b).

Il reste à étudier le comportement asymptotique de X_n quand $m > 1$. On va considérer le compactifié $\overline{I\!\!N}$ de $I\!\!N$. Dans cet espace compact, chaque trajectoire $n \mapsto X_n(\omega)$ admet au moins un point d'accumulation, et pour $j \in \overline{I\!\!N}$ on note Ω_j l'ensemble des ω tels que $n \mapsto X_n(\omega)$ admette j pour point d'accumulation: ainsi $\Omega = \cup_j \Omega_j$ (attention: les Ω_j ne sont a priori pas deux-à-deux disjoints). Comme chaque $j \in I\!\!N$ est isolé, l'ensemble Ω_j pour $j \in I\!\!N$ est aussi l'ensemble des ω tels que $X_n(\omega)$ passe une infinité de fois en j. Comme j est transient dès que $j \in I\!\!N^\star$ on en déduit que $I\!\!P_i(\Omega_j) = 0$. Par ailleurs, comme 0 est un état absorbant, les ensembles Ω_0 et $\{T < \infty\}$ sont égaux $I\!\!P_i$-p.s., et donc en particulier $I\!\!P_i(\Omega_0 \cap \Omega_\infty) = 0$: en rassemblant tous ces résultats, on voit qu'à des ensembles $I\!\!P_i$–négligeables près, les deux ensembles Ω_0 et Ω_∞ constituent une partition de Ω: donc $I\!\!P_i$–presque toutes les trajectoires de X_n tendent vers une limite X_∞ qui ne prend que les valeurs 0 et $+\infty$, et de plus on a $\{\mathbf{X}_\infty = 0\} = \{T < \infty\}$ $I\!\!P_i$–p.s.: on a donc (c). □

Noter que la formule (2.2.9) nous donne en fait la loi de la variable T sous $I\!\!P_1$ (donc aussi sous $I\!\!P_i$, d'après (2.2.3)), à savoir

$$I\!\!P_1(T = n) = \begin{cases} 0 & \text{si } n = 0 \\ \varphi^{(n)}(0) - \varphi^{(n-1)}(0) & \text{si } n \in I\!\!N^* \\ 1 - s_0 & \text{si } n = +\infty. \end{cases} \tag{2.2.10}$$

Lorsque $m \leq 1$, cette formule décrit complètement le comportement asymptotique de X_n. Dans le cas sus–critique il est intéressant de préciser "à quelle vitesse" X_n tend vers l'infini, sur l'ensemble $\{T = \infty\}$ (qui est de probabilité strictement positive). On peut traiter le cas général, mais nous nous limiterons au cas où les $Y_{n,q}$ ont un moment d'ordre 2 fini:

Théorème 2.2.4. *Si $m_2 < \infty$ et $m > 1$, la suite X_n/m^n converge $I\!\!P_i$–p.s. vers une variable alétoire Y qui est p.s. strictement positive sur l'ensemble $\{T = \infty\}$ et telle que $I\!\!E_i(Y) = i$.*

Ce résultat dit alors que, sur l'ensemble où il n'y a pas extinction, la croissance de X_n vers $+\infty$ est exponentielle (en m^n): c'est l'expression mathématique de la "loi de Malthus".

Preuve. Posons $U_n = X_n/m^n$. Considèrons la filtration $(\mathcal{F}_n)$ décrite après (2.1.1). Comme m et m_2 sont les deux premiers moments de toutes les variables $Y_{n,q}$, on a d'après (2.2.1):

$$\mathbb{E}_i(X_{n+1}|\mathcal{F}_n) = mX_n, \qquad \mathbb{E}_i(X_{n+1}^2|\mathcal{F}_n) = m_2X_n + m^2X_n(X_n-1),$$

et donc

$$\mathbb{E}_i(U_{n+1}|\mathcal{F}_n) = U_n, \qquad \mathbb{E}_i(U_{n+1}^2|\mathcal{F}_n) \leq \frac{m_2}{m^{n+2}}\, U_n + U_n^2.$$

Par suite $\mathbb{E}_i(U_n) = i$ et donc

$$\mathbb{E}_i(U_{n+1}^2) \leq \mathbb{E}_i(U_n^2) + \frac{im_2}{m^{n+2}} \leq i^2 + \frac{im_2}{m^2}\sum_{j=0}^n m^{-j} \leq i^2 + \frac{im_2}{m(m-1)}.$$

On en déduit finalement que U_n est une martingale uniformément intégrable, qui converge donc $\mathbb{P}_i$–p.s. et dans $\mathbb{L}^1(\mathbb{P}_i)$ vers $U = \liminf_n U_n$, et en particulier $\mathbb{E}_i(U) = i$.

Il reste à montrer que $U > 0$ $\mathbb{P}_i$–p.s. sur $\{T = \infty\}$. On pose $q_i = \mathbb{P}_i(U = 0)$. En vertu de la propriété de branchement, la loi de U sous $\mathbb{P}_i$ est la même que celle de $U^{(1)}+\ldots+U^{(i)}$ où les $U^{j)}$ sont des v.a. indépendantes, de même loi que U sous $\mathbb{P}_1$: par suite on a $q_i = q_1^i$. Par ailleurs si θ est la translation associée à la chaîne (X_n) (rappelons que $X_n\circ\theta = X_{n+1}$), on a évidemment $U = U\circ\theta$. Donc d'après la propriété de Markov,

$$\mathbb{P}_1(U=0|\mathcal{F}_1) = \mathbb{P}_1(U\circ\theta = 0|\mathcal{F}_1) = \mathbb{P}_{X_1}(U=0) = q_1^{X_1}.$$

En prenant l'espérance, on obtient $q_1 = \mathbb{E}_1(q^{X_1}) = \varphi(q_1)$. Donc q_1 est solution de l'équation $\varphi(s) = s$, équation qui dans $[0,1]$ admet les deux solutions $s = 1$ et $s = s_0$. Or $\mathbb{E}_1(U) = 1$ implique que $q_1 = \mathbb{P}_1(U = 0) < 1$, donc nécessairement $q_1 = s_0$ et $q_i = s_0^i$ c'est-à-dire $\mathbb{P}_i(U=0) = \mathbb{P}_i(T<\infty)$. Mais on a à l'évidence $\{T<\infty\} \subset \{U=0\}$, donc nécessairement $U > 0$ $\mathbb{P}_i$–p.s. sur $\{T=\infty\}$. □

2.3 Chaîne BGW avec immigration

Ainsi qu'on vient de le voir, la chaîne BGW a un comportement limite "dégénéré" en un certain sens, et il n'y a pas de probabilité invariante hormis la masse de Dirac en 0. La situation devient différente, avec notamment l'existence possible de probabilités invariantes non triviales, lorsqu'on ajoute une immigration.

Soit donc une chaîne de vie et de mort densité–indépendante avec immigration. On a $\nu(m) = \nu$ et $\eta(m) = \eta$ pour tout m, et on suppose bien–sûr que $\eta_0 < 1$ sinon on retrouve la chaîne BGW. Au lieu de (2.2.1), on a

$$X_{n+1} = \sum_{q=1}^m Y_{n,q} + Z_n \quad \text{si} \quad X_n = m: \; m = 0,1,\ldots, \tag{2.3.1}$$

où les $Y_{n,q}$ et Z_n sont indépendantes entre elles et de X_0, les $Y_{n,q}$ sont de loi ν et les Z_n de loi η.

La différence essentielle avec la chaîne BGW simple (sans immigration) est que 0 n'est plus un état absorbant, puisque $p_{0i} = \eta_i$ est strictement positif pour au moins un $i \geq 1$. A nouveau, certains cas sont inintéressants:

a) Si $\nu_0 = 1$: on a $X_{n+1} = Z_n$, et la chaîne se réduit à une suite de v.a. indépendantes et de même loi.

b) Si $\nu_0 = 0$: on a $X_{n+1} > X_n$ et X_n tend en croissant vers $+\infty$.

c) Si $\nu_1 = 1$: on a $X_{n+1} = X_n + Z_n$ et la chaîne est une marche aléatoire croissante, qui tend p.s. vers $+\infty$.

Dans la suite on suppose donc que

$$0 < \nu_0 < 1, \qquad \nu_1 < 1, \qquad \eta_0 < 1. \tag{2.3.2}$$

On utilisera les notations φ, m et m_2 de (2.2.5) et (2.2.6), ainsi que les itérées $\varphi^{(n)}$. On note aussi $m' = \sum_{i\geq 0} i\eta_i$ le nombre moyen d'immigrants à chaque instant. On se contentera du résultat suivant, en supposant si besoin est l'existence de moments.

Théorème 2.3.1. *Supposons (2.3.2).*

a) Si $m > 1$, ou si $m = 1$ et $m_2 < \infty$, il n'y a pas de probabilité invariante, et les états sont donc tous nuls.

b) Si $m < 1$ et $m' < \infty$ il existe une probabilité invariante et une seule, et la classe positive correspondante contient le point $r = \inf(i : \eta_i > 0)$.

La description précise de la classe positive est un peu compliquée en général. Lorsque $\nu_1 > 0$, c'est simplement $C = \{r, r+1, \ldots\}$, et dans tous les cas les singletons $\{i\}$ pour $i < r$ (quand $r \geq 1$) sont des classes transientes.

Preuve. Soit $\Phi(s) = \sum_{i\geq 0} \eta_i s^i$ (fonction génératrice des Z_n). Si $\psi_{i,n}(s) = \mathbb{E}_i(s^{X_n})$, on déduit de (2.3.1) que

$$\begin{aligned} \psi_{i,n+1}(s) &= \sum_{m=0}^{\infty} \mathbb{E}_i\left(s^{Y_{n,1}+\ldots+Y_{n,m}+Z_n} 1_{\{X_n=m\}}\right) \\ &= \sum_{m=0}^{\infty} \Phi(s)\varphi(s)^m \; \mathbb{P}_i(X_n = m) \;=\; \Phi(s)\psi_{i,n}(\varphi(s)). \end{aligned}$$

En itérant cette relation, et comme $\psi_{i,0}(s) = s^i$, on voit que

$$\psi_{i,n} = \prod_{q=0}^{n-1} \Phi\circ\varphi^{(q)} \; (\varphi^{(n)})^i. \tag{2.3.3}$$

Si de plus il existe une probabilité invariante $\mu = (\mu_i)$, dont la fonction génératrice est notée $R(s) = \sum_i \mu_i s^i$, il vient alors pour tout $n \geq 1$:

$$R = R\circ\varphi^{(n)} \prod_{q=0}^{n-1} \Phi\circ\varphi^{(q)}. \tag{2.3.4}$$

On sait $\varphi^{(n)}(s) \to s_0$ quand $n \to \infty$ (cf. Lemme 2.2.2) Si $0 \le s < 1$ on a $0 \le \Phi(s) < 1$, donc la suite $\prod_{q=0}^{n-1} \Phi \circ \varphi^{(q)}(s)$ converge vers une limite $\Psi(s) \in [0,1[$ si $s \in [0,1[$, et on a

$$s \in [0,1[\quad \Longrightarrow \quad R(s) = R(s_0)\ \Psi(s). \tag{2.3.5}$$

Nous allons alors distinguer diverses situations:

(i) Supposons $m > 1$, de sorte que $s_0 \in]0,1[$. On a $R(s_0) = R(s_0)\Psi(s_0)$ et $\Psi(s_0) < 1$, donc nécessairement $R(s_0) = 0$, donc R ne peut pas être une fonction génératrice et il n'y a pas de probabilité invariante.

(ii) Supposons $m = 1$ et $m_2 < \infty$. On a $1 - \Phi(s) \ge (1-\eta_0)(1-s)$ et le lemme 2.2.2(c) entraine que $\varphi^{(q)}(s) \sim a/q$ quand $q \to \infty$, pour tout $s \in [0,1[$. Par suite $1 - \Phi \circ \varphi^{(q)}(s) \ge b_s/q$ pour tout $q \ge 1$, où $b_s > 0$ (si $s < 1$). On en déduit que le produit infini définissant $\Psi(s)$ diverge, donc $\Psi(s) = 0$: donc $R(s) = 0$, et R ne peut pas être une fonction génératrice et il n'y a pas de probabilité invariante.

(iii) Supposons $m < 1$ et $m' < \infty$. On a $1 - \Phi(s) \le m'(1-s)$, et on sait que $1 - \varphi^{(q)}(s) \le m^q(1-s)$ (lemme 2.2.2(d)). Par suite $1 - \Phi \circ \varphi^{(q)}(s) \le m'm^q(1-s)$, tandis que $\varphi^{(q)}(s) > 0$ pour $s \ge 0$ et $\Phi(s) > 0$ si $s > 0$: donc le produit infini définissant $\Psi(s)$ converge lorsque $s \in]0,1]$: on a $0 < \Psi(s) < 1$ pour tout $s \in]0,1[$. Comme par ailleurs $\varphi^{(n)}(s) \to 1$ si $n \to \infty$, on déduit de (2.3.3) que

$$0 < s \le 1 \qquad \Longrightarrow \qquad \psi_{i,n}(s) \to \Psi(s) > 0. \tag{2.3.6}$$

Si on avait $p_{ij}^{(n)} \to 0$ (quand $n \to \infty$) pour tout j, le théorème de Lebesgue entrainerait que $\psi_{i,n}(s) \to 0$ pour tout $s \in [0,1[$, ce qui contredirait (2.3.6). Par suite il existe au moins un j tel que $p_{ij}^{(n)}$ ne tende pas vers 0. D'après les résultats ergodiques sur les chaînes de Markov, il en découle que j est un état récurrent positif, et il existe donc au moins une probabilité invariante. Cette probabilité invariante est unique, car sa fonction génératrice est $R = \Psi$ d'après (2.3.5) (rappelons que $s_0 = 1$ ici). Enfin, on a $p_{ir} = \nu_0^i > 0$, donc tous les états mènent à r qui doit donc appartenir à la classe positive. $\square$

2.4 Les probabilités quasi–stationnaires

La chaîne BGW a été introduite historiquement en vue de modéliser l'évolution du nombre de personnes ayant un patronyme donné, dans la noblesse anglaise. Il s'agissait d'une part d'étudier l'extinction éventuelle d'un patronyme, d'autre part d'évaluer, à l'instant courant, la loi du nombre de personnes (ou de familles) ayant un patronyme donné.

Si cette chaîne modélise correctement l'évolution de l'effectif (génération par génération) d'un patronyme donné, ce dont on peut évidemment douter, l'extinction a été étudiée ci–dessus. Quant au second problème, "l'instant courant" signifie qu'on regarde une sorte de phénomène "stationnaire" pour un patronyme ancien; mais bien-sûr on ne considère que les patronymes existant (c'est-à-dire, non éteints) à l'instant courant. Mathématiquement, on s'intéresse donc aux quantités $I\!P_i(X_n = j/T > n) = I\!P_i(X_n = j/X_n \neq 0)$ pour tout $j = 1, 2, \ldots$, et ceci quand n est grand.

Nous allons voir que sous des hypothèses adéquates, les $\mathbb{P}_i(X_n = j/T > n)$ convergent pour tout $j \geq 1$ vers des limites π_j, bien-sûr positives, et vérifiant $\sum_j \pi_j = 1$: on a donc une sorte de “théorème ergodique conditionnel”. Ce phénomène n'est pas propre aux chaînes BGW, mais concerne une large classe de chaînes de Markov ayant un état absorbant.

Nous nous plaçons donc dans le cadre du chapitre précédent: on a une chaîne de Markov (X_n) de transition $P = (p_{ij})$, avec un état absorbant. Ce n'est pas une restriction de supposer que l'espace d'état est $\mathbb{N}$ et que l'état absorbant est 0, donc notre hypothèse revient à dire que

$$p_{00} = 1, \qquad i \geq 1 \quad \Longrightarrow \quad p_{ii} < 1. \tag{2.4.1}$$

Le temps d'absorption est $T = \inf(n : X_n = 0)$ et $X_n = 0$ si et seulement si $n \geq T$. Comme d'habitude on note $\mathbb{P}_i$ la probabilité sur l'espace de probabilité, pour laquelle $X_0 = i$ presque surement. Supposons aussi qu'il existe une fonction réelle g sur $\mathbb{N}$ et un réel $\lambda > 0$ tels que

$$Pg = \lambda g, \qquad g(0) = 0, \qquad \inf_{i\geq 1} g(i) > 0. \tag{2.4.2}$$

Lemme 2.4.1. *La formule $q_{ij} = \frac{p_{ij}g(j)}{\lambda g(i)}$ pour $i, j \geq 1$ définit une probabilité de transition $Q = (q_{ij})$ sur $\mathbb{N}^\star$, dont la nième puissance $Q^n = (q_{ij}^{(n)}$ est donnée par*

$$q_{ij}^{(n)} = \frac{p_{ij}^{(n)}\ g(j)}{\lambda^n\ g(i)}. \tag{2.4.3}$$

Preuve. On a $q_{ij} \geq 0$, et puisque $g(0) = 0$,

$$\sum_{j\geq 1} q_{ij} = \frac{1}{\lambda g(i)} \sum_{j\geq 1} p_{ij} g(j) = \frac{1}{\lambda g(i)} \sum_{j\geq 0} p_{ij} g(j) = \frac{1}{\lambda g(i)}\ (Pg)(i) = \frac{\lambda g(i)}{\lambda g(i)} = 1,$$

où on a utilisé $Pg = \lambda g$. Cela prouve la première assertion. La relation (2.4.3) est vraie pour $n = 1$. Si on la suppose vraie pour $n - 1$, il vient

$$\begin{aligned} q_{ij}^{(n)} &= \sum_{k\geq 1} q_{ik} q_{kj}^{(n-1)} = \sum_{k\geq 1} \frac{p_{ik} g(k)}{\lambda g(i)} \frac{p_{kj}^{(n-1)} g(j)}{\lambda^{n-1} g(k)} \\ &= \frac{g(j)}{\lambda^n g(i)} \sum_{k\geq 1} p_{ik} p_{kj}^{(n-1)} = \frac{g(j)}{\lambda^n g(i)} \sum_{k\geq 0} p_{ik} p_{kj}^{(n-1)} = \frac{p_{ij}^{(n)} g(j)}{\lambda^n g(i)}, \end{aligned}$$

où on a utilisé $0 < g(k) < \infty$ pour $k \geq 1$, ainsi que le fait que $p_{0j}^{(n-1)} = 0$ si $j \geq 1$. On a donc montré (2.4.3) par récurrence. $\square$

On déduit en particulier de (2.4.2) et (2.4.3) que $p_{ij}^{(n)} > 0 \Leftrightarrow q_{ij}^{(n)} > 0$ si $i, j \geq 1$. Par suite, si $C_0 = \{0\}$ et $C_1, C_2, \ldots$ désignent les classes de la chaîne (X_n), il est immédiat que:

$$\text{La période de l'état } i \geq 1 \text{ est la même pour les chaînes de transitions } P \text{ et } Q, \tag{2.4.4}$$

$$\text{Les classes de la chaîne de transition } Q \text{ sont } C_1, C_2, \ldots, \tag{2.4.5}$$

Théorème 2.4.2. *a) Supposons (2.4.1) et (2.4.2) avec un $\lambda < 1$. On a alors $I\!P_i(T < \infty) = 1$ pour tout i, et les états $i \geq 1$ sont transients pour la chaîne (X_n).*

b) Supposons en outre que la chaîne de transition Q admet une unique classe positive C, apériodique, et si $\pi = (\pi_i)_{i\geq 1}$ désigne l'unique probabilité invariante pour Q (donc $\pi_i > 0$ si et seulement si $i \in C$), on a pour tous $i, j \geq 1$:

$$\lim_{n\to\infty} I\!P_i(X_n = j/X_n \neq 0) \;=\; \mu_j \;:=\; \frac{\pi_j/g(j)}{\sum_{k\geq 1}\pi_k/g(k)}, \tag{2.4.6}$$

et $\mu = (\mu_j)_{j\geq 1}$ est aussi une probabilité, portée par C.

La probabilité μ s'appelle la *probabilité quasi–stationnaire* de la chaîne (X_n). Il existe une version plus générale de ce théorème, lorsqu'il y a plusieurs classes positives et/ou lorsque la période diffère de 1: cette version générale est laissée au lecteur.

Preuve. (a) Si $a = \inf_{i\geq 1} g(i)$, on a

$$aI\!P_i(T \geq n) = a\sum_{j\geq 1} p_{ij}^{(n)} \leq P^n g(i) = \lambda^n g(i),$$

qui tend vers 0 quand $n \to \infty$ puisque $\lambda < 1$. Comme $a > 0$, on en déduit la première assertion. De plus $I\!P_i(T < \infty) = 1$ entraine que la chaîne ne passe $I\!P_i$–p.s. qu'un nombre fini de fois en $i \geq 1$, puisque 0 est absorbant, donc i est transient.

(b) Sous l'hypothèse supplémentaire, on a $q_{ij}^{(n)} \to \pi_j$, et même (d'après le théorème de Lebesgue) $\sum_{j\geq 1} q_{ij}^{(n)} h(j) \to \sum_{j\geq 1}\pi_j h(j)$ pour toute fonction bornée h. Mais si $i, j \geq 1$, (2.4.3) implique

$$I\!P_i(X_n = j/X_n \neq 0) \;=\; \frac{p_{ij}^{(n)}}{\sum_{k\geq 1} p_{ik}^{(n)}} \;=\; \frac{q_{ij}^{(n)}/g(j)}{\sum_{k\geq 1} q_{ik}^{(n)}/g(k)},$$

(car la fonction $1/g$ est bornée sur $I\!N^\star$), et la convergence (2.4.6) est alors immédiate. La propriété $\sum_{j\geq 1}\mu_j$ est également évidente. $\square$

Ce résultat nous donne le comportement limite (en loi) de X_n, sachant $X_n \neq 0$. Il est également intéressant de déterminer le comportement limite de la chaîne $(X_l)_{0\leq l\leq N}$ à “horizon fini” N (arbitrairement grand, mais fixé), sachant que $X_n \neq 0$, lorsque $n \to \infty$.

Théorème 2.4.3. *Soit les hypothèses du théorème 2.4.2(b). Soit $N \geq 1$ fixé. La chaîne $(X_l)_{1\leq l\leq N}$ converge en loi, sous les probabilité conditionnelles $I\!P_i(./X_n \neq 0)$ et lorsque $n \to \infty$, vers une chaîne de Markov à valeurs dans $I\!N^\star$, partant de i et de transition Q.*

Remarquer que la chaîne $(X_l)_{1\leq l\leq N}$ *n'est pas une chaîne de Markov* sous les $I\!P_i./X_n \neq 0)$ (puisqu'il y a un conditionnement par rapport au futur). Remarquer que l'événement conditionnant $\{X_n \neq 0\}$ a une probabilité qui tend vers 0. Enfin, une manière équivalent d'énoncer le résultat consiste à dire que

$$I\!P_i(X_0 = i_0, X_1 = i_1, \ldots, X_N = i_N/X_n \neq 0) \;\to\; \delta_{ii_0} q_{i_0 i1} q_{i_1 i_2} \ldots q_{i_{N-1} i_N} \tag{2.4.7}$$

pour tous $i_j \geq 1$, lorsque $n \to \infty$, et où δ_{ij} est le symbole de Kronecker.

Preuve. On va en fait montrer (2.4.7). Notons (pour $i_0, \dots i_N$ fixés) les membres de gauche et de droite de (2.4.7) par γ_n et γ respectivement. En utilisant (2.4.3), on obtient

$$\begin{aligned}
\gamma_n &= \delta_{ii_0} p_{i_0 i_1} p_{i_1 i_2} \dots p_{i_{N-1} i_N} \frac{\sum_{k\geq 1} p_{i_N k}^{(n-N)}}{\sum_{k\geq 1} p_{i_0 k}^{(n)}} \\
&= \delta_{ii_0} \frac{\lambda q_{i_0 i_1} g(i_0)}{g(i_1)} \frac{\lambda q_{i_1 i_2} g(i_1)}{g(i_2)} \dots \frac{\lambda q_{i_{N-1} i_N} g(i_N)}{g(i_{N-1})} \frac{\sum_{k\geq 1} \lambda^{n-N} q_{i_N k}^{(n-N)} g(i_N)/g(k)}{\sum_{k\geq 1} \lambda^n q_{i_N k}^{(n)} g(i_0)/g(k)} \\
&= \gamma \frac{\sum_{k\geq 1} q_{i_N k}^{(n-N)}/g(k)}{\sum_{k\geq 1} q_{i_N k}^{(n)}/g(k)}.
\end{aligned}$$

On a vu dans la preuve précédente que $\sum_{k\geq 1} q_{ik}^{(n)}/g(k) \to \sum_{k\geq 1} \pi_k g(k)$, de sorte que l'expression précédente tend vers γ, et on a le résultat. □

Revenons à la chaîne BGW, pour laquelle nous utilisons les notations m, m_2 et φ du paragraphe 2.2. Nous supposons (2.2.4), et C est la classe transiente décrite dans la proposition 2.2.1.

Théorème 2.4.4. *Si $m < 1$ (cas sous–critique) et $m_2 < \infty$, il existe une (unique) probabilité quasi–stationnaire $\mu = (\mu_i)_{i\geq 1}$: pour $i \geq 1$, on a*

$$\mathbb{P}_i(X_n = j/X_n \neq 0) = \mathbb{P}_i(X_n = j/T > n) \ \to \ \mu_i,$$

et $\mu_i > 0$ si et seulement si $i \in C$. De plus pour tout N fixé, la chaîne $(X_l)_{1\leq l\leq N}$ converge en loi, sous la probabilité conditionnelle $\mathbb{P}_i(./X_n \neq 0)$ avec $i \geq 1$ et quand $n \to \infty$, vers une chaîne de Markov à valeurs dans $\mathbb{N}^\star$ partant de i et de transition $q_{ij} = jp_{ij}/im$.

Preuve. D'après la preuve du théorème 2.2.4 on a $\mathbb{E}_i(X_1) = im$, de sorte que (2.4.2) est satisfait avec la fonction identité $g(i) = i$ et $\lambda = m < 1$, tandis que (2.4.1) est trivial ici. On associe à P les matrices de transition Q et Q^n par (2.4.3). D'après la preuve de la proposition 2.2.1 on a $p_{ij}^{(n)} = 0$ si $n \geq 1$ et $j \notin C \cup \{0\}$; par suite les singletons $\{j\}$ pour $j \notin C$ et $j \geq 1$ sont aussi des classes transientes pour la chaîne de transition Q. D'après (2.4.5), C est alors l'unique classe de la chaîne associée à Q qui peut être récurrente positive. Enfin, si $i = \inf(j \geq 1, \nu_j > 0$, on a $p_{jj} = \nu_j \nu_O^{j-1} > 0$, de sorte que la période de j, donc de la classe C, est 1 pour la chaîne (X_n), donc aussi pour celle associée à Q (par (2.4.4)).

Pour pouvoir appliquer les théorèmes 2.4.2 et 2.4.3, il nous suffit alors de montrer que $q_{1i}^{(n)}$ ne tend pas vers 0 pour tout $i \geq 1$. Pour cela, on opère comme dans la preuve du théorème 2.3.1 (cas (iii)): si $\Phi_n(s) = \sum_{i\geq 1} q_{1i}^{(n)} s^i$ converge vers une limite non nulle pour un $s \in]0,1[$, alors il existe $i \geq 1$ tel que $q_{1i}^{(n)}$ ne converge pas vers 0. On a (rappelons $\lambda = m$ et $g(k) = k$):

$$m^{n+1}\Phi_{n+1}(s) \ = \ \sum_{i\geq 1} p_{1i}^{(n+1)} \, i \, s^i \ = \ \mathbb{E}_1\left(X_{n+1} s^{X_{n+1}}\right)$$

$$
\begin{aligned}
&= \sum_{q\geq 1} \mathbb{E}_1\left((Y_{n,1}+\ldots+Y_{n,q})s^{Y_{n,1}+\ldots+Y_{n,q}} 1_{\{X_n=q\}}\right) \\
&= \sum_{q\geq 1} \varphi(s)^{q-1}\ s\ \varphi'(s)\ \mathbb{E}_1(X_n 1_{\{X_n=q\}}) \\
&= \frac{s\varphi'(s)}{\varphi(s)}\ \mathbb{E}_1\left(X_n\ \varphi(s)^{X_n}\right) \ = \ \frac{s\varphi'(s)}{\varphi(s)}\ m^n \Phi_n \circ \varphi(s),
\end{aligned}
$$

où la troisième ligne vient de ce que $\mathbb{E}(Y_{n,r}s^{Y_{n,r}}) = s\varphi'(s)$ et de ce que les $Y_{n,r}$ pour $r = 1, 2, \ldots$ sont indépendantes entre elle et de X_n. Par suite, comme $\Phi_0(s) = s$, on a

$$\Phi_n(s) = s \prod_{q=0}^{n-1} \frac{\varphi' \circ \varphi^{(q)}(s)}{m}.$$

Noter que $0 < \varphi'(s) < m$ pour tout $s \in [0,1[$, à cause de (2.2.4), donc $0 < \Phi_n(s) < 1$ si $0 < s < 1$. Il nous suffit alors de prouver que le produit infini de terme général $\varphi' \circ \varphi^{(q)}(s)/m$ est convergent (ce qui signifie que $\Phi(s)$ converge vers une limite $\Phi(s) > 0$), et ceci est réalisé si la série de terme général $u_q := 1 - \varphi' \circ \varphi^{(q)}(s)/m$ est absolument convergente. En vertu des hypothèses, il existe une constante a telle qu'on ait

$$0 \leq 1 - \frac{\varphi'(s)}{m} = a(1-s)$$

On sait aussi que $0 \leq 1 - \varphi^{(q)}(s) \leq m^q(1-s)$. Donc $0 \leq u_q \leq am^q(1-s)$, et on a le résultat puisque $m < 1$. □

2.5 Les chaînes densité–dépendantes

Nous allons considérer maintenant, de manière un peu superficielle, les chaînes BGW qui sont densité–dépendantes, mais toujours sans immigration. Cela revient à considérer la chaîne donnée par (2.1.1) avec $Z_{n,m} = 0$. L'état 0 est toujours un état absorbant, et le temps d'extinction est encore $T = \inf(n : X_n = 0)$. En revanche, la "propriété de branchement" n'est plus valide.

Typiquement, une telle chaîne modélise l'évolution d'une population de type BGW, mais avec des "ressources limitées" qui diminuent le taux de reproduction lorsque l'effectif de la population augmente. Ainsi, il est naturel de supposer les conditions suivantes sur les moyennes $m(r)$ des lois de reproduction $\nu(r)$ et les probabilités de mort sans descendance:

$$m(r+1) \leq m(r), \qquad \nu(r)_0 \leq \nu(r+1)_0. \tag{2.5.1}$$

On pourrait démontrer des résultats analogues à ceux des chaînes BGW standards, mais nous nous contenterons d'un résultats très partiels sur l'extinction. Rappelons que $m(r) < 1$ implique $\nu(r)_0 > 0$, puisque $m(r) = \sum_{j\geq 1} j\nu(r)_j \geq \sum_{j\geq 1} \nu(r)_j = 1 - \nu(r)_0$.

Théorème 2.5.1. *Supposons que $\nu(r)_0 > 0$ pour tout $r \geq 1$ et que $m(r) < \infty$ pour tout $r \geq 1$ et $m(r) \leq a$ sauf pour un nombre fini de r, avec un $a \in]0,1[$. On a alors $\mathbb{P}_i(T < \infty) = 1$ (extinction p.s.).*

Noter que sous (2.5.1) les hypothèses de ce théorème se réduisent à $\nu(1)_0 > 0$ et $\lim_{r\to\infty} m(r) < 1$.

Preuve. Comme $\nu(r)_0 \geq 1 - m(r) \geq 1 - a$ si $m(r) \leq a$, on a $b = \inf_r \nu(r)_0 > 0$, d'après les hypothèses.

Le même calcul que dans le théorème 2.2.4 montre que $\mathbb{E}_i(X_{n+1}|\mathcal{F}_n) = X_n m(X_n)$ (avec $m(0) = 0$). Ainsi, la suite

$$Z_n = \begin{cases} X_0, & \text{si } \ n = 0 \\ X_n / \prod_{l=0}^{n-1} m(X_l) & \text{si } \ n \geq 1 \end{cases}$$

est une martingale positive, d'espérance $\mathbb{E}_i(Z_n) = i$. Elle converge alors $\mathbb{P}_i$–p.s. vers une limite finie Z. Par suite $M := \sup_n Z_n < \infty$ $\mathbb{P}_i$–p.s., et on a

$$X_n \leq M U_n, \qquad \text{où} \quad U_n = \prod_{l=0}^{n+1} m(X_l). \tag{2.5.2}$$

Comme 0 est un état absorbant, le nombre de fois où X_n passe en $j \geq 1$ entre les instants 0 et N égale le nombre de fois où $X_n = j$ et $X_{n+1} > 0$ pour n entre 0 et $N-1$, plus éventuellement 1. Donc si $j \geq 1$,

$$\alpha_N(i,j) = \mathbb{E}_i(\sum_{n=0}^{N} 1_{\{X_n=j\}}) \leq 1 + \mathbb{E}_i(\sum_{n=0}^{N-1} 1_{\{X_n=j, X_{n+1}>0\}}).$$

Mais $p_{j0} = \nu(j)_0^j$, donc d'après la propriété de Markov,

$$\mathbb{P}_i(X_n = j, X_{n+1} > 0) = (1 - \nu(j)_0^j)\mathbb{E}_i(1_{\{X_n=j\}}) \leq (1 - b^j)\mathbb{E}_i(1_{\{X_n=j\}})$$

et on déduit que

$$\alpha_N(i,j) \leq 1 + (1 - b^j)\alpha_{N-1}(i,j).$$

En itérant cette relation et comme $\alpha_0(i,j) \leq 1$, on arrive à $\alpha_N(i,j) \leq 1 + (1-b^j) + \ldots + (1-b^j)^N \leq 1/b^j$. En faisant tendre N vers l'infini, on voit que $\mathbb{E}_i(N_j) < \infty$, où N_j est le nombre de visites de la chaîne en j.

Si C désigne l'ensemble (fini) des r tels que $m(r) > a$, on a de manière évidente $U_n \leq a^n V$, où $V = \prod_{j\in C} m(j)^{N_j}$, et comme chaque N_j est p.s. fini on a $V < \infty$ p.s. Par suite $U_n \to 0$ p.s., donc (2.5.2) entraîne que $X_n \to 0$ p.s. Comme X_n est à valeurs entières et que 0 est abosrbant, cela donne le résultat. $\square$

La preuve précédente montre aussi que, sous la seule condition $\inf_{r>0} \nu(r)_0 >$, la suite X_n tend vers 0 ou vers $+\infty$ presque sûrement, exactement comme dans le théorème 2.2.3. On pourrait montrer que $a \leq 1$ au lieu de $a < 1$ est suffisant dans le théorème ci–dessus. En revanche, les critères pour que $\mathbb{P}_i(T < \infty) < 1$ (et donc $\mathbb{P}_i(X_n \to +\infty) > 0$) sont plus délicats à exprimer.

Chapitre 3

Les processus de Markov

3.1 Processus de Markov généraux

Dans ce chapitre on va aborder les processus à temps continu, indicés par $\mathbb{R}_+$.

3.1.1 Définitions

Soit $(E,\mathcal{E})$ est espace mesurable quelconque. Exactement comme pour les chaînes de Markov, il y a plusieurs définitions pour les processus de Markov; mais nous allons d'emblée nous situer dans un cadre comparable à celui des définitions 1.3.1 et 1.3.2, dans le cas homogène.

D'abord, on suppose donnée une famille $(P_t)_{t\geq 0}$ de probabilités de transition de $(E,\mathcal{E})$ dans lui-même, avec $P_0 = I$ (l'identité). Cette famille sera appelée un *semi-groupe* si elle vérifie

$$P_{t+s} = P_t P_s \qquad \forall s,t \geq 0. \tag{3.1.1}$$

Ensuite, soit $(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\geq 0},\mathbb{P})$ un espace probabilisé filtré (les $\mathcal{F}_t$ sont des sous–tribus de $\mathcal{F}$ avec $\mathcal{F}_t \subset \mathcal{F}_s$ si $t \leq s$). Un *processus adapté* est une famille $(X_t)_{t\geq 0}$ de variables aléatoires à valeurs dans $(E,\mathcal{E})$, telle que chaque X_t soit $\mathcal{F}_t$–mesurable. L'analogue, dans le cas homogène, de la définition 1.3.1 est alors:

Définition 3.1.1. Le processus (X_t) sur $(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\geq 0},\mathbb{P})$ est dit vérifier la $(\mathcal{F}_t)$–*propriété de Markov*, s'il existe un semi-groupe de transitions (P_t) tel que pour tous $s,t\geq 0$ et toute fonction mesurable f, positive ou bornée,

$$\mathbb{E}(f(X_{t+s})|\mathcal{F}_t) = P_s f(X_t). \tag{3.1.2}$$

Si $\mathcal{F}_t = \sigma(X_s : s \leq t)$, on parle simplement de la *propriété de Markov.*

Pour l'analogue de la définition 1.3.2 on part de l'espace Ω muni du processus (X_t), on pose $\mathcal{F}_t = \sigma(X_s : s \leq t)$ et $\mathcal{F} = \bigvee_t \mathcal{F}_t$, et on suppose donnée une famille $(\theta_t)_{t\geq 0}$ d'applications de Ω dans lui-même, telle que

$$\theta_0 = I, \qquad \theta_{s+t} = \theta_s \circ \theta_t, \qquad X_{s+t} = X_s \circ \theta_t, \quad \forall s,t\geq 0. \tag{3.1.3}$$

On se donne enfin une famille $(I\!P_x)_{x\in E}$ de probabilités sur $(\Omega,\mathcal{F})$ telle que $x\mapsto I\!P_x(A)$ soit $\mathcal{E}$–mesurable pour tout $A\in\mathcal{F}$, et vérifiant $I\!P_x(X_0=x)=1$ (cf. (1.3.4)) . On définit $I\!P_\mu$, pour toute probabilité μ sur $(E,\mathcal{E})$, par la formule (1.3.6): $I\!P_\mu(A)=\int I\!P_x(A)\mu(dx)$.

Définition 3.1.2. Le terme $\mathbf{X}=(\Omega,\mathcal{F},(\mathcal{F}_t),(\theta_t),(X_t),(I\!P_x))$ est un *processus de Markov* s'il existe un semi-groupe de transitions (P_t) tel que pour tout $x\in E$, tous $s,t\geq 0$ et toute fonction mesurable f, positive ou bornée,

$$I\!E_x(f(X_{t+s})|\mathcal{F}_t)=P_sf(X_t). \tag{3.1.4}$$

On a alors l'analogue de la proposition 1.3.3:

Proposition 3.1.3. *Si* $\mathbf{X}$ *est un processus de Markov, pour toute probabilité* μ *sur* $(E,\mathcal{E})$*, la loi de* X_0 *sous* $I\!P_\mu$ *(loi initiale) est* μ*. De plus pour tout* $t\geq 0$ *et toute variable aléatoire réelle* Y *sur* $(\Omega,\mathcal{F})$*, positive ou bornée, la variable* $Y\circ\theta_t$ *est* $\mathcal{F}$*–mesurable et on a:*

$$I\!E_\mu(Y\circ\theta_t|\mathcal{F}_t)=I\!E_{X_t}(Y). \tag{3.1.5}$$

Preuve. La première propriété se montre comme dans le cas discret. D'après le théorème des classes monotones 1.2.2, pour la seconde assertion il suffit de montrer le résultat quand Y est de la forme $Y=\prod_{i=0}^m f_i(X_{t_i})$, pour des fonctions bornées mesurables f_i et des temps $0=t_0<\ldots<t_m$. Dans ce cas $Y\circ\theta_t=\prod_{i=0}^m f_i(X_{t+t_i})$, donc on voit que $Y\circ\theta_t$ est $\mathcal{F}$–mesurable. Si Z est une variable bornée $\mathcal{F}_t$–mesurable, on voit (comme dans le cas discret encore) qu'il suffit de montrer que pour tout $x\in E$,

$$I\!E_x(Z\ Y\circ\theta_t)=I\!E_x(Z\ I\!E_{X_t}(Y)). \tag{3.1.6}$$

Cela se montre par récurrence sur m. C'est évident pour $m=0$. Supposons (3.1.6) vraie pour tout t et toute variable de la forme $Y=\prod_{i=0}^{m-1} g_i(X_{s_i})$, et montrons qu'elle est vraie pour $Y=\prod_{i=0}^m f_i(X_{t_i})$. Posons $U=\prod_{i=0}^{m-1} f_i(X_{t_i})$ et $V=U\ P_{t_m-t_{m-1}}f_m(X_{t_{m-1}})$. On a

$$\begin{aligned} E_x(Z\ Y\circ\theta_t) &= I\!E_x(Z\ U\circ\theta_t\ f_m(X_{t+t_m}))\\ &= I\!E_x\left(I\!E_x(Z\ U\circ\theta_t\ f_m(X_{t+t_m})|\mathcal{F}_{t+t_{m-1}})\right)\\ &= I\!E_x\left(Z\ U\circ\theta_t\ P_{t_m-t_{m-1}}f_m(X_{t+t_{m-1}})\right)\\ &= I\!E_x\left(Z\ V\circ\theta_t\right)\ =\ I\!E_x\left(Z\ I\!E_{X_t}(V)\right), \end{aligned}$$

où la dernière égalité provient de l'hypothèse de récurrence. En appliquant ce qui précède à $t=0$ et $Z=1$, on obtient aussi $I\!E_y(Y)=I\!E_y(V)$ pour tout y, d'où (3.1.6). □

La formule (3.1.5) admet une extension simple, mais utile dans certaines circonstances. Si f une fonction sur $\Omega\times E$ et Y une application de Ω dans E, on note $f(.,Y\circ\theta_t)$ la fonction $\omega\mapsto f(\omega,Y\circ\theta_t(\omega))$.

Proposition 3.1.4. *Soit* $\mathbf{X}$ *un processus de Markov, et* μ *une probabilité sur* $(E,\mathcal{E})$*. Pour tout* $t\geq 0$*, toute fonction mesurable* Y *de* $(\Omega,\mathcal{F})$ *dans un espace* $(E,\mathcal{E})$*, et toute fonction* f *positive ou bornée sur* $\Omega\times E$*, qui est* $\mathcal{F}_t\otimes\mathcal{E}$*–mesurable, une version de l'espérance conditionnelle* $I\!E_\mu(f(.,Y\circ\theta_t)|\mathcal{F}_t)$ *est donnée par* $\omega\mapsto\int I\!P_\mu(d\omega')f(\omega,T(\omega')$*. On écrit de manière plus concise:*

$$I\!E_\mu(f(.,Y\circ\theta_t|\mathcal{F}_t)=\int I\!P_{X_t}(d\omega')f(.,Y(\omega')). \tag{3.1.7}$$

Preuve. On sait que la tribu $\mathcal{F}_t \otimes \mathcal{E}$ est engendrée par les ensemble de la forme $A \times B$, avec $A \in \mathcal{F}_t$ et $B \in \mathcal{E}$. Donc, exactement comme pour le théorème de Fubini, la linéarité des espérances conditionnelles et un argument de classes monotones montre qu'il suffit de prouver le résultat lorsque f est de la forme $f(\omega, x) = g(\omega)h(x)$, avec g et h des fonctions mesurables sur $(\Omega, \mathcal{F}_t)$ et $(E, \mathcal{E})$ respectivement. Mais dans ce cas le membre de gauche de (3.1.7) est $\mathbb{E}_\mu(g\ h(Y \circ \theta_t)|\mathcal{F}_t) = g\mathbb{E}_\mu(h(Y \circ \theta_t)|\mathcal{F}_t)$ puisque g est $\mathcal{F}_t$–mesurable; le membre de droite vaut $g \int \mathbb{P}_{X_t}(d\omega')h(Y(\omega')) = g\mathbb{E}_{X_t}(h(Y))$, et donc l'égalité provient de (3.1.5) appliqué à $h(Y)$. $\square$

Construction d'un processus de Markov: Etant donné un semi–groupe (P_t) de transitions sur E, on peut se poser le problème de la construction d'un processus de Markov $\mathbf{X}$ au sens de la définition 3.1.2 admettant ce semi–groupe pour transitions.

Comme dans le cas discret, il est naturel de considérer la construction canonique suivante: on prend $\Omega^{\mathbb{R}_+}$, qui est l'espace de toutes les fonctions $t \mapsto x(t)$ sur $\mathbb{R}_+$, à valeurs dans E. Puis, pour toute fonction $x(.)$, on pose

$$X_t(x(.)) = x(t), \qquad \theta_t(x)(.) = x(t + .),$$

de sorte qu'on a (3.1.3). Les tribus $\mathcal{F}$ et $\mathcal{F}_t$ sont comme avant la définition 3.1.2. Il reste à définir les probabilités $\mathbb{P}_x$: malheureusement le théorème de Ionescu–Tulcea n'est valide que pour une suite dénombrable *strictement ordonnée et avec un plus petit élément* de temps, comme $\mathbb{N}$: cela ne marche si l'ensemble des temps est $\mathbb{R}_+$ (ni non plus, d'ailleurs, si l'ensemble des temps était l'ensemble des entiers négatifs). Il nous faut nous référer à un théorème d'une autre nature, appelé *théorème de Kolmogorov*, et qui nécessite une certaine structure de l'espace E. Voici (énoncée seulement) la version du théorème de Kolmogorov relative aux processus de Markov:

Théorème 3.1.5. *Supposons que E soit un espace polonais (i.e. métrique, complet, séparable), et $\mathcal{E}$ sa tribu borélienne. A tout semi–groupe $(P_t)_{t\geq 0}$ de probabilités de transition de $(E, \mathcal{E})$ dans lui-même on associe une famille $\mathbb{P}_x$ de probabilités sur l'espace canonique $(\Omega, \mathcal{F})$ ci–dessus, telle que $\mathbf{X} = (\Omega, \mathcal{F}, (\mathcal{F}_t), (\theta_t), (X_t), (\mathbb{P}_x))$ soit un processus de Markov de transitions (P_t).*

3.1.2 La propriété forte de Markov

Soit $\mathbf{X}$ un processus de Markov, au sens de la définition 3.1.2. Rappelons qu'un temps d'arrêt, dans le cadre présent, est une application T de Ω dans $[0, \infty]$ telle que $\{T < t\} \in \mathcal{F}_t$ pour tout $t \in \mathbb{R}_+$ (**Attention**: on ne suppose pas ici que la filtration $(\mathcal{F}_t)$ est "continue à droite", donc la définition des temps d'arrêt n'est pas tout–à–fait standard; la même remarque s'applique à ce qui suit et explique la notation $\mathcal{F}_{T+}$). On lui associe sa "tribu antérieure":

$$\mathcal{F}_{T+} = \{A : \ A \in \mathcal{F}, A \cap \{T < t\} \in \mathcal{F}_t \ \ \forall t \in \mathbb{R}_+\}.$$

Ceci définit bien une tribu, et si $T(\omega) = t$ pour tout ω alors T est un temps d'arrêt et $\mathcal{F}_{T+} = \mathcal{F}_{t+} = \cap_{s>t}\mathcal{F}_s$. Ensuite, sur l'ensemble $\{T < \infty\}$, on pose

$$\theta_T(\omega) = \theta_{T(\omega)}(\omega), \qquad X_T(\omega) = X_{T(\omega)}(\omega).$$

On dira que le processus vérifie *la propriété forte de Markov* si, pour toute loi initiale μ, tout temps d'arrêt T et toute variable aléatoire Y positive ou bornée, on a

$$\mathbb{E}_\mu(Y \circ \theta_T | \mathcal{F}_{T+}) = E_{X_T}(Y) \qquad \text{sur l'ensemble } \{T < \infty\}. \tag{3.1.8}$$

Comme dans le cas discret, l'ensemble $\{T < \infty\}$ est $\mathcal{F}_{T+}$–mesurable, et donc l'égalité ci–dessus ne fait en fait intervenir $Y \circ \theta_T$ que sur cet ensemble. Ceci dit, (3.1.8) suppose, pour avoir un sens, que X_T et $Y \circ \theta_T$ soient, en restriction à $\{T < \infty\}$, mesurables respectivement par rapport à $\mathcal{F}_{T+}$ et à $\mathcal{F}$ (au moins à des ensembles $\mathbb{P}_x$–négligeables près pour tout x). Ceci peut être faux, mais, même si c'est vrai, la propriété (3.1.8) *n'est pas toujours vraie.*

Il existe un certain nombre de conditions assurant cette propriété. Nous en explicitons une ci–dessous:

Théorème 3.1.6. *Soit* $\mathbf{X}$ *un processus de Markov de transitions* (P_t)*. Si* E *est un espace polonais, la propriété de Markov forte (3.1.8) est valide (et en particulier pour tout* x*,* X_T *est* $\mathbb{P}_x$*–p.s. égale à une variable* $\mathcal{F}_{T+}$*–mesurable, et* $Y \circ \theta_T$ *est* $\mathbb{P}_x$*-p.s. égale à une variable* $\mathcal{F}$*-mesurable) dès qu'on a les deux propriétés suivantes:*

1) pour toute fonction continue bornée f *sur* E *les fonctions* $P_t f$ *sont également continues (on dit que le semi–groupe* (P_t) *est* fellérien*),*

2) pour tout x *et* $\mathbb{P}_x$*–presque tout* ω*, la fonction* $t \mapsto X_t(\omega)$ *est continue à droite.*

Preuve. On peut supposer (par un argument de classe monotone, et puisque E est polonais) que $Y = \prod_{i=0}^m f_i(X_{t_i})$ pour des f_i bornées et *continues* sur E et $0 = t_0 < \ldots < t_m$.

a) Soit T un temps d'arrêt, et posons

$$T_n = \begin{cases} \frac{k+1}{2^n} & \text{si } \frac{k}{2^n} \leq T < \frac{k+1}{2^n}, \quad k \in \mathbb{N}, \\ \infty & \text{si } T = \infty. \end{cases}$$

On a $B_{n,k} := \{T_n = k/2^n\} = \{(k-1)/2^n \leq T < k/2^n\}$, donc $B_{n,k} \in \mathcal{F}_{k/2^n}$ pour tout $k \in \mathbb{N}^\star$. On pose aussi $B = \{T < \infty\}$, qui égale $\cup_{k \geq 1} B_{n,k}$ pour tout n. Enfin, il est clair que T_n décroit vers T. Dans la suite "p.s." veut dire $\mathbb{P}_x$–p.s. pour tout x.

b) Montrons d'abord les propriétés de mesurabilité de X_T et $Y \circ \theta_T$ sur B. D'abord, on a $\{X_{T_n + t_i} \in A\} \cap B = \cup_{k \geq 1}(B_{n,k} \cap \{X_{k/2^n + t_i} \in A\})$, qui est dans $\mathcal{F}$ pour tout $A \in \mathcal{E}$: donc $(Y \circ \theta_{T_n}) 1_B$ est $\mathcal{F}$–mesurable et comme $Y \circ \theta_{T_n} 1_B \to Y \circ \theta_T 1_B$ p.s. (à cause de (2) et de la forme spéciale de Y), on a la $\mathcal{F}$–mesurabilité p.s. de $Y \circ \theta_T 1_B$.

Par ailleurs fixons $t > 0$. Posons $V_n = X_{T_n}$ sur $\{T_n < t\}$ et $V_n = \Delta$ ailleurs (où Δ est un point extérieur à E, isolé topologiquement dans $E_\Delta = E \cup \{\Delta\}$), et de même $V = X_T$ sur $\{T < t\}$ et $V = \Delta$ ailleurs. Comme $\{V_n \in A\} = \cup_{1 \leq k < t2^n}(B_{n,k} \cap \{X_{k/2^n} \in A\}) \in \mathcal{F}_t$ pour tout $A \in \mathcal{E}$, les variables V_n sont $\mathcal{F}_t$–mesurable, donc V, qui par (2) et $T_n \downarrow T$ est la limite p.s. des V_n, est p.s. égale à une variable $\mathcal{F}_t$–mesurable. Comme $\{X_T \in A, T < t\} = \{V \in A\} \in \mathcal{F}_t$ pour $A \in \mathcal{E}$, on voit que X_T en restriction à B est p.s. égale à une variable $\mathcal{F}_{T+}$–mesurable.

c) Soit Z bornée et $\mathcal{F}_{T+}$–mesurable. La variable $Z1_{B_{n,k}}$ est $\mathcal{F}_{k/2^n}$–mesurable, donc

$$\begin{aligned}
\mathbb{E}_\mu\left(Z\ 1_B\ Y \circ \theta_{T_n}\right) &= \sum_{k=1}^{\infty} \mathbb{E}_\mu\left(Z\ 1_{B_{n,k}}\ Y \circ \theta_{k/2^n}\right) \\
&= \sum_{k=1}^{\infty} \mathbb{E}_\mu\left(Z\ 1_{B_{n,k}}\ E_{X_{k/2^n}}(Y)\right) \qquad \text{(par (3.1.5))} \\
&= \mathbb{E}_\mu(Z\ 1_B\ \mathbb{E}_{X_{T_n}}(Y)),
\end{aligned} \tag{3.1.9}$$

Quand $n \to \infty$, le membre de gauche tend vers $\mathbb{E}_\mu\left(Z1_B Y \circ \theta_T\right)$ à cause de (2) et de la forme spéciale de Y. Par ailleurs le même raisonnement que dans la proposition 3.1.3 montre que $\mathbb{E}_x(Y) = g_0(x)$, où on définit les g_i par récurrence descendante en partant de $g_m = f_m$ et en posant $g_i(x) = f_i(x) P_{t_{i+1}-t_i} g_{i+1}(x)$. Etant donné (1) et la continuité des f_i, on voit que les g_i sont également continues et bornées: donc le membre de droite de (3.1.9) converge vers $\mathbb{E}_\mu(Z1_B E_{X_t}(Y))$, ce qui achève la preuve. □

De même que (3.1.5) admet l'extension (3.1.7), la propriété forte de Markov admet l'extension suivante: si on a une fonction mesurable Y de $(\Omega, \mathcal{F})$ dans un espace $(E, \mathcal{E})$, et une fonction f positive ou bornée sur $\Omega \times E$, qui est $\mathcal{F}_{T+} \otimes \mathcal{E}$–mesurable pour un temps d'arrêt T, alors

$$\mathbb{E}_\mu(f(.,Y \circ \theta_T | \mathcal{F}_{T+}) = \int \mathbb{P}_{X_T}(d\omega') f(.,Y(\omega')) \qquad \text{sur l'ensemble } \{T < \infty\}. \tag{3.1.10}$$

Corollaire 3.1.7. (Loi 0–1) *Si on a la propriété de Markov forte, alors pour tout $x \in E$ et tout $A \in \mathcal{F}_{0+}$ le nombre $\mathbb{P}_x(A)$ vaut 0 ou 1.*

Preuve. Soit $A \in \mathcal{F}_{0+}$ et $f(x) = \mathbb{P}_x(A)$. Comme θ_0 est l'identité, en appliquant (3.1.8) à $T \equiv 0$ on obtient

$$f(x) = \mathbb{E}_x(1_A\ 1_A \circ \theta_0) = \mathbb{E}_x(1_A\ \mathbb{E}_x(1_A \circ \theta_0 | \mathcal{F}_{0+})) = \mathbb{E}_x(1_A\ f(X_0)) = f(x)^2,$$

puisque $\mathbb{P}_x(X_0 = x) = 1$. □

La condition (1) du théorème 3.1.6 est agréable, puisque qu'elle s'exprime directement en fonction du semi–groupe (P_t). Il n'en est pas de même de (2), qui est évidemment fort difficile à vérifier dans le cas général. Nous énonçons ci–dessous (sans la démonstration – difficile) un critère impliquant (2), et même un peu plus.

Théorème 3.1.8. *Soit (P_t) un semi-groupe de probabilités de transition sur l'espace $(E, \mathcal{E})$, espace localement compact de type dénombrable muni de ses boréliens. On peut construire un processus de Markov $\mathbf{X}$ de transitions (P_t), dont les trajectoires $t \mapsto X_t(\omega)$ sont continues à droite et admettent des limites à gauche, dès que le semi–groupe est* fortement fellérien*, ce qui signifie qu'il vérifie les deux propriétés suivantes:*

(a) pour toute fonction continue nulle à l'infini (pour le compactifié d'Alexandrov de E) f, les fonctions $P_t f$ sont également continues nulles à l'infini,

(b) pour toute fonction comme ci–dessus, $P_t f$ converge simplement vers f quand $t \to 0$.

On remarquera les conditions sur E: pour les théorèmes 3.1.5 et 3.1.6 on a besoin de E polonais, ce qui est le cas si E est un borélien de $E = I\!R^d$ ou de $I\!R^{I\!N}$ (pour la topologie produit). Pour le dernier résultat, la condition est pratiquement que E est un compact séparable, ou est un ouvert de $I\!R^d$ pour d fini. Noter aussi, à l'inverse, que si (X_t) est à trajectoires continues à droite, la condition (b) ci–dessus est nécessairement satisfaite.

3.1.3 Stationnarité

Soit $\mathbf{X}$ un processus de Markov à valeurs dans $(E, \mathcal{E})$. Comme dans le cas discret, on peut se poser la question de savoir s'il existe une ou plusieurs lois initiales $I\!P_\mu$ rendant le processus stationnaire. Rappelons que (X_t) est *stationnaire* sous la probabilité $I\!P_\mu$ si pour tous $t_1, \dots, t_m$ la loi des variables $(X_{t+t_1}, \dots, X_{t+t_m})$ sous $I\!P_\mu$ ne dépend pas de $t \geq 0$.

De même que dans la définition 1.4.3, on dit qu'une mesure η sur $(E, \mathcal{E})$ est *invariante*, resp. *sous–invariante*, si $\eta P_t = \eta$, resp. $\eta P_t \leq \eta$, pour tout $t \geq 0$.

Proposition 3.1.9. *Le processus (X_t) est stationnaire sous la probabilité $I\!P_\mu$ si et seulement si la probabilité μ est invariante.*

Preuve. On reproduit essentiellement la preuve de la proposition 1.4.6. Si le processus est stationnaire sous $I\!P_\mu$, la loi μP_t de X_t égale la loi μ de X_0 pour tout $t \geq 0$, donc μ est invariante.

Inversement supposons μ invariante. Soit $m \geq 0$ et $0 = t_0 < \ldots < t_m$ et $A \in \mathcal{E}^{\otimes(m+1)}$. On pose $Y = 1_A(X_{t_0}, \dots, X_{t_m})$. La loi η_t de $(X_{t+t_0}, \dots, X_{t+t_m})$ est caractérisée par les nombres $\eta_t(A) = I\!E_\mu(Y \circ \theta_t)$. D'après (3.1.5), et si $f(x) = I\!E_x(Y)$, on a alors

$$\eta_t(A) = I\!E_\mu(f(X_t)) = \mu P_t f,$$

qui vaut $\mu(f) = \eta_0(A)$ à cause de l'invariance de μ. $\square$

3.2 Un exemple: les processus de Lévy

Un exemple fondamental de processus de Markov est constitué des "processus de Lévy", ou "processus à accroissements indépendants et stationnaires", étudiés intensivement par Paul Lévy dans les années 1930–1940.

Définition 3.2.1. Soit $(X_t)_{t \geq 0}$ un processus à valeurs dans $I\!R^d$, défini sur un espace probabilisé filtré $(\Omega, \mathcal{F}, (\mathcal{F}_t), I\!P)$. On dit que c'est un processus *à accroissements* $(\mathcal{F}_t)$*–indépendants et stationnaires* si:

(i) chaque X_t est $\mathcal{F}_t$–mesurable;

(ii) pour tous $s, t \geq 0$, la variable $X_{t+s} - X_t$ est indépendante de la tribu $\mathcal{F}_t$, et de même loi que X_s.

On écrit en abrégé $(\mathcal{F}_t)$–PAIS, et simplement PAIS si $\mathcal{F}_t = \sigma(X_s : s \leq t)$ (dans ce cas (i) est automatique).

On dit que X est un $(\mathcal{F}_t)$*–processus de Lévy* (resp. *processus de Lévy*), si c'est un $(\mathcal{F}_t)$–PAIS (resp. PAIS) qui vérifie en outre:

(iii) $X_0 = 0$ et les trajectoires $t \mapsto X_t(\omega)$ sont continues à droites et avec des limites à gauche. □

Proposition 3.2.2. *Soit X un $(\mathcal{F}_t)$–PAIS et μ_t la loi commune des accroissements $X_{s+t} - X_s$. La famille $(\mu_t)_{t\geq 0}$ est un semi–groupe de convolution, ce qui signifie que pour tout $s, t \geq 0$ on a*

$$\mu_{t+s} = \mu_t \star \mu_s. \tag{3.2.1}$$

De plus le processus X satisfait la $(\mathcal{F}_t)$–propriété de Markov (au sens de la définition 3.1.1) avec le semi–groupe de transition (P_t) défini par

$$P_t(x, A) = \int 1_A(x + y)\mu_t(dy). \tag{3.2.2}$$

Preuve. Comme $X_{t+s} = X_t + (X_{t+s} - X_t)$ et comme les variables X_t et $X_{t+s} - X_t$ sont indépendantes et de lois respectives μ_t et μ_s, (3.2.1) est évident. La formule (3.2.2) définit clairement une probabilité de transition P_t de $\mathbb{R}^d$ dans lui-même. De plus (3.2.1) implique

$$\begin{aligned} P_{t+s}(x, A) &= \int 1_A(x + y)\mu_{t+s}(dy) = \int 1_A(x + u + v)\mu_t(du)\mu_s(dv) \\ &= \int \mu_t(du) \left(\int 1_A(x + u + v)\mu_s(dv) \right) \\ &= \int \mu_t(du)P_s(x + v, A) \;=\; (P_t P_s)(x, A), \end{aligned}$$

donc (P_t) est une semi–groupe. Enfin on a

$$\mathbb{E}(f(X_{t+s})|\mathcal{F}_t) = \mathbb{E}(f(X_t + (X_{t+s} - X_t))|\mathcal{F}_t) = \int \mu_s(dy)f(X_t + y) = P_s f(X_t),$$

où la seconde égalité vient de ce que $X_{t+s} - X_t$ est indépendant de $\mathcal{F}_t$. □

A l'inverse, on peut associer à tout semi–groupe de convolution (μ_t) de probabilités sur $\mathbb{R}^d$ un PAIS:

Théorème 3.2.3. *Soit (μ_t) un semi–groupe de convolution sur $\mathbb{R}^d$.*

a) La formule (3.2.2) définit un semi–groupe de transitions, auquel on peut associer un processus de Markov $\mathbf{X}$ au sens de la définition 3.1.2, et sous chaque probabilité $\mathbb{P}_\nu$ le processus (X_t) est un PAIS, la loi de $X_{t+s} - X_t$ étant μ_s.

b) Pour qu'on puisse définir un processus de Markov $\mathbf{X}$ comme ci–dessus, tel que les trajectoires de (X_t) soient $\mathbb{P}_\nu$–p.s. continues à droite et avec des limites à gauche pour toute probabilité ν sur $\mathbb{R}^d$, il faut et il suffit qu'on ait:

$$\mu_t \quad \text{converge étroitement vers la masse de Dirac en } 0 \text{ quand } \; t \to 0. \tag{3.2.3}$$

Dans ce cas, le processus $\mathbf{X}$ a la propriété de Markov forte.

Preuve. a) La première assertion a été prouvée ci–dessus. D'après le théorème 3.1.5 on peut construire un processus de Markov $\mathbf{X} = (\Omega, \mathcal{F}, (\mathcal{F}_t), (\theta_t), (X_t), (I\!\!P_x))$ de transitions (P_t). Si f est borélienne bornée et ν une probabilité sur $I\!\!R^d$, on a alors d'après (3.1.5):

$$I\!\!E_\nu(f(X_{t+s} - X_t)|\mathcal{F}_t) = I\!\!E_\nu(f(X_s \circ \theta_t - X_t)|\mathcal{F}_t) = \int P_s(X_t, dy) f(y - X_t) = \mu_s(f),$$

où on a utilisé (3.1.7) pour la seconde égalité, et (3.2.1) pour la dernière: on en déduit la dernière assertion de (a).

b) On a $\mu_t = P_t(0,.)$, donc $\mu_t(f) = I\!\!E_0(f(X_t))$. Si X est $I\!\!P_0$–p.s. à trajectoires continues à droite, dès que f est continue on a $f(X_t) \to f(X_0) = 0$ $I\!\!P_0$–p.s. Si de plus f est bornée, le théorème de Lebesgue implique alors $\mu_t(f) \to f(0)$, et on a (3.2.3).

Pour la condition suffisante et la propriété de Markov forte, et comme $E = I\!\!R^d$ est un espace localement compact de type dénombrable, il suffit en vertu des théorèmes 3.1.6 et 3.1.8 de montrer que pour toute fonction continue bornée f sur $I\!\!R^d$ on a

$$P_t f \quad \text{est continue,} \tag{3.2.4}$$

$$\text{si } f(x) \to 0 \text{ quand } |x| \to \infty, \text{ alors } \quad P_t f(x) \to 0 \quad \text{quand} \quad |x| \to \infty, \tag{3.2.5}$$

$$\text{si } f(x) \to 0 \text{ quand } |x| \to \infty, \text{ alors } \quad P_t f(x) \to f(x) \quad \text{quand} \quad t \to 0. \tag{3.2.6}$$

On a $P_t f(x) = \int f(x+y)\mu_t(dy)$, donc (3.2.4) et (3.2.5) découlent de manière évidente du théorème de Lebesgue, tandis que (3.2.6) vient de (3.2.3) (et d'ailleurs $P_t f \to f$ simplement dès que f est continue bornée). □

Les trois exemples les plus simples de processus de Lévy sont:

- La "translation pure", qui est le processus (déterministe) $X_t = at$, où $a \in I\!\!R^d$.
- Le mouvement brownien (linéaire si $d = 1$, d–dimensionnel dans le cas $d \geq 2$).
- Le processus de Poisson (pour $d = 1$, et plus généralement les "processus de Poisson composé", que nous étudierons plus tard.

On peut "combiner" les exemples précédents, en additionnant des procesus de Lévy indépendants: on obtient encore un processus de Lévy. On peut aussi montrer (mais c'est difficile) que tout processus de Lévy est limite, en un certain sens, de processus qui sont des sommes des trois exemples ci–dessus.

Remarque 3.2.4. *:* Supposons que X soit un processus déterministe, donc de la forme $X_t = f(t)$ pour une fonction f de $I\!\!R_+$ dans $I\!\!R^d$, et supposons aussi $f(0) = 0$. Bien entendu, X est à accroissements indépendants (des variables "déterministes" sont toujopurs indépendantes entre elles), donc dire que X est un PAIS revient à dire que

$$f(t+s) = f(t) + f(s) \qquad \forall s, t \geq 0. \tag{3.2.7}$$

Si la fonction f est continue, ou continue à droite, ou même simplement borélienne, on sait que (3.2.7) entraîne que f est linéaire, i.e. $f(t) = at$ pour un $a \in I\!\!R^d$, mais il existe des solutions de l'équation fonctionnelle (3.2.7) qui ne sont pas boréliennes (donc pas linéaires non plus).

Dans le même ordre d'idées, si on a un semi–groupe de convolution (μ_t) et si on note $u \mapsto \widehat{\mu}_t(u)$ la fonction caractéristique de μ_t, l'équation (3.2.1) est équivalente à

$$\widehat{\mu}_{t+s}(u) = \widehat{\mu}_t(u)\widehat{\mu}_s(u) \qquad \forall s,t \geq 0, \forall u \in I\!R^d.$$

Dès que $\widehat{\mu}_t(u)$ est mesurable en t, l'équation fonctionnelle ci–dessus implique que $\widehat{\mu}_t(u) = \exp t\psi(u)$ pour un certain nombre complexe $\psi(u)$: dans ce cas on a évidemment que $\widehat{\mu}_t(u) \to 1$ quand $t \to 1$, et d'après le théorème de Lévy cela implique (3.2.3) si cette propriété est vraie pour tout u: on peut donc considérablement affaiblir (3.2.3): par exemple, il suffit que $\mu_t(f)$ soit borélienne, pour tout fonction continue bornée f sur $I\!R^d$, sur un intervalle $[0,\varepsilon]$. A l'inverse, si le PAIS X associé a des trajectoires boréliennes, on a nécessairement (3.2.3) et donc il existe une "version" de X qui a des trajectoires continues à droite et avec des limites à gauche. $\square$

Revenons sur les rapports entre PAIS et processus de Lévy. Un processus de Lévy est un PAIS continu à droite et limité à gauche (nous avons vu un critère pour cela), *qui de plus vérifie* $X_0 = 0$.

Soit alors $\mathbf{X} = (\Omega, \mathcal{F}, (\mathcal{F}_t), I\!P_x)$ un processus de Markov (définition 3.1.2) qui est un PAIS à trajectoires continues à droite et limitées à gauche. Alors, quitte à ôter de Ω l'ensemble $I\!P_0$–négligeable $\{X_0 \neq 0\}$, on voit que (X_t) est un processus de Lévy sous $I\!P_0$ (mais pas sous $I\!P_x$ pour $x \neq 0$. On peut aussi poser

$$Y_t = X_t - X_0. \tag{3.2.8}$$

Le processus Y est un processus de Lévy sous chaque $I\!P_x$ (et même chaque $I\!P_\nu$), associé bien–sûr au même semi–groupe de convolution. Noter qu'on a $Y_{t+s} - Y_t = Y_s \circ \theta_t$ ici, et non pas $Y_{t+s} = Y_s \circ \theta_t$.

On peut se poser le problème inverse: à partir d'un processus de Lévy Y, comment construire un processus de Markov $\mathbf{X}$ qui est un PAIS associé au même semi–groupe ? Habituellement on suppose en plus que l'espace $(\Omega', \mathcal{F}', (\mathcal{F}'_t), I\!P')$ sur lequel est défini Y est muni d'un semi–groupe de translations (θ'_t), qui vérifie

$$\theta'_0 = I, \qquad \theta'_{s+t} = \theta'_s \circ \theta_t, \qquad Y_{s+t} - Y_s = Y_t \circ \theta'_s, \quad \forall s,t \geq 0 \tag{3.2.9}$$

(en accord avec ce qui est dit plus haut, comparer avec (3.1.3)). Par exemple, on peut prendre pour Ω' l'espace canonique, qui ici est l'espace de toutes les fonctions $t \mapsto x(t)$ de $I\!R_+$ dans $I\!R^d$ vérifiant $x(0) = 0$ et continues à droite avec des limites à gauche. Puis, pour toute fonction $x(.)$ on pose

$$Y_t(x) = x(t), \qquad \theta'_t(x)(.) = x(t+.) - x(t),$$

de sorte qu'on a (3.2.9). Enfin $\mathcal{F}'_t = \sigma(Y_s : s \leq t)$ et $\mathcal{F}' = \bigvee_t \mathcal{F}'_t$. Dans ce cas, on dit qu'on a un processus de Lévy *canonique*.

On peut alors poser

$$\Omega = \Omega' \times I\!R^d, \qquad X_t(\omega, x) = x + Y_t(\omega), \qquad \theta'_t(\omega, x) = (\theta'_t(\omega), x + Y_t(\omega)),$$

puis

$$\mathcal{F} = \mathcal{F}' \otimes \mathcal{R}^d, \qquad \mathcal{F}_t = \mathcal{F}'_t \otimes \mathcal{R}^d$$

(où $\mathcal{R}^d$ est la tribu borélienne de $I\!R^d$), et enfin

$$I\!P_y(d\omega, dx) = I\!P(d\omega) \otimes \varepsilon_y(dx) \quad \forall y \in I\!R^d.$$

On vérifie immédiatement la

Proposition 3.2.5. *Le terme* $\mathbf{X} = (\Omega, \mathcal{F}, (\mathcal{F}_t), (\theta_t), (X_t), (I\!P_y)_{y \in I\!R^d})$ *ci–dessus est un processus de Markov (PAIS), de semi–groupe de transition* (P_t) *donné par (3.2.2), où* (μ_t) *est le semi–groupe de convolution associé au processus de Lévy* Y*.*

Enfin, les processus de Lévy admettent une version de la propriété forte de Markov qui mérite d'être énoncée. Elle découle de manière quasi–évidente de la propriété forte de Markov pour le processus de Markov **X** associé ci–dessus (propriété satisfaite par **X**, dont les trajectoires sont continues à droite par construction, en vertu du théorème 3.1.6).

Théorème 3.2.6. *Soit* (Y_t) *un* $(\mathcal{F}'_t)$*–processus de Lévy sur l'espace* $(\Omega', \mathcal{F}', (\mathcal{F}'_t), (\theta'_t), I\!P')$*. Si* T *est un temps d'arrêt, pour toute variable aléatoire* Z *positive ou bornée on a*

$$I\!E'(Z \circ \theta'_T | \mathcal{F}'_{T+}) = I\!E'(Z) \qquad \textit{sur l'ensemble} \ \{T' < \infty\}. \tag{3.2.10}$$

En particulier, si $T < \infty$ *identiquement, le processus translaté* $Y \circ \theta'_T = (Y_{T+t} - Y_T)_{t \geq 0}$ *est indépendant de la tribu* $\mathcal{F}'_{T+}$ *et a même loi que* Y *lui–même.*

3.3 Générateur infinitésimal et résolvante

Ce paragraphe est consacré à l'étude des semi–groupes de probabilités de transition (P_t), et pour l'essentiel est indépendant des propriétés des processus de Markov associés. L'espace d'états $(E, \mathcal{E})$ est a priori arbitraire, bien que de temps en temps on suppose que c'est un espace topologique, voire que c'est $I\!R^d$.

L'idée de ce qu'on va faire ici est très simple: de manière très informelle, supposons que $t \mapsto P_t$ soit "dérivable en 0", et notons A la dérivée. Si on "dérive" l'équation $P_{t+s} = P_t P_s$ en s ou en t, on obtient que $t \mapsto P_t$ est aussi dérivable, et que $P'_t = AP_t = P_t A$, de sorte que (P_t) est solution d'une équation différentielle linéaire, et donc s'écrit sous la forme $P_t = e^{tA}$. Ce qui précède n'a pour le moment aucun sens (sauf si E est un ensemble fini, puisqu'alors chaque P_t se représente par une matrice carrée). Mais l'objectif de ce paragraphe est d'essayer de donner un contenu mathématique précis à ces élucubrations informelles.

3.3.1 Le générateur

Commençons par quelques notations. Nous noterons B l'espace de toutes les fonctions mesurables bornées sur $(E,\mathcal{E})$ et, quand E est un espace topologique, C désigne l'ensemble des fonctions *continues* bornées. On munit B de la topologie "simple–bornée", pour laquelle une suite $(f_n)_{n\in\mathbb{N}}$ ou une famille $(f_t)_{t\geq 0}$ de fonctions converge vers une limite f quand $n\to\infty$ ou $t\to 0$ si elle converge simplement, et si en plus $\sup_{x\in E,n\in\mathbb{N}}|f_n(x)|<\infty$ ou $\sup_{x\in E,t\geq 0}|f_t(x)|<\infty$. On écrit alors $f_n \stackrel{\text{s−b}}{\longrightarrow} f$ ou $f_t \stackrel{\text{s−b}}{\longrightarrow} f$.

Le semi–groupe (P_t) de probabilités de transition sur $(E,\mathcal{E})$ est supposé fixé. On lui associe l'espace B_0 suivant

$$B_0=\{f\in B:\ \lim_{t\to 0} P_tf(x)=f(x)\ \ \forall x\} \tag{3.3.1}$$

(on a bien–sûr $P_tf \stackrel{\text{s−b}}{\longrightarrow} f$ quand $t\to 0$, si $f\in L$). Puis on définit l'opérateur linéaire à domaine $(A,\mathcal{D}_A)$, sur B, par

$$\left.\begin{array}{l} \mathcal{D}_A=\{f\in B:\ \ \exists g\in B_0\ \text{ avec }\ \frac{P_tf-f}{t}\stackrel{\text{s−b}}{\longrightarrow} g\ \text{ quand }\ t\downarrow 0\} \\ f\in\mathcal{D}_A\quad\Longrightarrow\quad Af(x)=\lim_{t\downarrow 0}\frac{P_tf(x)-f(x)}{t}. \end{array}\right\} \tag{3.3.2}$$

($\mathcal{D}_A$ est évidemment un sous–espace vectoriel de B_0, lui-même un sous-espace vectoriel de B, et si $f\in\mathcal{D}_A$ on a $Af\in B$).

Par définition, cet opérateur s'appelle le *générateur infinitésimal faible* (ou simplement le "générateur infinitésimal", ou même le "générateur") du semi–groupe. Les deux théorèmes suivants sont connus sous le nom de théorème(s) de Hille–Yoshida.

Théorème 3.3.1. *a) Si $f\in B_0$ alors $P_tf\in B_0$ et $t\mapsto P_tf(x)$ est continu à droite.*

b) L'espace $\mathcal{D}_A$ est dense dans B_0 pour la convergence simple–bornée.

c) Si $f\in\mathcal{D}_A$, les fonctions $t\mapsto P_tf(x)$ sont continues, et dérivables à droite, et si $\frac{d^+}{dt}\,P_tf(x)$ désigne les dérivées à droite on a $P_tf\in\mathcal{D}_A$ et

$$\frac{d^+}{dt}\,P_tf(x)=P_tAf(x)=AP_tf(x) \tag{3.3.3}$$

et

$$P_tf(x)=f(x)+\int_0^t P_sAf(x)ds=f(x)+\int_0^t AP_sf(x)ds. \tag{3.3.4}$$

d) L'opérateur $(A,\mathcal{D}_A)$ est, dans l'espace B_0 muni de la convergence simple–bornée, un opérateur fermé.

Preuve. a) Soit $f\in B_0$. Comme $P_sf\stackrel{\text{s−b}}{\longrightarrow} f$ si $s\downarrow 0$, d'après le théorème de Lebesgue on a $P_sP_tf=P_{t+s}f=P_tP_sf\to P_tf$, d'où d'une part $P_tf\in B_0$, et d'autre part la continuité à droite.

b) Soit $f\in B_0$. D'après (a), on peut faire l'intégrale $g_a(x)=\int_0^a P_tf(x)dt$. Mais alors on obtient pour $s<a$, en appliquant Fubini:

$$P_sg_a(x)=\int_0^a P_sP_tf(x)dt=\int_s^{a+s} P_tf(x)dx=g_a(x)+\int_a^{a+s} P_tf(x)dt-\int_0^s P_tf(x)dt,$$

$$\frac{P_s g_a(x) - g_a(x)}{s} = \frac{1}{s}\int_a^{a+s} P_t f(x)dt - \frac{1}{s}\int_0^s P_t f(x)dt.$$

le membre de droite ci–dessus est borné uniformément en x et s, et à cause de (a) il converge vers $P_a f(x) - f(x)$ quand $s \downarrow 0$, et enfin $P_a f - f \in B_0$ à cause de (a) encore, de sorte que $g_a \in \mathcal{D}_A$. Comme par ailleurs $ng_{1/n} \stackrel{\text{s−b}}{\longrightarrow} f$ (pour les mêmes raisons), on a le résultat.

c) Soit $f \in \mathcal{D}_A$. D'après le théorème de Lebesgue et (3.3.2), on a

$$\frac{P_s(P_t f) - P_t f}{s} = \frac{P_{t+s} f - P_t f}{s} = P_t\left(\frac{P_s f - f}{s}\right) \stackrel{\text{s−b}}{\longrightarrow} P_t A f.$$

Cela montre la dérivabilité à droite, la propriété $P_t f \in A$ (rappelons que $Af \in B_0$ donc $P_t A f \in B_0$ par (a)) et la formule (3.3.3). En particulier le second et le troisième membres de (3.3.4) sont égaux. Notons alors, pour x fixé, la différence du premier et du second membres de (3.3.4) par $g(t)$. On a $g(0) = 0$ et g est continue, et dérivable à droite, de dérivée à droite nulle. Cela implique classiquement[1] que $g(t) = 0$ pour tout t et on a donc (3.3.4).

d) Rappelons que l'opérateur $(A, \mathcal{D}_A)$ sur B_0 est dit "fermé" si, pour toute suite (f_n) d'éléments de $\mathcal{D}_A$ convergeant vers une limite $f \in B_0$, et telle que les $g_n = Af_n$ convergent vers une limite $g \in B_0$ (ici, au sens de la convergence simple–bornée), alors $f \in \mathcal{D}_A$ et $Af = g$. D'après (3.3.4), on a

$$P_t f_n(x) = f_n(x) - \int_0^t P_s g_n(x)ds.$$

On peut passer à la limite (au sens de la convergence simple–bornée) dans chacun des deux membres ci–dessus, et on obtient

$$P_t f(x) = f(x) - \int_0^t P_s g(x)ds.$$

Comme $g \in B_0$, on en déduit d'abord que $\frac{P_t f - f}{t} \stackrel{\text{s−b}}{\longrightarrow} g$ (utiliser (a)), puis que $f \in \mathcal{D}_A$ et $Af = g$. $\square$

Ensuite, on définit la "résolvante" du semi–groupe. Informellement, la *résolvante* est la famille de mesures de transition U^λ, définies pour tout $\lambda > 0$ par $U^\lambda = \int_0^\infty e^{-\lambda t} P_t dt$. Donner un sens précis à cette intégrale est toutefois un peu délicat. On voudrait poser pour toute fonction f mesurable positive ou bornée:

$$U^\lambda f(x) = \int_0^\infty e^{-\lambda t}\, P_t f(x)\, dt. \tag{3.3.5}$$

Mais la fonction qu'on cherche à intégrer ci–dessus n'est pas nécessairement mesurable (en t), et il se peut donc que (3.3.5) n'ait pas de sens. Toutefois, si $f \in B_0$, alors l'intégrand

[1] pour le voir, on peut fixer $a > 0$ et poser $s = \inf(t : g(t) > at)$. Si $s < \infty$ on a $g(s) \leq as$ (à cause de la continuité) et puisque la dérivée à droite est nulle on a $g(s+u) - g(s) \leq ua$ et donc $g(s+u) \leq a(s+u)$ pour tout $u > 0$ assez petit: cela contredit la définition de s, donc $s = \infty$, donc $g(t) \leq at$ pour tout t. Ceci est vrai pour tout $a > 0$, donc $g \geq 0$; on montre de même que $g \leq 0$, donc $g \sim 0$.

est mesurable (et même continu à droite), et aussi majoré par le produit d'une constante par la fonction intégrable $t \mapsto e^{-\lambda t}$: donc (3.3.5) a un sens, et définit un réel.

En d'autres termes, la formule (3.3.5) définit pour chaque $\lambda > 0$ un opérateur U^λ, évidemment linéaire, de B_0 dans B. On a même bien plus, puisque le résultat suivant implique notamment que $U^\lambda f \in \mathcal{D}_A$ lorsque $f \in B_0$:

Théorème 3.3.2. *Si $g \in B_0$ et si $\lambda > 0$, l'équation $(\lambda I - A)f = g$ (où I est l'opérateur identité) admet une solution et seule dans $\mathcal{D}_A$, qui est donnée par $f = U^\lambda g$.*

Preuve. On va d'abord montrer que $h = U^\lambda g$ est solution. D'après Fubini, on a

$$P_t h(x) = \int_0^\infty e^{-\lambda s}\ P_{t+s} g(x)\ ds,$$

et comme $t \mapsto P_{t+s}g(x)$ est continu à droite le théorème de Lebesgue permet d'obtenir que $P_t h(x) \to h(x)$ quand $t \downarrow 0$. On a aussi

$$\begin{aligned}\frac{P_t h(x) - h(x)}{t} &= \frac{1}{t}\left(e^{\lambda t}\int_t^\infty e^{-\lambda s}\ P_s g(x)\ ds - \int_0^\infty e^{-\lambda s}\ P_s g(x)\ ds\right) \\ &= \frac{1}{t}\left((e^{\lambda t}-1)h(x) - e^{\lambda t}\int_0^t e^{-\lambda s}\ P_s g(x)\ ds\right) \ \xrightarrow{\text{s−b}}\ \lambda h(x) - g(x)\end{aligned}$$

puisque $g \in B_0$. On en déduit a fortiori que $h \in B_0$, donc finalement $h \in \mathcal{D}_A$ et $Ah = \lambda h - g$: ainsi, h est solution de notre équation.

Il reste à montrer l'unicité. Soit f et f' deux solutions. La fonction $k = f - f'$ vérifie alors $k \in \mathcal{D}_A$ et $Ak = \lambda k$. Par suite, pour chaque x fixé, la fonction $\psi(t) = P_t k(x)$ est continue, et dérivable à droite, et sa dérivé à droite $\psi'_+(t)$ est (cf. (3.3.3)) $P_t Ak(x) = \lambda P_t k(x) = \lambda\psi(t)$. On en déduit, comme pour la preuve de (3.3.4), que

$$\psi(t) = \psi(0) + \lambda\int_0^t \psi(s)ds.$$

Par suite ψ, qui est continue, est aussi dérivable et vérifie l'équation différentielle $\psi' = \lambda\psi$: comme $\psi(0) = k(x)$, on obtient $P_t k(x) = \psi(t) = k(x)e^{\lambda t}$. Mais $e^{\lambda t} \uparrow \infty$ quand $t \to \infty$, tandis que $P_t k(x)$ reste borné (puisque f et f', donc k, sont des fonctions bornées). On arrive à une contradiction, à moins que $k(x) = 0$, et cela nous donne finalement $f' = f$. $\square$

Corollaire 3.3.3. *Si deux semi–groupes (P_t) et (P'_t) sur E admettent le même générateur infinitésimal, les deux espaces B_0 et B'_0 associés par (3.3.1) sont égaux, et pour toute fonction f dans la fermeture de $B_0 = B'_0$ pour la convergence simple–bornée et tout $t \geq 0$ on a $P_t f = P'_t f$.*

Preuve. Notons $(A, \mathcal{D}_A)$ le générateur commun, et $(U^\lambda)_{\lambda>0}$ et $(U'^\lambda)_{\lambda>0}$ les deux résolvantes. Si $g \in B''_0 := B_0 \cap B'_0$, l'unique solution de l'équation $(\lambda I - A)f = g$ est $U^\lambda g$ et aussi $U'^\lambda g$, de sorte que $U^\lambda g = U'^\lambda g$. Cela est vrai pour tout $\lambda > 0$, et en vertu de (3.3.5) on en déduit que les transformées de Laplace des deux fonctions $t \mapsto P_t g(x)$ et $t \mapsto P'_t g(x)$ coïncident, pour tout $x \in E$ fixé. Ces deux fonctions sont continues à droite (par le théorème 3.3.1(a)):

il est alors bien connu que ces deux fonctions sont identiques, ce qui prouve que $P_t g = P'_t g$ pour tous $t \geq 0$ si $g \in B''_0$.

Si g est dans la fermeture $\overline{B''_0}$ de B''_0 pour la convergence simple–bornée, il existe $g_n \in B''_0$ avec $g_n \stackrel{\text{s−b}}{\longrightarrow} g$. D'après $P_t g_n = P'_t g_n$ et le théorème de Lebesgue on voit que $P_t g = P'_t g$. Enfin le théorème 3.3.1(b) et le fait que $\mathcal{D}_A \subset B''_0$ montrent que B_0 et B'_0 sont contenus dans $\overline{B''_0}$. Si alors $g \in B_0$ on a $P'_t g = P_t g \to 0$ quand $t \to 0$, donc $g \in B'_0$, et de même toute $g \in B'_0$ est dans B_0: cela achève la preuve. □

3.3.2 Générateur pour un processus de Markov continu à droite

Le corollaire (3.3.3 montre que le générateur "caractérise" d'une certaine manière le semi–groupe, mais il n'est pas entièrement satisfaisant, puisqu'il ne dit rien de ce qui se passe pour les fonctions qui ne sont pas dans la fermeture de B_0. On peut faire bien mieux dans certains cas.

Dans ce paragraphe on suppose que E est un espace polonais, et que (P_t) est le semi–groupe d'un processus de Markov à trajectoires continues à droite. Le générateur du semi–groupe est alors aussi appelé, par abus de langage, le générateur du processus de Markov. L'hypothèse de continuité à droite implique que si $f \in C$ (continue bornée), alors $P_t f(x) = I\!\!E_x(f(X_t)) \to f(x)$: donc $C \subset B_0$. Le corollaire précédent admet alors la version améliorée suivante, qui est fondamentale pour les applications:

Théorème 3.3.4. *Soit (P_t) et (P'_t) deux semi–groupes de probabilités de transition sur l'espace polonais E, ademttant le même générateur infinitésimal. Si de plus (P_t) est associé à un processus de Markov à trajectoires continues à droite, alors on a $P'_t = P_t$ pour tout $t \geq 0$.*

Preuve. On a vu que $P'_t f(x) = P_t f(x)$ pour toute $f \in C$. Il suffit alors d'appliquer la propriété suivante: si μ et ν sont deux probabilités sur un espace polonais E (en l'occurence $\mu = P_t(x,.)$ et $\nu = P'_t(x,.)$), et si les intégrales des fonctions de C sont les mêmes relativement à μ et à ν, on a alors $\mu = \nu$. □

Toujours sous les mêmes hypothèses concernant E et (P_t), nous allons revenir sur la définition de la résolvante. La fonction $t \mapsto P_t f(x)$ est continue à droite, donc borélienne, dès que que $f \in C$. Par un argument classe monotone (par exemple) elle est donc aussi borélienne et bornée (resp. positive) pour f borélienne et bornée (resp. positive): donc l'intégrale (3.3.5) est bien définie dans ces cas.

On peut voir les choses un peu autrement: $Q((x,t),dy) = P_t(x,dy)$ est une probabilité de transition de $(E \times I\!\!R_+, \mathcal{E} \otimes \mathcal{R}_+)$ dans $(E, \mathcal{E})$ (la fonction $(x,t) \mapsto Qf(x,t) = P_t f(x)$ est mesurable en x et continue à droite en t, donc borélienne en (x,t), si $f \in C$, et cette mesurabilité est préservée si on prend $f \in B$). Par ailleurs la mesure de densité $\lambda e^{-\lambda t}$ sur $I\!\!R_+$ est une probabilité (la loi exponentielle de paramètre λ). Donc une modification triviale du théorème 1.2.5 nous dit que $\int_0^\infty \lambda\ e^{-\lambda t}\ Q((x,t),.)\ dt$ est une probabilité de transition de $(E, \mathcal{E})$ dans lui-même.

En d'autres termes, la formule (3.3.5) définit *une mesure de transition* U^λ de $(E, \mathcal{E})$ dans lui-même, de masse $1/\lambda$.

Proposition 3.3.5. (L'équation résolvante) *Sous les hypothèses de ce paragraphe, les mesures de transition U^λ vérifient pour tous $\lambda, \mu > 0$ l'équation suivante:*

$$U^\mu - U^\lambda = (\lambda - \mu) U^\lambda U^\mu = (\lambda - \mu) U^\mu U^\lambda. \tag{3.3.6}$$

Preuve. Si $f \in B$ est positive, le théorème de Fubini et l'équation des semi–groupes impliquent si $\lambda \neq \mu$:

$$\begin{aligned} U^\lambda U^\mu f(x) &= \int_0^\infty e^{-\lambda t}\, dt \int_0^\infty e^{-\mu s}\ P_{t+s} f(x)\ ds \\ &= \int_0^\infty e^{-\lambda t + \mu t}\ dt \int_t^\infty e^{-\mu r}\ P_r f(x)\ dr \\ &= \int_0^\infty e^{-\mu r}\ P_r f(x)\ dr \int_0^r e^{-\lambda t + \mu t}\ dt \\ &= \frac{1}{\lambda - \mu} \int_0^\infty \left(e^{-\mu r} - e^{-\lambda r} \right)\ P_r f(x) dr \ = \ \frac{1}{\lambda - \mu} \left(U^\mu f(x) - U^\lambda f(x) \right). \end{aligned}$$

On a donc la première égalité (3.3.6) quand $\lambda \neq \mu$, tandis que $U^\mu U^\lambda = U^\lambda U^\mu$ est évident. Enfin, (3.3.6) est évidente si $\lambda = \mu$. □

Les mesures U^λ décroissent quand λ croît, au sens où $U^\mu - U^\lambda$ est une mesure de transition positive si $\mu < \lambda$: cela découle de (3.3.5) ou aussi de (3.3.6). Par suite si $\lambda \downarrow 0$, les mesures U^λ croissent vers une limite $U := U^0$, qui est encore une mesure de transition, et qu'on appelle le *potentiel* du semi–groupe, ou du processus de Markov. Noter que $U(x, .)$ est une mesure infinie, et qu'elle n'est pas nécessairement σ–finie. Pour toute fonction positive f, on a

$$Uf(x) = \int_0^\infty P_t f(x)\ dt. \tag{3.3.7}$$

3.3.3 Générateur et martingales

Le générateur infinitésimal et la résolvante sont naturellement associés à certaines martingales ou surmartingales faisant intervenir le processus de Markov $\mathbf{X}$ associé au semi–groupe. On se place sous les mêmes hypothèses que dans le paragraphe précédent: E est un espace polonais, et le processus de Markov $\mathbf{X}$ est à trajectoires continues à droites.

Soit d'abord $(A, \mathcal{D}_A)$ le générateur du semi–groupe. Si $g \in \mathcal{D}_A$ on définit le processus

$$M_t^g = g(X_t) - g(X_0) - \int_0^t Ag(X_s) ds. \tag{3.3.8}$$

Remarquons que $Ag \in B$, donc $(\omega, s) \mapsto Ag(X_s(\omega))$ est $\mathcal{F} \otimes \mathcal{R}_+$–mesurable et borné, donc l'intégrale ci–dessus existe st définit (par Fubini) une variable aléatoire.

Théorème 3.3.6. *Si $g \in \mathcal{D}_A$, le processus M^g défini ci–dessus est une martingale sur $(\Omega, \mathcal{F}, (\mathcal{F}_t), I\!\!P_\mu)$ pour toute probabilité μ sur E.*

Preuve. Non seulement on a la $\mathcal{F}$–mesurabilité de l'intégrale $\int_0^t Ag(X_s)ds$, mais cette intégrale est $\mathcal{F}_t$–mesurable: en effet, si t est fixé, la fonction $(\omega, s) \mapsto X_s(\omega)$ sur $\Omega \times [0,t]$ est mesurable par rapport au produit de $\mathcal{F}_t$ par la tribu borélienne de $[0,t]$. Par suite la variable M_t^g est $\mathcal{F}_t$–mesurable, et elle est aussi bornée (donc $I\!P_\mu$–intégrable) pour tout t.

Si Y est une variable $\mathcal{F}_t$–mesurable bornée, on a alors (en utilisant Fubini, puis la propriété de Markov):

$$\begin{aligned}
I\!E_\mu(Y(M_{t+s}^g - M_t^g)) &= I\!E_\mu\left(Y\left(g(X_{t+s}) - g(X_t) - \int_t^{t+s} Ag(X_r)\ dr\right)\right) \\
&= I\!E_\mu(Y(g(X_{t+s})) - I\!E_\mu(Yg(X_t)) - \int_t^{t+s} I\!E_\mu(Y\ Ag(X_r))\ dr \\
&= I\!E_\mu(Y\ P_s g(X_t)) - I\!E_\mu(Yg(X_t)) - \int_0^s I\!E_\mu(Y\ P_y Ag(X_t))\ du \\
&= I\!E_\mu\left(Y\left(P_s g(X_t) - g(X_t) - \int_0^s P_y Ag(X_t)\ du\right)\right),
\end{aligned}$$

et cette quantité est nulle à cause de (3.3.4): la propriété de martingale en découle. □

Théorème 3.3.7. *Si $g \in B_0$ est positive et si $\lambda > 0$, le processus $e^{-\lambda t}\ U^\lambda g(X_t)$ est une surmartingale sur $(\Omega, \mathcal{F}, (\mathcal{F}_t), I\!P_\mu)$ pour toute probabilité μ sur E.*

Preuve. Soit $Z_t = e^{-\lambda t}\ U^\lambda g(X_t)$. Chaque variable Z_t est $\mathcal{F}_t$–mesurable et bornée. Remarquer que

$$P_s U^\lambda g(x) = e^{\lambda s} \int_s^\infty e^{-\lambda r}\ P_r g(x)\ dr \leq e^{\lambda s}\ U^\lambda g(x)$$

puisque $g \geq 0$. Par suite

$$I\!E_\mu(Z_{t+s}|\mathcal{F}_t) = e^{-\lambda(t+s)}\ P_s U^\lambda g(X_t) \leq e^{-\lambda t}\ U^\lambda g(X_t) = Z_t,$$

et on a le résultat. □

Le générateur infinitésimal est assez difficile à utiliser, souvent, car la détermination de son domaine est délicate, voire impossible. En plus, sa définition même fait intervenir $P_t f$, donc une espérance, et les espérances sont difficiles à manipuler (et n'existent pas toujours...) En revanche, les martyingales permettent l'utilisation du calcul stochastique, et sont donc plus souples d'utilisation.

C'est pourquoi le contenu du théorème 3.3.6 suggère une autre définition du générateur.

Définition 3.3.8. On appellera *générateur étendu* du processus de Markov tout opérateur linéaire $(A', \mathcal{D}_{A'})$ sur l'espace des fonctions mesurables, tel que si $g \in \mathcal{D}_{A'}$ la fonction $A'g$ soit mesurable, et la formule

$$M_t^g = g(X_t) - g(X_0) - \int_0^t A'g(X_s)ds \tag{3.3.9}$$

ait un sens et définisse un processus qui est une *martingale locale* sur $(\Omega, \mathcal{F}, (\mathcal{F}_t), I\!P_\mu)$ pour toute probabilité μ sur E.

Comme nous n'utiliserons pas vraiment cette notion plus loin, nous ne rappellerons pas la notion de martingale locale. Mais bien-sûr toute martingale est une martingale locale, donc le "vrai" générateur $(A, \mathcal{D}_A)$ est un générateur étendu. Il existe bien d'autres générateurs étendus pour le même processus de Markov: on perd ainsi l'unicité, mais on gagne énormément en souplesse d'utilisation.

Signalons enfin que le théorème 3.3.7 est l'outil essentiel pour montrer les théorèmes de régularité du genre 3.1.8.

3.3.4 Générateur et stationnarité

Nous avons vu que la possibilité de rendre le processus de Markov stationnaire grâce à une loi initiale adéquate revient à chercher les probabilités μ invariantes pour le semi–groupe. Heuristiquement, en "dérivant" la relation d'invariance $\mu P_t = \mu$ en temps nous conduit à $\mu A P_t = 0$, donc (pour $t = 0$) à "$\mu A = 0$". On peut donner un sens précis à cela:

Proposition 3.3.9. *a) Si μ est une probabilité invariante pour le semi–groupe (P_t) de générateur $(A, \mathcal{D}_A)$, on a $\mu(Af) = 0$ pour tout $f \in \mathcal{D}_A$.*

b) Si μ est une probabilité vérifiant $\mu(Af) = 0$ pour tout $f \in \mathcal{D}_A$, alors $\mu P_t f = \mu(f)$ pour tout $f \in \overline{B_0}$ (fermeture de B_0 pour la convergence simple–bornéee) et $t \geq 0$.

c) Si μ est une probabilité vérifiant $\mu(Af) = 0$ pour tout $f \in \mathcal{D}_A$, et si E est un espace polonais et si $\mathbf{X}$ est à trajectoires continues à droite, alors μ est une probabilité invariante.

Preuve. Ci–dessous, μ est toujours une probabilité sur $(E, \mathcal{E})$, et si $f \in \mathcal{D}_A$ on note $g(t)$ le nombre $g(t) = \mu P_t f$. En utilisant (3.3.4) et le théorème de Lebesgue, on voit que g est continue, donc aussi dérivable, et sa dérivée est $g'(t) = \mu(AP_t f)$.

Si μ est invariante, on a bien-sûr $g(t) = g(0) = \mu(f)$ ci–dessus, donc $g'(t) = 0$, et en particulier pour $t = 0$ cela donne $\mu(Af) = 0$: on a donc (a).

Supposons inversement que $\mu(Af) = 0$ pour tout $f \in \mathcal{D}_A$. On sait qu'on a aussi $P_t f \in \mathcal{D}_A$, donc $\mu(AP_t f) = 0$, donc la fonction g est dérivable, de dérivée identiquement nulle, donc elle est constante: par suite $\mu P_t f = \mu(f)$ pour tout $f \in \mathcal{D}_A$. Comme $\mathcal{D}_A$ est dense dans $\overline{B_0}$, on obtient (b) en utilisant le théorème de Lebesgue.

Enfin, sous les hypothèses de (c) on a $\mu P_t f = \mu(f)$ en particulier pour toute fonction f continue bornée, et on sait que cela entraîne que les deux mesures μP_t et μ sont égales. $\square$

Ce résultat est malheureusement moins utile qu'il n'y paraît, car encore une fois il est rare qu'on connaisse complètement le domaine de A.

Chapitre 4

Processus de Markov de saut pur

4.1 Définitions et propriétés élémentaires

Dans ce chapitre on va considérer une classe particulière de processus de Markov, qui sont à valeurs dans un espace E fini ou dénombrable. Comme dans le chapitre 1, les points de E sont notés i, j, k. On a vu au chapitre précédent que les propriétés topologiques de E jouent un rôle important, et nous munissons ici E de la *topologie discrète*, pour laquelle tous les points sont isolés.

Définition 4.1.1. Un processus de Markov *de saut pur* est un processus de Markov $\mathbf{X} = (\Omega, \mathcal{F}, (\mathcal{F}_t), (\theta_t), (I\!\!P_i))$ au sens de la définition 3.1.2, à valeurs dans l'espace fini ou dénombrable E, et dont toutes les trajectoires $t \mapsto X_t(\omega)$ sont continues à droite.

Comme dans le chapitre 1, une mesure est un "vecteur ligne" $\mu = (\mu_i)$, et la probabilité de transition P_t de $\mathbf{X}$ est une matrice $P_t = (p(t)_{ij})_{i,j \in E}$. Bien entendu $p(0)_{ij} = \delta_{ij}$, et comme $\mathbf{X}$ est à trajectoires continues à droite on a

$$t \mapsto p(t)_{ij} \quad \text{est continue à droite} \tag{4.1.1}$$

(car les fonctions $1_{\{j\}}$, comme toutes les fonctions sur E, sont continues). On verra en fait bien mieux plus tard.

Proposition 4.1.2. *Tout processus de Marov de saut pur* $\mathbf{X}$ *vérifie la propriété de Markov forte.*

Preuve. L'espace E muni de la toplogie discrète est polonais, donc il suffit de vérifier les deux conditions du théorème 3.1.6: la première est satisfaite car toute fonction sur E est continue, et la seconde l'est par hypothèse. $\square$

On va maintenant étudier de plus près la structure de $\mathbf{X}$. D'abord, comme la toplogie de E est la topologie discrète, la continuité à droite entraine que les trajectoires sont "constantes par morceaux", au sens où pour tout $t \geq 0$ et tout ω on a $X_s(\omega) = X_t(\omega)$ pour tout s dans un intervalle $[t, t'(\omega, t)[$, où $t'(\omega, t)$ est un réel strictement plus grand que

t. En particulier, si on définit par récurrence les variables:

$$T_0 = 0, \qquad T_{n+1} = \inf(t:\ t > T_n, X_t \neq X_{T_n}) \tag{4.1.2}$$

(comme d'habitude l'inf d'un ensemble vide vaut $+\infty$), on obtient une suite croissante de temps d'arrêt vérifiant $T_{n+1} > T_n$ sur $\{T_n < \infty\}$. On pose aussi $T_\infty = \lim_n T_n$, qu'on appelle le *premier temps d'explosion*. Il se peut bien sûr que $T_\infty < \infty$, auquel cas X_t prend une certaine valeur i en $t = T_\infty$ et y reste un temps positif, puis "saute" de nouveau, etc... Si $\mathbb{P}_i(T_\infty = \infty) = 1$ pour tout $i \in E$, on dit que le processus *est sans explosion*.

En utilisant les translations (θ_t), on remarque aussi que

$$T_n - T_{n-1} = T_1 \circ \theta_{T_{n-1}} \qquad \text{sur l'ensemble } \{T_{n-1} < \infty\}. \tag{4.1.3}$$

Le fait que X_{T_n} n'est a priori pas défini si $T_n = \infty$ nous incite à considérer un point Δ extérieur" à E et à poser $E_\Delta = E \cup \{\Delta\}$, comme on l'a déjà fait plusieurs fois. On peut alors poser pour $n \geq 0$

$$\left.\begin{array}{lll} S_n = T_{n+1} - T_n, & \xi_n = X_{T_n} & \text{si} \quad T_n < \infty \\ S_n = \infty & \xi_n = \Delta & \text{si} \quad T_n = \infty. \end{array}\right\} \tag{4.1.4}$$

Le théorème de structure suivant est fondamental (rappelons que la loi exponentielle de paramètre $a \in]0, \infty[$ est la loi sur $\mathbb{R}_+$ de densité $x \mapsto ae^{-ax}$; par convention, si $a = 0$ c'est la masse de Dirac ε_∞ en l'infini. Une variable exponentielle de paramètre a admet l'espérance $1/a$, avec la convention $1/0 = \infty$):

Théorème 4.1.3. *Soit* $\mathbf{X}$ *un processus de Markov de saut pur.*

(1) Sous $\mathbb{P}_i$, les variables S_0 et ξ_1 sont indépendantes, et S_0 suit une loi exponentielle dont on note $a_i \in \mathbb{R}_+$ le paramètre; si on pose $\pi_{ij} = \mathbb{P}_i(\xi_1 = j)$ pour $j \in E_\Delta$ on a aussi

$$\pi_{ii} = 0, \qquad \left.\begin{array}{lll} a_i > 0 & \Longrightarrow & \sum_{j\in E} \pi_{ij} = 1, \quad \pi_{i\Delta} = 0 \\ a_i = 0 & \Longrightarrow & \sum_{j\in E} \pi_{ij} = 0, \quad \pi_{i\Delta} = 1 \end{array}\right\} \tag{4.1.5}$$

pour tout $i \in E$. On pose aussi par convention $a_\Delta = 0$, et $\pi_{\Delta\Delta} = 1$ et $\pi_{\Delta i} = 0$ pour $i \in E$ (ce qui est cohérent avec (1.2.10)).

(2) Sous chaque $\mathbb{P}_\mu$ la chaîne (ξ_n) à valeurs dans E_Δ a la propriété de Markov relativement à la filtration discrète $(\mathcal{F}_{T_n})_{n\geq 0}$, avec la transition $\Pi = (\pi_{ij})$.

(3) Sous chaque $\mathbb{P}_\mu$, conditionnellement à $\mathcal{F}_{T_n}$, les variables S_n et ξ_{n+1} sont indépendantes, la première est de loi exponentielle de paramètre a_{ξ_n} et la seconde de loi $(\pi_{\xi_n,j})_{j\in E_\Delta}$.

Preuve. Noter que (1) est un cas particulier de (3), mais nous allons commencer par prouver (1).

En observant que $S_0 = t + S_0 \circ \theta_t$ et $\xi_1 = \xi_1 \circ \theta_t$ et $X_t = X_0$ sur l'ensemble $\{S_0 > t\}$, on obtient pour tous $s, t \geq 0$, $i \in E$ et $A \subset E_\Delta$, par la propriété de Markov en t:

$$\begin{aligned} \mathbb{P}_i(S_0 > t + s, \xi_1 \in A) &= \mathbb{P}_i(S_0 > t, S_0 \circ \theta_t > s, \xi_1 \circ \theta_t \in A) \\ &= \mathbb{E}_i(1_{\{S_0 > t\}}\ \mathbb{P}_{X_t}(S_0 > s, \xi_1 \in A)) \\ &= \mathbb{P}_i(S_0 > t)\ \mathbb{P}_i(S_0 > s, \xi_1 \in A). \end{aligned}$$

En prenant $A = E_\Delta$, cela donne $I\!P_i(S_0 > t+s) = I\!P_i(S_0 > t) I\!P_i(S_0 > s)$, relation caractéristique des lois exponentielles de paramètre dans $[0,\infty]$; comme $S_0 > 0$, le paramètre a_i de la loi de S_0 sous $I\!P_i$ est dans $I\!R_+$. En prenant $s = 0$ ci–dessus, on obtient aussi $I\!P_i(S_0 > t, \xi_1 \in A) = I\!P_i(S_0 > t) I\!P_i(\xi_1 \in A)$: cela entraine l'indépendance de S_0 et ξ_1 sous $I\!P_i$. Enfin si $a_i > 0$ on a $S_0 < \infty$ $I\!P_i$–p.s., donc $\xi_1 = X_{S_0} \in E$ $I\!P_i$–p.s., tandis que si $a_i = 0$ on a $I\!P_i(S_0 = \infty) = 1$, donc $I\!P_i(\xi_1 = \Delta) = 1$ et on en déduit (4.1.5).

Sur l'ensemble $\{T_n < \infty\}$ on a $S_n = S_0 \circ \theta_{T_n}$ et $\xi_{n+1} = \xi_1 \circ \theta_{T_n}$, de sorte que d'après la propriété forte de Markov au temps T_n on obtient pour tout borélien B de $I\!R_+ \times E_\Delta$:

$$I\!P_\mu\left((S_n, \xi_{n+1}) \in B | \mathcal{F}_{T_n}\right) = I\!P_{X_{T_n}}\left((S_0, \xi_1) \in B\right).$$

Donc conditionnellement à $\mathcal{F}_{T_n}$ et sur l'ensemble $\mathcal{F}_{T_n}$–mesurable $\{T_n < \infty\}$, on a (3). Sur l'ensemble $\{T_n = \infty\}$ on a $S_n = \infty$ et $\xi_n = \xi_{n+1} = \Delta$, donc comme $a_\Delta = 0$ et $\pi_{\Delta\Delta} = 1$ on a aussi (3) sur cet ensemble. En particulier pour $A \subset E_\Delta$ il vient $I\!P_\mu(\xi_{n+1} \in A | \mathcal{F}_{T_n}) = \sum_{j\in A} \pi_{\xi_n, j}$, et on a donc (2) également. $\square$

Corollaire 4.1.4. *Soit* $\mathbf{X}$ *un processus de Markov de saut pur. Conditionnellement à la tribu* $\mathcal{G} = \sigma(\xi_n : n \geq 0)$*, et sous chaque probabilité* $I\!P_\mu$*, les variables* $(S_n)_{n\geq 0}$ *sont indépendantes, chaque* S_n *étant de loi exponentielle de paramètre* a_{ξ_n}*.*

Preuve. Notons η_i la loi exponentielle de paramètre a_i (rappelons que $a_\Delta = 0$). Il suffit de montrer que pour tout entier N, tous boréliens A_n de $[0,\infty]$ et tous $i_n \in E_\Delta$, on a

$$I\!P_\mu\left((\cap_{n=0}^N \{S_n \in A_n\}) \cap (\cap_{n=0}^{N+1} \{\xi_n = i_n\} \right) = I\!E_\mu \left(\prod_{n=0}^{N+1} 1_{\{i_n\}}(\xi_n) \prod_{n=0}^{N} \eta_{i_n}(A_n) \right). \quad (4.1.6)$$

Comme $I\!P_\mu(S_n \in A_n, \xi_{n+1} = i_{n+1} | \mathcal{F}_{T_n}) = \eta_{\xi_n}(A_n)\ \pi_{\xi_n, i_{n+1}}$ par le théorème précédent, on vérifie alors immédiatement (4.1.6) par récurrence descendante sur N. $\square$

Le processus "minimal": Dans la suite, on va essayer de reconstruire le processus de Markov à partir des suites (ξ_n) et (S_n). En effet, comme $X_t = X_{T_n}$ sur chaque intervalle $[T_n, T_{n+1}[$, on voit d'après (4.1.4) que

$$X_t = \xi_n \qquad \text{si} \quad T_n = S_0 + \ldots + S_{n-1} \leq t < S_0 + \ldots + S_n = T_{n+1}, \qquad (4.1.7)$$

Ainsi, les deux suites (ξ_n) et (S_n) caractérisent entièrement X_t sur l'ensemble $\{t < T_\infty\}$. On est donc naturellement conduit à introduire un nouveau processus $(\widetilde{X}_t)$ à valeurs dans E_Δ, en complétant (4.1.7):

$$\widetilde{X}_t = \begin{cases} \xi_n & \text{si } \ S_0 + \ldots + S_{n-1} \leq t < S_0 + \ldots + S_n \\ \Delta & \text{si } \ t \geq S_0 + \ldots + S_n + \ldots. \end{cases} \qquad (4.1.8)$$

De manière équivalente, on a aussi:

$$\widetilde{X}_t = \begin{cases} X_t & \text{si } \ t < T_\infty \\ \Delta & \text{si } \ t \geq T_\infty. \end{cases} \qquad (4.1.9)$$

Proposition 4.1.5. *Soit* $\mathbf{X}$ *un processus de Markov de saut pur, et* $\widetilde{X}_t$ *défini ci–dessus. Pour tout* $i \in E$ *le processus* $(\widetilde{X}_t)$ *sur* $(\Omega, \mathcal{F}, (\mathcal{F}_t), I\!\!P_i)$ *vérifie la* $(\mathcal{F}_t)$*–propriété de Markov (définition 3.1.1) avec le semi–groupe de transition* $(\widetilde{P}_t = (\widetilde{p}(t)_{ij}))$ *défini par*

$$\widetilde{p}(t)_{ij} = \begin{cases} I\!\!P_i(X_t = j, t < T_\infty) & si \;\; i,j \in E \\ I\!\!P_i(t \geq T_\infty) & si \;\; i \in E, \;\; j = \Delta \\ 1 & si \;\; i,j = \Delta \\ 0 & si \;\; i = \Delta, \;\; j \in E. \end{cases} \tag{4.1.10}$$

De plus, les variables ξ_n *et* S_n *associées à* $\widetilde{X}$ *sont les mêmes que celles associées à* X.

Le processus $(\widetilde{X}_t)$ s'appelle le *processus minimal* associé à X.

Preuve. La dernière assertion est triviale. Pour la propriété de semi–groupe de $(\widetilde{P}_t)$ et la propriété de Markov de $\widetilde{X}$, il nous suffit de montrer que si $k \in E$ et $s, t \geq 0$ on a:

$$I\!\!P_k(\widetilde{X}_{s+t} = j | \mathcal{F}_s) = \widetilde{p}(t)_{\widetilde{X}_s, j} \,. \tag{4.1.11}$$

Sur l'ensemble $\mathcal{F}_s$–mesurable $\{s \geq T_\infty\}$ on a $\widetilde{X}_s = \widetilde{X}_{s+t} = \Delta$, de sorte que (4.1.11) est évident. Sur le complémentaire $\{s < T_\infty\}$ on a $\widetilde{X}_{s+t} = \widetilde{X}_t \circ \theta_s$, et aussi $\widetilde{X}_s = X_s$, donc la proposition 3.1.3 entraine que le membre de gauche de (4.1.11) égale $I\!\!P_{\widetilde{X}_s}(\widetilde{X}_t = j)$, qui vaut bien $\widetilde{p}(t)_{\widetilde{X}_s, j}$ lorsque $j \in E$ et aussi lorsque $j = \Delta$. □

4.2 Construction d'un processus de saut pur

Pour construire un processus de Markov de saut pur on peut appliquer le théorème 3.1.5, mais cela nécessite d'une part de connaitre le semi–groupe, d'autre part de vérifier que le processus de Markov ainsi construit est effectivement à trajectoires continues à droite, ce qui n'est pas évident du tout. On peut aussi utiliser le théorème 4.1.3 et son corollaire, qui d'une certaine manière fournissent de façon explicite la "dynamique" du processus, ou au moins celle du processus minimal associé. C'est ce que nous allons faire maintenant.

On part d'une famille $(a_i)_{i \in E_\Delta}$ de réels positifs ou nuls, avec $a_\Delta = 0$, et d'une probabilité de transition $\Pi = (\pi_{ij})_{i,j \in E_\Delta}$ vérifiant (4.1.5) (et donc $\pi_{\Delta\Delta} = 1$).

Intuitivement, la démarche consiste à construire une chaîne de Markov (ξ_n) de transition Π, et une suite de variables (S_n) dont la loi, conditionnellement à $(\xi_n)_{n \geq 0}$, est celle décrite au corollaire 4.1.4, puis à définir $\widetilde{X}$ par (4.1.8). Formellement, les choses sont évidemment un peu plus compliquées !

On considère d'abord la chaîne de Markov "canonique" $\Xi = (\Omega', \mathcal{F}, (\xi_n), \theta', (I\!\!P'_i)_{i \in E_\Delta})$ de transition Π (cf. la fin du paragraphe 1.3). Il est évident que $I\!\!P'_i$–p.s. on a $\xi_{n+m} = \Delta$ pour tout $m \geq 0$ si $\xi_n = \Delta$, et aussi que $I\!\!P'_\Delta(\xi_0 = \Delta) = 1$. Ensuite on pose $\Omega'' =]0, \infty]^{I\!\!N}$, qu'on munit des variables $S_n(s_0, s_1, \ldots) = s_n$ et de la tribu produit $\mathcal{F}'' = \sigma(S_n : n \geq 0)$. Puis, pour toute suite $\omega' = (x_0, x_1, \ldots)$ dans E_Δ on note $I\!\!P''(\omega', d\omega'')$ l'unique probabilité sur $(\Omega'', \mathcal{F}'')$ sous laquelle les (S_n) sont indépendantes, chaque S_n étant de loi exponentielle de paramètre a_{x_n}. Par un argument de classe monotone on vérifie immédiatement que $I\!P''$ est une probabilité de transition de $(\Omega', \mathcal{F}')$ dans $(\Omega'', \mathcal{F}'')$. D'après le théorème 1.2.4, pour tout $i \in E_\Delta$ on peut munir l'espace produit

$$\Omega = \Omega' \times \Omega'', \qquad \mathcal{F}''' = \mathcal{F}' \otimes \mathcal{F}''$$

d'une unique probabilité $I\!P_i$ telle que $I\!P_i(A \times B) = \int I\!P'_i(d\omega') I\!P''(\omega', B) 1_A(\omega')$ pour tous $A \in \mathcal{F}'$ et $B \in \mathcal{F}''$, et sous $I\!P_i$ on a les deux propriétés suivantes:

1. la suite (ξ_n) a la propriété de Markov avec la transition Π;
2. conditionnellement lorsque la suite (ξ_n) est fixée, les variables $(S_n)_{n\geq 0}$ sont indépendantes, et S_n est de loi exponentielle de paramètre a_{ξ_n};

(ci–dessus, ξ_n ou S_n, définis respectivement sur Ω' et Ω'', sont considérés également comme des fonctions sur Ω).

Il nous reste alors à définir le processus $(\widetilde{X}_t)$ sur Ω par la formule (4.1.8), la filtration $\mathcal{F}_t = \sigma(\widetilde{X}_s : s \leq t)$, la tribu $\mathcal{F} = \bigvee_t \mathcal{F}_t$, et le semi–groupe des translations (θ_t) comme étant les uniques applications de Ω dans lui-même vérifiant identiquement (3.1.3) (avec $\widetilde{X}$ au lieu de X, bien–sûr; l'existence des θ_t est très facile à montrer). On a à l'évidence la troisième propriété:

3. si $\widetilde{X}_t(\omega) = \Delta$ pour un t, on a aussi $\widetilde{X}_s(\omega) = \Delta$ pour tout $s \geq t$.

Théorème 4.2.1. *Le terme* $\widetilde{\mathbf{X}} = (\Omega, \mathcal{F}, (\mathcal{F}_t), (\theta_t), (\widetilde{X}_t), (I\!P_i)_{i \in E_\Delta})$ *est un processus de Markov de saut pur, qui est un processus minimal, et qui est associé aux données* $\Pi = (\pi_{ij})$ *et* (a_i) *par le théorème 4.1.3.*

Preuve. La seule chose non évidente est la propriété de Markov (3.1.4) du terme $\widetilde{\mathbf{X}}$. Pour cela, il suffit de vérifier (3.1.5) lorsque $\mu = \varepsilon_i$ pour $i \in E_\Delta$: en effet, en posant $P_t(i, A) = I\!P_i(\widetilde{X}_t \in A)$ on aura (3.1.4), et le fait que (P_t) soit un semi-groupe découle immédiatement d'une itération de la propriété (3.1.5).

On pose $T_0 = 0$ et $T_n = S_0 + \ldots + S_{n-1}$ pour $n \geq 1$, et $T_\infty = \lim_n T_n$. Remarquons que toute variable $\mathcal{F}$–mesurable Y vérifie $Y(\omega) = \alpha$ pour une constante α sur l'ensemble $\{\omega : \widetilde{X}_0(\omega) = \Delta\}$, en vertu de la propriété 3 ci–dessus. Donc sur l'ensemble $\mathcal{F}_t$–mesurable $\{T_\infty \leq t\}$ on a $Y \circ \theta_t = \alpha$, et aussi $\widetilde{X}_t = \Delta$, de sorte que (3.1.5) est satisfaite sur cet ensemble.

Il nous reste à montrer la propriété suivante: considérons les ensembles $\mathcal{F}_t$–mesurables $A(n,t) = \{T_n \leq t < T_{n+1}\}$; si alors Z est une variable $\mathcal{F}_t$–mesurable bornée et Y est $\mathcal{F}$–mesurable et bornée, alors

$$I\!E_i(Z\, Y \circ \theta_t\, 1_{A(n,t)}) = I\!E_i\left(Z\, 1_{A(n,t)}\, I\!E_{\widetilde{X}_t}(Y)\right), \tag{4.2.1}$$

et, par un argument de classe monotone, il suffit même de montrer ceci lorsque $Y = f(S_0, \ldots, S_m; \xi_0, \ldots, \xi_m)$ pour une fonction mesurable bornée f. Le résultat est intuitivement "évident" par application de la propriété de non–vieillissement de la loi exponentielle, mais formellement c'est un peu compliqué...

En restriction à l'ensemble $A(n,t)$ chaque variable $\widetilde{X}_s$ pour $s \leq t$ est une fonction de $(S_0, \ldots, S_{n-1}; \xi_0, \ldots, \xi_n)$, donc il est est de même de Z, de sorte qu'on peut écrire $Z\ 1_{A(n,t)} = g(S_0, \ldots, S_{n-1}; \xi_0, \ldots, \xi_n)\ 1_{A(n,t)}$ pour une fonction mesurable bornée g.

En utilisant les propriétés 1 et 2 ci–dessus, il est facile de voir que

$$\mathbb{E}_{j_0}(Y) = \sum_{j_1,\ldots,j_m \in E_\Delta} \left(\prod_{k=0}^{m-1} \pi_{j_k,j_{k+1}} \right)$$
$$\int_{]0,\infty]^{m+1}} \left(\prod_{k=0}^{m} a_{j_k} e^{-a_{j_k} r_k} \right) f(r_0, \ldots, r_m; j_0, \ldots, j_m) dr_0 \ldots dr_m,$$

avec la convention que la mesure $a_i e^{-a_i r} dr$ est ε_∞ si $a_i = 0$. Il s'ensuit que le membre de droite de (4.2.1) vaut, avec la convention $i_0 = i$:

$$\sum_{i_1,\ldots,i_n \in E} \left(\prod_{k=0}^{n-1} \pi_{i_k,i_{k+1}} \right) \int_{]0,\infty[^{n+1}} \left(\prod_{k=0}^{n} a_{i_k} e^{-a_{i_k} s_k} \right) g(s_0, \ldots, s_{n-1}; i_0, \ldots, i_n)$$
$$1_{\{s_0+\ldots+s_{n-1} \leq t < s_0+\ldots+s_n\}} ds_0 \ldots ds_n \sum_{i_{n+1},\ldots,i_{n+m} \in E_\Delta} \left(\prod_{l=n}^{n+m-1} \pi_{i_l,i_{l+1}} \right)$$
$$\int_{]0,\infty]^{m+1}} \left(\prod_{l=n}^{n+m} a_{i_l} e^{-a_{i_l} r_l} \right) f(r_n, \ldots, r_{n+m}; i_n, \ldots, i_{n+m}) dr_n \ldots dr_{n+m}.$$

Mais l'intégrale de $a_{i_n} e^{-a_{i_n} s_n}$ par rapport à ds_n, sur l'ensemble $\{s_0 + \ldots + s_{n-1} \leq t < s_0 + \ldots + s_n\}$, vaut $\exp(-a_{i_n}(t - (s_0 + \ldots + s_{n-1}))\ 1_{\{t \geq s_0+\ldots+s_{n-1}\}}$, de sorte que l'expression précédente égale

$$\sum_{i_1,\ldots,i_n \in E;\ i_{n+1},\ldots,i_{n+m} \in E_\Delta} \left(\prod_{k=0}^{n+m-1} \pi_{i_k,i_{k+1}} \right) \int_{]0,\infty[^n} \left(\prod_{k=0}^{n-1} a_{i_k} e^{-a_{i_k} s_k} \right)$$
$$g(s_0, \ldots, s_{n-1}; i_0, \ldots, i_n)\ e^{-a_{i_n}(t-(s_0+\ldots+s_{n-1})}\ 1_{\{s_0+\ldots+s_{n-1} \leq t\}} ds_0 \ldots ds_{n-1}$$
$$\int_{]0,\infty]^{m+1}} \left(\prod_{l=n}^{n+m} a_{i_l} e^{-a_{i_l} r_l} \right) f(r_n, \ldots, r_{n+m}; i_n, \ldots, i_{n+m}) dr_n \ldots dr_{n+m}.$$

Dans la dernière intégrale on fait les changement de variables $r_k = s_k$ si $k \geq n+1$ et $r_n = s_0 + \ldots + s_n - t$ (pour $s_0, \ldots, s_{n-1}$ fixés), de jacobien 1, et l'expression précédente devient

$$\sum_{i_1,\ldots,i_n \in E;\ i_{n+1},\ldots,i_{n+m} \in E_\Delta} \left(\prod_{k=0}^{n+m-1} \pi_{i_k,i_{k+1}} \right) \int_{]0,\infty]^{n+m+1}} \left(\prod_{k=0}^{n+m} a_{i_k} e^{-a_{i_k} s_k} \right)$$
$$g(s_0, \ldots, s_{n-1}; i_0, \ldots, i_n) 1_{\{s_0+\ldots+s_{n-1} \leq t < s_0+\ldots+s_n\}}$$
$$f(s_0 + \ldots + s_n - t, s_{n+1}, \ldots, s_{n+m}; i_n, \ldots, i_{n+m}) ds_0 \ldots ds_{n+m}.$$

Comme sur l'ensemble $A(n,t)$ on a $(S_k, \xi_k) \circ \theta_t = (S_{k+n}, \xi_{n+k})$ pour $k \geq 1$ et $(S_0, \xi_0) \circ \theta_t = (S_0 + \ldots + S_n - t, \xi_n)$, on voit que cette expression égale le premier membre de (4.2.1), et la preuve est terminée. $\square$

Condition de "non–explosion": Comme il est naturel, la donnée de Π et des a_i permet de reconstruire le processus jusqu'au temps T_∞ (temps d'explosion), tandis qu'après cet instant on le pose égal à Δ de manière arbitraire. Si on veut un processus à valeurs dans E lui-même, il est nécessaire (et possible, de diverses manières) de "prolonger" le processus $\widetilde{X}$ au delà de T_∞, ce que nous ne ferons pas ici.

Un cas important est celui où il n'y a pas d'explosion, ce qui veut dire que $\mathbb{P}_i(T_\infty = \infty) = 1$ pour tout $i \in E$: en effet dans ce cas le processus de Markov $\mathbf{X}$ est son propre processus minimal, et le processus $\widetilde{X}$ ci–dessus ne prend $\mathbb{P}_i$–p.s. que des valeurs dans E, lorsque $i \in E$. La propriété $\mathbb{P}_i(T_\infty = \infty) = 1$ ne dépend que de Π et des a_i, et il est utile de disposer de conditions impliquant cette propriété.

A cet effet, pour tout $\lambda > 0$ on note $Q(\lambda) = (q(\lambda)_{ij})_{i,j\in E}$ la matrice donnée par (avec la convention $e^{-\infty} = 0$):

$$q(\lambda)_{ij} = \mathbb{E}_i\left(e^{-\lambda S_0}\ 1_{\{j\}}(\xi_1)\right) = \begin{cases} \frac{a_i}{a_i+\lambda}\ \pi_{ij} & \text{si } \ a_i > 0 \\ 0 & \text{si } \ a_i = 0 \end{cases} \tag{4.2.2}$$

La seconde égalité ci–dessus provient du théorème 4.1.3-(1) et du fait que la transformée de Laplace d'une variable exponentielle S de paramètre $a > 0$ est $\mathbb{E}(e^{-\lambda S}) = \frac{a}{a+\lambda}$.

Lemme 4.2.2. *Les éléments de la puissance* nième $Q(\lambda)^n$ *de la matrice* $Q(\lambda)$ *sont donnés par* $q(\lambda)^{(n)}_{ij} = \mathbb{E}_i\left(e^{-\lambda T_n}\ 1_{\{j\}}(\xi_n)\right)$.

Preuve. Comme $S_0 = T_1$ c'est évident pour $n = 1$. Supposons la propriété vraie pour n. On a $T_{n+1} = T_n + S_0 \circ \theta_{T_n}$ et $\xi_{n+1} = \xi_1 \circ \theta_{T_n}$ et $X_{T_n} = \xi_n$ si $T_n < \infty$, donc d'après la propriété de Markov forte et la convention $e^{-\infty} = 0$:

$$\begin{aligned} \mathbb{E}_i\left(e^{-\lambda T_{n+1}}\ 1_{\{j\}}(\xi_{n+1})\right) &= \mathbb{E}_i\left(e^{-\lambda T_n}\ 1_{\{T_n<\infty\}}\ \left(e^{-\lambda S_0}\ 1_{\{j\}}(\xi_1)\right) \circ \theta_{T_n}\right) \\ &= \mathbb{E}_i\left(e^{-\lambda T_n}\ 1_{\{T_n<\infty\}}\ q(\lambda)_{X_{T_n},j}\right) \\ &= \sum_{k\in E} \mathbb{E}_i\left(e^{-\lambda T_n}\ 1_{\{k\}}(X_{T_n})\right)\ q(\lambda)_{kj} \\ &= \sum_{k\in E} q(\lambda)^{(n)}_{ik}\ q(\lambda)_{kj}\ =\ q(\lambda)^{(n+1)}_{ij}. \end{aligned}$$

On obtient donc le résultat par récurrence sur n. □

Théorème 4.2.3. *Soit* $\mathbf{X}$ *un processus de Markov de saut pur, et* $\lambda > 0$ *arbitraire.*

(1) Il y a équivalence entre $\mathbb{P}_i(T_\infty = \infty) = 1$ *et* $\lim_{n\to\infty} \sum_{j\in E} q(\lambda)^{(n)}_{ij} = 0$.

(2) Il y a équivalence entre la propriété $\mathbb{P}_i(T_\infty = \infty) = 1$ *pour tout* $i \in E$ *(i.e.,* $\mathbf{X}$ *est sans explosion), et le fait que* $Q(\lambda)f = f$ *n'admette que* $f \equiv 0$ *comme solution positive bornée.*

Preuve. D'après le lemme on a $Q(\lambda)^n 1(i) = \sum_{j\in E} q(\lambda)^{(n)}_{ij} = \mathbb{E}_i(e^{-\lambda T_n})$, qui décroit vers $g(i) = \mathbb{E}_i(e^{-\lambda T_\infty})$, d'où (1).

Soit f positive bornée, avec $Q(\lambda)f = f$, et $b = \sup_i f(i)$. D'une part, $f = Q(\lambda)^n f \leq bQ(\lambda)^n 1$ pour tout n, donc $f \leq bg$, de sorte que l'implication $\Rightarrow$ de (2) découle de (1). On a aussi $0 \leq g \leq 1$, et $Q(\lambda)g = \lim_n Q(\lambda)^{n+1} g = g$, d'où l'implication $\Leftarrow$ de (2). □

Corollaire 4.2.4. *Si* $\sup_{i\in E} a_i < \infty$*, le processus* $\mathbf{X}$ *est sans explosion.*

Preuve. Soit $a = \sup_i a_i$. On a $\frac{a_i}{a_i+\lambda} \le \frac{a}{a+\lambda}$, donc $Q(\lambda)1 \le \frac{a}{a+\lambda}\,\Pi 1 = \frac{a}{a+\lambda}$ et par suite $Q(\lambda)^n 1 \le \left(\frac{a}{a+\lambda}\right)^n \to 0$. □

Terminons par un critère de nature assez différente. On dit qu'un point i est *absorbant* si $a_i = 0$ (en effet, dans ce cas on a $I\!\!P_i(X_t = i \;\; \forall t) = 1$).

Proposition 4.2.5. *Si tous les points de E sont soit absorbants, soit récurrents pour la chaîne de Markov* (ξ_n) *de transition* Π*, le processus* $\mathbf{X}$ *est sans explosion.*

Preuve. Soit $i \in E$. Si i est absorbant on a $I\!\!P_i(S_0 = \infty) = 1$, donc *a fortiori* $I\!\!P_i(T_\infty = \infty) = 1$. Supposons donc i non–absorbant, mais récurrent pour la chaîne (ξ_n), et notons N_k le $k$ième passage de la chaîne (ξ_n) en i, i.e. $N_1 = \inf(l \ge 1 : \xi_l = i)$ et $N_{k+1} = \inf(l \ge N_k + 1 : \xi_l = i)$. L'hypothèse implique $I\!\!P_i(N_k < \infty) = 1$ pour tout k, et le corollaire 4.1.4 implique que les S_{N_k} sont des variables indépendantes de loi exponentielle de paramètre $a_i > 0$, sous $I\!\!P_i$. On a bien–sûr $T_\infty \ge \sum_{k\ge 1} S_{N_k}$, et cette somme est $I\!\!P_i$–p.s. infinie d'après la loi des grands nombres (par exemple): d'où le résultat. □

4.3 Les équations de Kolmogorov

On vient de voir qu'on peut reconstruire un processus de Markov de saut pur à partir des quantités Π et a_i, au moins lorsqu'il n'y a pas d'explosion. On doit donc pouvoir aussi calculer le semi–groupe P_t à partir des mêmes données, ce qui est l'objet de ce paragraphe. Ci-dessous on part donc d'un processus $\mathbf{X}$ auquel sont associés $\Pi = (\pi_{ij})$ et les $a_i \ge 0$.

Rappelons que $p(t)_{ij} = I\!\!P_i(X_t = j)$. Sous $I\!\!P_i$, il n'y a eu aucun saut entre 0 et t avec la probabilité $e^{-a_i t}$, qui est approximativement égale à $1 - a_i t$ lorsque t est "petit"; il y a eu exactement un saut entre 0 et t, arrivant en $j \ne i$, avec la probabilité $\int_0^t a_i e^{-a_i s}\pi_{ij} e^{-a_j(t-s)} ds$, approximativement égale à $a_i\pi_{ij}t$ pour t petit, et la probabilité pour qu'il y ait eu au moins 2 sauts est de l'ordre de t^2. En se rappelant que $p(0)_{ii} = 1$ et $p(0)_{ij} = 0$ pour $j \ne i$, on s'attend donc à ce que $t \mapsto p(t)_{ij}$ soit dérivable (à droite) en 0, avec la dérivée suivante:

$$p'(0)_{ij} = \begin{cases} -a_i & \text{si } \; j = i \\ a_i\pi_{ij} & \text{si } \; j \ne i. \end{cases} \tag{4.3.1}$$

On a en fait bien mieux:

Théorème 4.3.1. *Les fonctions* $t \mapsto p(t)_{ij}$ *sont dérivables en tout point de* $I\!\!R_+$*, et vérifient les systèmes d'équations différentielles et intégrales suivants:*

$$p'(t)_{ij} = -a_i p(t)_{ij} + a_i \sum_k \pi_{ik}\; p(t)_{kj}, \tag{4.3.2}$$

$$p(t)_{ij} = \delta_{ij} e^{-a_i t} + \sum_k \pi_{ik} \int_0^t a_i e^{-a_i s}\; p(t-s)_{kj}\; ds. \tag{4.3.3}$$

Ces équations s'appellent les *premières équations de Kolmogorov.*

Preuve. Nous allons appliquer l'extension (3.1.10) de la propriété de Markov forte à la fonction $f(\omega,\omega') = 1_{\{j\}}(X_{t-S_0(\omega)}(\omega'))1_{\{S_0(\omega)\le t\}}$ (avec l'espace "auxiliaire" $(E,\mathcal{E})$ de cette formule pris égal à $(\Omega,\mathcal{F})$ lui-même), et la fonction identité $Y(\omega)=\omega$, et avec le temps d'arrêt $T=T_1=S_0$. On a alors $V(\omega) := f(\omega, Y\circ\theta_T(\omega)) = 1_{\{S_0(\omega)\le t\}}1_{\{j\}}(X_t(\omega))$ et $\int \mathbb{P}_k(d\omega')f(\omega,Y(\omega')) = 1_{\{S_0(\omega)\le t\}}\ p(t-S_0(\omega))_{kj}$, de sorte qu'il vient

$$\begin{aligned} p(t)_{ij} \;=\; \delta_{ij}\mathbb{P}_i(S_0>t)+\mathbb{E}_i(V) &\;=\; \delta_{ij}\mathbb{P}_i(S_0>t)+\mathbb{E}_i\left(1_{\{S_0\le t\}}\ \mathbb{E}_i(V|\mathcal{F}_{S_0+})\right)\\ &\;=\; \delta_{ij}\mathbb{P}_i(S_0>t)+\mathbb{E}_i\left(1_{\{S_0\le t\}}\ p(t-S_0)_{X_{S_0},j}\right). \end{aligned}$$

Il suffit alors d'appliquer le théorème 4.1.3-(1) pour obtenir (4.3.3).

Le changement de variable $t-s=u$ dans (4.3.3) donne

$$p(t)_{ij} = \delta_{ij}e^{-a_it} + e^{-a_it}\ \sum_k \pi_{ik}\int_0^t a_ie^{a_iu}\ p(u)_{kj}\ du. \tag{4.3.4}$$

On en déduit (utiliser le théorème de convergence dominée et le fait que $\sum_k \pi_{ik}\le 1$) que les fonctions $t\mapsto p(t)_{ij}$ sont continues, puis (par dérivation des intégrales d'intégrands continus par rapport à la borne supérieure) que leurs dérivées satisfont (4.3.2). □

Noter que (4.3.1) se déduit de (4.3.2), qu'on peut d'ailleurs réécrire sous la forme plus compacte suivante:

$$P'_t = P'_0P_t. \tag{4.3.5}$$

La formule ci–dessus est à prendre au sens "matriciel", mais elle est évidemment à rapprocher de l'égalité des membres extrêmes de (3.3.3): si f appartient au domaine $\mathcal{D}_A$ du générateur infinitésimal A du semi–groupe (P_t), la dérivée à droite de $t\mapsto P_tf(x)$ est $AP_tf(x)$. L'égalité de cette expression et de $P_tAf(x)$ suggère la propriété "matricielle" suivante:

$$P'_t = P_tP'_0. \tag{4.3.6}$$

Ceci revient à écrire

$$p'(t)_{ij} = -p(t)_{ij}\ a_j + \sum_k\ p(t)_{ik}\ a_k\ \pi_{kj}, \tag{4.3.7}$$

ou, sous forme intégrale:

$$p(t)_{ij} = \delta_{ij}\ e^{-a_jt} + \sum_k\int_0^t p(s)_{ik}\ a_k\ \pi_{kj}\ e^{-a_j(t-s)}\ ds \tag{4.3.8}$$

(comparer à (4.3.4)). Bien entendu, (4.3.7) découle de (4.3.8) par dérivation, et ces formules sont connues sous le nom de *secondes équations de Kolmogorov.*

En fait, les secondes équations de Kolmogorov ne sont pas toujours vraies, et on a seulement le résultat suivant:

Théorème 4.3.2. *Si le processus de Markov de saut pur* $\mathbf{X}$ *est sans explosion (c'est-à-dire* $\mathbb{P}_i(T_\infty=\infty)=1$ *pour tout* $i\in E$*), alors on a (4.3.7) et (4.3.8).*

Preuve. Supposons qu'on ait montré l'égalité (4.3.8) pour presque tout t. La fonction $\alpha(s) = \sum_k p(s)_{ik} a_k \pi_{kj} e^{a_j s}$ est positive et vérifie $\int_0^t \alpha(s) ds < \infty$ pour presque tout t, donc pour tout t, et par suite $t \mapsto \int_0^t \alpha(s) ds$ est continue, ce qui entraine que (4.3.8) a lieu pour tout t. Il suffit donc de montrer que les transformées de Laplace des deux membres de (4.3.8), en tant que fonctions de t, sont identiques.

Si $\lambda > 0$, en utilisant Fubini, le lemme 4.2.2 et la propriété de Markov forte en chaque T_n, et comme il n'y a pas d'explosion,

$$\begin{aligned}
(\lambda + a_j) \int_0^\infty e^{-\lambda t} p(t)_{ij} dt &= (\lambda + a_j) I\!E_i \left(\sum_{n \geq 0} 1_{\{j\}}(\xi_n) \int_{T_n}^{T_{n+1}} e^{-\lambda t} dt \right) \\
&= (\lambda + a_j) \sum_{n \geq 0} I\!E_i \left(1_{\{j\}}(\xi_n) e^{-\lambda T_n} \int_0^{S_0 \circ T_n} e^{-\lambda s} ds \right) \\
&= (\lambda + a_j) \sum_{n \geq 0} I\!E_i \left(1_{\{j\}}(\xi_n) e^{-\lambda T_n} \int_0^\infty a_j e^{-a_j u} du \int_0^u e^{-\lambda s} ds \right) \\
&= \sum_{n \geq 0} I\!E_i \left(1_{\{j\}}(\xi_n) e^{-\lambda T_n} \right) \;=\; \sum_{n \geq 0} q(\lambda)_{ij}^{(n)} \qquad (4.3.9)
\end{aligned}$$

La transformée de Laplace du membre de droite de (4.3.8), miltipliée par $\lambda + a_j$, s'écrit

$$\begin{aligned}
&(\lambda + a_j) \int_0^\infty e^{-\lambda t} dt \left(\delta_{ij}\ e^{-a_j t} + \sum_k \int_0^t p(s)_{ik}\ a_k\ \pi_{kj}\ e^{-a_j (t-s)}\ ds \right) \\
&= \delta_{ij} + (\lambda + a_j) \sum_k a_k\ \pi_{kj} \int_0^\infty e^{a_j s} p(s)_{ik} ds \int_s^\infty e^{-(\lambda + a_j) t} dt \\
&= \delta_{ij} + \sum_k a_k\ \pi_{kj} \int_0^\infty e^{-\lambda s} p(s)_{ik} ds \;=\; \delta_{ij} + \sum_k \sum_{n \geq 0} q(\lambda)_{ik}^{(n)} \frac{a_k\ \pi_{kj}}{\lambda + a_k}
\end{aligned}$$

en vertu de (4.3.9) appliqué à $p(s)_{ik}$. D'après (4.2.2) cette expression vaut

$$\delta_{ij} + \sum_{n \geq 0} q(\lambda)_{ij}^{(n+1)} = \sum_{n \geq 0} q(\lambda)_{ij}^{(n)}.$$

Donc les trasformées de Laplace des deux memebres de (4.3.8), multipliées par $\lambda + a_j$, sont identiques, et par suite (4.3.8) est vraie.

Passons à la preuve de (4.3.7). On pose $g(t)_{ij} = \sum_k p(t)_{ik} a_k \pi_{kj}$, de sorte qu'on peut réécrire (4.3.8) ainsi:

$$p(t)_{ij} = e^{-a_j t} \left(\delta_{ij} + \int_0^t g(s)_{ij} e^{a_j s}\ ds \right). \qquad (4.3.10)$$

Par ailleurs, en appliquant (4.3.4) et Fubini, on voit que si $h(t)_{ij} = \sum_k \pi_{ik} g(t)_{kj}$,

$$g(t)_{ij} = a_i e^{-a_i t} \left(\pi_{ij} + \int_0^t e^{a_i u} h(u)_{ij}\ du \right). \qquad (4.3.11)$$

D'abord, (4.3.10) implique que $s \mapsto g(s)_{ij}$ est intégrable par rapport à la mesure de Lebesgue sur tout compact, donc fini pour presque tout s; par suite (4.3.11) montre que $u \mapsto h(u)_{ij}$ est aussi intégrable sur chaque intervalle $[0,t]$ tel que $g(t)_{ij} < \infty$, donc en fait sur tout compact; par suite (4.3.11) encore montre qu'en fait $t \mapsto g(t)_{ij}$ est continue (donc finie partout): il suffit alors de dériver (4.3.10) pour obtenir (4.3.7). □

Equations de Kolmogorov et générateur. Revenons sur l'analogie entre les équations (3.3.3) d'une part, les équations (4.3.4) et (4.3.5) d'autre part. Cette analogie conduit à "identifier" A avec $P'(0)$, et on a effectivement le résultat suivant:

Proposition 4.3.3. *Si* $\sup_{i\in E} a_i < \infty$*, le générateur infinitésimal* $(A, \mathcal{D}_A)$ *du semi–groupe* (P_t) *est l'opérateur de domaine* $\mathcal{D}_A = B$ *(ensemble des fonctions bornées sur* E*), et qui pour* $f \in B$ *prend la valeur:*

$$Af(i) = \sum_j p'(0)_{ij} f(j) = \sum_j a_i \pi_{ij} f(j) - a_i f(i). \tag{4.3.12}$$

Preuve. Si $f \in B$, on a $\sum_{j,k} \pi_{ik} p(u)_{kj} |f(j)| \leq C$, où $C = \sup_i |f(i)|$. Donc on déduit de (4.3.4) et de Fubini que

$$P_t f(i) = e^{-a_i t} \left(f(i) + \int_0^t a_i e^{a_i u} \, \Pi P_u f(i) \, du \right), \tag{4.3.13}$$

de sorte que $|P_t f(i) - f(i)| \leq 2CC't$, où $C' = \sup_i a_i$. On en déduit d'abord que $B_0 = B$, avec la notation (3.3.1). On en déduit ensuite que $t \mapsto \Pi P_t f(i)$ est continu, de sorte que le membre de droite de (4.3.13) est dérivable en t, et un calcul simple montre que sa dérivée en $t = 0$ égale le membre de droite de (4.3.12). Par suite si Af est défini par (4.3.13), on a $\frac{P_t f - f}{t} \stackrel{\text{s–b}}{\longrightarrow} Af$, et $Af \in B = B_0$ est évident: cela montre que $B \subset \mathcal{D}_A$, et comme l'inclusion inverse est toujours vraie on a en fait $\mathcal{D}_A = B$. La formule (4.3.13) est alors évidente. □

Attention !! Le résultat précédent *est faux* lorsque que $\sup_i a_i = \infty$: dans ce cas, en effet, la formule (4.3.12) définit une fonction Af qui n'est pas bornée pour certaines fonctions bornées f. On peut cependant montrer que la formule (4.3.12) est vraie pour tout $f \in \mathcal{D}_A$, mais cette affirmation est à manier avec précaution: en effet, le générateur caractérise le semi–groupe, tandis que la formule (4.3.12) permet au mieux de récupérer les a_i et la matrice Π, et donc pas le semi–groupe lorsqu'il y a explosion. Il semble y avoir une contradiction, mais cette apparence est dûe à une mauvaise lecture de (4.3.12). En effet, le générateur est le "couple" $(A, \mathcal{D}_A)$, et l'information manquante dans (4.3.12) pour reconstruire le semi–groupe est en fait "cachée" dans l'ensemble $\mathcal{D}_A$.

Par exemple supposons que $E = \mathbb{N}^\star$ et $a_i = 1/i^2$ et $\pi_{ij} = 1$ si $j = i+1$ et $= 0$ sinon: le "processus minimal" associé parcourt tous les entiers successivement, et reste dans i un temps exponentiel de paramètre $1/i^2$; comme la série $\sum_i 1/i^2$ converge, il y a donc explosion. Si après chaque explosion le processus repart de 1, on obtient un semi–groupe (P_t); s'il repart chaque fois de 2, on obtient un *autre* semi–groupe (P'_t) (les deux processus correspondants sont différents: le premier ne retourne jamais en 1 après l'avoir quité, le second y retourne une infinité de fois). Au vu de (4.3.12), les deux générateurs $(A, \mathcal{D}_A)$ et $(A', \mathcal{D}_{A'})$ vérifient $Af = A'f$ pour tout $f \in \mathcal{D}_A \cap \mathcal{D}_{A'}$. Mais la fonction indicatrice $f = 1_{\{1\}}$ est dans $\mathcal{D}_A$ (puisque $p(t)_{11} = e^{-t}$ et $p(t)_{i1} = 0$ si $i \geq 2$), mais n'est pas dans $\mathcal{D}_{A'}$ (puisqu'on peut vérifier que $\liminf_{i \to \infty} p(t)_{i1} \geq e^{-t}$, et donc $\sup_i |(P_t f(i) - f(i))/t| = \infty$). Par suite $\mathcal{D}_{A'} \neq \mathcal{D}_A$.

4.4 Classification et Stationnarité

On considère toujours un processus de Markov de saut pur $\mathbf{X}$. De plus, et quoique cette hypothèse ne soit pas toujours nécessaire, on le supposera *sans explosion.*

4.4.1 Première classification

Exactement comme pour les chaînes de Markov, on peut diviser l'espace d'état E en "classes", à condition de bien définir les temps de passage dans les états. On pose

$$R_i = \inf(t : t > S_0, X_t = i), \qquad R_i^1 = R_i, \qquad R_i^{n+1} = \begin{cases} R_i^n + R_i \circ \theta_{R_i^n} & \text{si } R_i^n < \infty \\ \infty & \text{si } R_i^n = \infty. \end{cases}$$

(Nous utilisons la notation R_i au lieu de T_i comme dans le cas des chaînes, puisque ici T_n désigne le $n$ième temps de saut de $\mathbf{X}$).

Définition 4.4.1. Si $i, j \in E$ on dit que i mène à j, et on écrit $i \mapsto j$, si on a soit $i = j$, soit $\mathbb{P}_i(R_j < \infty) > 0$. On écrit $i \sim j$ si $i \mapsto j$ et $j \mapsto i$.

On a un critère analogue à la proposition 1.5.2, pour lequel nous devons aussi introduire le potentiel $U = (u_{ij})$ (c'est la même notion que (3.3.7)):

$$u_{ij} = \int_0^\infty p(t)_{ij} dt = \mathbb{E}_i \left(\int_0^\infty 1_{\{j\}}(X_t) dt \right). \tag{4.4.1}$$

Proposition 4.4.2. *Supposons* $\mathbf{X}$ *sans explosion. Il y a équivalence entre les quatre conditions*

(a) $i \mapsto j$,

(b) $u_{ij} > 0$,

(c) Il existe $t > 0$ *avec* $p(t)_{ij} > 0$,

(d) pour tout $t > 0$ *on a* $p(t)_{ij} > 0$,

et la relation $i \sim j$ *est une relation d'équivalence.*

Preuve. Si $j = i$, on a $p(t)_{ij} \geq e^{-a_i t} > 0$ par (4.3.3) (par exemple), donc on a clairement (a), (b), (c) et (d).

Supposons maintenant $j \neq i$. (d) $\Rightarrow$ (c) $\Rightarrow$ (b) $\Rightarrow$ (a) sont évidents (car $t \mapsto p(t)_{ij}$ est continue). Comme $\mathbf{X}$ n'explose pas, on peut écrire $R_j = T_N$, où N est une variable aléatoire à valeurs dans $\{1, 2, \ldots, \infty\}$, et $R_j < \infty \Leftrightarrow N < \infty$. Sous $I\!P_i$ et conditionnellement à $\sigma(\xi_n : n \geq 0)$, chaque variable T_m est une somme de m variables exponentielles ou infinies, donc le support de sa loi est soit $\{\infty\}$, soit $I\!R_+$. Donc le support de la loi de T_m sous $I\!P_i$ est soit $[0, \infty[$, soit $[0, \infty]$, soit $\{\infty\}$. Il en est évidemment de même pour la variable R_j, de sorte que sous (a) le support de la loi de R_j sous $I\!P_i$ est soit $[0, \infty[$, soit $[0, \infty]$. Comme $X_t = j$ si $R_j \leq t < R_j + S_0 \circ \theta_{R_j}$ et comme $S_0 \circ \theta_{R_j}$ est indépendante de R_j et exponentielle de paramètre a_j, il est alors clair que (a)$\Rightarrow$ (d).

Enfin, la dernière assertion se montre comme dans la proposition 1.5.2, en utilisant $p(t+s)_{ik} \geq p(t)_{ij} p(s)_{jk}$. □

On peut alors considérer les classes d'équivalence associées à la relation $\sim$ ci-dessus, qu'on appelle simplement les *classes* du processus $\mathbf{X}$. Elles se comparent aux classes de la chaîne de transition Π de la manière suivante (*attention*: cette dernière chaîne est à valeurs dans E_Δ; mais $\{\Delta\}$ en constitue évidemment une classe, et les autres sont toutes entièrement contenues dans E):

Proposition 4.4.3. *Supposons* $\mathbf{X}$ *sans explosion. Les classes de* $\mathbf{X}$ *sont exactement les classes associées à* Π *et contenues dans* E.

Preuve. En reprenant la preuve précédente, dans l'égalité $R_j = T_N$ la variable N représente le premier instant (entier, supérieur ou égal à 1) où la chaîne (ξ_n) atteint j: donc la relation $i \mapsto j$ signifie la même chose pour $\mathbf{X}$ et pour la chaîne associée à Π, d'où le résultat. □

Enfin, vu la proposition 4.4.2, la notion de période n'a ici aucune signification: de ce point de vue, les processus de Markov sont un peu plus simples que les chaînes de Markov.

4.4.2 Seconde classification

Définition 4.4.4. Un état i est dit *récurrent* si $I\!P_i(R_i < \infty) = 1$ ou s'il est absorbant ($a_i = 0$), et *transient* sinon.

Théorème 4.4.5. *Supposons* $\mathbf{X}$ *sans explosion.*

(1) On a les équivalences

$$i \text{ récurrent} \Leftrightarrow u_{ii} = \infty \Leftrightarrow I\!P_i(\cap_n \{R_i^n < \infty\}) = 1 \Leftrightarrow I\!P_i(\forall t, \exists s > t \text{ avec } X_s = i) = 1. \tag{4.4.2}$$

Dans ce cas si i *est non-absorbant il est aussi récurrent pour la chaîne* (ξ_n) *et si de plus* $i \sim j$ *alors* j *est aussi récurrent et on a* $I\!P_i(R_j < \infty) = 1$ *et* $u_{ij} = \infty$ *et* $I\!P_i(\forall t, \exists s > t \text{ avec } X_s = j) = 1$.

(2) On a les équivalences

$$i \text{ transient} \Leftrightarrow u_{ii} < \infty \Leftrightarrow I\!P_i(\cup_n \{R_i^n = \infty\}) = 1 \Leftrightarrow I\!P_i(\exists t, \forall s > t \text{ on a } X_s \neq i) = 1. \tag{4.4.3}$$

Dans ce cas i *est non-absorbant et si* $i \sim j$ *alors* j *est transient et* $u_{ij} < \infty$ *et* $I\!P_i(\exists t, \forall s > t \text{ on a } X_s \neq j) = 1$.

Preuve. Si i est absorbant, il est évident qu'on a les quatre assertions de (4.4.2), et tout $j \sim i$ verifie $j = i$.

Supposons donc $i \sim j$ non–absorbants. Comme $\mathbf{X}$ est non–explosif, on a $\lim_n R_i^n = \infty$, donc clairement

$$\cap_n\{R_i^n < \infty\} = \{\forall t, \exists s > t \text{ avec } X_s = i\}, \quad \cup_n\{R_i^n = \infty\}) = \{\exists t, \forall s > t \text{ on a } X_s \neq i\}, \tag{4.4.4}$$

Noter aussi que

$$\int_0^\infty 1_{\{i\}}(X_s)ds = S_0 1_{\{X_0=i\}} + \sum_{n\geq 1} S_0 \circ \theta_{R_i^n}\ 1_{\{R_i^n<\infty\}}. \tag{4.4.5}$$

Posons $N_i^0 = 0$ et $N_i^{k+1} = \inf(n > N_i^k : \xi_n = i)$ (temps de passage successifs de (ξ_n) en i). Il est clair que $R_i^n = T_{R_i^n}$ et R_i^n est fini si et seulement si N_i^n est fini (les N_i^n ici correspondent, pour la chaîne (ξ_n), aux T_i^n du paragraphe 1.5.1). Ainsi on a

$$f_{ij} := I\!\!P_i(R_j < \infty) = I\!\!P_i(N_i^1 < \infty),$$

et d'après le lemme 1.5.8 il vient

$$f_{ij}^{(n+1)} := I\!\!P_i(R_j^{n+1} < \infty) = I\!\!P_i(N_j^{n+1} < \infty) = f_{ij}(f_{jj})^n. \tag{4.4.6}$$

Comme $S_0 \circ \theta_{R_i^n}$ est exponentielle de paramètre a_i, donc d'espérance $1/a_j$, conditionnellement à $R_i^n < \infty$, il découle de (4.4.1), (4.4.5) et (4.4.6) que

$$\begin{aligned} u_{ij} &= \delta_{ij}\ I\!\!E_i(S_0) + \sum_{n\geq 1} I\!\!E_i(S_0 \circ \theta_{R_j^n}\ 1_{\{R_j^n<\infty\}}) \\ &= \frac{1}{a_j}\left(\delta_{ij} + \sum_{n\geq 1} I\!\!P_i(R_j^n < \infty)\right) = \frac{1}{a_j}\left(\delta_{ij} + f_{ij}\sum_{n\geq 1}(f_{jj})^{n-1}\right). \end{aligned} \tag{4.4.7}$$

Compte tenu de (4.4.4), (4.4.6) et (4.4.7), on obtient les équivalences dans (1) et (2), ainsi que le fait que i est récurrent pour $\mathbf{X}$ si et seulement s'il est récurrent pour (ξ_n) (rappelons que i est non–absorbant, donc $a_i > 0$). Au vu de (4.4.4) et (4.4.7) et du théorème 1.5.9, les dernières assertions de (1) et (2) sont aussi immédiates. □

4.4.3 Propriétés ergodiques

On va maintenant étudier le comportement ergodique du processus $\mathbf{X}$, c'est-à-dire l'existence de limites pour les $p(t)_{ij}$ lorsque $t \to \infty$, au moins dans le cas non–explosif. Les résultats sont plus simples que dans le cas discret, puisqu'il n'y a pas de notion de période. Nous allons nous contenter d'un résultat partiel, qui n'explicite pas la limite obtenue. Pour simplifier on supposera aussi $\mathbf{X}$ sans explosion.

Théorème 4.4.6. *Supposons $\mathbf{X}$ sans explosion. Pour tous $i, j \in E$ les quantités $p(t)_{ij}$ convergent vers une limite α_{ij} lorsque $t \to \infty$. De plus:*

(1) Si j est transient on a $\alpha_{ij} = 0$ pour tout i.

(2) Si j est récurrent on a $\alpha_{ij} = I\!\!P_i(R_j < \infty)\alpha_{jj}$; si de plus $i \sim j$, on a $\alpha_{ij} = \alpha_{jj}$.

(3) On a $\sum_j \alpha_{ij} \leq 1$.

Preuve. On fera la démonstration seulement dans le cas où on a l'hypothèse supplémentaire suivante, dont on peut en fait se dispenser: à savoir que $A = \sup_i a_i < \infty$.

Soit $s > 0$. La matrice P_s est la probabilité de transition de la chaîne de Markov (à temps discret) $(X_{ns})_{n\in\mathbb{N}}$, et évidemment $(P_s)^n = P_{ns}$. De plus $p(s)_{jj} \geq e^{-a_j s} > 0$, donc la période de j pour cette chaîne est 1. En vertu du théorème 1.6.6 et de son corollaire 1.6.7, la suite $p(ns)_{ij}$ converge vers une limite $\gamma(s)_{ij}$ quand $n \to \infty$.

Soit maintenant $t > 0$; il existe un entier n tel que $ns \leq t < (n+1)s$, et $P_t = P_{ns}P_{t-ns}$. Avec la notation $g(u)_{ij}$ de la preuve du théorème 4.3.2, il vient alors par (4.3.8):

$$p(t)_{ij} = p(ns)_{ij}e^{-a_j(t-ns)} + \sum_k p(ns)_{ik} \int_0^{t-ns} g(u)_{kj}\ e^{-a_j(t-ns-u)}\ du.$$

D'une part on en déduit que $p(t)_{ij} \geq p(ns)_{ij}e^{-a_j s}$; donc pour tout $s > 0$ fixé, et comme $n \to \infty$ lorsque $t \to \infty$, on obtient

$$b_{ij} := \liminf_t\ p(t)_{ij} \geq \gamma(s)_{ij}e^{-a_j s}.$$

D'autre part notre hypothèse additionnelle implique que $g(u)_{kj} \leq A$ pour tout k, donc on a aussi $p(t)_{ij} \leq p(ns)_{ij} + sA$, et par suite

$$c_{ij} := \limsup_t\ p(t)_{ij} \leq \gamma(s)_{ij} + sA.$$

En rassemblant ces deux résultats, on voit que $c_{ij} - b_{ij} \leq sA + 1 - e^{a_i j s}$ pour tout s, et comme cette dernière quantité tend vers 0 quand $s \to 0$ on en déduit que $c_{ij} = b_{ij}$: cela entraine que $p(t)_{ij}$ converge vers une limite α_{ij} $(= b_{ij})$ quand $t \to \infty$, et évidemment on a en fait $\gamma(s)_{ij} = \alpha_{ij}$ pour tout s.

Comme $\sum_j p(t)_{ij} = 1$, (3) découle de ce qui précède et du lemme de Fatou. (2) provient immédiatement du fait que si j est transient on a $u_{ij} = \int_0^\infty p(t)_{ij}dt < \infty$ (cf. théorème 4.4.5). Supposons enfin j récurrent. En utilisant la propriété (3.1.10) avec le temps d'arrêt R_j on voit que

$$p(t)_{ij} = \mathbb{P}_i(R_j \leq t, X_t = j) = \mathbb{E}_i\left(1_{\{R_j \leq t\}}\ p(t - R_j)_{jj}\right).$$

Lorsque $t \to \infty$, la variable $1_{\{R_j\leq t\}}p(t-R_j)_{jj}$ converge simplement vers $1_{\{R_j<\infty\}}\alpha_{jj}$, donc la première partie de (2) provient du théorème de Lebesgue, et la seconde de la première et du fait que $\mathbb{P}_i(R_j < \infty) = 1$ si $i \sim j$ (cf. théorème 4.4.5-(1)). $\square$

4.4.4 Probabilités invariantes

Nous nous contenterons d'examiner le problème de l'existence et de la détermination des probabilités invariantes (les lois initiales qui rendent le processus $\mathbf{X}$ stationnaire, d'après la proposition 3.1.9). On pourrait utiliser le théorème ergodique 4.4.6, mais outre le fait qu'on n'a pas explicité les limites dans ce théorème (et qu'il n'a été montré que lorsque $\sup_i a_i < \infty$), il est intéressant de relier les probabilités invariantes aux termes Π et a_i.

Théorème 4.4.7. *Supposons* $\mathbf{X}$ *sans explosion. Si* $\mu = (\mu_i)$ *est une probabilité sur* E*, il y a équivalence entre les assertions suivantes:*

(i) μ *est invariante.*

(ii) On a l'équation matricielle $\mu P'_0 = 0$.

(iii) On a $\mu_i a_i = \sum_j \mu_j a_j \pi_{ji}$ *pour tout* i.

Bien qu'il n'en découle pas au sens mathématique, ce théorème est à rapprocher de la proposition 3.3.9.

Preuve. La matrice P'_0 étant donnée par (4.3.1), il est clair que (ii) = (iii).

On va utiliser les matrices $Q(\lambda) = (q(\lambda)_{ij})$ de (4.2.2) pour $\lambda > 0$, et le lemme 4.2.2. D'après Fubini, la propriété de Markov forte et le fait que $\mathbb{E}_j(1 - e^{-\lambda S_0}) = \frac{\lambda}{a_j+\lambda}$, on obtient (rappelons que $e^{-\infty} = 0$):

$$\begin{aligned}
\lambda \int_0^\infty e^{-\lambda t}\, p(t)_{ij}\, dt &= \mathbb{E}_i\left(\lambda \int_0^\infty e^{-\lambda t}\, 1_{\{j\}}(X_t)\, dt\right)\\
&= \mathbb{E}_i\left(\sum_{n\geq 0} 1_{\{j\}}(X_{T_n})\, 1_{\{T_n<\infty\}}\, \lambda \int_{T_n}^{T_{n+1}} e^{-\lambda t}\, dt\right)\\
&= \sum_{n\geq 0} \mathbb{E}_i\left(1_{\{j\}}(X_{T_n})\, e^{-\lambda T_n}\left(1 - e^{-\lambda S_0\circ\theta_{T_n}}\right)\right)\\
&= \sum_{n\geq 0} \mathbb{E}_i\left(1_{\{j\}}(X_{T_n})\, e^{-\lambda T_n}\, \frac{\lambda}{a_j+\lambda}\right) \;=\; \sum_{n\geq 0} q(\lambda)_{ij}^{(n)}\, \frac{\lambda}{a_j+\lambda}.
\end{aligned}$$

Si on a (i) il est clair que

$$\int_0^\infty e^{-\lambda t} \sum_j \mu_j p(t)_{ji}\, dt = \frac{1}{\lambda}\,\mu_i \tag{4.4.8}$$

pour tout $\lambda > 0$. Supposons à l'inverse (4.4.8); le membre de gauche (resp. de droite) est la transformée de Laplace en α de la fonction positive continue $t \mapsto \sum_j \mu_j p(t)_{ji}$ (resp. $t \mapsto \mu_i$), donc par un résultat classique ces deux fonctions sont égales: ainsi, (i) équivaut à (4.4.8) pour tout $\lambda > 0$ et tout i.

D'après le calcul précédent et Fubini, (4.4.8) et donc (i) équivalent à

$$\mu_i(a_i + \lambda) = \lambda \sum_j \mu_j \sum_{n\geq 0} q(\lambda)_{ji}^{(n)} \tag{4.4.9}$$

pour tout i. Mais $\sum_{n\geq 0} Q(\lambda)^n = I + \left(\sum_{n\geq 0} Q(\lambda)^n\right) Q(\lambda)$, donc (i) équivaut aussi à

$$\mu_i(a_i + \lambda) = \mu_i\lambda + \sum_k \left(\lambda \sum_j \mu_j \sum_{n\geq 0} q(\lambda)_{jk}^{(n)}\right) q(\lambda)_{ki} \tag{4.4.10}$$

pour tout i. En injectant (4.4.9) dans (4.4.10), cela conduit à

$$\mu_i(a_i+\lambda)=\mu_i\lambda+\sum_k \mu_k(a_k+\lambda)q(\lambda)_{ki} \tag{4.4.11}$$

pour tout i. Vu (4.2.2), ceci n'est autre que (iii), et donc (i) implique (iii).

A l'inverse, si on a (iii), on a aussi (4.4.11), et en itérant on voit facilement que pour tout $q\geq 1$, on a

$$\mu_i(a_i+\lambda)=\lambda\sum_k\mu_k\sum_{n=0}^{q-1}q(\lambda)^{(n)}_{ki}+\sum_k \mu_k(a_k+\lambda)q(\lambda)^{(q)}_{ki}.$$

Mais d'après le théorème (4.2.3, notre hypothèse de non–explosion entraîne que le dernier terme ci–dessus tend vers 0 quand $q\to\infty$. En passant à la limite, on obtient donc (4.4.9), et on a donc (i). □

En d'aures termes, si on asocie à toute mesure μ la mesure μ' définie par $\mu'_i=\mu_ia_i$, la mesure μ est invariante pour le semi–groupe de $\mathbf{X}$ si et seulement si la mesure μ' est invariante pour la transition Π (en restriction à E). Mais il faut faire attention: si μ est une probabilité il se peut que μ' n'en soit pas une, et même soit infinie; inversement si Π admet une probabilité invariante μ', la mesure associée μ n'est pas nécessairement une probabilité, et peut être infinie.

Une manière de trouver les probabilités invariantes est alors la suivante: d'abord, si i est absorbant, la masse de Dirac ε_i est évidemment invariante. Ensuite, on détermine les mesures invariantes μ' pour Π (un problème *a priori* plus simple...) portées par une classe d'équivalence de Π (donc vérifiant $\mu'_i=0$ si i est absorbant); on pose alors $\mu''_i=\mu_i/a_i$ (avec $0/0=0$); si $C=\sum_i\mu''_i<\infty$, la mesure $\mu_i=\mu''_i/C$ est une probabilité invariante pour $\mathbf{X}$. Enfin, toute combinaison linéaire convexe de probabilités invariantes est encore une probabilité invariante.

Corollaire 4.4.8. *Supposons $\mathbf{X}$ sans explosion. Si la chaîne (ξ_n) associée à Π est irréductible récurrente, le processus $\mathbf{X}$ est aussi irréductible récurrent, et admet au plus une probabilité invariante.*

Preuve. Dans ce cas on sait que les états sont tous non–absorbants. De plus le théorème 4.4.5 entraine que $\mathbf{X}$ possède une seule classe, dont tous les états sont récurrents. Soit μ et ν deux probabilités invariantes pour (P_t). Alors μ' et ν' sont invariantes pour Π, et d'après le théorème 1.6.2 on a $\mu'_i=\alpha\nu'_i$ pour une constante $\alpha\geq 0$, et comme $a_i>0$ et $\sum_i\mu_i=\sum_i\nu_i=1$ on a nécessairement $\mu=\nu$. □

Remarque: Rappelons encore que si la chaîne (ξ_n) est irréductible récurrente positive on n'est pas sûr de l'existence d'une probabilité invariante pour $\mathbf{X}$. Si à l'inverse $\mathbf{X}$ est irréductible et possède une probabilité invariante, tous les états sont récurrents pour $\mathbf{X}$ (car en passant à la limite dans $\mu P_t=\mu$ on obtient $\sum_j\mu_i\alpha_{ij}=\mu_j$, donc $\alpha_{ij}>0$ pour au moins un couple (i,j) et on applique le théorème 4.4.6), donc tous les états sont aussi récurrents pour (ξ_n); mais ils ne sont pas nécessairement positifs pour cette chaîne.

Chapitre 5

Processus de Poisson

5.1 Le processus de Poisson

5.1.1 Les processus de comptage

Définition 5.1.1. Un *processus de comptage* est un processus $N = (N_t)_{t\geq 0}$ dont les trajectoires $t \mapsto N_t(\omega)$ sont à valeurs dans $\overline{I\!N} = \{0, 1, \ldots, \infty\}$, croissantes, continues à droite, nulles en 0, et de sauts de taille 1.

On associe au processus de comptage N ses temps de saut successifs T_n et les intervalles inter–sauts S_n, par les formules suivantes (avec la convention $\infty - \infty = \infty$):

$$T_n = \inf(t : N_t = n), \qquad S_n = T_{n+1} - T_n, \qquad \forall n \in I\!N, \tag{5.1.1}$$

et on pose aussi $T_\infty = \lim_n T_n$, le "temps d'explosion". Les temps T_0 et T_∞ ne sont *pas* des temps de saut. On a $N_t = \infty$ si et seulement si $t \geq T_\infty$, et le processus N est dit *non–explosif* si $T_\infty = \infty$ identiquement (ou au moins p.s., si on a une probabilité).

Remarquer que $T_n < T_{n+1}$ si $T_n < \infty$, et en particulier $T_1 > 0$. On peut imaginer que les T_n pour $n \geq 1$ sont les temps d'occurrence d'un certain événement, et N_t est le nombre d'occurrences sur l'intervalle $]0, t]$: cela explique la terminologie "processus de comptage".

Si (5.1.1) donne les T_n et les S_n en fonction de N, on peut à l'inverse reconstruire le processus N à partir des T_n (une suite de variables avec $T_0 = 0$, $T_{n+1} > T_n$ si $T_n < \infty$ et $T_{n+1} = \infty$ si $T_n = \infty$), ou des variables S_n strictement positives, grâce aux formules

$$N_t = \sum_{n\geq 1} 1_{\{T_n \leq t\}}, \qquad T_n = S_0 + \ldots + S_{n-1} \quad \text{si } n \geq 1. \tag{5.1.2}$$

(l'indexation des S_n, en partant de S_0, est la même que pour les processus de Markov de saut pur).

Si enfin on a une filtration $(\mathcal{F}_t)$, en remarquant que $\{N_t = n\} = \{T_n \leq t < T_{n+1}\}$ et que $\{T_n \leq t\} = \{N_t \geq n\}$, on voit qu'on a:

$$N \text{ est adapté} \qquad \Longleftrightarrow \qquad \text{les } T_n \text{ sont des temps d'arrêt.} \tag{5.1.3}$$

5.1.2 Trois définitions équivalentes

Il existe plusieurs définitions possibles – équivalentes – du (ou des) processus de Poisson. La première n'est pas la plus "élémentaire", mais elle est la plus facile à énoncer, quand on connaît les processus de Lévy.

Définition 5.1.2. On appelle $(\mathcal{F}_t)$–*processus de Poisson* un processus de comptage N sur un espace probabilisé filtré $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ qui est aussi un $(\mathcal{F}_t)$–processus de Lévy. Lorsque $\mathcal{F}_t = \sigma(N_s : s \leq t)$ on dit simplement *processus de Poisson.*

Ainsi, un processus de Poisson prend ses valeurs dans $\overline{\mathbb{N}}$ (comme processus de comptage) et dans $\mathbb{R}^d$ (comme processus de Lévy), donc $d = 1$ et en fait il prend ses valeurs dans $\mathbb{N}$. Il est donc non–explosif. Noter aussi qu'un $(\mathcal{F}_t)$–processus de Poisson est *a fortiori* un processus de Poisson.

Donnons maintenant, sous forme de propriétés équivalentes à la définition 5.1.1, deux autres définitions des processus de Poisson.

Théorème 5.1.3. *Soit N est un processus de comptage adapté à une filtration $(\mathcal{F}_t)$. Il y a équivalence entre*

(1) N est une $(\mathcal{F}_t)$–processus de Poisson,

(2) Il existe $\lambda \geq 0$ tel que le processus $N_t - \lambda t$ soit une martingale relativement à $(\mathcal{F}_t)$.

Si de plus $\mathcal{F}_t = \sigma(N_s : s \leq t)$, ces deux propriétés sont aussi équivalentes à

(3) les S_n sont des variables indépendantes de loi exponentielle de même paramètre $\lambda \geq 0$ (rappelons qu'une variable exponentielle de paramètre 0 est identiquement infinie).

Enfin dans ce cas, le nombre λ est le même dans (2) et dans (3), et est appelé le paramètre *du processus de Poisson.*

Une conséquence importante de ce théorème est que pour tout $\lambda \geq 0$ il existe un processus de Poisson de paramètre λ, et un seul en loi: il suffit de construire une suite (S_n) comme dans (3), et de poser (5.1.2).

La preuve sera faite en plusieurs étapes, en même temps que nous démontrerons quelques autres propriétés intéressantes.

Lemme 5.1.4. *On a (1) $\Rightarrow$ (3).*

Preuve. Si on a (1), N est de saut pur et vérifie la propriété de Markov, donc d'après le théorème 4.1.3 la variable S_0 est exponentielle avec un certain paramètre $\lambda \geq 0$. Si $\lambda = 0$ on a $S_0 = \infty$ p.s., donc aussi $S_n = \infty$ p.s. d'après (5.1.1), et on a (3).

Supposons alors $\lambda > 0$. En utilisant les translation θ_t associées à N comme dans (3.2.9), on voit que $S_n = S_0 \circ \theta_{T_n}$. La propriété (3.2.10) entraine alors que S_n est indépendante de $\mathcal{F}_{T_n+}$ et de même loi que S_0, conditionnellement à $\{T_n < \infty\}$: on en déduit d'abord par récurrence sur n que $T_n < \infty$ p.s., puis donc que S_n est indépendante de $\mathcal{F}_{T_n+}$ et *a fortiori* de $S_0, \ldots, S_{n-1}$: on a donc (3). $\square$

D'après cette preuve, on peut considèrer le processus de Poisson comme un processus de Markov de saut pur à valeurs dans $\mathbb{N}$; avec ce point de vue, les caractéristiques Π et

a_i qui lui sont associées dans le théorème 4.1.3 sont

$$a_i = \lambda, \qquad \pi_{ij} = \begin{cases} 1 & \text{si } j = i+1 \\ 0 & \text{sinon.} \end{cases} \tag{5.1.4}$$

Pour le résultat suivant, rappelons que la loi gamma de paramètre $\lambda > 0$ et d'indice n est la loi de densité

$$f_{n,\lambda}(s) = \frac{\lambda^n \ s^{n-1}}{(n-1)!} \ e^{-\lambda s} \ 1_{I\!R_+}(s). \tag{5.1.5}$$

Donc si $n = 1$ c'est la loi exponentielle de paramètre λ. La transformée de Laplace en est

$$\phi_{n,\lambda}(v) := \int e^{-vs} f_{n,\lambda}(s)ds = \left(\frac{\lambda}{\lambda + v}\right)^n.$$

Lemme 5.1.5. *Si N est un processus de Poisson de paramètre $\lambda > 0$, la loi de N_t est la loi de Poisson de paramètre λt (i.e., $I\!P(N_t = n) = \frac{(\lambda t)^n}{n!} \ e^{-\lambda t}$ pour $n \in I\!N$), et la loi de T_n est la loi gamma de paramètre λ et d'indice n.*

Preuve. Comme T_n est la somme de n variables exponentielles de paramètre λ et indépendantes, sa transformée de Laplace est $v \mapsto (\phi_{1,\lambda}(v))^n = \phi_{n,\lambda}(v)$, et on a la seconde assertion. On a aussi

$$I\!P(N_t = n) = I\!P(T_n \le t < T_{n+1}) = I\!P(T_n \le t) - I\!P(T_{n+1} \le t).$$

Une intégration par parties utilisant les fonctions $u(s) = e^{-\lambda s}$ et $v(s) = \frac{\lambda^n s^n}{n!}$ donne

$$\int_0^t f_{n,\lambda}(s)ds = u(t)v(t) - u(0)v(0) + \int_0^t f_{n+1,\lambda}(s)ds,$$

et $u(t)v(t) = \frac{(\lambda t)^n}{n!} \ e^{-\lambda t}$ et $u(0)v(0) = 0$, d'où la première assertion. $\square$

Lemme 5.1.6. *On a (1) $\Rightarrow$ (2) et le paramètre λ dans (2) est le même que dans (3).*

Preuve. Supposons (1). D'après les lemmes précédents on a (3) avec un paramètre $\lambda \ge 0$ et si $\lambda > 0$ la variable N_t est intégrable et d'espérance λt; si $\lambda = 0$ on a $S_0 = \infty$ p.s., donc $N_t = 0$ p.s., donc en fait $I\!E(N_t) = \lambda t$ dans tous les cas.

Posons $M_t = N_t - \lambda t$, qui est $\mathcal{F}_t$–mesurable et d'intégrale nulle. En utilisant le fait que $N_{t+s} - N_t$ est indépendante de $\mathcal{F}_t$ et de même loi que N_s, on voit que

$$I\!E(M_{t+s} - M_t|\mathcal{F}_t) = I\!E(N_{t+s} - N_t|\mathcal{F}_t) - \lambda s = I\!E(N_s) - \lambda s = 0,$$

d'où (3) avec le paramètre λ. $\square$

Lemme 5.1.7. *Si $\mathcal{F}_t = \sigma(N_s : s \le t)$, on a (3) $\Rightarrow$ (1).*

Preuve. Il s'agit de montrer la propriété suivante: étant donnée une suite $(S_n)_{n \in I\!N}$ de variables indépendantes et exponentielles de paramètre λ, le processus N défini par (5.1.2) est un processus de Poisson, relativement à la filtration $\mathcal{F}_t = \sigma(N_s : s \le t)$. Si $\lambda = 0$ on a en fait $N \equiv 0$ et il n'y a rien à montrer, donc nous supposons $\lambda > 0$.

D'abord, d'après le loi des grands nombres $T_n/n \to 1/\lambda > 0$ p.s., donc $T_n \to \infty$ p.s. et le processus N est non–explosif.

On peut voir le résultat comme un cas particulier du théorème 4.2.1, avec les données (5.1.4), mais il est aussi simple (ou compliqué...!) de reprendre la même démonstration plutôt que d'adapter ce théorème. Comme toute variable $\mathcal{F}_t$–mesurable est égale, en restriction à chaque ensemble $\{N_t = n\}$, à une fonction de $(S_0, \ldots, S_{n-1})$, il suffit (à cause du caractère non–explosif) de montrer que pour tous $m, n \in \mathbb{N}$, tous $s, t \geq 0$ et toute fonction borélienne bornée g sur $\mathbb{R}_+^n$,

$$\mathbb{E}(Z\ 1_{\{N_t=n\}}\ 1_{\{N_{t+s}-N_t=m\}}) = \mathbb{E}(Z\ 1_{\{N_t=n\}})\ \mathbb{P}(N_s = m). \tag{5.1.6}$$

On a calculé déjà $\mathbb{P}(N_s = m)$ dans le lemme 5.1.5), mais comme cette quantité est $\mathbb{P}(T_m \leq s < T_{m+1})$ on peut aussi l'écrire ainsi:

$$\mathbb{P}(N_s = m) = \int_{]0,\infty[^{m+1}} \left(\prod_{k=0}^{m} \lambda\ e^{-\lambda r_k}\right)\ f(r_0, \ldots, r_m) dr_0 \ldots dr_m,$$

avec $f(r_0, \ldots, r_m) = 1_{\{r_0+\ldots+r_{m-1} \leq s < r_0+\ldots+r_m\}}$. Donc le membre de droite de (5.1.6) vaut

$$\int_{]0,\infty[^{n+1}} \left(\prod_{k=0}^{n} \lambda\ e^{-\lambda s_k}\right) g(s_0, \ldots, s_{n-1})\ 1_{\{s_0+\ldots+s_{n-1} \leq t < s_0+\ldots+s_n\}} ds_0 \ldots ds_n$$
$$\int_{]0,\infty[^{m+1}} \left(\prod_{l=0}^{m} \lambda\ e^{-\lambda r_l}\right) f(r_0, \ldots, r_m)\ dr_0 \ldots dr_m.$$

L'intégrale de $s_n \mapsto \lambda e^{-\lambda s_n}$ sur l'ensemble $\{s_0 + \ldots + s_{n-1} \leq t < s_0 + \ldots + s_n\}$ égale $\exp(-\lambda(t - (s_0 + \ldots + s_{n-1}))\ 1_{\{t \geq s_0+\ldots+s_{n-1}\}}$, donc l'expression précédente vaut

$$\int_{]0,\infty[^{n+m+1}} \lambda^n\ e^{-\lambda t} \left(\prod_{l=0}^{m} \lambda\ e^{-\lambda r_l}\right) g(s_0, \ldots, s_{n-1}) 1_{\{s_0+\ldots+s_{n-1} \leq t\}} f(r_0, \ldots, r_m)$$
$$ds_0 \ldots ds_{n-1} dr_0 \ldots dr_m.$$

On fait les changement de variables $s_{n+k} = r_k$ si $k \geq 1$ et $s_n = r_0 + t - (s_0 + \ldots + s_{n-1})$, de jacobien 1, et l'expression précédente devient

$$\int_{]0,\infty]^{n+m+1}} \left(\prod_{k=0}^{n+m} \lambda\ e^{-\lambda s_k}\right) g(s_0, \ldots, s_{n-1})\ 1_{\{s_0+\ldots+s_{n-1} \leq t < s_0+\ldots+s_n\}}$$
$$f(s_0 + \ldots + s_n - t, s_{n+1}, \ldots, s_{n+m}) ds_0 \ldots ds_{n+m},$$

qui égale clairement le premier membre de (5.1.6), et la preuve est terminée. □

La preuve de l'implication (2) $\Rightarrow$ (1) nécessite une version élémentaire du "calcul stochastique" discontinu. On se place sur une espace probabilisé filtré. Un processus X est dit à variation finie s'il est la différence de deux processus croissant positifs, continus à droite, nuls en 0, soit $X_t = X'_t - X''_t$. La "variation" de X est le processus $X'_t + X''_t$.

On peut intégrer par rapport à un tel processus X tout processus $H = (H_t)$ borné et à trajectoires continues à gauche, par exemple:

$$\int_0^t H_s dX_s = \int_0^t H_s dX'_s - \int_0^t H_s dX''_s,$$

où les deux intégrales de droite sont des "intégrales de Stieltjes" (la première, par exemple, est l'intégrale de $s \mapsto H_s$ par rapport à la mesure dont la fonction de répartition est $t \mapsto X'_t$); tout ceci se fait "ω par ω".

Lorsque de plus H s'écrit $H_t = \sum_{n\geq 0} \alpha_n 1_{]T_n,T_{n+1}]}(t)$, pour une suite de variables α_n uniformément bornées, et une suite strictement croissante (T_n) tendant vers l'infini, on a évidemment

$$\int_0^t H_s dX_s = \sum_{n\geq 0} \alpha_n (X_{T_{n+1}\wedge t} - X_{T_n \wedge t}), \tag{5.1.7}$$

où la somme ci–dessus ne comporte qu'un nombre fini (aléatoire) de termes non nuls.

Lemme 5.1.8. *En plus des notations et hypothèses ci–dessus, supposons que les T_n soient des temps d'arrêt, que les α_n soient $\mathcal{F}_{T_n+}$–mesurables, que X soit une martingale et que sa variation $X'_t + X''_t$ soit intégrable pour tout t. Alors le processus $H \bullet X_t = \int_0^t H_s dX_s$ donné par (5.1.7) est également une martingale continue à droite.*

Preuve. Posons $Y = H \bullet X$. Si C est une borne pour H, on a $|Y_t| \leq C(X'_t + X''_t)$, qui est intégrable, et (5.1.7) montre que Y_t est $\mathcal{F}_t$–mesurable (cf. la partie (b) de la preuve du théorème 3.1.6), et la continuité à droite de Y est évidente.

Il reste à montrer que si $s < t$ on a $\mathbb{E}(Y_t - Y_s|\mathcal{F}_s) = 0$. Vu (5.1.7), il suffit de montrer que $\mathbb{E}(U_n|\mathcal{F}_s) = 0$ pour chaque n, où

$$U_n = \alpha_n \left(X_{T_{n+1}\wedge t} - X_{T_n \wedge t} - X_{T_{n+1}\wedge s} + X_{T_n \wedge s} \right).$$

Il suffit même de le montrer séparément sur les trois éléments $A_n = \{T_{n+1} < s\}$, $B_n = \{T_n < s \leq T_{n+1}\}$, $C_n = \{s \leq T_n\}$. Sur A_n on a $U_n = 0$, d'où le résultat. Sur B_n on a $U_n = \alpha_n(X_{T_{n+1}\wedge t} - X_s)$ et $\alpha_n 1_{B_n}$ est $\mathcal{F}_s$–mesurable, donc $\mathbb{E}(U_n 1_{B_n}|\mathcal{F}_s) = 0$ par le théorème d'arrêt pour les martingales à temps continu, continues à droite. Sur C_n on a $U_n = \alpha_n(X_{T_{n+1}\wedge t} - X_{T_n \wedge t})1_{\{T_n<t\}}$, et $\alpha_n 1_{\{T_n<t\}}$ est $\mathcal{F}_{(T_n \wedge t)+}$–mesurable, donc une nouvelle application du théorème d'arrêt montre que $\mathbb{E}(U_n 1_{C_n}|\mathcal{F}_s) = 0$. □

Lemme 5.1.9. *On a (2) ⇒ (1).*

Preuve. On part d'un processus de comptage N, tel que le processus $M_t = N_t - \lambda t$ soit une martingale relativement à la filtration $(\mathcal{F}_t)$, pour un certain nombre $\lambda \geq 0$. En particulier $\mathbb{E}(M_t) = \mathbb{E}(M_0) = 0$, donc M_t et aussi N_t sont à valeurs finies p.s.: par suite N est non–explosif.

Si $\lambda = 0$, on a aussi $\mathbb{E}(N_t) = \mathbb{E}(M_t) = 0$. Donc en fait $N \equiv 0$ et le résultat est évident. On va donc supposer $\lambda > 0$ dans la suite. L'idée consiste à appliquer le lemme précédent avec la martingale M (qui vérifie toutes les hypothèses, et en particulier sa variation $M'_t + M''_t = N_t + \lambda t$ est intégrable), et avec le processus $H_t = (u-1)u^{N_{t-}}$, où

$u \in [0,1]$ et N_{t-} désigne la limite à gauche de N au temps t. Il est clair que $|H_t| \leq 1$, et on a aussi $H_t = \sum_{n\geq 0} \alpha_n 1_{]T_n,T_{n+1}]}(t)$ avec les T_n qui sont les temps de saut de N et $\alpha_n = (u-1)u^n$. Par suite le processus

$$Y_t = \int_0^t H_s dM_s = \int_0^t H_s dN_s - \lambda \int_0^t H_s ds$$

est une martingale.

D'autre part le processus $t \mapsto \int_0^t H_s dN_s$ est constant sur chaque intervalle $[T_n, T_{n+1}[$, et au temps T_n (avec $n \geq 1$) il saute de la quantité $\alpha_{n-1} = (u-1)u^{n-1} = u^n - u^{n-1}$. Le processus u^{N_t} a exactement les mêmes propriétés, donc ces deux processus ne diffèrent que par leurs valeurs à l'instant 0, et donc

$$u^{N_t} = 1 + \int_0^t H_s dN_s.$$

En rassemblant ces deux résultats, on voit que le processus $Z_t = u^{N_t} - 1 - \lambda \int_0^t H_s ds$ est une martingale nulle en 0. Donc en utilisant le théorème de Fubini pour les espérances conditionnelles, pour tous $s, t \geq 0$ on arrive à

$$\begin{aligned}
\mathbb{E}\left(u^{N_{t+s}-N_t}|\mathcal{F}_t\right) &= \mathbb{E}\left(1 + u^{-N_t}(u^{N_{t+s}} - u^{N_t})|\mathcal{F}_t\right) \\
&= 1 + u^{-N_t}\ \mathbb{E}\left(Z_{t+s} - Z_t + \lambda \int_t^{t+s} H_r dr|\mathcal{F}_t\right) \\
&= 1 + \lambda(u-1)\ e^{-N_t} \int_0^s \mathbb{E}\left(u^{N_{t+r}}|\mathcal{F}_t\right)\ dr \\
&= 1 + \lambda(u-1) \int_0^s \mathbb{E}\left(u^{N_{t+r}-N_t}|\mathcal{F}_t\right)\ dr. \qquad (5.1.8)
\end{aligned}$$

Cela prouve que $\phi_u(s) = \mathbb{E}(u^{N_{t+s}-N_t}|\mathcal{F}_t)$ (pour t, u fixés) vérifie l'équation intégrale

$$\phi_u(s) = 1 + \lambda(u-1) \int_0^s \phi_u(r) dr. \qquad (5.1.9)$$

Par ailleurs ϕ_u est une fonction continue à droite de s, donc nécessairement $\phi_u(s) = e^{\lambda(u-1)s}$, c'est-à-dire

$$\mathbb{E}(u^{N_{t+s}-N_t}|\mathcal{F}_t) = e^{\lambda(u-1)s}. \qquad (5.1.10)$$

Ci–dessus à droite, et en tant que fonction de u, on obtient la fonction génératrice de la loi de Poisson de paramètre λs, et comme la fonction génératrice caractérise la loi on en déduit que la loi conditionnelle de $N_{t+s} - N_t$ sachant $\mathcal{F}_t$ est la loi de Poisson de paramètre λs: cela prouve (1), et aussi la seconde partie du lemme 5.1.5.

Nous avons été trop vite, car (5.1.8) et la fonction ϕ_u elle-même (qui est aléatoire) ne sont vraies ou définies qu'à des ensembles négligeables près. Mais cela peut se justifier: on peut considérer la loi conditionnelle si $\mathcal{F}_t$ du processus $(N'_s = N_{t+s} - N_t)_{s\geq 0}$, soit $\mathbb{P}' = \mathbb{P}'(\omega; d\omega')$. Ensuite on prend la version de ϕ_u donnée par $\phi_u(s)(\omega) = \int \mathbb{P}'(\omega, d\omega') u^{N'_s(\omega')}$, qui d'une part vérifie (5.1.9) presque sûrement, pour chaque s fixé, et d'autre part est continue à droite en s: par suite elle vérifie (5.1.9) pour tout s, en dehors d'un ensemble négligeable, et on en déduit (5.1.10), donc $\phi_u(s) = e^{\lambda(u-1)s}$, en dehors d'un ensemble négligeable. Enfin les deux membres de cette égalité sont (pour s fixé) continus en u: donc on a $\phi_u(s) = e^{\lambda(u-1)s}$ pour tout $u \in [0,1]$, en dehors d'un ensemble négligeable, ce qui prouve que pour presque tout ω, sous $I\!P'(\omega,.)$ la loi de N'_s est la loi de Poisson de paramètre λs: à ce point, la démonstration est correcte et achevée. $\square$

En rassemblant tous ces lemmes, on obtient le théorème 5.1.3.

5.1.3 Trois autres caractérisations

Il existe une autre manière de considérer un processus de comptage, comme étant la “fonction de répartition” d'une mesure aléatoire sur $I\!R_+$. En effet au processus N défini sur l'espace probabilisé $(\Omega, \mathcal{F}, I\!P)$ on peut associer la mesure $\mu = \mu(\omega, dt)$ sur $I\!R_+$ par

$$\mu(\omega, dt) = \sum_{n \geq 1} 1_{\{T_n(\omega) < \infty\}} \, \varepsilon_{T_n(\omega)}(dt), \tag{5.1.11}$$

et on a $N_t(\omega) = \mu(\omega, [0,t])$. Une telle mesure (aléatoire), somme de masses de Dirac en des points T_n deux-à-deux distincts, s'appelle une “répartition ponctuelle” su $I\!R_+$ (ou plutôt, sur $]0,\infty[$, puisqu'on a exclu $T_0 = 0$ de la somme). Clairement, $\omega \mapsto \mu(\omega, A)$ est mesurable pour tout borélien A.

Les trois définitions précédentes pour les processus de Poisson sont adaptées aux processus indicés par $I\!R_+$ (ou éventuellement $I\!R$) mais n'ont aucun sens lorsque l'espace des “temps” est $I\!R^q$, ou un espace “abstrait”. Nous proposons ci–dessous trois autres définitions équivalentes aux précédentes, et qui s'étendront aux répartitions ponctuelles sur un espace plus général que $I\!R_+$. Là encore, ces définitions supplémentaires seront formulées comme des propriéés caractéristiques.

Pour formuler ce théorème nous avons encore besoin de notations. On note m la mesure de Lebesgue sur $I\!R_+$. Pour tout borélien A on note μ_A la restriction de μ à A, i.e. $\mu_A(B) = \mu(A \cap B)$. On considère alors la tribu

$$\mathcal{F}(A) = \sigma(\mu_A(B) : B \in \mathcal{R}_+) \tag{5.1.12}$$

qui est “engendrée” par la restriction μ_A ($\mathcal{R}_+$ est la tribu borélienne de $I\!R_+$). Il est clair que $\mathcal{F}([0,t]) = \sigma(N_s, s \leq t)$.

Pour toute fonction borélienne positive sur $I\!R_+$ on note $\mu(f) = \mu(f)(\omega)$ l'intégrale de f par rapport à la mesure $\mu(\omega,.)$, c'està-dire

$$\mu(f) = \sum_{n \geq 1} f(T_n) 1_{\{T_n < \infty\}}.$$

Il est donc évident que $\omega \mapsto \mu(f)(\omega)$ est mesurable.

Enfin, si $E \in \mathcal{R}_+$ vérifie $0 < m(E) < \infty$, on dit qu'une variable S est “uniformément répartie sur E” si sa loi est la restriction de m à E, normalisée par $m(E)$ (i.e., $I\!P(S \in A) = \frac{m(A \cap E)}{m(E)}$). Si $S_1, \ldots, S_m$ sont m variables indépendantes uniformément réparties sur E, on appelle “réarrangement croissant” des S_i la suite $R_1, \ldots, R_m$ de variables telles que

$R_1 \le \ldots \le R_m$ et que les deux ensembles $\{R_i : 1 \le i \le m\}$ et $\{S_i : 1 \le i \le m\}$ soient les mêmes. On voit facilement qu'il existe une "version mesurable" de ces variables, et en dehors d'un ensemble négligeable les S_i sont toutes différentes, donc $R_1 < \ldots < R_m$ p.s.

Théorème 5.1.10. *Soit N un processus de comptage sur un espace $(\Omega, \mathcal{F}, I\!\!P)$, et μ la mesure aléatoire (répartition ponctulle) associée par (5.1.11). Fixons également une partition borélienne $(E_n)_{n\ge1}$ de $I\!\!R_+$ avec $m(E_n) < \infty$ pour tout n. Il y a équivalence entre*

(1) N est un processus de Poisson de paramètre $\lambda \ge 0$.

(4) Pour toute famille $(A_i)_{1\le i\le n}$ de boréliens de $I\!\!R_+$ deux-à-deux disjoints, les variables $\mu(A_i)$ $(= \mu(\omega, A_i))$ sont indépendantes, et $I\!\!E(\mu(A)) = \lambda m(A)$ pour tout borélien A.

(5) Les tribus $\mathcal{F}(E_n)$ sont indépendantes; de plus, pour chaque n la loi de $\mu(E_n)$ est de Poisson de paramètre $\lambda m(E_n)$; enfin, conditionnellement au fait que $\mu(E_n) = m$ pour un $m \ge 1$, si on note $T_{n_1} < \ldots < T_{n_m}$ les temps successifs T_n appartenant à E_m, la loi de $(T_{n_1}, \ldots, T_{n_m})$ est la même que la loi du réarrangement croissant de m variables indépendantes uniformément réparties sur E_n.

(6) Pour toute fonction borélienne positive f l'application $\omega \mapsto \mu(f)(\omega)$ est mesurable, et on a (avec $e^{-\infty} = 0$):

$$I\!\!E(e^{-\mu(f)}) = \exp\left(-\lambda \int_0^\infty (1 - e^{-f(t)})dt\right). \tag{5.1.13}$$

Noter que dans (5), la partition (E_n) est totalement arbitraire. Cette propriété fournit une nouvelle méthode de construction d'un processus de Poisson: on prend par exemple $E_n =]n-1, n]$; puis on construit des suites $(M_n)_{n\ge1}$ et $(R_{n,m})_{n\ge1,m\ge1}$ de variables globalement indépendantes, telles que chaque M_n suive la loi de Poisson de paramètre λ et chaque $R_{n,m}$ suive la loi uniforme sur E_n. Il reste alors à poser

$$\mu = \sum_{n\ge1} \sum_{m=1}^{M_n} \varepsilon_{R_{n,m}}, \qquad \text{où une somme "vide" vaut 0,} \tag{5.1.14}$$

et $N_t = \mu(]0, t])$.

Dans (6), l'application $f \mapsto I\!\!E(e^{-\mu(f)})$ s'appelle la *fonctionnelle de Laplace* de la mesure aléatoire μ.

Là encore la preuve sera faite en une série de lemmes. Noter que si $\lambda = 0$, les quatre conditions (1), (4), (5) et (6) impliquent chacune que $N \equiv 0$, et il n'y a donc rien à montrer. Donc dans la suite on suppose toujours $\lambda > 0$.

Lemme 5.1.11. *On a (1) $\Rightarrow$ (6).*

Preuve. Soit $f_A(t) = f(t)1_{\{t\le A\}}$. D'une part $\mu(f_A)$ croit vers $\mu(f)$ quand $A \to \infty$, donc $e^{-\mu(f_A)}$ décroit vers $e^{\mu(f)}$ en restant borné par 1, donc $I\!\!E(e^{-\mu(f_A)}) \to I\!\!E(e^{-\mu(f)})$. D'autre part $\int_0^\infty (1 - e^{-f_A(t)})dt \to \int_0^\infty (1 - e^{-f(t)})dt$, toujours d'après le théorème de Lebesgue. Par suite il suffit de montrer (5.1.13) pour chaque f_A, donc en fait pour une fonction borélienne positive f vérifiant $f(t)=0$ pour tout $t>A$, où A est un réel arbitraire. Dans ce cas,

$$\mathbb{E}(e^{-\mu(f)})=\sum_{n=0}^{\infty}a_n, \qquad \text{où} \quad a_n=\mathbb{E}\left(e^{-\sum_{j=1}^n f(T_j)}\ 1_{\{N_A=n\}}\right).$$

D'après la propriété (3) du théorème 5.1.3, il vient

$$\begin{aligned} a_n &= \int_{\mathbb{R}_+^{n+1}}\left(\prod_{i=0}^n \lambda\ e^{-\lambda s_i}\right)\ e^{-\sum_{j=1}^n f(s_0+\ldots+s_{j-1})}\ 1_{\{s_0+\ldots+s_{n-1}\le A<s_0+\ldots+s_n\}}\ ds_0\ldots ds_n \\ &= \lambda^n\ e^{-\lambda A}\int_{\mathbb{R}_+^n} e^{-\sum_{j=1}^n f(s_0+\ldots+s_{j-1})}\ 1_{\{s_0+\ldots+s_{n-1}\le A\}}\ ds_0\ldots ds_{n-1} \\ &= \lambda^n e^{-\lambda A}\alpha, \qquad \text{où} \quad \alpha=\int_{\mathbb{R}_+^n} e^{-\sum_{j=1}^n f(\ t_j)}\ 1_{\{t_1<t_2<\ldots<t_n\le A\}}\ dt_1\ldots dt_n, \end{aligned}$$

où la seconde égalité est obtenue en intégrant $s_n\mapsto \lambda e^{-\lambda s_n}$ sur l'ensemble $\{t-(s_0+\ldots+s_{n-1}),\infty[$, et la seconde en faisant les changements de variables $t_i=s_0+\ldots+s_{i-1}$.

Soit $\mathcal{S}_n$ l'ensemble des permutations de $\{1,\ldots,n\}$. Si $\sigma\in\mathcal{S}_n$, on a $\sum_{i=1}^n f(t_{\sigma(i)})=\sum_{i=1}^n f(t_i)$. Par suite en faisant les changements de variables $t'_i=t_{\sigma(i)}$ dans l'intégrale définissant α, on voit qu'on a

$$\alpha=\int_{\mathbb{R}_+^n} e^{-\sum_{j=1}^n f(\ t_j)}\ 1_{\{t_{\sigma(1)}<t_{\sigma(2)}<\ldots<t_{\sigma(n)}\le A\}}\ dt_1\ldots dt_n.$$

Comme $\cup_{\sigma\in\mathcal{S}_n}\{(t_1,\ldots,t_n): t_{\sigma(1)}<t_{\sigma(2)}<\ldots<t_{\sigma(n)}\le A\}=]0,A]^n$ à un ensemble de mesure nulle près pour $dt_1\ldots dt_n$, on en déduit que

$$\alpha=\frac{1}{n!}\int_{]0,A]^n} e^{-\sum_{i=1} f(t_i)}\ dt_1\ldots dt_n=\frac{1}{n!}\ \left(\int_0^A e^{-f(t)}dt\right)^n.$$

Il vient alors

$$\mathbb{E}(e^{-\mu(f)})=\sum_{n=0}^{\infty}\frac{\lambda^n}{n!}\ e^{-\lambda A}\ \left(\int_0^A e^{-f(t)}dt\right)^n=\exp\left(-\lambda A+\lambda\int_0^A e^{-f(t)}dt\right),$$

et comme $f(t)=0$, donc $e^{-f(t)}=1$, si $t>A$, on obtient (5.1.13). □

Lemme 5.1.12. *(6) implique (4) et aussi que pour tout borélien A la variable $\mu(A)$ suit une loi de Poisson de paramètre $\lambda m(A)$ si $m(A)<\infty$ et est p.s. infinie si $m(A)=\infty$.*

Preuve. Soit A_i des boréliens disjoint, et $u_i\ge 0$. Une application de (5.1.13) à la fonction $f(t)=\sum_{i=1}^n u_i 1_{A_i}(t)$ donne

$$\mathbb{E}\left(\prod_{i=1}^n e^{-u_i\mu(A_i)}\right)=\prod_{i=1}^n \exp\left(-\lambda\mu(A_i)(1-e^{-u_i})\right).$$

Cela prouve d'abord que la transformée de Laplace multi–dimensionnelle de la variable vectorielle $(\mu(A_i))$ est le produit des transformées de Laplace uni–dimensionnelles, donc les variables $\mu(A_i)$ sont indépendantes. Cela prouve aussi que $I\!E(e^{-u\mu(A)}) = e^{-\lambda\mu(A)(1-e^{-u})}$ pour tout $u \geq 0$, d'où la dernière assertion: lorsque $m(A) < \infty$, car la transformée de Laplace de la loi de Poisson de paramètre θ est $u \mapsto \exp(-\theta(1-e^{-u}))$; lorsque $m(A) = 0$, car on a alors $I\!E(e^{-u\mu(A)}) = 0$, donc $e^{-u\mu(A)} = 0$ p.s., donc $\mu(A) = \infty$ p.s.

En particulier, on a $I\!E(\mu(A)) = \lambda m(A)$, ce qui achève de prouver (4). □

Lemme 5.1.13. *On a (4) ⇒ (1).*

Preuve. La propriété (4) implique que pour tous $s, t \geq 0$ la variable $N_{t+s} - N_t = \mu(]t, t+s])$ est indépendante des variables $N_r = \mu(]0, r])$ pour $r \leq t$, donc de la tribu $\mathcal{F}_t = \sigma(N_r : r \leq t)$, et par ailleurs elle est d'espérance λs. Donc $M_r = N_r - \lambda r$ vérifie

$$I\!E(M_{t+s} - M_t | \mathcal{F}_t) = I\!E(N_{t+s} - N_t | \mathcal{F}_t) - \lambda s = I\!E(N_{t+s} - N_t) - \lambda s = 0.$$

Donc M est une martingale, et il suffit d'appliquer l'implication (2) ⇒ (1) du théorème 5.1.3. □

Lemme 5.1.14. *On a (1) ⇔ (5).*

Preuve. On a vu que (1) ⇔ (6). Par ailleurs la propriété (5) est une description de la construction de la mesure μ, comme nous l'avons vu en (5.1.14). Il suffit donc de montrer que si on construit μ par (5.1.14), alors cette mesure vérifie (6). Soit f une fonction borélienne positive. A cause de l'indépendance des variables $(M_n, R_{n,m})$ et au vu de leurs lois, il est clair que

$$\begin{aligned}
I\!E(e^{-\mu(f)}) &= I\!E\left(\prod_{n\geq 1}\prod_{m=1}^{M_n} e^{-f(R_{n,m})}\right) \\
&= \prod_{n\geq 1}\left(I\!P(M_n = 0) + \sum_{j\geq 1} I\!P(M_n = j) I\!E\left(\prod_{m=1}^{j} e^{-f(R_{n,m})}\right)\right) \\
&= \prod_{n\geq 1}\left(I\!P(M_n = 0) + \sum_{j\geq 1} I\!P(M_n = j)\prod_{m=1}^{j} I\!E\left(e^{-f(R_{n,m})}\right)\right) \\
&= \prod_{n\geq 1}\sum_{j\geq 0}\frac{\lambda^j m(E_n)^j}{j!} e^{-\lambda m(E_n)}\left(\frac{1}{m(E_n)}\int_{E_n} e^{-f(t)}dt\right)^j \\
&= \prod_{n\geq 1} e^{-\lambda\int_{E_n}(1-e^{-f(t)})dt} = e^{-\lambda\int_0^\infty(1-e^{-f(t)})dt},
\end{aligned}$$

d'où le résultat. □

Ce lemme achève la preuve du théorème 5.1.10.

5.2 Répartitions ponctuelles de Poisson

Nous allons maintenant considérer la généralisation des processus de Poisson au cas où l'espace des "temps" est remplacé par un espace abstrait. Les bonnes définitions seront alors des généralisations des propriétés (4), (5) ou (6) du théorème 5.1.10.

On considère un espace mesurable $(E, \mathcal{E})$, et une mesure aléatoire $\mu = \mu(\omega, dx)$ qui est la somme d'un nombre fini ou infini dénombrable (éventuellement aléatoire) de masses de Dirac. Une telle mesure s'appelle une *répartition ponctuelle* et peut évidemment s'écrire

$$\mu(\omega, dx) = \sum_{n \geq 1}^{M(\omega)} \varepsilon_{T_n}(\omega), \tag{5.2.1}$$

où M est à valeurs dans $\overline{I\!N}$, et les T_n sont à valeurs dans E. Si de plus les points sont tous distincts, i.e. $T_n(\omega) \neq T_m(\omega)$ pour tous $1 \leq n < m \leq M(\omega)$, on dit que la répartition ponctuelle est *simple*.

Il faut préciser maintenant les propriétés de mesurabilité: on suppose que M et les T_n sont $\mathcal{F}$–mesurables. Dans ce cas,

$$\mu(f) = \sum_{n=1}^{M} f(T_n) \tag{5.2.2}$$

est aussi $\mathcal{F}$–mesurable pour toute fonction positive mesurable f sur $(E, \mathcal{E})$, de sorte que μ est une mesure de transition de $(\Omega, \mathcal{F})$ dans $(E, \mathcal{E})$. On peut alors définir sa *fonctionnelle de Laplace* par

$$f \ \mapsto \ I\!E(e^{-\mu(f)}) \qquad \text{pour } f \text{ mesurable positive sur } (E, \mathcal{E}). \tag{5.2.3}$$

et sa *mesure intensité* (qui est une mesure sur $(E, \mathcal{E})$) par

$$A \ \mapsto \ I\!E(\mu(A)) \qquad \forall A \in \mathcal{E}. \tag{5.2.4}$$

Remarque 1: On pourrait à l'inverse partir de μ, supposée être d'une part une répartition ponctuelle, d'autre part une mesure de transition de $(\Omega, \mathcal{F})$ dans $(E, \mathcal{E})$. Peut-on écrire μ comme (5.2.1) avec M et les T_n mesurables ? C'est évident pour M, qui vaut $\mu(E)$, mais pour les T_n une première difficulté tient à ce que chaque T_n n'est défini que sur l'ensemble où $M \geq n$; et même si $M(\omega) = \infty$ pour tout ω il reste difficile de prouver qu'on peut trouver des T_n mesurables, et c'est même faux si on ne fait pas des hypothèses supplémentaires sur $\mathcal{E}$ (à savoir, que $\mathcal{E}$ est séparable et contient les singletons $\{x\}$). Ici, nous supposerons *a priori* la mesurabilité des T_n. $\square$

Définition 5.2.1. Une répartition ponctuelle μ est dite

(i) *complètement aléatoire* si pour tous $A_i \in \mathcal{E}$ deux–à–deux disjoints les variables $(\mu(A_i))$ sont indépendantes,

(ii) *de Poisson* si sa fonctionnelle de Laplace s'écrit, pour une mesure ν sur $(E, \mathcal{E})$:

$$I\!E(e^{-\mu(f)}) = \exp\left(-\int_E (1 - e^{-f(x)})\nu(dx)\right). \tag{5.2.5}$$

Théorème 5.2.2. *Si μ est une répartition ponctuelle de Poisson associée à la mesure ν dans (5.2.5), alors elle est complètement aléatoire. De plus pour tout $A \in \mathcal{E}$ la variable $\mu(A)$ est p.s. infinie si $\nu(A) = \infty$, et suit une loi de Poisson de paramètre $\nu(A)$ si $\nu(A) < \infty$, et ν est aussi la mesure intensité de μ.*

Preuve. La preuve est exactement la même que celle du lemme 5.1.12 (où λm joue le rôle de ν). □

Ainsi, un processus de Poisson de paramètre λ est une répartition ponctuelle de Poisson sur $I\!R_+$, de mesure intensité $\nu = \lambda m$. Au vu du théorème 5.1.10 on peut penser que la réciproque est vraie aussi, et que tout cela est équivalent à la construction décrite dans la propriété (5) et adaptée à la situation présente. Voici une formalisation de cette constrution:

Propriété (5'): La mesure μ satisfait cette propriété si elle s'écrit comme dans (5.1.14), c'est–à–dire $\mu = \sum_{n\geq 1}\sum_{m=1}^{M_n} \varepsilon_{R_{n,m}}$, où

(i) les variables M_n et $R_{m,n}$ sont toutes indépendantes;

(ii) les variables M_n suivent des lois de Poisson de paramètres $\nu(E_n)$, pour une certaine mesure ν sur $(E,\mathcal{E})$ et une partition $\mathcal{E}$–mesurable $(E_n)_{n\geq 1}$ de E avec $\nu(E_n) < \infty$;

(iii) les variables $R_{n,m}$ sont à valeurs dans E_n et de loi ν_n, où on a posé $\nu_n(A) = \frac{\nu(E_n\cap A)}{\nu(E_n)}$. □

Cette propriété dépend *a priori* de la partition choisie. Elle implique évidemment que ν est σ–finie. Le résultat suivant montre que dans ce cas, si la propriété est vraie pour une partition (E_n), elle est vraie pour toute autre partition (E'_n) satisfaisant bien–sûr $\nu(E'_n) < \infty$.

Théorème 5.2.3. *Si la répartition ponctuelle μ vérifie (5') avec une cetaine mesure σ–finie ν sur $(E,\mathcal{E})$, elle est de Poisson et ν est sa mesure intensité.*

Réciproquement si μ est une répartition ponctuelle de Poisson de mesure intensité ν σ–finie, et si $(E,\mathcal{E})$ est un espace polonais, on a (5') pour toute partition (E_n) vérifiant $\nu(E_n) < \infty$.

Preuve. La preuve de la partie directe est ici encore exactement la même que la preuve de (5) $\Rightarrow$ (6) dans le lemme 5.1.14, en remplaçant partout la mesure λm par la mesure ν.

La réciproque est plus délicate, et en rapport avec la remarque 1. On fixe une partition arbitraire vérifiant les conditions requises, et on pose $M_n = \mu(E_n)$. En fait, si on définit les tribus $\mathcal{F}(A)$ comme dans (5.1.12) (pour $A \in \mathcal{E}$), (5') revient à dire que

a) les M_n sont de lois de Poisson de paramètres $\nu(E_n)$ et les tribus $\mathcal{F}(E_n)$ sont indépendantes,

b) et, conditionnellement à $\{M_n = p\}$, on peut trouver des variables $R_{n,1},\dots,R_{n,p}$ indépendantes et de loi ν_n, telles que la restriction de μ à E_n s'écrive $\mu_n = \sum_{m=1}^{p} \varepsilon_{R_{n,m}}$.

Il est clair que le théorème 5.2.2 entraine (a). Mais (b) n'est pas évident. On résoud la difficulté en notant que la donnée des $R_{n,m}$ (ordonnés de manière arbitraire) satisfaisant $\mu_n = \sum_{m=1}^{p} \varepsilon_{R_{n,m}}$ est équivalente à la donnée des $\mu(A)$ pour tout $A \subset E_n$, donc si μ et μ' sont deux répartitions ponctuelles telles que les lois de $(\mu(A_1),\dots,\mu(A_q))$ et $(\mu'(A_1),\dots,\mu'(A_q))$ coïncident pour tout choix de q et des $A_i \subset E_n$, alors si μ' vérifie la propriété (b) ci–dessus il en est de même de μ.

Mais la loi de $(\mu(A_1),\ldots,\mu(A_q))$ est entièrement caractérisée par sa transformée de Laplace q-dimensionnelle

$$(u_1,\ldots,u_q) \;\mapsto\; \mathbb{E}(\exp(-\sum_{i=1}^q u_i\mu(A_i)))$$

(pour $u_i \geq 0$), qui est en fait la fonctionnelle de Laplace pour la fonction $\sum_{i=1}^q u_i 1_{A_i}$. En d'autres termes la propriété (5') *ne dépend que de la fonctionnelle de Laplace.* Par suite la réciproque découle de la partie directe.

(*Remarque:* le fait que $(E,\mathcal{E})$ est polonais est utilisé – de manière implicite – quand on écrit que la donnée des $R_{n,m}$ équivaut à celle des $\mu(A)$; le lecteur vérifiera que, quand $E = \mathbb{R}_+$, ce résultat est essentiellement trivial, dû aux faits que les T_n sont bien définis partout et qu'on peut utiliser la "fonction de répartition" $N_t = \mu(]0,t])$: cf. lemme 3.2.9).) $\square$

En ce qui concerne la réciproque du théorème 5.2.2, elle n'est pas toujours vraie non plus: il faut des hypothèses supplémentaires.

Théorème 5.2.4. *Supposons que $(E,\mathcal{E})$ soit un espace polonais muni de ses boréliens. Soit μ une répartition ponctuelle* simple *et complètement aléatoire, dont la mesure intensité ν est σ-finie et n'a pas d'atome (i.e. $\nu(\{x\}) = 0$ pour tout $x \in E$). Alors, μ est une répartition de Poisson.*

L'hypothèse sur E est de nature technique. En revanche le fait que ν n'ait pas d'atome est primordial (si $a \in E$, la mesure $\mu(\omega,dx) = \varepsilon_a(dx)$ – déterministe – est complètement aléatoire mais n'est pas de Poisson car $\mu(E)$ n'est ni infinie ni de loi de Poisson). La "simplicité" de μ est aussi essentielle (si μ est une répartition de Poisson, la mesure 2μ est aussi une répartition ponctuelle complètement aléatoire, mais pas de Poisson).

Preuve. i) On va d'abord montrer qu'il suffit de prouver le résultat quand ν est une mesure finie.

Supposons le résultat vrai pour toute répartition simple complètement aléatoire de mesure intensité finie et sans atome. Soit une partition mesurable (E_n) de E telle que $\nu(E_n) < \infty$, et μ_n la répartition ponctuelle définie par $\mu_n(A) = \mu(A\cap E_n)$, dont la mesure intensité ν_n est clairement $\nu_n(A) = \nu(A\cap E_n)$. La mesure ν_n est finie et sans atome, et la répartition μ_n est simple et évidemment complètement aléatoire, donc on a

$$\mathbb{E}(e^{-\mu_n(f)}) = \exp\left(-\int(1-e^{-f(x)})\nu_n(dx)\right)$$

pour toute f borélienne positive. Mais l'hypothèse sur μ entraine que les $\mu_n(f)$ sont indépendantes si f est étagée, donc pour f positive mesurable quelconque par le théorème de limite monotone. Donc

$$\mathbb{E}(e^{-\mu(f)}) = \prod_n \exp\left(-\int(1-e^{-f(x)})\nu_n(dx)\right) = \exp\left(-\int(1-e^{-f(x)})\nu(dx)\right)$$

et donc μ est de Poisson.

ii) On suppose donc dans la suite que ν est une mesure finie. Soit $A \in \mathcal{E}$. Nous allons montrer que $\mu(A)$ suit une loi de Poisson de paramètre $\nu(A)$.

L'espace E est muni d'une distance d et contient une suite $(x_k)_{k\geq 1}$ qui est dense. Si $B(x,\varepsilon) = \{y : d(x,y) < \varepsilon\}$ on a $\nu(B(x,\varepsilon)) \to 0$ lorsque $\varepsilon > 0$ puisque $\nu(\{x\}) = 0$, donc $\varepsilon_{n,k} = (1/n) \bigwedge \sup(\varepsilon : \nu(B(x_k,\varepsilon) \leq 1/n)$ vérifie $\varepsilon_{n,k} > 0$ et $\nu(B(x_k, \varepsilon(n,k)) \leq 1/n$. Pour tout x il existe une suite x_{n_m} tendant vers x, et on a $\nu(B(x,\varepsilon)) \leq 1/n$ pour un $\varepsilon \in]0, 1/2n[$. Si $d(x, x_{n_m}) \leq \varepsilon/2$, on a donc $B(x_{n_m}, \varepsilon/2) \subset B(x,\varepsilon)$, de sorte que $\varepsilon_{n,n_m} > \varepsilon/2$, et donc $x \in B(x_{n_m}, \varepsilon_{n,n_m})$. Par suite on voit que $\cup_{k\geq 1} B(x_k, \varepsilon_{n,k}) = E$.

On pose alors $A_{n,1} = A \cap B(x_0, \varepsilon_{n,0})$ et, par récurrence sur k, $A_{n,k} = A \cap B(x_k, \varepsilon_{n,k}) \cap (\cup_{1\leq i\leq k-1} A_{n,i})^c$ pour $k \geq 2$. On obtient ainsi une partition $(A_{n,i})_{i\geq 1}$ de A, chaque $A_{n,i}$ étant de diamètre $\leq 2/n$, mesurable, et avec $\nu(A_{n,i}) \leq 1/n$.

Avec la notation (5.2.1), on pose

$$\Delta = \inf(d(T_n, T_m) : 1 \leq n < m \leq M).$$

Comme $I\!\!E(M) = I\!\!E(\mu(E)) = \nu(E) < \infty$, on a $M < \infty$ p.s., et la répartition étant simple on a aussi $T_n \neq T_m$ si $1 \leq n < m \leq M$. Par suite $\Delta > 0$ p.s. et, pour tout $u \geq 0$, on a

$$I\!\!E\left(e^{-u\mu(A)}\ 1_{\{\Delta > 2/n\}}\right) \ \to\ I\!\!E(e^{-u\mu(A)}) \tag{5.2.6}$$

si $n \to \infty$. Si $\Delta > 2/n$, chaque ensemble $A_{n,i}$ contient au plus un seul point T_k, puisqu'il est de diamètre $\leq 2/n$, et on a $\mu(A_{n,i}) = 1 \bigwedge \mu(A_{n,i})$, tandis que $\mu(A) = \sum_{i\geq 1} \mu(A_{n,i})$. Par suite

$$\alpha_n := I\!\!E\left(e^{-u\mu(A)}\ 1_{\{\Delta>2/n\}}\right) = I\!\!E\left(e^{-u\sum_{i\geq 1} 1 \bigwedge \mu(A_{n,i})}\ 1_{\{\Delta>2/n\}}\right),$$

et comme $\Delta > 0$ p.s., on obtient aussi lorsque $n \to \infty$:

$$\left|\alpha_n - I\!\!E\left(e^{-u\sum_{i\geq 1} 1 \bigwedge \mu(A_{n,i})}\right)\right| \leq I\!\!P(\Delta > \frac{2}{n}) \to 0. \tag{5.2.7}$$

Comme μ est complètement aléatoire, les variables $1 \bigwedge \mu(A_{n,i})$ sont indépendantes pour $i = 1, \ldots$, et prennent la valeur 1 avec la probabilité $\delta_i^n = I\!\!P(\mu(A_{n,i}) \geq 1) \leq I\!\!E(\mu(A_{n,i})) = \nu(A_{n,i}) \leq \frac{1}{n}$, et 0 avec la probabilité $1 - \delta_i^n$. Par suite

$$\beta_n := I\!\!E\left(e^{-u\sum_{i\geq 1} 1 \bigwedge \mu(A_{n,i})}\right) = \prod_{i\geq 1}\left(1 - \delta_i^n(1 - e^{-u})\right).$$

Si $0 \leq x \leq 1/n$ on a $-x(1+1/n) \leq \log(1-x) \leq -x$ si $n \geq 2$, donc comme $\delta_i^n(1-e^{-u}) \leq 1/n$ il vient

$$\exp\left(-(1+\frac{2}{n})(1-e^{-u})\sum_i \delta_i^n\right) \leq \beta_n \leq \exp\left(-(1-e^{-u})\sum_i \delta_i^n\right).$$

En combinant ceci avec (5.2.6) et (5.2.7), on obtient que $\sum_i \delta_i^n$ converge vers une limite finie δ et que $I\!\!E(e^{-\mu(A)}) = \exp\left(-(1-e^{-u})\delta\right)$. Ceci étant vrai pour tout $u \geq 0$, on en déduit que $\mu(A)$ suit une loi de Poisson de paramètre δ, et comme de plus $\mathbb{E}(\mu(A)) = \nu(A)$ on a nécessairement $\delta = \nu(A)$.

iii) Si f est mesurable positive étagée, de la forme $f = \sum_{i=1}^m u_i 1_{A_i}$ pour des $A_i \in \mathcal{E}$ deux–à–deux disjoints et $u_i \geq 0$, les variables $\mu(A_i)$ sont indépendantes (puisque μ est complètement aléatoire) et de lois de Poisson de paramètres $\nu(A_i)$ d'après ii), donc

$$\begin{aligned} \mathbb{E}(e^{-\mu(f)}) &= \prod_{i=1}^m \mathbb{E}\left(e^{-u_i\mu(A_i)}\right) = \prod_{i=1}^m e^{-(1-e^{-u_i})\nu(A_i)} \\ &= e^{-\sum_{i=1}^m \int (1-e^{-f(x)})1_{A_i}(x)\nu(dx)} = e^{-\int (1-e^{-f(x)})\nu(dx)}. \end{aligned}$$

Toute fonction mesurable positive f étant limite croissante de fonctions étagées, Par limite monotone les deux membres extrêmes ci–dessus sont égaux pour f, et on a le résultat. $\square$

5.3 Processus de Poisson composé

Définition 5.3.1. On appelle $(\mathcal{F}_t)$*–processus de Poisson composé* un processus défini sur un espace filtré $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$, à valeurs dans $\mathbb{R}^d$, qui est un $(\mathcal{F}_t)$–processus de Lévy et dont les trajectoires $t \mapsto X_t(\omega)$ sont constantes par morceaux. Si $\mathcal{F}_t = \sigma(X_s : s \leq t)$ on dit simplement que X est un *processus de Poisson composé.*

On associe à $\mathbf{X}$ ses temps de saut successifs T_n

$$T_0 = 0, \quad T_{n+1} = \begin{cases} \inf(t : t > T_n, X_t \neq X_{T_n}) & \text{si } T_n < \infty \\ \infty & \text{si } T_n = \infty, \end{cases} \tag{5.3.1}$$

(T_0 n'est pas un temps de saut; comme d'habitude $T_\infty = \lim_n T_n$) et les tailles Y_n des sauts correspondants, c'est–à–dire

$$Y_n = \begin{cases} X_{T_n} - X_{T_n-} = X_{T_n} - X_{T_{n-1}} & \text{si } T_n < \infty \\ 0 & \text{si } T_n = \infty, \end{cases} \tag{5.3.2}$$

Par ailleurs, on dit qu'un processus (X_t) vérifie la *propriété A* si on peut l'écrire

$$X_t = \sum_{n \geq 1} Z_n 1_{\{U_n \leq t\}}, \tag{5.3.3}$$

où les $(U_n)_{n\geq 1}$ sont les points d'un processus de Poisson de paramètre $\lambda \geq 0$ et où les variables Z_n sont indépendantes, de même loi G et indépendantes de la suite (T_n)

Théorème 5.3.2. *(i) Si le processus (X_t) vérifie la propriété A avec λ et G, c'est un processus de Poisson composé. De plus la fonction caractéristique de la variable X_t est*

$$\mathbb{E}(e^{i<u,X_t>}) = \exp -\lambda t \int (1 - e^{i<u,x>})G(dx), \tag{5.3.4}$$

et la répartition ponctuelle $\mu = \sum_{n\geq 1} 1_{\{U_n<\infty\}}\ \varepsilon_{(U_n,Z_n)}$ *est de Poisson (sur* $I\!R_+ \times I\!R^d$*), simple, et de mesure intensité* $\nu(dt,dx) = \lambda dt \otimes G(dx)$*.*

(ii) Si le processus (X_t) *est de Poisson composé, il vérifie la propriété A avec* $U_n = T_n$ *et* $Z_n = Y_n$*. Dans ce cas on a* $G(\{0\}) = 0$ *si* $\lambda > 0$*, et* $G = \varepsilon_0$ *si* $\lambda = 0$*.*

Le lecteur remarquera la différence entre (i) et (ii): si on a (i), les U_n ne sont pas forcément des temps de saut du processus X, puisqu'on peut avoir $Z_1 = 0$ avec une probabilité positive. On peut donc "réaliser" le même processus de Poisson composé sous la forme (5.3.3) de différentes manières. Lorsque $T_n = U_n$ et $Z_n = Y_n$, on dit qu'on a la *réalisation minimale.* Le premier temps de saut, par exemple, du processus défini par (5.3.3) est $T_1 = \inf(U_n : n \geq 1, Z_n \neq 0)$.

Lorsque le processus de Poisson composé est à valeurs réelles et a des sauts d'amplitude 1, i.e. si $Y_n = 1$ sur $\{T_n < \infty\}$, c'est un processus de Poisson: le théorème précédent découle alors des théorèmes 5.1.3 et 5.1.10.

Noter que pour un processus de Poisson composé avec sa réalisation minimale, soit on a $\lambda = 0$ et alors $X \equiv 0$, soit on a $\lambda > 0$ et $T_n < \infty$ pour tout n, et $T_\infty = \infty$: les variables Y_n sont alors bien définies partout (ou au moins presque partout).

La démonstration du théorème suit encore d'une série de lemmes simples.

Lemme 5.3.3. *Si* X *est un processus de Poisson composé, les variables* T_1 *et* Y_1 *sont indépendantes, et* T_1 *est de loi exponentialle de paramètre* $\lambda \geq 0$ *(rappelons que si* $\lambda = 0$*, cela signifie que* $T_1 = \infty$ *p.s.). De plus si* $\lambda > 0$ *on a* $G(\{0\}) = 0$*, et si* $\lambda = 0$ *on a* $G = \varepsilon_0$*, où* G *désigne la loi de* Y_1*.*

Preuve. La preuve est essentiellement la même que celle du théorème 4.1.3-(1). On peut toujours supposer que l'espace Ω est muni de translations (θ_t) vérifiant (3.2.9), donc $T_1 = t + T_1 \circ \theta_t$ et $Y_1 = Y_1 \circ \theta_t$ sur l'ensemble $\{T_1 > t\}$. Par suite d'après (3.2.10) on a pour $A \in \mathcal{R}^d$ et $s, t \geq 0$:

$$I\!P(T_1 > t+s, Y_1 \in A) = I\!P(T_1 > t, T_1\circ\theta_t > s, Y_1\circ\theta_t \in A) = I\!P(T_1 > t)\ I\!P(T_1 > s, Y_1 \in A).$$

Si $s = 0$ on obtient $I\!P(T_1 > t, Y_1 \in A) = I\!P(T_1 > t)I\!P(Y_1 \in A)$ (puisque $T_1 > 0$ identiquement), donc T_1 et Y_1 sont indépendants. Si $A = I\!R^d$ on obtient $I\!P(T_1 > t+s) = I\!P(T_1 > t)I\!P(T_1 > s)$, donc T_1 suit une loi exponentielle dont on note λ le paramètre. Soit aussi G la loi de Y_1.

Enfin si $\lambda = 0$ on a $T_1 = \infty$ p.s., donc $Y_1 = 0$ p.s. (cf. (5.3.2)). Si $\lambda > 0$ on a $T_1 < \infty$ p.s., donc $Y_1 \neq 0$ p.s. et $G(\{0\}) = 0$. □

Lemme 5.3.4. *On a la partie (ii) du théorème 5.3.2.*

Preuve. Reprenons les notations du lemme précédent. Si $\lambda = 0$, il n'y a rien à montrer puisque $T_n = \infty$ et $Y_n = 0$ pour tout $n \geq 1$. On suppose donc $\lambda > 0$, donc $T_1 < \infty$.

On a $T_{n+1} - T_n = T_1 \circ \theta_{T_n}$ et $Y_{n+1} = Y_1 \circ \theta_{T_n}$ sur $\{T_n < \infty\}$. D'après la propriété (3.2.10) on a donc que $(T_{n+1} - T_n, Y_{n+1})$ a même loi que (T_1, Y_1), conditionnellement par rapport à $\sigma(T_i, Y_i : i \leq n)$ et en restriction à $\{T_n < \infty\}$. On en déduit d'abord par récurrence sur n que $T_n < \infty$ p.s. pour tout n, puis que la suite $(T_n)_{n\geq 1}$ et les variables Y_i

sont toutes indépendantes, que les Y_i ont toutes même loi G, et que les (T_n) sont les temps de saut d'un processus de Poisson de paramètre λ (appliquer (3) $\Rightarrow$ (1) dans le théorème 5.1.3). $\square$

Lemme 5.3.5. *On a la partie (i) du théorème 5.3.2.*

Preuve. Soit X vérifiant la propriété A avec λ et G, et soit la répartition ponctuelle $\mu = \sum_{n\geq 1} 1_{\{U_n<\infty\}}\ \varepsilon_{(U_n,Z_n)}$. Cette répartition est évidemment simple. On va voir qu'elle satisfait la propriété (5') avec $\nu(dt,dx) = \lambda dt \otimes G(dx)$. On prend la partition de $I\!\!R_+ \times I\!\!R^d$ constituée des $E_n = [n-1, n[\times I\!\!R^d$, et $\nu_n(A) = \frac{\nu(A\cap E_n)}{\nu(E_n)}$ (on a $\nu(E_n) = \lambda$ ici).

D'après le théorème 5.1.10 (ou le théorème 5.2.3 appliqué à la répartition ponctuelle de Poisson $\sum_{n\geq 1} 1_{\{U_n<\infty\}}\varepsilon_{U_n}$ sur $I\!\!R_+$), les U_n peuvent être construites ainsi: on se donne des variables indépendantes M_n de loi de Poisson de paramètre λ, et des variables $(V_{n,m})_{n,m\geq 1}$ indépendantes entre elles et des M_n, telles que chaque $V(n,m)$ soit uniformément répartie sur $]n-1, n]$. On a alors

$$\sum_{n\geq 1} 1_{\{U_n<\infty\}}\varepsilon_{U_n} = \sum_{n\geq 1}\sum_{m=1}^{M_n} \varepsilon_{V_{n,m}}.$$

Ensuite on considère une double suite $(Z'_{n,m})$ de variables de loi G, indépendantes entre elles et indépendantes des M_n et des $V_{n,m}$. Une manière de construire la mesure μ consiste alors à poser

$$\mu = \sum_{n\geq 1}\sum_{m=1}^{M_n} \varepsilon_{(V_{n,m},Z'_{n,m})}.$$

Il est clair que les variables $((V_{n,m}, Z'_{n,m}))_{n,m\geq 1}$ sont indépendantes entre elles et indépendantes des M_n, et chaque variable $(V_{n,m}, Z'_{n,m})$ est de loi ν_n. Cela démontre que μ satisfait la propriété (5') avec la mesure ν, et par suite μ est de Poisson d'intensité ν. Enfin, si $u \in I\!\!R^s$ on a

$$\begin{aligned}
I\!\!E(e^{i<u,X_t>}) &= \sum_{n\geq 0} I\!\!E\left(e^{i\sum_{m=1}^n <u,Y_m>}\ 1_{\{U_n\leq t<U_{n+1}\}}\right) \\
&= \sum_{n\geq 0}\left(\int e^{i<u,x}G(dx)\right)^n\ I\!\!P(U_n \leq t < U_{n+1}) \\
&= \sum_{n\geq 0}\left(\int e^{i<u,x}G(dx)\right)^n\ \frac{\lambda^n t^n}{n!}\ e^{-\lambda t} \\
&= e^{-\lambda t\int (1-e^{i<u,x>})G(dx)},
\end{aligned}$$

d'où (5.3.4). $\square$

Ces trois lemmes nous donnent le théorème 5.3.1.

Il existe aussi une caractérisation des processus de Poisson composé analogue à la propriété (2) du théorème 5.1.3 pour les processus de Poisson. Soit (X_t) un processus sur $(\Omega, \mathcal{F}, (\mathcal{F}_t), I\!\!P)$ à valeurs dans $I\!\!R^d$, continu à droite et dont les trajectoires sont constantes par morceaux. On lui associe les T_n et les Y_n par (5.3.1) et (5.3.2). Pour tout $A \in \mathcal{R}^d$ on pose

$$N_t^A = \sum_{n \geq 1} 1_A(Y_n)\ 1_{\{T_n \leq t\}}. \tag{5.3.5}$$

Proposition 5.3.6. *Si (X_t) est un $(\mathcal{F}_t)$–processus de Poisson composé, auquel on associe λ et G comme dans le théorème 5.3.2-(ii), alors pour tout $A \in \mathcal{R}^d$ le processus $M_t^A = N_t^A - \lambda G(A)t$ est une $(\mathcal{F}_t)$–martingale.*

Preuve. Encore une fois, on peut supposer qu'il existe des translations (θ_t) vérifiant (3.2.9). Le processus N^A est un processus de comptage qui vérifie $N_{t+s}^A - N_t^A = N_s^A \circ \theta_t$ et qui est clairement adapté à la filtration $(\mathcal{F}_t)$. Donc le théorème 5.1.3 implique que c'est un $(\mathcal{F}_t)$–processus de Poisson dont on note λ_A le paramètre, et $N_t^A - \lambda_A t$ est une martingale. Il suffit donc de vérifier que $\lambda_A = \lambda G(A)$. Mais $\lambda_A = \mathbb{E}(N_1^A)$, et si μ est la répartition ponctuelle $\mu = \sum_{n \geq 1} 1_{\{T_n < \infty\}}\ \varepsilon_{(T_n, Y_n)}$ on a $N_1^A = \mu(]0,1] \times A)$, qui est d'espérance $\nu(]0,1] \times A) = \lambda G(A)$. □

En fait la propriété précédente caractérise les processus de Poisson composé, et on a le résultat suivant, qui se montre comme le lemme 5.1.9 (en un peu plus compliqué):

Théorème 5.3.7. *Soit (X_t) un processus adapté sur $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ dont les trajectoires sont nulles en 0, continues à droite et constantes par morceaux, et N^A les processus associés par (5.3.5). S'il existe $\lambda \geq 0$ et une probabilité G sur $\mathbb{R}^d$ tels que pour tout $A \in \mathcal{R}^d$ le processus $N_t^A - \lambda G(A)t$ soit une $(\mathcal{F}_t)$–martingale, alors le processus (X_t) est un $(\mathcal{F}_t)$–processus de Poisson composé.*