

MAT2717 – Processus stochastiques

Élise Davignon

Email address: `elise.davignon@umontreal.ca`

RÉSUMÉ. Ces notes de cours accompagnent le cours MAT2717 – Processus stochastiques, enseigné au département de Mathématiques et Statistiques de la faculté des arts et sciences de l'Université de Montréal.

Le cours de processus stochastiques est une suite quasi-directe du cours de probabilités (MAT1720). Il est possible de trouver les notes de cours de probabilités [ici](#).

Création le : 23 mai 2022.

Dernière compilation le : 4 mai 2023

Lien permanent vers la version la plus à jour :

<https://www.dms.umontreal.ca/~davignon/files/notes/MAT2717-ndc.pdf>



Léa Stréliski 🌳 @LeaStreliski · 20 min

Le lol > les mathématiques.

[Afficher cette discussion](#)

« Le lol > les mathématiques. »

– Léa Stréliski, 3 octobre 2022, via Twitter.

Capture d'écran faite le 3 octobre 2022.

Table des matières

Introduction	1
partie 1. Théorie générale des processus stochastiques	3
Chapitre 1. Les chaînes de Markov	5
1.1. <i>Quoi ?</i> – L’état du système	5
1.2. <i>Pourquoi ?</i> – La quête de sens	8
1.3. La propriété de Markov	10
1.4. Les probabilités de transition	12
1.5. La représentation en graphes et les classes d’équivalence.	24
1.6. Temps d’atteinte et de retour, récurrence, transience, périodicité	30
1.7. Le théorème ergodique et la distribution stationnaire.	46
1.8. <i>Bonus</i> : Du point de vue de l’algèbre et de l’analyse.	68
1.9. Exercices	75
Chapitre 2. Chaînes de Markov à temps discret : études.	85
2.1. Le processus de Galton-Watson.	85
2.2. La marche aléatoire sur \mathbb{Z}^d	96
2.3. Exercices	110
Chapitre 3. Chaînes de Markov à temps continu	115
3.1. Le chapitre 1 à nouveau.	115
3.2. Une description qualitative de X_t .	116
3.3. Probabilités de transitions sur un intervalle de temps.	119
3.4. Taux générateurs et temps d’arrivée exponentiels.	125
3.5. Espérance de temps d’atteinte.	128
3.6. Distribution stationnaire et théorème ergodique.	130
3.7. Exercices	139
Chapitre 4. Processus de naissances et de mort	143
4.1. Le processus de Poisson.	143
4.2. Les processus de naissances	150
4.3. Les processus de naissances et de morts généraux.	153
4.4. Exercices	164
partie 2. Structures de dépendance particulières	167
Chapitre 5. Processus de renouvellements	169
Chapitre 6. Les martingales	171
Un peu de contexte	171

6.1. Définition et exemples	172
6.2. Temps d'arrêt	177
6.3. Le théorème d'arrêt.	180
6.4. Exercices	186
partie 3. Sujets avancés	189
Chapitre 7. Le mouvement Brownien	191
7.1. Définition axiomatique du processus de Wiener	192
7.2. Le processus de Wiener avec dérive et variance.	196
7.3. Le processus de Wiener géométrique	198
partie 4. Annexes	199
Annexe A. Lexique de la théorie des graphes	201
A.1. Description de la structure des graphes	201
A.2. Connexité, topologie et métrique des graphes	204
Annexe B. Résultats d'analyse	207
B.1. Convergence dominée et limites de séries	207
B.2. Convergence de suites	208
B.3. Convexité	210
B.4. Mesures, fonctions généralisées et distributions.	210
Annexe C. Notions avancées en probabilités	213
C.1. Tribus et sous-tribus	213
C.2. Filtrations, processus adaptés	214
C.3. Résultats sur les limites.	215

Introduction

Vous sortez de chez vous. Il fait beau, et vous avez en tête la seule idée d'aller marcher, mais sans objectif précis. Avec vous, une bouteille d'eau pour la route. Pour choisir dans quelle direction vous partez, vous la faites tourner sur elle-même jusqu'à ce qu'elle s'arrête. Vous empruntez alors la direction suggérée par le goulot.

Arrivée à une intersection, vous ressortez votre bouteille, puis vous la faites tourner à nouveau sur elle-même afin de choisir dans quelle direction vous diriger ensuite. Vous répétez ces étapes chaque fois que vous avez à choisir parmi plusieurs chemins alternatifs.

Où irez-vous ? Jusqu'où irez-vous ? Reviendrez-vous un jour chez vous ? Ou errerez-vous ainsi jusqu'aux confins de l'univers ? Et si vous revenez, combien de temps cela vous prendra-t-il ? Et si vous ne revenez pas, à quel rythme vous éloignerez-vous tranquillement de votre point d'origine ?

Le processus décrit plus haut est modélisé par ce que l'on appelle une marche aléatoire. Il s'agit d'un des plus importants exemples de *processus stochastiques* – des processus dont l'évolution est gouvernée par des phénomènes aléatoires. Ces notes de cours se penchent sur la théorie mathématique qui décrit, à l'aide de la théorie des probabilités, le comportement des processus stochastiques. C'est avec ces outils que nous pourrions répondre aux questions posées plus haut.

Bonne lecture !

REMARQUE. *Ce document contient les notes de cours pour le cours MAT2717 – Processus stochastiques. Il s'agit d'un cours de deuxième année au Baccalauréat en Mathématiques au département de l'université de Montréal.*

Le contenu est inspiré de celui présenté dans l'ouvrage Processus stochastiques : cours et exercices corrigés, de Sabin Lessard (éditions Ellipses).

Veuillez noter que ce document est un chantier en cours, et sera probablement modifié au cours de la session ; la version la plus à jour sera toujours sur StudiUM, mais peut aussi être récupérée, en tout temps, à l'adresse <https://www.dms.umontreal.ca/~davignon/files/notes/MAT2717-ndc.pdf>.

Date de compilation : 4 mai 2023

Première partie

Théorie générale des processus stochastiques

Les chaînes de Markov



Andreï Andreïevitch Markov, 1856 – 1922

Andreï Markov n'avait même pas vingt ans lorsqu'il entra à l'université impériale de Saint-Petersbourg. C'est avec l'encouragement de Pafnouti Tchebychev (un autre nom qui devrait être familier...) que Markov fait ses premiers travaux, d'abord en théorie des nombres. Ce n'est qu'après sa retraite que Markov s'intéresse de près à la théorie des probabilités et qu'il fait une première description rigoureuse des processus stochastiques.

Contrairement à la croyance populaire, Andreï Markov n'a jamais été défenseur pour les Canadiens de Montréal.

Image : domaine public.

Tout au long de ces notes, notre discussion portera sur les *processus stochastiques*. Grossièrement parlant, les processus stochastiques sont des suites de variables aléatoires. Le nom de « processus » provient du fait qu'on utilise très souvent de telles suites pour décrire l'évolution dans le temps d'*objets* quelconques. C'est vague – par exprès. Le modèle est très général, et permet de décrire autant la propagation d'un feu de forêt sur un territoire que les déplacements de requins au fond des océans, ou l'achallandage d'un réseau de transport.

Dans ce chapitre, on se penche sur les processus stochastiques *en général*, c'est à dire qu'on va décrire – de façon très vague – un « système » qui évolue de façon aléatoire. Ça peut être *n'importe quoi*.¹

1.1. *Quoi ?* – L'état du système

La propriété fondamentale d'un tel système, c'est son *état* – c'est encore délibérément vague puisque, encore une fois, ça peut être n'importe quoi.

- EXEMPLE 1.1. (a) Un interrupteur peut se trouver dans deux *états* : ouvert ou fermé.
- (b) Sur un échiquier, une pièce peut se trouver dans l'un de 64 *états* – correspondant bien sûr aux 64 cases de l'échiquier.
- (c) Dans une partie de *Serpents et échelles*, une joueuse peut se trouver dans 100 *états* différents, correspondant aux cent cases de la planche de jeu.
- (d) Un américain peut se trouver dans 50 *états* : Hawaï, l'Alaska, Washington, l'Oregon, la Californie, l'Arizona, le Nevada, l'Utah, l'Idaho, le Montana, le Wyoming, le Colorado, le Nouveau-Mexique, le Texas, l'Oklahoma, le Kansas, le Nebraska, le Dakota du Sud, le Dakota du Nord, le Minnesota, le Wisconsin, l'Iowa, l'Illinois, le Missouri, l'Arkansas, la Louisiane, le Mississippi, l'Alabama, la Floride, la Georgie, la Caroline du Sud, la Caroline du Nord, le Tennessee, la Virginie, la Virginie Occidentale, le Kentucky, l'Indiana, le

1. Oui, oui. N'importe quoi.

Michigan, l'Ohio, la Pennsylvanie, le Maryland, le Delaware, le New-Jersey, New-York, le Connecticut, le Rhode-Island, le Massachussetts, le Vermont, le New-Hampshire ou le Maine.²

- (e) Un incendie de forêt peut se trouver dans une infinité d'*états* – ici, par exemple, chaque état correspond à « la région qui brûle ». Par exemple, ça pourrait être l'ensemble des coordonnées géographiques où il y a le feu.
- (f) Dans un processus qui décrit la position d'un requin, l'*état* du système correspond à la position du requin. Il y a donc une infinité d'états possibles, correspondant à chaque paires de coordonnées géographiques où le requin pourrait se trouver (voir figure 1.1).

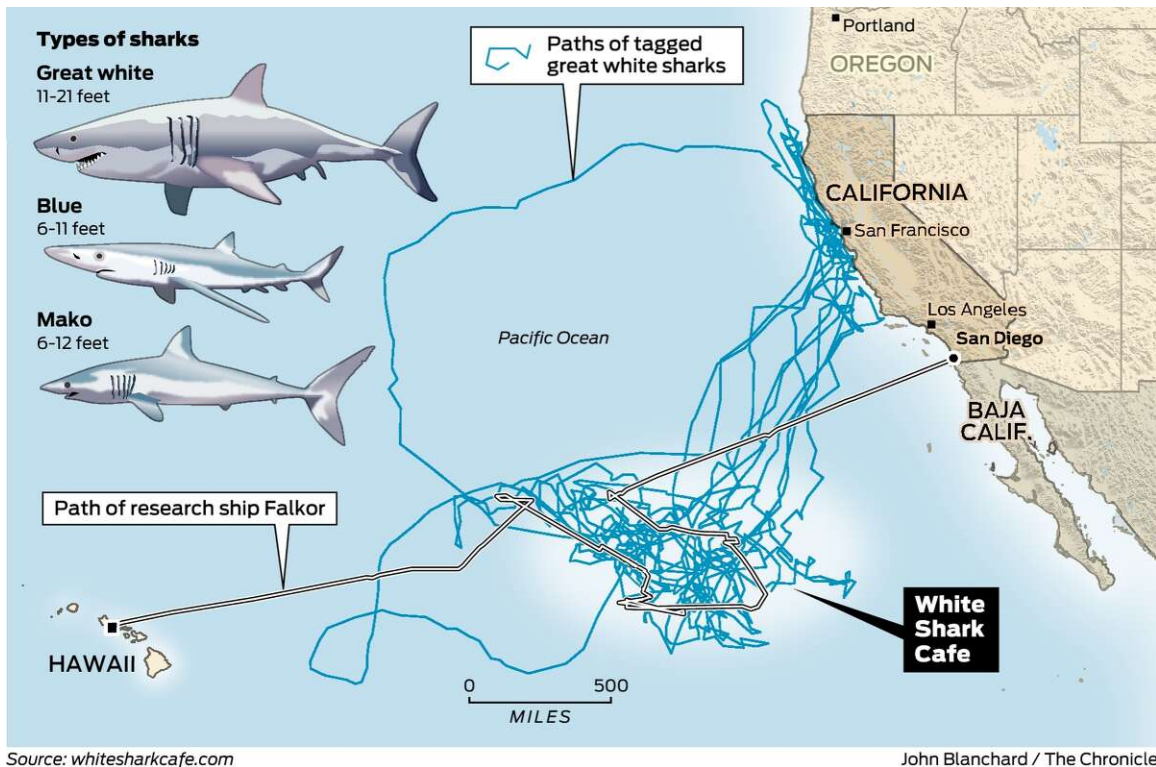


FIGURE 1.1 – Les déplacements de requins suivis par des scientifiques au large des côtes de Californie.

Image : John Blanchard / San Francisco Chronicle / whitesharkcafe.com

- (g) Dans un processus qui décrit l'achallandage d'une ligne d'autobus, l'*état* du système pourrait correspondre au nombre de voyageurs/ses desservi-e-s par l'autobus jusqu'à présent.
- (h) etc.

EXERCICE. Trouver d'autres exemples de processus qui pourraient être modélisés par un processus stochastiques. Quels seraient les états possibles ?

2. Il est laissé en exercice de vérifier qu'hormis les deux premiers, les 48 états subséquents sont listés dans un ordre tel qu'il serait possible de les parcourir en voiture sans jamais décoller de Terre.

En ce qui nous concerne, la toute première caractéristique qui nous servira à décrire un processus sera donc **l'espace des états** – c'est à dire l'ensemble des états possibles dans lequel notre *système* peut se trouver.

DÉFINITION 1.1 (L'espace des états.). L'**espace des états** correspond à l'ensemble des états possibles pour un processus stochastique.

On le dénotera souvent par S . Il peut être fini, infini dénombrable ou même indénombrable, et il peut également porter des structures supplémentaires – par exemple, ça peut être un espace vectoriel, ou un espace topologique, ou une variété riemannienne, ou un groupe, ou les sommets d'un graphe, etc.

On dira que S est **discret** si il est fini ou dénombrable. Sinon, S sera **continu**³.

REMARQUE. Attention de bien distinguer l'espace des états S de l'ensemble fondamental d'une expérience aléatoire Ω . Ici, l'expérience aléatoire correspond à *toute l'évolution du système dans le temps*. Son ensemble fondamental est donc beaucoup plus riche que simplement S .

DÉFINITION 1.2 (Processus stochastique). Un **processus stochastique** sur S est une famille de variables aléatoires⁴ $\mathbf{X} = (X_t : \Omega \rightarrow S)_{t \geq 0}$ à valeurs dans S .

EXEMPLE 1.2. Soit $(X_i)_{i \in \mathbb{Z}^+}$ une suite de variables aléatoires i.i.d. à valeurs dans \mathbb{R} . Cette suite est *techniquement* un processus stochastique.

Toutefois, bien sûr, les X_i étant tous indépendants, ça ne servira pas à grand chose de l'étudier comme un processus stochastique. L'intérêt de l'étude des processus stochastiques, c'est de se pencher sur des processus où il y a une certaine dépendance entre les variables...

EXEMPLE 1.3. On lance une pièce de monnaie de manière répétée. À chaque lancer, on gagne un point si la pièce retombe sur « pile ». Soit X_n le nombre de points total que l'on a accumulé après n lancers.

La suite $(X_n)_{n \in \mathbb{Z}^+}$ est un processus stochastique sur l'ensemble des états $S = \mathbb{Z}^+$. La figure 1.2 montre l'évolution d'une réalisation de ce processus pour les cent premiers lancers.

Quand ? – Le temps. Dans un processus stochastique $(X_t)_{t \geq 0}$, on réfère souvent à l'indice t comme au **temps** – c'est d'ailleurs pour cette raison qu'on aime employer le symbole t . Les valeurs possibles pour t sont des nombres positifs – on remarque que dans la définition, on précise simplement que $t \geq 0$. C'est, encore une fois, volontairement vague.

On va globalement distinguer deux types de processus :

- les processus **à temps discret** ;

Ce sont les processus de la forme $(X_t)_{t \in \mathbb{Z}^+}$. Les processus à temps discrets sont donc des processus stochastiques indicés par des entiers positifs ; ce sont donc des suites de variables aléatoires : $X_0, X_1, X_2, X_3, \dots$

3. Les termes *discret* et *continu* font respectivement référence aux topologie discrète et du « continuum » ; pas forcément au cardinal. Toutefois, pour nous, le terme *discret* sera systématiquement synonyme de *fini* ou *dénombrable*.

4. Il convient ici de rappeler qu'une *variable aléatoire* est en fait une fonction qui dépend du résultat d'une expérience aléatoire. Dans le cours de probabilités, on s'est particulièrement intéressés au cas particulier où les valeurs de ces variables étaient des nombres, mais ce n'est pas forcément le cas. Ici, nos variables aléatoires prennent des valeurs dans l'ensemble S .

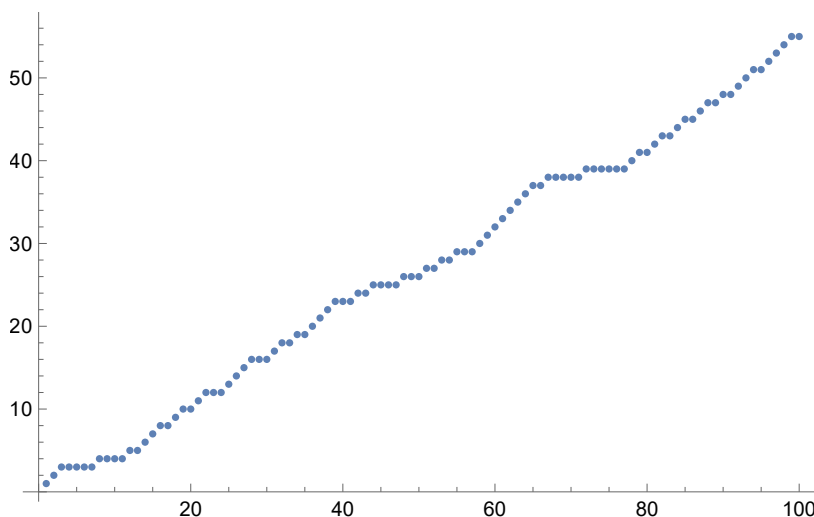


FIGURE 1.2 – Le nombre de fois qu'on a obtenu « Pile » en n lancers.

– les processus **à temps continu**.

Ce sont les processus de la forme $(X_t)_{t \in \mathbb{R}^+}$. Ils sont indicés par $\mathbb{R}^+ = [0, +\infty)$ – ce sont donc en quelques sortes des « fonctions aléatoires » dont le domaine est \mathbb{R}^+ et l'image est incluse dans S . Pour cette raison, on notera souvent les processus stochastiques à temps continu, non pas avec des indices, mais plutôt comme des fonctions dont t serait l'argument, comme ceci : $(X(t))_{t \in \mathbb{R}^+}$.

Dans le chapitre qui suit, nous nous intéressons en particulier aux *processus à temps discret* sur des *espaces d'états discrets*.

1.2. Pourquoi ? – La quête de sens

C'est le moment d'une bonne crise existentielle : *pourquoi fait-on tout ça ?* Nous allons passer toutes ces notes à se pencher sur les processus stochastiques. Qu'est-ce qu'on cherche ?

La réponse évidente, c'est bien sûr que l'on cherche à prédire le comportement de systèmes dynamiques imprévisibles.⁵ Bien sûr, il s'agit de processus *stochastiques* – ce sont des processus aléatoires. On ne peut donc pas prédire le comportement exact de tels processus. Il existe néanmoins des quantités auxquelles on peut s'intéresser.

1.2.1. La distribution de X_t . Ce sera bien sûr la quantité d'intérêt centrale tout au long de notre étude : la probabilité que le processus se trouve dans la région $A \subseteq S$ de l'espace d'états S au temps t :

$$\mathbb{P}\{X_t \in A\}.$$

Plus spécifiquement, on pourra s'intéresser à la probabilité conditionnelle *sachant que* $X_0 = x$ – et étudier l'influence des conditions initiales sur la dynamique :

$$\mathbb{P}\{X_t \in A \mid X_0 = x\}.$$

5. Nous reparlerons plus loin des parallèles entre la théorie de processus stochastique et celle des systèmes dynamiques « déterministes »...

Dans le cas d'un espace d'états discret, il nous suffira d'étudier la fonction de masse⁶ – soit la probabilité que $X_t = y$ sachant que $X_0 = x$:

$$\mathbb{P}\{X_t = y \mid X_0 = x\}.$$

Nous verrons comment, en utilisant des outils d'analyse et d'algèbre linéaire, nous parviendrons à calculer ces quantités d'intérêt.

Naturellement, la prochaine question à considérer sera celle de...

1.2.2. La distribution-limite de X_t . Il s'agira en fait simplement de considérer la limite de la distribution de X_t lorsque t tend vers l'infini, comme ceci :

$$\lim_{t \rightarrow \infty} \mathbb{P}\{X_t = y \mid X_0 = x\}.$$

À condition, bien sûr, que cette limite existe ! Ça ne sera pas forcément toujours le cas. En fait, la détermination-même de l'existence de cette limite fera l'objet d'une partie importante de notre étude. On portera une attention toute particulière aux raisons spécifiques qui font en sorte qu'elle pourrait ne pas exister. On s'intéressera aussi à d'autres façon de la déterminer, sans passer par le calcul direct de la limite (ça pourrait s'avérer laborieux).

L'intérêt d'identifier si il existe une distribution-limite, c'est de comprendre si le comportement du système « se stabilise à long terme » – c'est-à-dire, de déterminer si, après un moment, l'influence de la condition initiale (la valeur de X_0 , l'état initial du système) sera complètement dissipée.

1.2.3. La proportion moyenne de temps passée dans une région $A \subseteq S$ à long terme. Une autre façon d'approcher le problème de savoir « où le processus se retrouve à long terme » est de s'intéresser à la proportion de temps passée en moyenne, à long terme, dans divers régions de notre espace d'états S .

Par exemple, dans le cas d'un processus à temps discret sur des états discrets, on pourrait noter $N_t(y)$ le temps passé (en nombre d'incrément, excluant le temps 0) dans l'état $y \in S$, jusqu'au temps t inclusivement :

$$N_t(y) = \sum_{k=1}^t \mathbb{1}_{\{X_k=y\}}.$$

La proportion de temps passée dans l'état y serait alors $N_t(y)/t$, et en moyenne, à long terme, on s'intéresserait donc à :

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{1}{t} \mathbb{E}[N_t(y) \mid X_0 = x] &= \lim_{t \rightarrow \infty} \frac{1}{t} \mathbb{E}\left[\sum_{k=1}^t \mathbb{1}_{\{X_k=y\}} \mid X_0 = x\right] \\ &= \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{k=1}^t \mathbb{P}\{X_k = y \mid X_0 = x\}. \end{aligned}$$

Cette dernière limite correspond en fait à la *limite de Cesaró* pour la distribution-limite qui nous intéressait plus haut.

6. Vous aurez deviné que dans le cas d'un espace d'états continu, il faudra plutôt s'intéresser à une fonction de densité sur S . Nous en reparlerons à la fin des notes lorsque nous aborderons le mouvement Brownien.

1.2.4. Les temps d'atteinte, les temps de visite, les temps de retour ... On s'intéressera également à une panoplie d'autres quantités pertinentes ; en voici une petite liste non-exhaustive :

- le temps moyen entre deux visites au même état ;
- en partant de l'état x , la probabilité de visiter l'état y avant l'état z ,
- et le cas particulier lorsque $x = z$;
- la probabilité de ne jamais revenir à l'état x ;
- le nombre moyen de visites à l'état x ;
- lorsqu'on aura une notion de *distance* dans l'espace des états : la vitesse moyenne à laquelle le système se déplace ;
- etc.

Pour l'instant, ces quantités ne sont pas bien définies ; ce n'est pas important. L'objectif de cette section était surtout de citer des exemples du genre de mystères que nous chercherons (et que parfois nous réussirons) à élucider.

1.3. La propriété de Markov

Jusqu'à présent, nous avons parlé de processus stochastiques en général. Toutefois, dans ces notes, nous nous concentrerons sur un type de processus en particulier : les chaînes de Markov.

Pour définir ces dernières de façon rigoureuse, nous aurons besoin d'introduire un peu de notation :

NOTATION (Sous-familles). Soit $\mathbf{x} = (x_i)_{i \in I}$ une famille d'éléments x_i où les indices i sont dans l'ensemble I .

Pour n'importe quel ensemble S et n'importe quel ensemble d'indices I , l'ensemble S^I est l'ensemble des familles $\mathbf{x} = (x_i)_{i \in I}$ où $x_i \in S$ pour tout i .

Pour $J \subseteq I$, on notera $\mathbf{x}_J = (x_i)_{i \in J}$ la **sous-famille** d'éléments qui correspond à seulement la famille des éléments dont les indices sont dans l'ensemble J .

Dans le cas d'un processus stochastique (à temps discret ou continu) $\mathbf{X} = (X_t)_{t \geq 0}$, on emploiera les notations suivantes ;

- $\mathbf{X}_{>t} := (X_u)_{u>t}$, le **futur** du processus strictement après le temps t .
- $\mathbf{X}_{<t} := (X_u)_{0 \leq u < t}$ le **passé** du processus strictement avant le temps t .
- $\mathbf{X}_{\geq t}$ et $\mathbf{X}_{\leq t}$ les futurs et passés du processus, respectivement après et avant le temps t inclusivement, définis de façon analogue.

DÉFINITION 1.3 (Propriété de Markov faible). Soit $\mathbf{X} = (X_t)_{t \geq 0}$ un processus stochastique quelconque (à temps discret ou continu, sur un espace d'états S discret ou continu).

On dit que le processus \mathbf{X} possède la **propriété de Markov faible**⁷ si, pour tout temps $t \geq 0$, on a que

$$\mathbb{P}\{\mathbf{X}_{>t} \in A \mid X_t = i, \mathbf{X}_{<t} \in B\} = \mathbb{P}\{\mathbf{X}_{>t} \in A \mid X_t = i\},$$

pour tous $A \subseteq S^{\{u:u>t\}}$ et $B \subseteq S^{\{u:0 \leq u < t\}}$.

7. La propriété de Markov *forte* requiert la notion plus avancée de *temps d'arrêt*, un temps aléatoire ayant certaines propriétés utiles. Les temps d'arrêt seront introduits lorsque nous parlerons de martingales.

En mots (et c'est la partie importante à retenir) :

*Un processus stochastique a la propriété de Markov faible si et seulement si, pour tout temps t , sachant l'état du processus au temps t , le **passé** (strictement avant t) et le **futur** (strictement après t) sont indépendants.*

On dit qu'un processus qui a la propriété de Markov faible est une **chaîne de Markov**.

EXEMPLE 1.4. Voici quelques exemples de processus stochastiques. Dire quel est l'espace des états S . Dire aussi si ce sont des chaînes de Markov.

- (a) On lance une pièce de monnaie à répétition. X_i est le résultat de l' i ème lancer. Soit $\mathbf{X} = (X_i)_{i \geq 1}$.
- (b) On lance une pièce de monnaie à répétition. Soit X_i le nombre de lancers effectués depuis la dernière fois qu'on a obtenu « Pile », après i lancers (inclusivement). Soit $\mathbf{X} = (X_i)_{i \geq 1}$.
(Par exemple, si on a obtenu Pile, Face, Face, Pile, Pile, Face, Pile, ..., on aurait $\mathbf{X} = (0, 1, 2, 0, 0, 1, 0, \dots)$.)
- (c) On lance une pièce de monnaie à répétition. Après i lancers, soit X_i le nombre de lancers restants avant d'obtenir Pile de nouveau (incluant le lancer où on obtient finalement Pile). Soit $\mathbf{X} = (X_i)_{i \geq 0}$.
(Par exemple, si on a obtenu Pile, Face, Face, Pile, Pile, Face, Pile, ..., on aurait $\mathbf{X} = (1, 3, 2, 1, 1, 2, \dots)$.)
- (d) Soit $k \in \mathbb{N}$. On lance une pièce de monnaie à répétition. Après $i + k$ lancers (pour $i \geq 1$), on note X_i le nombre de fois qu'on a obtenu « Pile » entre les i ème et $(i + k)$ èmes lancers (inclusivement). Soit $\mathbf{X} = (X_i)_{i \geq 1}$.
(Par exemple, avec $k = 1$, si on a obtenu Pile, Face, Face, Pile, Pile, Face, Pile, ..., on aurait $\mathbf{X} = (1, 0, 1, 2, 1, 1, \dots)$)

SOLUTION. (a) L'espace des états serait $S = \{\text{Pile}, \text{Face}\}$. Le processus \mathbf{X} a la propriété de Markov – en effet, puisque tous les lancers sont indépendants, il est trivialement vrai que le futur du processus après le temps t est indépendant du passé avant le temps t .

- (b) L'espace des états serait $S = \mathbb{Z}^+$ – en effet, il peut s'écouler un nombre entier de lancers depuis la dernière fois qu'on a eu « Pile », mais aussi 0, lorsqu'on vient tout juste d'avoir Pile.

C'est une chaîne de Markov – si je sais combien de lancers j'ai fait depuis la dernière fois que j'ai eu « Pile », ce qui s'est passé avant n'importe pas – je sais qu'au prochain lancer, j'aurai soit Pile (et on retombera à 0), ou Face (et le compteur augmentera encore).

- (c) L'espace des états serait $S = \mathbb{N}$ – à chaque lancer, soit le prochain sera Pile (alors il faut un lancer pour avoir Pile de nouveau), ou bien ce sera Face (alors il faudra plus d'un lancer pour avoir Pile de nouveau).

Malgré que ça puisse paraître contre-intuitif, le processus \mathbf{X} a bel et bien la propriété de Markov. En effet, même si la valeur de X_t dépend du futur *des lancers*, ça ne correspond pas à la définition d'une chaîne de Markov !

Dans les faits, une fois qu'on sait que $X_i = 5$, par exemple, peu importe ce qui s'est passé avant, on sait immédiatement que $X_{i+1} = 4$, $X_{i+2} = 3$, etc. Peu importe, donc, ce qui s'est passé avant.

De même, si on sait que $X_i = 1$, on sait qu'on obtiendra « Pile » au prochain lancer. Sachant cela, X_{i+1} sera donc simplement le nombre de lancers nécessaires pour obtenir Pile de nouveau immédiatement après celui qu'on vient d'obtenir au lancer $i + 1$. C'est donc une variable aléatoire géométrique de paramètre $1/2$, peu importe ce qui s'est passé avant.

- (d) Ici, l'espace des états serait $S = \{0, 1, 2, 3, \dots, k + 1\}$, correspondant au nombre de Pile qu'on a obtenu dans les $k + 1$ derniers lancers.

Le processus \mathbf{X} n'a PAS la propriété de Markov faible. En effet, avec $k = 2$, supposons qu'on sait que $X_1 = 2$ et $X_2 = 1$. Alors, on sait que les trois premiers lancers ont donné respectivement Pile, Pile et Face. Donc, on sait qu'il est impossible d'avoir $X_3 = 2$, puisqu'il faudrait que les troisième et quatrième lancers aient tous les deux donné Pile, ce qui est impossible.

Par contre, si on avait plutôt que $X_1 = 0$ et $X_2 = 1$, alors la probabilité d'avoir $X_3 = 2$ serait non-nulle.

On vient de voir que, sachant X_2 , des informations sur le passé du processus \mathbf{X} peuvent influencer les probabilités pour le futur du processus. Donc ce processus stochastique n'a pas la propriété de Markov faible.

1.4. Les probabilités de transition

Comme mentionné plus haut, pour la suite des choses, nous allons travailler seulement avec des chaînes de Markov – des processus stochastiques qui ont la propriété de Markov. Plus particulièrement, nous nous concentrons sur les chaînes de Markov sur des espaces d'états discrets.

Nous introduisons maintenant formellement les *probabilités de transition* :

DÉFINITION 1.4 (Probabilités de transition). Soit $\mathbf{X} = (X_t)_{t \geq 0}$ un processus stochastique sur l'espace d'états discret S . Soient $x, y \in S$ deux états, et soient $0 \leq s < t$.

On définit la **probabilité de transition** de x vers y entre les temps s et t :

$$P_{x,y}(s, t) = \mathbb{P}\{X_t = y \mid X_s = x\}.$$

Il s'agit de la probabilité que le processus se retrouve à l'état y au temps t , sachant qu'il était à l'état x au temps s .

Nous allons maintenant montrer quelques conséquences de la propriété de Markov sur les probabilités de transition.

PROPOSITION 1.1. Soit $\mathbf{X} = (X_t)_{t \geq 0}$ une chaîne de Markov sur l'espace d'états discret S . Les probabilités de transition ont les propriétés suivantes :

- i. Pour tout $t > 0$ et pour tous $x, y \in S$:

$$(1.4.1) \quad P_{x,y}(t, t) = \delta_{x,y} = \begin{cases} 1 & \text{si } x = y; \\ 0 & \text{si } x \neq y. \end{cases}$$

- ii. Pour tous $0 \leq s \leq s' \leq t$, $x \in S$:

$$(1.4.2) \quad \sum_{y \in S} P_{x,y}(s, t) = 1.$$

iii. Pour tous $0 \leq s \leq t$ et pour tous $x, y \in S$:

$$(1.4.3) \quad P_{x,y}(s, t) = \sum_{z \in S} P_{x,z}(s, s') P_{z,y}(s', t).$$

DÉMONSTRATION. La preuve de **i** est triviale.

ii est simplement un énoncé de la condition de normalisation pour la fonction de masse de X_t sachant que $X_s = x$.

Montrons maintenant **iii**.

On a que par définition, $P_{x,y}(s, t) = \mathbb{P}\{X_t = y \mid X_s = x\}$. Or, par la formule des probabilités totales, on a :

$$\begin{aligned} \mathbb{P}\{X_t = y \mid X_s = x\} &= \sum_{z \in S} \mathbb{P}\{X_t = y \mid X_{s'} = z, X_s = x\} \mathbb{P}\{X_{s'} = z \mid X_s = x\} \\ &= \sum_{z \in S} \mathbb{P}\{X_t = y \mid X_{s'} = z\} \mathbb{P}\{X_{s'} = z \mid X_s = x\}. \end{aligned}$$

La dernière égalité s'obtient par la propriété de Markov ; si on sait que $X_{s'} = z$, on peut ignorer l'information que $X_s = x$, puisqu'on détient déjà de l'information « plus à jour. »

Dès lors, en ré-utilisant la définition des probabilités de transition, on trouve forcément :

$$P_{x,y}(s, t) = \sum_{z \in S} P_{x,z}(s, s') P_{z,y}(s', t).$$

□

Avant de continuer, nous introduisons finalement une autre notation pratique :

NOTATION (Point de départ). Soit $\mathbf{X} = (X_t)_{t \geq 0}$ un processus stochastique sur l'espace d'états S , et soit $x \in S$ un état. On notera dorénavant

$$\mathbb{P}_x\{\cdot\} := \mathbb{P}\{\cdot \mid X_0 = x\}$$

la mesure de probabilité correspondant au cas où le processus se trouve initialement dans l'état x .

En particulier, avec nos définitions des probabilités de transition, on trouve que pour tous $x, y \in S$, $0 \leq t$,

$$P_{x,y}(0, t) = \mathbb{P}_x\{X_t = y\}$$

On introduit maintenant la notion d'*homogénéité* d'une chaîne de Markov :

DÉFINITION 1.5 (Homogénéité). Soit $\mathbf{X} = (X_t)_{t \geq 0}$ une chaîne de Markov sur un espace discret d'états S . On dit que \mathbf{X} est une chaîne de Markov **homogène** si, pour toute paire d'états $x, y \in S$, et pour tous temps $0 \leq s < t$, la probabilité de transition $P_{x,y}(s, t)$ ne dépend que de la différence entre t et s .

Autrement dit, la chaîne de Markov \mathbf{X} est homogène si et seulement si pour tous $x, y \in S$, $0 \leq s < t$,

$$(1.4.4) \quad P_{x,y}(s, t) = P_{x,y}(0, t - s) = \mathbb{P}_x\{X_{t-s} = y\}.$$

NOTATION (Probabilité de transition sur un intervalle). Dans le cas où une chaîne de Markov est homogène, on notera simplement la probabilité de transition entre x et y en un intervalle de temps t par : $P_{x,y}^{(t)} = \mathbb{P}_x\{X_t = y\}$.

Nous verrons plus loin que la notation avec exposants n'est pas sans raison...

COROLLAIRE 1.1. Soit $\mathbf{X} = (X_t)_{t \geq 0}$ une chaîne de Markov homogène sur l'espace d'états discret S . Les probabilités de transition ont les propriétés suivantes :

i. Pour tous $x, y \in S$:

$$(1.4.5) \quad P_{x,y}^{(0)} = \delta_{x,y} = \begin{cases} 1 & \text{si } x = y \\ 0 & \text{si } x \neq y. \end{cases}$$

ii. Pour tous $t \geq 0, x \in S$:

$$(1.4.6) \quad \sum_{y \in S} P_{x,y}^{(t)} = 1.$$

iii. Pour tous $s, t > 0$, et tous $x, y \in S$:

$$(1.4.7) \quad P_{x,y}^{(s+t)} = \sum_{z \in S} P_{x,z}^{(s)} P_{z,y}^{(t)}.$$

1.4.1. Le temps discret et les probabilités de transition en un pas. Jusqu'ici, nous avons défini les probabilités de transition en termes très généraux, de telle sorte que les définitions et les résultats obtenus s'appliquent peu importe que l'on ait une chaîne de Markov à temps discret ou continu.

Toutefois, nous allons maintenant examiner une conséquence pratique des processus en temps discret : le fait qu'il existe « une plus petite unité de temps ». Nous appellerons cette unité de temps un *pas*.

DÉFINITION 1.6 (Probabilité de transition en un pas.). Soit $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ une chaîne de Markov à temps discret sur l'espace d'états discret S . Soient $x, y \in S$ deux états. Alors la **probabilité de transition de x vers y en un pas au temps t** est donnée par

$$P_{x,y}(t) := P_{x,y}(t, t+1).$$

Si la chaîne \mathbf{X} est homogène, on note simplement la **probabilité de transition de x vers y en un pas** par :

$$P_{x,y} := P_{x,y}(0) = P_{x,y}^{(1)} = \mathbb{P}_x \{X_1 = y\}.$$

En particulier, dans le cas homogène, cette notation est très pratique. En effet, elle permet d'obtenir le résultat suivant :

PROPOSITION 1.2. Soit $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ une chaîne de Markov à temps discret homogène sur un espace d'états discret S . Soit $P = (P_{x,y})_{(x,y) \in S \times S}$ la famille des probabilités de transition en un pas.

Alors, cette famille caractérise entièrement la loi de \mathbf{X} .

DÉMONSTRATION. Soient $t_1 < t_2 < \dots < t_n$ un nombre fini de temps (entiers non-négatifs), et soient $x_1, x_2, x_3, \dots, x_n$ autant d'états.

Alors, par une propriété élémentaire des probabilités, ainsi qu'avec la propriété de Markov, on a :

$$\begin{aligned}\mathbb{P}\{X_{t_1} = x_1, \dots, X_{t_n} = x_n\} &= \prod_{k=2}^n \mathbb{P}\{X_{t_k} = x_k \mid X_{t_{k-1}} = x_{k-1}, \dots, X_{t_1} = x_1\} \\ &= \prod_{k=2}^n \mathbb{P}\{X_{t_k} = x_k \mid X_{t_{k-1}} = x_{k-1}\} \\ &= \prod_{k=2}^n P_{x_{k-1}, x_k}^{(t_k - t_{k-1})}.\end{aligned}$$

Autrement dit, ces probabilités ne dépendent que des probabilités de transition.

Or, pour tout $x, y \in S$, et pour tout $m \in \mathbb{N}$, par (1.4.7) on peut écrire :

$$P_{x,y}^{(m+1)} = \sum_{z \in S} P_{x,z}^{(m)} P_{z,y}.$$

Et par récurrence, on a que pour tout $x, y \in S$, et pour tout $m \in \mathbb{N}$, $P_{x,y}^{(m)}$ peut être écrit comme une somme de produits d'éléments de P .

Par conséquent, si $P = (P_{x,y})_{(x,y) \in S \times S}$ est donnée, cela fixe la valeur de $\mathbb{P}\{X_{t_1} = x_1, \dots, X_{t_n} = x_n\}$ pour toute famille finie de temps $t_1 < \dots < t_n$ et toute famille finie d'états x_1, \dots, x_n .

En somme, donc, la loi de toute sous-famille finie de \mathbf{X} est caractérisée par P , donc la loi de \mathbf{X} est également caractérisée par P . \square

1.4.2. Les espaces d'états finis et la représentation matricielle. Nous venons de voir que la famille de probabilités de transition en un pas $P = (P_{x,y})_{(x,y) \in S \times S}$ est très importante pour les chaînes de Markov à temps discret sur des espaces d'états discrets : elle caractérise complètement la loi de notre processus stochastique (proposition 1.2).

Par abus de langage, nous appellerons P la *matrice de transition* ; toutefois, dans le cas où S est un espace d'états finis, on pourra rendre ce langage parfaitement rigoureux – et ce sera extrêmement utile.

Numérotation des états. Avant d'aller plus loin, il faut se mettre d'accord sur un « ordre » pour les états. Celui-ci est arbitraire, mais il doit être fixé, et – si on veut pouvoir s'y retrouver – il faut l'énoncer explicitement. Ça revient à donner à chaque état un « numéro ».

EXEMPLE 1.5. On considère un processus stochastique indiquant pour le jour t quel temps il fera. Les états possibles sont $S = \{\text{clair, nuageux, pluvieux, orageux}\}$.

On peut alors associer un numéro à chaque état, par exemple comme ceci :

- (1) clair ;
- (2) nuageux ;
- (3) pluvieux ;
- (4) orageux.

De cette façon, pour n'importe quel processus stochastique sur un espace fini d'états S' avec $|S'| = n$, on peut considérer un processus de loi identique sur l'espace d'états $S = \{1, 2, 3, \dots, n\}$ (ou $\{0, 1, 2, \dots, n-1\}$). Si l'espace d'états est dénombrable – c'est-à-dire que $|S'| = |\mathbb{N}|$ – alors on peut carrément considérer plutôt le processus de loi identique sur l'espace d'états $S = \mathbb{N}$ (ou \mathbb{Z}^+).

Cette numérotation sera pratique parce qu'elle permettra de noter sans ambiguïté la famille de probabilités de transition $P = (P_{x,y})_{(x,y) \in S' \times S'}$ comme une matrice $P \in \mathbb{R}^{n \times n}$.

EXEMPLE 1.6. On considère une chaîne de Markov homogène indiquant pour le jour t quel temps il fera. Les états possibles sont $S = \{\text{clair, nuageux, pluvieux}\}$.

- Si le temps est clair, la probabilité qu'il fasse nuageux le lendemain est de $2/5$, et la probabilité qu'il pleuve, $1/10$. Sinon il fera encore beau.
- Si le temps est nuageux, la probabilité qu'il fasse nuageux le lendemain est de $1/2$, et la probabilité qu'il pleuve est de $1/3$. Sinon, il fera beau.
- Si le temps est pluvieux, la probabilité qu'il fasse beau (clair) le lendemain est de $3/5$, et la probabilité qu'il fasse nuageux est de $1/10$. La probabilité qu'il pleuve encore est de $3/10$.

On associe des numéros aux états de la façon suivante :

- (1) clair ;
- (2) nuageux ;
- (3) pluvieux ;

Alors, la matrice de transition en un pas $P = (P_{i,j})_{1 \leq i,j \leq 3} \in \mathbb{R}^{3 \times 3}$ est donnée par :

$$P = \begin{pmatrix} 1/2 & 2/5 & 1/10 \\ 1/6 & 1/2 & 1/3 \\ 3/5 & 1/10 & 3/10 \end{pmatrix}.$$

Dans cette matrice, l'entrée $P_{i,j}$ (i ème ligne, j ème colonne) donne la probabilité de transition de l'état i vers l'état j en un pas.

De façon complètement analogue, on peut définir des matrices pour tous les types de probabilités de transition que nous avons vu jusqu'ici. Nous utiliserons en particulier ces notations :

- $P(s, t) = (P_{i,j}(s, t))_{1 \leq i,j \leq n} \in \mathbb{R}^{n \times n}$, la matrice de transition entre les temps s et t , pour $0 \leq s \leq t$;
- $P^{(t)} = (P_{i,j}^{(t)})_{1 \leq i,j \leq n} \in \mathbb{R}^{n \times n}$, la matrice de transition sur un intervalle de temps t pour les chaînes de Markov homogènes ;
- $P(t) = P(t, t+1) = (P_{i,j}(t))_{1 \leq i,j \leq n} \in \mathbb{R}^{n \times n}$, la matrice de transition en un pas au temps t pour les chaînes de Markov inhomogènes.

Avec cette notation, on obtient directement les propriétés suivantes :

PROPOSITION 1.3. Soit $\mathbf{X} = (X_t)_{t \geq 0}$ une chaîne de Markov sur un espace fini d'états $S = \{1, 2, 3, \dots, n\}$, et soient $P(s, t)$ ses matrices de transition entre les temps s et t , pour $0 \leq s \leq t$. Les matrices de transition ont les propriétés suivantes :

- i. Pour tout $t \geq 0$,

$$(1.4.8) \quad P(t, t) = \mathbb{I}_n$$

où \mathbb{I}_n est l'identité $n \times n$.

- ii. Pour tous $0 \leq s \leq t$ et tout $1 \leq i \leq n$,

$$(1.4.9) \quad \sum_{j=1}^n P_{i,j}(s, t) = 1.$$

Autrement dit, la somme de chaque ligne est de 1.

iii. Pour tous $0 \leq s \leq s' \leq t$, on a que :

$$(1.4.10) \quad P(s, t) = P(s, s')P(s', t),$$

où le produit matriciel est le produit matriciel usuel.

DÉMONSTRATION. La propriété **i** est obtenue en comparant simplement les éléments de $P(t, t)$ (on a $P_{i,j}(t, t) = \delta_{i,j}$) à ceux de \mathbb{I}_n – ce sont les mêmes.

La propriété **ii** est exactement la même que celle énoncée dans la proposition 1.1 ; il suffit de considérer le cas où $S = \{1, 2, 3, \dots, n\}$.

La propriété **iii** est plus surprenante – en effet, elle nous montre qu’au-delà de la façon pratique et synthétique d’organiser l’information sous forme d’un « tableau », la représentation en matrices est carrément *utile*. Cela pointe vers une vérité plus profonde sur les matrices de transition.

Pour voir qu’elle est vraie, il suffit de voir que, élément par élément de la matrice, on a que

$$\begin{aligned} [P(s, s')P(s', t)]_{i,j} &= \sum_{k=1}^n P_{i,k}(s, s')P_{k,j}(s', t) \\ &= P_{i,j}(s, t). \end{aligned}$$

Ici la première égalité est simplement par définition du produit matriciel ; la seconde est vraie par (1.4.3). \square

REMARQUE. Si on introduit le vecteur-colonne à n lignes $\mathbf{1}_n = (1, 1, 1, \dots, 1)$, alors l’équation (1.4.9) peut être réécrite simplement comme :

$$P(s, t)\mathbf{1}_n = \mathbf{1}_n.$$

Qu’est-ce qu’on remarque ?

Pour les chaînes de Markov homogènes, on obtient le corollaire suivant :

COROLLAIRE 1.2. Soit $\mathbf{X} = (X_t)_{t \geq 0}$ une chaîne de Markov homogène sur un espace fini d’états $S = \{1, 2, 3, \dots, n\}$ et $P^{(t)}$ ses matrices de transition en un intervalle de temps t , pour $t \geq 0$. Alors les matrices de transition ont les propriétés suivantes :

i.

$$(1.4.11) \quad P^{(0)} = \mathbb{I}_n.$$

ii. Pour tous $t \geq 0$ et pour tous $1 \leq i \leq n$, on a :

$$(1.4.12) \quad \sum_{j=1}^n P_{i,j}^{(t)} = 1. \quad (\Leftrightarrow P^{(t)}\mathbf{1}_n = \mathbf{1}_n)$$

Autrement dit, pour chaque ligne de $P^{(t)}$, la somme est 1.

iii. Soient $s, t \geq 0$. Alors, on a :

$$(1.4.13) \quad P^{(s+t)} = P^{(s)}P^{(t)}.$$

La démonstration de ce corollaire est laissée en exercice.

Tous ces résultats sont valides en temps discret ou continu. Toutefois, Ils ont des conséquences intéressantes en temps discret. Notamment, la proposition suivante :

PROPOSITION 1.4. Soit $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ une chaîne de Markov homogène à temps discret sur un espace d'états fini $S = \{1, 2, 3, \dots, n\}$, et soit P sa matrice de transition en un pas, et $P^{(t)}$ ses matrices de transition en un intervalle de temps $t \in \mathbb{Z}^+$.

Alors,

$$(1.4.14) \quad P^{(t)} = P^t.$$

DÉMONSTRATION. On voit immédiatement que cela est vrai pour $t = 1$. Supposons que ce soit vrai pour $t = m$:

$$P^{(m)} = P^m.$$

Alors,

$$P^{(m+1)} = P^{(m)}P^{(1)} = P^m P = P^{m+1},$$

où toutes la première égalité est par (1.4.10) ; les autres sont par hypothèse d'induction, par définition et par associativité du produit.

On a donc vérifié que si c'est vrai pour $t = m$, c'est aussi vrai pour $t = m + 1$. Par induction, puisque c'est vrai pour tout $t = 1$, c'est donc vrai pour tout $t \in \mathbb{N}$. On accepte également par convention que $P^0 = \mathbb{I}_n = P^{(0)}$. \square

EXEMPLE 1.7. En reprenant l'exemple 1.6, calculer les matrices de transition

- (a) après deux jours ;
- (b) après trois jours ;
- (c) après une semaine.
- (d) S'il fait beau aujourd'hui, quelle est la probabilité qu'il fera beau la semaine prochaine ?
- (e) Comparer $P_{1,1}^{(7)}$, $P_{2,1}^{(7)}$ et $P_{3,1}^{(7)}$. Après une semaine, la probabilité qu'il fasse beau dépend-elle beaucoup des conditions initiales ?

SOLUTION. (a) La matrice de transition en un jour est

$$P = \begin{pmatrix} 1/2 & 2/5 & 1/10 \\ 1/6 & 1/2 & 1/3 \\ 3/5 & 1/10 & 3/10 \end{pmatrix}.$$

La matrice de transition en deux jours est donc

$$\begin{aligned} P^{(2)} = P^2 &= \begin{pmatrix} 1/2 & 2/5 & 1/10 \\ 1/6 & 1/2 & 1/3 \\ 3/5 & 1/10 & 3/10 \end{pmatrix} \begin{pmatrix} 1/2 & 2/5 & 1/10 \\ 1/6 & 1/2 & 1/3 \\ 3/5 & 1/10 & 3/10 \end{pmatrix} \\ &= \begin{pmatrix} 113/300 & 41/100 & 16/75 \\ 11/30 & 7/20 & 17/60 \\ 149/300 & 8/25 & 11/60 \end{pmatrix} \end{aligned}$$

(b) La matrice de transition en trois jours est

$$\begin{aligned} P^{(3)} = P^{(2)}P &= \begin{pmatrix} 113/300 & 41/100 & 16/75 \\ 11/30 & 7/20 & 17/60 \\ 149/300 & 8/25 & 11/60 \end{pmatrix} \begin{pmatrix} 1/2 & 2/5 & 1/10 \\ 1/6 & 1/2 & 1/3 \\ 3/5 & 1/10 & 3/10 \end{pmatrix} \\ &= \begin{pmatrix} 577/1500 & 377/1000 & 143/600 \\ 247/600 & 7/20 & 143/600 \\ 247/600 & 377/1000 & 317/1500 \end{pmatrix} \end{aligned}$$

(c) La matrice de transition en six jours est :

$$\begin{aligned} P^{(6)} = (P^{(3)})^2 &= \begin{pmatrix} 577/1500 & 377/1000 & 143/600 \\ 247/600 & 7/20 & 143/600 \\ 247/600 & 377/1000 & 317/1500 \end{pmatrix} \begin{pmatrix} 577/1500 & 377/1000 & 143/600 \\ 247/600 & 7/20 & 143/600 \\ 247/600 & 377/1000 & 317/1500 \end{pmatrix} \\ &= \begin{pmatrix} 601921/1500500 & 366821/1000000 & 139139/600000 \\ 240331/600000 & 7351/20000 & 139139/600000 \\ 240331/600000 & 366821/1000000 & 348941/1500000 \end{pmatrix} \end{aligned}$$

Donc, la matrice de transition après une semaine est :

$$\begin{aligned} P^{(7)} = P^{(6)}P &= \begin{pmatrix} 601921/1500500 & 366821/1000000 & 139139/600000 \\ 240331/600000 & 7351/20000 & 139139/600000 \\ 240331/600000 & 366821/1000000 & 348941/1500000 \end{pmatrix} \begin{pmatrix} 1/2 & 2/5 & 1/10 \\ 1/6 & 1/2 & 1/3 \\ 3/5 & 1/10 & 3/10 \end{pmatrix} \\ &= \begin{pmatrix} 2405497/6000000 & 1835563/5000000 & 6959137/30000000 \\ 2404039/6000000 & 734371/2000000 & 87053/375000 \\ 1503709/3750000 & 3668939/10000000 & 6963511/30000000 \end{pmatrix} \end{aligned}$$

Avec des approximations décimales, on trouve :

$$P^{(7)} \approx \begin{pmatrix} 0,4009 & 0,3671 & 0,2320 \\ 0,4007 & 0,3672 & 0,2321 \\ 0,4010 & 0,3669 & 0,2321 \end{pmatrix}.$$

(d) La probabilité qu'il fera beau la semaine prochaine sachant qu'il fait beau aujourd'hui correspond à $P_{1,1}^{(7)} \approx 40,09\%$.

(e) On remarque que les lignes sont très semblables après une semaine – cela traduit le fait que, à long terme, les conditions sont « plus ou moins indépendantes » des conditions initiales.

Si c'est le cas, ça veut dire qu'en moyenne, il fera beau environ 40% des dimanches (en moyenne, à long terme), tandis qu'environ 37% d'entre eux seront nuageux, et environ 23% seront pluvieux. En appliquant ce même raisonnement aux autres jours de la semaine, on voit qu'en moyenne, à long terme, il fera beau 40% du temps, il y aura des nuages 37% du temps, et de la pluie 23% du temps !

Nous verrons comment formaliser ce raisonnement plus loin...

EXEMPLE 1.8. On va régler ça tout de suite : quand une chaîne de Markov a deux états possible, notre matrice de transition ne peut que ressembler à ceci :

$$P = \begin{pmatrix} (1-a) & a \\ b & (1-b) \end{pmatrix};$$

Trouver $P^{(n)}$ pour un n arbitraire ; calculer $\lim_{n \rightarrow \infty} P^{(n)}$.

SOLUTION. On veut $P^{(n)} = P^n$. Pour trouver cette valeur, il faudra *diagonaliser* P afin de pouvoir l'écrire comme

$$P = QDQ^{-1},$$

où

$$D = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$$

est une matrice diagonale. λ_1 et λ_2 sont les valeurs propres de P ; ce sont les zéros du polynôme caractéristique $\det(P - \lambda \mathbb{I})$.

Étape 1 : calculer les valeurs propres. Pour ce faire, on cherche d'abord le polynôme caractéristique. C'est le déterminant de $P - \lambda \mathbb{I}$:

$$\begin{aligned}\det(P - \lambda \mathbb{I}) &= (1 - a - \lambda)(1 - b - \lambda) - ab \\ &= (1 - a)(1 - b) - \lambda(1 - b) - \lambda(1 - a) + \lambda^2 - ab \\ &= (1 - a)(1 - b) - (2 - a - b)\lambda + \lambda^2 - ab \\ &= 1 - a - b - (2 - a - b)\lambda + \lambda^2\end{aligned}$$

On trouve :

$$\lambda_1 = \frac{(2 - a - b) + \sqrt{(2 - a - b)^2 - 4(1 - a - b)}}{2} = 1,$$

et

$$\lambda_2 = \frac{(2 - a - b) - \sqrt{(2 - a - b)^2 - 4(1 - a - b)}}{2} = 1 - a - b.$$

Pour la suite, on présume que $a, b \in (0, 1)$ parce que sinon le comportement à long terme est assez prévisible.

Étape 2 : calculer les vecteurs propres. Il nous faut un vecteur propre par valeur propre ; ce sera un vecteur \mathbf{v} qui solutionne

$$P\mathbf{v} = \lambda \mathbf{v}$$

pour la valeur propre λ de notre choix.

Pour $\lambda_1 = 1$, on cherche donc une solution $P\mathbf{v}_1 = \mathbf{v}_1$; on peut prendre $\mathbf{v}_1 = (1, 1)$, et la condition de normalisation nous garantit que ça fonctionne !

Pour $\lambda_2 = 1 - a - b$, il faut résoudre. On cherche une solution à

$$P\mathbf{v}_2 = (1 - a - b)\mathbf{v}_2,$$

, ou :

$$\begin{cases} (1 - a)v_{12} + av_{22} = (1 - a - b)v_{12} \\ bv_{12} + (1 - b)v_{22} = (1 - a - b)v_{22}. \end{cases}$$

En triturant ce système d'équations, on trouve entre autres que

$$bv_{12} = -av_{22},$$

qui peut avoir pour solution $v_{12} = a$ et $v_{22} = -b$. On peut donc prendre

$$\mathbf{v}_2 = (a, -b).$$

Étape 3 : trouver la matrice Q .

La matrice Q est celle dont les colonnes sont les vecteurs propres ; la colonne i un vecteur propre associé à la valeur propre λ_i . Donc, ici, on aurait

$$Q = \begin{pmatrix} 1 & a \\ 1 & -b \end{pmatrix},$$

par exemple.

On peut tester qu'on a trouvé la bonne matrice, en vérifiant que

$$PQ = QD.$$

$$\begin{aligned} PQ &= \begin{pmatrix} (1-a) & a \\ b & (1-b) \end{pmatrix} \begin{pmatrix} 1 & a \\ 1 & -b \end{pmatrix} \\ &= \begin{pmatrix} 1 & (1-a-b)a \\ 1 & -b(1-a-b) \end{pmatrix}. \end{aligned}$$

$$\begin{aligned} QD &= \begin{pmatrix} 1 & a \\ 1 & -b \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1-a-b \end{pmatrix} \\ &= \begin{pmatrix} 1 & a(1-a-b) \\ 1 & -b(1-a-b) \end{pmatrix}. \end{aligned}$$

Check !

Étape 4 : inverser Q . Pour ça, on peut utiliser le truc de l'échelonnage-réduisage (pour les plus grosses matrices c'est la façon la plus simple). Pour une matrice 2×2 , on peut se servir de la formule générale :

$$A = \begin{pmatrix} x & y \\ z & w \end{pmatrix} \Rightarrow A^{-1} = \frac{1}{xw - yz} \begin{pmatrix} w & -y \\ -z & x \end{pmatrix}.$$

Donc, dans notre cas, on trouve :

$$Q^{-1} = \frac{1}{ab} \begin{pmatrix} -b & -a \\ -1 & 1 \end{pmatrix};$$

on peut encore tester :

$$\begin{aligned} QQ^{-1} &= -\frac{1}{a+b} \begin{pmatrix} 1 & a \\ 1 & -b \end{pmatrix} \begin{pmatrix} -b & -a \\ -1 & 1 \end{pmatrix} \\ &= -\frac{1}{a+b} \begin{pmatrix} -(a+b) & 0 \\ 0 & -(a+b) \end{pmatrix} \\ &= \mathbb{I}. \end{aligned}$$

Check !

Étape 5 : trouver P^n . Pour faire ça, il faut juste tout mettre ensemble. On se souvient qu'on a

$$\begin{aligned} P^n &= QD^nQ^{-1} \\ &= -\frac{1}{a+b} \begin{pmatrix} 1 & a \\ 1 & -b \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & (1-a-b)^n \end{pmatrix} \begin{pmatrix} -b & -a \\ -1 & 1 \end{pmatrix} \\ &= -\frac{1}{a+b} \begin{pmatrix} 1 & a(1-a-b)^n \\ 1 & -b(1-a-b)^n \end{pmatrix} \begin{pmatrix} -b & -a \\ -1 & 1 \end{pmatrix} \\ &= -\frac{1}{a+b} \begin{pmatrix} -b-a(1-a-b)^n & -a+a(1-a-b)^n \\ -b+b(1-a-b)^n & -a-b(1-a-b)^n \end{pmatrix} \\ &= \frac{1}{a+b} \begin{pmatrix} b+a(1-a-b)^n & a-a(1-a-b)^n \\ b-b(1-a-b)^n & a+b(1-a-b)^n \end{pmatrix}. \end{aligned}$$

Étape 6 : trouver $\lim_{n \rightarrow \infty} P^n$. C'est simple : $(1 - a - b) \in (-1, 1)$, donc

$$\lim_{n \rightarrow \infty} (1 - a - b)^n = 0.$$

Il suit que

$$\lim_{n \rightarrow \infty} P^{(n)} = \frac{1}{a+b} \begin{pmatrix} b & a \\ b & a \end{pmatrix}.$$

REMARQUE. On remarque que les lignes sont exactement identiques – ce qu'on soupçonnait déjà à l'exemple précédent. On va voir plus tard que ce n'est pas un hasard !

Finalement, nous allons illustrer ici un dernier avantage de la représentation matricielle de nos probabilités de transition ; pour ce faire, nous allons introduire le vecteur de masse de la distribution de X :

DÉFINITION 1.7 (Vecteur de masse). Soit $\mathbf{X} = (X_t)_{t \geq 0}$ une chaîne de Markov sur un espace d'états fini $S = \{1, 2, 3, \dots, n\}$.

Pour tout $t \geq 0$, on définit le **vecteur de masse de la distribution de X_t** :

$$\mathbf{x}^{(t)} = (x_i^{(t)} = \mathbb{P}\{X_t = i\})_{1 \leq i \leq n} \in \mathbb{R}^{1 \times n}.$$

Attention ! Il s'agit d'un vecteur-ligne – c'est à dire une matrice d'une ligne à n colonnes, et *pas* une colonne à n lignes comme on a peut-être l'habitude...

Le vecteur $\mathbf{x}^{(0)} = (x_i^{(0)} = \mathbb{P}\{X_0 = i\})_{1 \leq i \leq n}$ donne la condition initiale ; dans le cas le plus simple où on sait qu'on commence à l'état k , par exemple, on aurait : $\mathbf{x}^{(0)} = (\delta_{k,j})_{1 \leq j \leq n}$.

Bien sûr, on doit avoir :

$$(1.4.15) \quad \sum_{j=1}^n x_j^{(t)} = 1 \quad (\Leftrightarrow \mathbf{x} \cdot \mathbf{1}_n = 1)$$

pour tout j .

On obtient alors le résultat suivant :

PROPOSITION 1.5. Soit $\mathbf{X} = (X_t)_{t \geq 0}$ une chaîne de Markov sur un espace d'états fini $S = \{1, 2, 3, \dots, n\}$. Soient $P(s, t) = (P_{i,j}(s, t))_{(i,j) \in S \times S} \in \mathbb{R}^{n \times n}$ les matrices de transitions des temps s à t pour tous $0 \leq s \leq t$

Alors, pour $0 \leq s \leq t$, on a :

$$(1.4.16) \quad \mathbf{x}^{(t)} = \mathbf{x}^{(s)} P(s, t)$$

où la multiplication est la multiplication matricielle habituelle.

Rappel : les vecteurs $\mathbf{x}^{(s)}$ et $\mathbf{x}^{(t)}$ sont des vecteurs-ligne de $\mathbb{R}^{1 \times n}$, et non des vecteurs-colonnes.

DÉMONSTRATION. Pour tout $1 \leq j \leq n$, nous avons :

$$\begin{aligned} x_j^{(t)} &= \mathbb{P}\{X_t = j\} \\ &= \sum_{i=1}^n \mathbb{P}\{X_s = i\} \mathbb{P}\{X_t = j \mid X_s = i\} \\ &= \sum_{i=1}^n x_i^{(s)} P_{i,j}(s, t) \\ &= [\mathbf{x}^{(s)} P(s, t)]_j. \end{aligned}$$

Ici, la seconde inégalité est simplement par la formule de probabilités totale ; la dernière égalité est due à la définition du produit matriciel. Les autres égalités sont simplement par définition des symboles. \square

EXEMPLE 1.9. En reprenant le processus des exemples 1.6 et 1.7, donner

- (a) La probabilité qu'il fera beau dans trois jours sachant qu'il fait beau aujourd'hui ;
- (b) la probabilité qu'il fera nuageux dans six jours sachant qu'il pleut aujourd'hui ;
- (c) la probabilité qu'il fera beau dans les cinq prochains jours sachant qu'il fait beau aujourd'hui.

SOLUTION. (a) On cherche $x_1^{(3)}$. Puisqu'on sait qu'il fait beau (aujourd'hui), on a bien sûr

$$\mathbf{x}^{(0)} = (1 \ 0 \ 0) ;$$

par conséquent, on a que

$$\begin{aligned} \mathbf{x}^{(3)} &= \mathbf{x}^{(0)} P^{(3)} \\ &= (1 \ 0 \ 0) \begin{pmatrix} 577/1500 & 377/1000 & 143/600 \\ 247/600 & 7/20 & 143/600 \\ 247/600 & 377/1000 & 317/1500 \end{pmatrix} \\ &= (577/1500 \ 377/1000 \ 143/600) , \end{aligned}$$

et

$$\mathbf{x}_1^{(3)} = \frac{577}{1500} \approx 38,47\%.$$

- (b) De même, on cherche $x_2^{(6)}$. Puisqu'on sait qu'il pleut aujourd'hui, on a bien sûr

$$\mathbf{x}^{(0)} = (0 \ 0 \ 1) ;$$

par conséquent, on a que

$$\begin{aligned} \mathbf{x}^{(6)} &= \mathbf{x}^{(0)} P^{(6)} \\ &= (0 \ 0 \ 1) \begin{pmatrix} 601921/1500500 & 366821/1000000 & 139139/600000 \\ 240331/600000 & 7351/20000 & 139139/600000 \\ 240331/600000 & 366821/1000000 & 348941/1500000 \end{pmatrix} \\ &= (240331/600000 \ 366821/1000000 \ 348941/1500000) , \end{aligned}$$

et

$$\mathbf{x}_2^{(6)} = \frac{366821}{1000000} \approx 36,68\%.$$

- (c) Ici, on cherche $(P_{1,1})^5 = \frac{1}{32}$.

1.5. La représentation en graphes et les classes d'équivalence.

Pour le reste du chapitre 1, nous allons nous concentrer plus particulièrement sur les chaînes de Markov à temps discret sur des espaces d'états finis – disons avec n états.

Nous venons de voir que, dans ce cas précis, il est possible

- i. d'énumérer les états en les numérotant de 1 à n ;
- ii. de caractériser la loi de la chaîne de Markov par ses probabilités de transition en un pas ; et
- iii. de représenter ces probabilités de transition sous la forme d'une matrice, et la fonction de masse de X_t pour tout t comme un vecteur-ligne ; le vecteur de masse $\mathbf{x}^{(t)}$.

Ce qu'on cherche à faire, maintenant, c'est de faire une description plus « intuitive » de nos chaînes de Markov, et de voir s'il est possible de déduire certaines choses simplement d'après la matrice de transition directement.

1.5.1. Graphe des transitions possibles.

DÉFINITION 1.8. Soit $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ une chaîne de Markov homogène à temps discret sur l'espace d'états fini $S = \{1, 2, 3, \dots, n\}$, avec matrice de transition en un pas $P = (P_{i,j})_{1 \leq i,j \leq n} \in \mathbb{R}^{n \times n}$.

Le **graphe des transitions possibles** pour \mathbf{X} est le graphe orienté⁸ $\Gamma = (S, E_P)$ où les sommets sont les états S , et les arêtes sont l'ensemble E_P des paires ordonnées de sommets (i, j) telles qu'une transition en un pas est possible de i vers j ; plus précisément :

$$(1.5.1) \quad E_P = \{(i, j) \in S^2 : P_{i,j} > 0\} \subseteq S^2.$$

EXEMPLE 1.10. Dans une usine de petits caramels durs, une machine indispensable à la production est monitorée chaque jour par le personnel. Si la machine fonctionne bien, la qu'elle soit défectueuse le lendemain est de $1/20$; alors, l'équipe technique tente de la réparer. Chaque jour, la probabilité de réussir à réparer la machine est de $4/5$. En revanche, la probabilité que la machine soit cassée de façon irréparable est de $1/100$, et dans ce cas, elle sera remplacée par une machine neuve le lendemain.

- (a) Décrire le processus stochastique à l'étude. Quels sont les états importants ?
- (b) Donner la matrice des probabilités de transition en un pas. À quoi un « pas » correspond-t-il dans notre mise en situation ?
- (c) Tracer le graphe des transitions possibles.

SOLUTION. (a) Ici, on s'intéresse à l'état de la machine. La machine peut être :

- (1) Fonctionnelle ;
- (2) Endommagée ;
- (3) En remplacement.

- (b) Dans le texte, on nous donne $P_{1,2} = 1/20$ et $P_{1,3} = 1/100$. Il suit que $P_{1,1} = 47/50$. On nous dit également que si la machine est en remplacement, elle sera remplacée par une machine fonctionnelle le lendemain. Donc, $P_{3,1} = 1$; bien sûr, on a donc $P_{3,2} = P_{3,3} = 0$. Finalement, si la machine est endommagée, elle est réparée (et donc fonctionne le lendemain) avec probabilité $4/5$. Donc, $P_{2,1} = 4/5$. Il est sous-entendu que si l'équipe

8. Consulter au besoin l'annexe A sur le vocabulaire de la théorie des graphes et les définitions de base.

technique n'a pas réussi un jour, la machine demeure endommagée (vu que l'équipe tentera encore de la réparer le lendemain). Donc, on a $P_{2,2} = 1/5$ et $P_{2,3} = 0$.

En mettant tout ceci ensemble, on a donc :

$$P = \begin{pmatrix} 47/50 & 1/20 & 1/100 \\ 4/5 & 1/5 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

(c) Le graphe associé à cette matrice est représenté à la figure 1.3

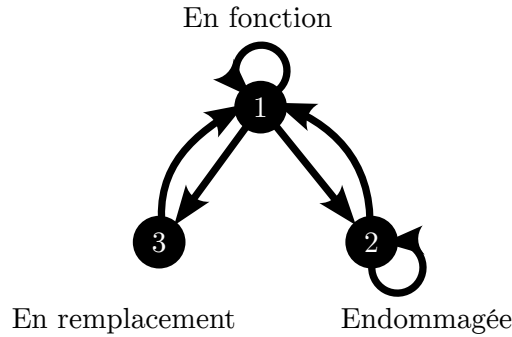


FIGURE 1.3 – Le graphe des transitions possibles pour la machine de l'exemple 1.10

Rappels : *les graphes orientés*. Le graphe des transitions possibles nous servira en quelque sorte de « carte » de notre espace d'états, et il sera très utile pour comprendre plus intuitivement la dynamique du processus à l'étude. Pour cette raison, ça vaut la peine de faire un petit rappel du vocabulaire de base de la théorie des graphes – en particulier pour les graphes orientés qu'on utilise ici.

DÉFINITION (Vocabulaire des graphes orientés). À noter que ces définitions sont des rappels ; pour des versions plus détaillées et complètes, consulter l'annexe A, qui présente les définitions importantes pour les graphes orientés et non-orientés, de façon très générale.

- i. Un **graphe orienté** $G = (V, E)$ est un objet mathématique constitué d'un ensemble de sommets V et d'un ensemble d'arêtes $E \subseteq V^2$.
- ii. Une **arête** $e \in E$ est représentée par une paire ordonnée de deux sommets (x, y) , avec $x, y \in V$. On dit que l'arête e **part de** x et qu'elle **se termine en** y , ou qu'elle est **incidente à** y .

On note $x \rightarrow e$ et $e \rightarrow y$ respectivement pour indiquer que l'arête e part de x et se termine en y .

Pour une arête $e \in E$, on note $e^- \in X$ le sommet d'où l'arête e part, et e^+ le sommet où l'arête e se termine.

- iii. On dit que le sommet $y \in V$ est **adjacent** au sommet $x \in V$ si et seulement si l'arête (x, y) est dans E . On note alors $x \rightarrow y$.

Attention : Dans les graphes orientés, l'adjacence n'est pas symétrique ; on peut avoir y adjacent à x sans que x soit adjacent à y .

On dit que l'arête f **suit** l'arête e si $e^+ = f^-$; c'est à dire que l'arête e se termine au même sommet où f commence. L'arête e est **incidente** sur l'arête f . On note $e \rightarrow f$. On dit également que les arêtes e et f se suivent, ou s'enchaînent.

- iv. Un **chemin fini de longueur** n est une suite d'arêtes $C = (e_1, e_2, e_3, \dots, e_n)$ où les arêtes consécutives s'enchaînent ; c'est-à-dire que $e_{i-1}^+ = e_i^-$ pour tout $i = 2, 3, 4, \dots, n$. Par abus de notation, on notera $|C|$ la **longueur** du chemin C (même si C n'est pas un ensemble).

Un tel chemin peut également être décrit par une famille de $n+1$ sommets $(x_0, x_1, x_2, \dots, x_n)$, à condition que les arêtes $(x_{i-1}, x_i) = e_i$ soient toutes incluses dans E .

- v. Un chemin fini passant par les sommets $(x_0, x_1, x_2, \dots, x_n)$ est un **cycle** si $x_0 = x_n$.
- vi. On dit que le sommet $y \in V$ est **accessible depuis** $x \in V$ si il existe un chemin fini passant par les sommets $(x_0, x_1, x_2, \dots, x_n)$ tel que $x_0 = x$ et $x_n = y$. On note cela $x \longrightarrow y$.

On dit que les sommets x et y sont **connectés** si x est accessible depuis y et y est accessible depuis x ; on note cela $x \longleftrightarrow y$. On admet également que le sommet x est accessible depuis lui-même par le chemin de longueur 0 ($x_0 = x$), et qu'il est donc connecté à lui-même.

- vii. Pour tous $x, y \in V$, la **distance** de x à y – notée $d(x, y)$ – est nulle lorsque $x = y$. Lorsque $x \neq y$, c'est la longueur du plus court chemin reliant x à y , s'il existe. *Attention* : dans les graphes orientés, la distance n'est pas symétrique ! On n'a pas forcément que $d(x, y) = d(y, x)$.

Ouf ! C'est beaucoup de vocabulaire, mais ça demeure relativement intuitif.

1.5.2. Classification des états. La suite de notre discussion repose sur un constat important concernant la relation de connexité entre deux sommets :

LEMME 1.1. *Soit $G = (V, E)$ un graphe orienté. Alors, la relation \longleftrightarrow sur les sommets S est une **relation d'équivalence** entre les sommets. C'est à dire que :*

- i. La relation est **réflexive** : $x \longleftrightarrow x$ pour tous $x \in V$;
- ii. La relation est **symétrique** : $x \longleftrightarrow y$ si et seulement si $y \longleftrightarrow x$, pour tous $x, y \in V$;
- iii. La relation est **transitive** : Pour tous $x, y, z \in V$, si $x \longleftrightarrow y$ et $y \longleftrightarrow z$, alors $x \longleftrightarrow z$.

DÉMONSTRATION. **i** Par l'item **vi** de la définition, c'est immédiat.

- ii** C'est également évident par la définition ; si $x \longleftrightarrow y$, alors $y \longrightarrow x$ et $x \longrightarrow y$, et il suit que $y \longleftrightarrow x$, et vice-versa.

- iii** Pour montrer la transitivité, on va « rabouter » des chemins ensemble ; ce qu'on appelle la *concaténation* :

Si $x \longleftrightarrow y$ et $y \longleftrightarrow z$, alors en particulier $x \longrightarrow y$ et $y \longrightarrow z$. Alors, il existe des chemins $C_1 = (e_1, e_2, e_3, \dots, e_m)$ et $C_2 = (f_1, f_2, f_3, \dots, f_n)$ tels que C_1 va de x à y , et C_2 va de y à z . Mais alors, on considère le chemin

$$C = (e_1, e_2, e_3, \dots, e_m, f_1, f_2, \dots, f_n)$$

C'est un chemin, puisque les arêtes consécutives s'enchaînent – on vérifie que $e_m^+ = y = f_1^-$; les autres paires d'arêtes consécutives s'enchaînent puisqu'elles sont à l'intérieur des chemins C_1 ou C_2 .

Or, le chemin C part de x et se termine en z ; donc, il existe un chemin de x à z , et $x \longrightarrow z$.

De façon similaire, on a aussi que $z \longrightarrow y$ et $y \longrightarrow x$, donc que $z \longrightarrow x$. On conclue finalement que $x \longleftrightarrow z$.

□

Et la propriété suivante des relations d'équivalence (citée sans preuve) :

PROPOSITION. Soit A un ensemble quelconque et soit \sim une relation d'équivalence entre les éléments de A .

Alors, le **quotient** de A par \sim , noté A/\sim , est une partition (non-indexée) de A en **classes d'équivalences**. Plus précisément, A/\sim est un ensemble avec les propriétés suivantes :

- i. Les éléments de A/\sim sont des parties de A appelées des **classes d'équivalence pour \sim** – si $C \in A/\sim$, alors $C \subseteq A$.
- ii. Pour toute classe d'équivalence $C \in A/\sim$, on a que tous les éléments de C sont équivalents entre eux : pour tous $x, y \in C$, $x \sim y$.
- iii. Pour tout élément $x \in A$, x appartient à exactement un (unique) $C \in A/\sim$; autrement dit, les éléments de A/\sim forment une partition de A .

En gros : aussitôt qu'on a une relation d'équivalence, on a tout de suite un genre de « découpage » de notre ensemble en différentes classes d'équivalence.

De retour à notre graphe des transitions possibles : si $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ est une chaîne de Markov homogène à temps discret sur un ensemble d'états fini $S = \{1, 2, 3, \dots, n\}$, avec sa matrice de transition en un pas $P = (P_{i,j})_{1 \leq i, j \leq n}$ et son graphe des transitions possibles $\Gamma = (S, E)$, alors on a le résultat suivant :

PROPOSITION 1.6. L'espace des états S se décompose en un nombre fini de classes d'équivalence par la relation de connexité \longleftrightarrow .

DÉMONSTRATION. C'est simplement une conséquence de la propriété générale des relations d'équivalence. □

EXEMPLE 1.11. Les états de la chaîne décrite à l'exemple 1.10 sont tous dans la même classe d'équivalence, puisqu'ils sont tous connectés entre eux dans le graphe des transitions possibles (figure 1.3).

EXEMPLE 1.12. On un dé tétraédrique (à quatre faces) à répétition. Soit X_i le maximum de tous les lancers observés depuis le début jusqu'au lancer i inclusivement. On considère $\mathbf{X} = (X_i)_{i \in \mathbb{N}}$.

- (a) Vérifier qu'il s'agit d'une chaîne de Markov homogène à temps discret. Donner l'espace des états et la matrice de transition en un pas.
- (b) Tracer le graphe des transitions possibles pour cette chaîne de Markov.
- (c) Identifier toutes les classes d'équivalence.

SOLUTION. (a) Les états possibles ici sont $S = \{1, 2, 3, 4\}$, correspondant aux valeurs possibles que le maximum pourra prendre (ce sont les valeurs que le dé peut afficher).

C'est bien une chaîne de Markov – si on connaît le maximum après i lancer, on n'a pas besoin de connaître les valeurs précédentes pour déterminer les probabilités de transition vers les autres valeurs. En effet :

- Si le maximum était 1, ça restera 1 si on roule encore un 1. Si on roule quoi que ce soit d'autre, cette nouvelle valeur deviendra le nouveau maximum. Donc,

$$P_{1,1} = P_{1,2} = P_{1,3} = P_{1,4} = \frac{1}{4};$$

- Si le maximum était 2, ça restera 2 si on roule un 1 ou un 2. Si on roule un 3 ou un 4, ça deviendra notre nouveau maximum. Il est impossible que le maximum devienne 1, puisque c'est déjà 2. Donc,

$$P_{2,1} = 0, \quad P_{2,2} = \frac{1}{2}, \quad P_{2,3} = P_{2,4} = \frac{1}{4};$$

- Si le maximum était 3, alors si on roule un 4, le maximum deviendra 4 ; sinon, le maximum restera 3. Donc :

$$P_{3,1} = P_{3,2} = 0, \quad P_{3,3} = \frac{3}{4}, \quad P_{3,4} = \frac{1}{4};$$

- Si le maximum était déjà 4, ça restera 4, quoi qu'on fasse.

$$P_{4,1} = P_{4,2} = P_{4,3} = 0, \quad P_{4,4} = 1.$$

Finalement, on a la matrice de transition suivante :

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

- (b) Le graphe des transitions possibles est présenté à la figure 1.4.
- (c) Ici, on remarque qu'aucun deux états ne sont connectés entre eux ; en effet, pour l'état j , tous les états $i < j$ sont inaccessibles ! Donc si j est accessible depuis i , c'est que $i \leq j$. Par conséquent, chaque état est seul dans sa propre petite classe d'équivalence.

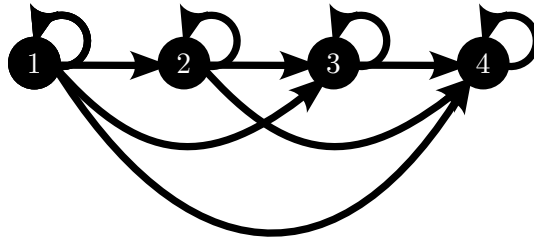


FIGURE 1.4 – Le graphe des transitions possibles pour la chaîne de Markov de l'exemple 1.12.

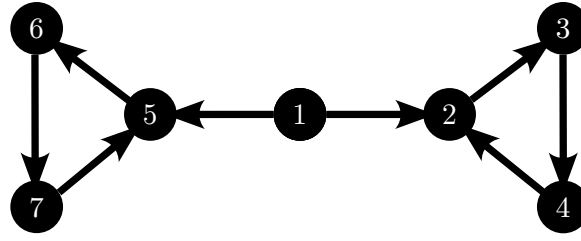


FIGURE 1.5 – Le graphe des transitions possibles pour la chaîne de Markov de l'exemple 1.13.

EXEMPLE 1.13. Soit $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ une chaîne de Markov homogène à temps discret sur l'espace fini des états $S = \{1, 2, 3, 4, 5, 6, 7\}$, avec la matrice de transition en un pas

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

- (a) Tracer le graphe des transitions possibles.
 (b) Identifier les classes d'équivalence.

SOLUTION. (a) Le graphe est présenté à la figure 1.5.

- (b) Il y a trois classes d'équivalence :

- L'état 1 est tout seul dans sa classe d'équivalence car il n'est connecté à aucun autre état, puisqu'il n'est accessible depuis aucun autre état.
- Les états 2, 3 et 4 sont dans une classe, puisqu'ils sont tous connectés entre eux (ils font tous partie d'un cycle).
- Les états 5, 6 et 7 sont dans une classe puisqu'ils sont tous connectés entre eux (ils font tous partie d'un cycle).

On introduit les définitions suivantes :

DÉFINITION 1.9 (Classe fermée, classe ouverte). Soit $G = (V, E)$ un graphe orienté, et $C \in V / \longleftrightarrow$ une classe d'équivalence de sommets connectés entre eux.

- i. C est une classe **ouverte** si il existe une arête $e \in E$ telle que $e^- \in C$ et $e^+ \notin C$.
- ii. Si C n'est pas une classe ouverte, c'est une classe **fermée**.

Autrement dit, une classe est *ouverte* si un chemin qui y commence peut en sortir ; elle est fermée si tout chemin qui y commence doit forcément y rester.

On a immédiatement le résultat suivant :

PROPOSITION 1.7. Soit $G = (V, E)$ un graphe orienté sur un nombre fini de sommets. Il existe au moins une classe d'équivalence $C \in G / \longleftrightarrow$ qui est fermée.

DÉMONSTRATION. *La preuve de cette proposition n'est pas nécessaire mais elle est amusante.*

On considère le graphe orienté $\Gamma = (G/\longleftrightarrow, \mathcal{E})$, où les sommets sont les classes d'équivalence du graphe G , et on choisit :

$$\mathcal{E} = \{(C, C') \in (G/\longleftrightarrow)^2 : \exists x \in C, y \in C' : (x, y) \in E, C \neq C'\}.$$

En mots : il y a une arête de C à C' si et seulement si il existe une arête dans G entre un sommet de C et un sommet de C' (et que C et C' sont différentes).

Le graphe Γ n'admet aucun cycle. Si il y avait un cycle passant par les classes C et C' , par exemple, tous les sommets de C et de C' seraient équivalents les uns aux autres – donc $C = C'$.

D'un autre côté, si on suppose que toutes les classes C sont ouvertes, alors il existe toujours au moins une arête dans le graphe Γ qui part de C et qui se termine à une autre classe. On commence avec une classe quelconque, que l'on appelle C_0 . Puis, pour tout $i \in \mathbb{N}$, on choisit un C_{i+1} parmi les classes adjacentes à C_i ; on construit ainsi une suite infinie de classes d'équivalence. Or, puisque V est fini, G/\longleftrightarrow est également fini ; il y a un nombre fini de classes d'équivalences, et donc forcément, il existe deux entiers i et j avec $i < j$ et tels que $C_i = C_j$. Le chemin dans Γ qui passe par les sommets $(C_i, C_{i+1}, C_{i+2}, \dots, C_j)$ est donc un cycle – ce qui est une contradiction, puisque comme nous venons de le montrer, le graphe Γ n'admet pas de cycles.

On est donc forcé-e-s d'admettre que le graphe Γ admet au moins un sommet C qui n'est le point d'origine d'aucune arête – c'est à dire qu'il existe au moins une classe d'équivalence C qui est fermée. \square

EXEMPLE 1.14. Pour les processus décrits dans les exemples 1.10, 1.12 et 1.13, identifier les classes d'équivalences et dire si elles sont ouvertes ou fermées.

SOLUTION. Pour le processus de l'exemple 1.10 (figure 1.3), il n'y a qu'une classe d'équivalence. Elle est donc fermée.

Pour le processus de l'exemple 1.12 (figure 1.4), la classe d'équivalence contenant l'état 4 est fermée ; toutes les autres ont un lien vers cette dernière, et sont par conséquent ouvertes.

Pour le processus de l'exemple 1.13 (figure 1.5), la classe d'équivalence contenant l'état 1 est ouverte ; les deux autres sont fermées.

Mais pourquoi diantre ces classes d'équivalences nous intéressent-elles ? C'est parce que, comme nous allons le voir sous peu, les états de la même classe d'équivalence partagent certaines propriétés...

1.6. Temps d'atteinte et de retour, récurrence, transience, périodicité

Nous venons de voir une façon simple de « cartographier » notre espace des états, et d'en faire une petite « topologie ». ⁹ Par contre, *a priori*, on se souvient que ce qui nous intéresse, ce sont les probabilités et les statistiques de la chaîne ; elles ont été presque entièrement évacuées lors de la dernière section, où on a principalement décrit des graphes.

On retourne donc à nos processus de Markov en introduisant les définitions suivantes :

9. Ici, j'emploie le mot *topologie* au sens « semi-mathématique » du terme ; c'est à dire dans un sens vaguement relié à la notion mathématique rigoureuse de topologie, compris comme « l'étude de la connexité des espaces ».

DÉFINITION 1.10 (Temps d'atteinte, temps de retour). Soit $\mathbf{X} = (X_t)_{t \geq 0}$ un processus stochastique quelconque sur un espace d'états quelconque S .

On introduit les variables aléatoires suivantes :

- i. Le **temps d'atteinte** (ou temps de frappe, de l'anglais *hitting time*) de la région $A \subseteq S$ – que nous noterons τ_A – correspond au « moment où le processus entre en A pour la toute première fois » (figure 1.6). On définit rigoureusement :

$$(1.6.1) \quad \tau_A = \inf \{t \geq 0 : X_t \in A\}.$$

Par abus de notation, pour un état $x \in S$, on admet $\tau_x := \tau_{\{x\}}$.

- ii. Le **temps d'atteinte positif** de la région A est défini de façon analogue, mais en excluant 0. On le note τ_A^+ :

$$(1.6.2) \quad \tau_A = \inf \{t > 0 : X_t \in A\}.$$

Et on fait le même abus de notation : pour $x \in S$, $\tau_x^+ := \tau_{\{x\}}^+$.

- iii. Le **temps de retour à l'origine** est le temps d'atteinte positif de X_0 ($\tau_{X_0}^+$) ; c'est le temps requis pour « revenir au point de départ ».

En pratique, comme on considère souvent la mesure de probabilités sachant que $X_0 = x$ pour un certain $x \in S$, le temps de retour à l'origine sera alors τ_x^+ .

Par convention, on admet que ces temps sont infinis si l'ensemble dont on prend l'infimum est vide (c'est-à-dire si on n'atteint jamais la région A ou l'état x , par exemple).

REMARQUE. Ces définitions sont très générales ; dans le cas précis où nous sommes en temps discret, les infimums seront en fait des *minimums*.

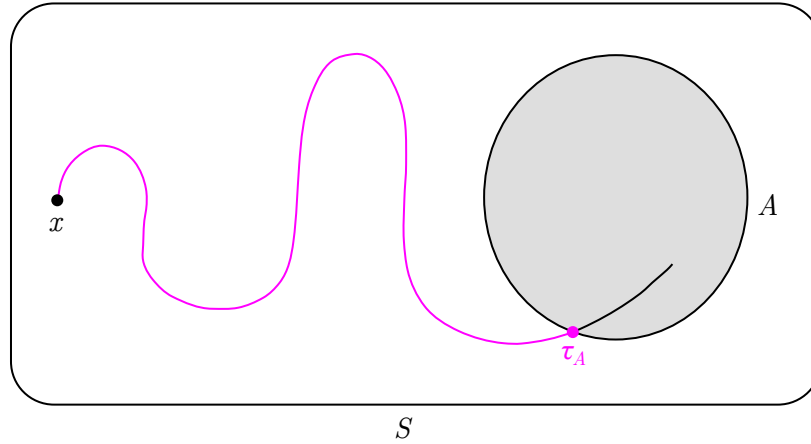


FIGURE 1.6 – Le temps d'atteinte de la région A correspond au premier moment où on touche A .

Pour un processus stochastique $\mathbf{X} = (X_t)_{t \geq 0}$ quelconque sur un espace d'états S quelconque, avec $x \in S$ un état et $A \subseteq S$ une région de S (on suppose $x \notin A$), on peut s'intéresser aux quantités suivantes¹⁰ :

- $\mathbb{P}_x \{\tau_A < \tau_B\}$: la probabilité qu'on atteigne A avant la région B ;

10. Il est à noter que dans l'ouvrage de Sabin Lessard, l'auteur utilise les symboles suivants (section 1.6.1) :

- $\mathbb{P}_x \{\tau_A < \tau_x^+\}$: la probabilité qu'on atteigne A *avant* de revenir en x ;
- $\mathbb{P}_x \{\tau_A < +\infty\}$: la probabilité qu'on atteigne éventuellement A sachant qu'on part de x ;
- $\mathbb{P}_x \{\tau_x^+ < +\infty\}$: la probabilité que le temps de retour soit fini sachant qu'on part de x – c'est à dire la probabilité qu'on revienne éventuellement à x .
- $\mathbb{E}_x [\tau_x^+], \mathbb{E}_x [\tau_A]$: Dans le cas où $\mathbb{P}_x \{\tau_x^+ < +\infty\}$, on peut s'intéresser à l'espérance de τ_x^+ ; c'est le « temps moyen pour revenir en x ». De même, on peut s'intéresser à l'espérance du temps pour atteindre la région A .
- etc.

1.6.1. La méthode du conditionnement par le premier pas. Toutes les définitions que nous avons introduites jusqu'à présent sont dans le contexte très général des processus stochastiques ; toutefois, comme nous allons le voir, dans le contexte particulier des chaînes de Markov, et plus spécifiquement des chaînes de Markov à temps discret sur des espaces d'états finis, il existe une méthode simple pour calculer ces quantités : c'est la méthode du conditionnement par le premier pas.

Afin de l'illustrer, nous allons commencer par un exemple.

EXEMPLE 1.15. Alexe se promène dans un labyrinthe constitué de neuf « pièces » carrées dans une configuration 3×3 , et où les pièces adjacentes communiquent entre elles par des portes. Alexe commence dans l'un des coins ; la sortie est dans le coin opposé, et un meurtrier sanguinaire se cache dans l'un des deux autres coins. (figure 1.7)

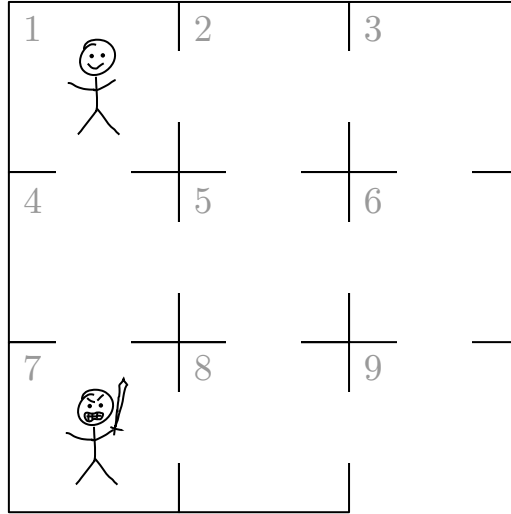


FIGURE 1.7 – Le labyrinthe. Alexe commence dans la pièce 1 ; le meurtrier est dans la pièce 7, et la sortie est dans la pièce 9.

Chaque fois qu'elle arrive dans une pièce, Alexe choisit aléatoirement l'une des portes dans la pièce de façon équiprobable. Si Alexe atteint la sortie, elle s'évade ; si elle croise le meurtrier sanguinaire, celui-ci la découpe en petits morceaux.

-
- $f_{i,j} = \mathbb{P}_i \{\tau_j^+\}$, la probabilité qu'on atteigne l'état j (ou qu'on revienne en i si $j = i$) sachant qu'on part de l'état i ;
 - $\mu_i = \mathbb{E}_i [\tau_i^+]$ l'espérance du temps du premier retour en i sachant qu'on part de l'état i .

- (a) Donner la matrice de transition en un pas pour ce processus.
 (b) Quelle est la probabilité qu'Alexe réussisse à s'échapper sans croiser le meurtrier sanguinaire ?

SOLUTION. (a) On a la matrice de transition suivante :

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

Notez que les états 7 et 9 sont davantage « Alexe est coupée en petits morceaux » et « Alexe s'est sauvée », plutôt que simplement les numéros des pièces...

- (b) Ce qu'on cherche, ici, c'est la probabilité d'atteindre éventuellement l'état 9 avant l'état 7, sachant que l'on part de l'état 1 ; $\mathbb{P}_1 \{ \tau_9 < \tau_7 \}$.

Cela peut être compliqué de calculer cette valeur directement ; il existe une infinité de trajectoires possibles qui permettraient de passer de 1 à 9 dans cette chaîne de Markov. Cependant, on peut employer une astuce :

Si on note $u_i = \mathbb{P}_i \{ \tau_9 < \tau_7 \}$ la probabilité d'atteindre l'état 9 sachant que l'on commence en i , alors pour tout i , on a l'équation suivante :

$$u_i = \sum_{j:i \rightarrow j} P_{i,j} u_j.$$

On obtient cette équation en *conditionnant sur le premier pas* :

$$\begin{aligned} u_i &= \mathbb{P}_i \{ \tau_9 < \tau_7 \} \\ &= \sum_{j \in S} \mathbb{P}_i \{ \tau_9 < \tau_7 \mid X_1 = j \} \mathbb{P}_i \{ X_1 = j \} \\ &= \sum_{j \in S} P_{i,j} \mathbb{P}_i \{ \tau_9 < \tau_7 \mid X_1 = j \}. \end{aligned}$$

La seconde égalité est donnée par la loi des probabilités totales, pour tous les j adjacents à l'état i (c'est à dire tous les états j accessibles en un pas depuis l'état i). La dernière égalité est par définition de $P_{i,j}$.

Or, par la propriété de Markov, on a bien sûr que

$$\mathbb{P}_i \{ \tau_9 < \tau_7 \mid X_1 = j \} = \mathbb{P}_j \{ \tau_9 < \tau_7 \};$$

en effet, une fois que nous savons que nous sommes à l'état j , nous savons également que la probabilité d'atteindre l'état 9 éventuellement est la même que si nous avions simplement commencé à l'état j .

Donc, on a finalement

$$\begin{aligned} u_i &= \sum_{j \in S} P_{i,j} \mathbb{P}_i \{ \tau_9 < \tau_7 \mid X_1 = j \} \\ &= \sum_{j \in S} P_{i,j} u_j. \end{aligned}$$

De plus, on doit avoir que $u_7 = 0$ et $u_9 = 1$, parce que si on commence découpée en petits morceaux, il n'y a aucune chance de s'échapper, et que si on commence déjà en dehors du labyrinthe, la probabilité de gagner est tout de suite 1.

En l'occurrence, ceci donne le système d'équations linéaire suivant : ¹¹

$$P\mathbf{u} = \mathbf{u},$$

où \mathbf{u} est le vecteur colonne donné par

$$\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \\ u_6 \\ u_7 \\ u_8 \\ u_9 \end{pmatrix}.$$

En extension, ce système d'équations est :

$$\left\{ \begin{array}{l} u_1 = \frac{1}{2}(u_2 + u_3) \\ u_2 = \frac{1}{3}(u_1 + u_3 + u_5) \\ u_3 = \frac{1}{2}(u_2 + u_6) \\ u_4 = \frac{1}{3}(u_1 + u_5 + u_7) \\ u_5 = \frac{1}{4}(u_2 + u_4 + u_6 + u_8) \\ u_6 = \frac{1}{3}(u_3 + u_5 + u_7) \\ u_7 = 0 \\ u_8 = \frac{1}{3}(u_5 + u_7 + u_9) \\ u_9 = 1. \end{array} \right.$$

11. Attention – ceci n'est pas forcément vrai en général, même si c'est le cas ici.

On peut simplifier ce système d'équations drastiquement en remarquant que, par symétrie, $u_2 = u_5 = u_8 = \frac{1}{2}$. Quoiqu'il en soit, on trouve que la solution est

$$\mathbf{u} = \begin{pmatrix} 2/5 \\ 1/2 \\ 3/5 \\ 3/10 \\ 1/2 \\ 7/10 \\ 0 \\ 1/2 \\ 1 \end{pmatrix}.$$

En observant ces résultats, on se convainc assez facilement que En particulier, on cherchait la probabilité d'atteindre 9 en partant de 1 ; c'est

$$u_1 = \mathbb{P}_1 \{ \tau_9 < \tau_7 \} = 2/5.$$

Le raisonnement que l'on fait ici n'est pas si spécifique à notre problème ; on peut donc en extraire facilement le résultat suivant (plus général) :

LEMME 1.2. *Soit $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ une chaîne de Markov homogène à temps discret sur l'espace d'états dénombrables $S = \{1, 2, 3, \dots\}$; soient $k, l \in S$ deux états de S distincts. Alors,*

i. *Alors on a le systèmes d'équations suivant :*

$$(1.6.3) \quad \begin{cases} \mathbb{P}_i \{ \tau_k < \tau_l \} = \sum_{j \in S} P_{i,j} \mathbb{P}_j \{ \tau_k < \tau_l \} & \forall i \neq k, l \\ \mathbb{P}_k \{ \tau_k < \tau_l \} = 1 \\ \mathbb{P}_l \{ \tau_k < \tau_l \} = 0 \end{cases}$$

ii. *De plus, $\mathbb{P}_l \{ \tau_k < \tau_l^+ \}$ est donné par*

$$(1.6.4) \quad \mathbb{P}_l \{ \tau_k < \tau_l^+ \} = \sum_{j \in S} P_{l,j} \mathbb{P}_j \{ \tau_k < \tau_l \}.$$

DÉMONSTRATION. i Clairement, par définition de τ_k, τ_l et \mathbb{P}_x , il est évident que

$$\mathbb{P}_k \{ \tau_k < \tau_l \} = 1, \quad \mathbb{P}_l \{ \tau_k < \tau_l \} = 0.$$

Pour tout autre état $i \in S$, on a :

$$\begin{aligned} \mathbb{P}_i \{ \tau_k < \tau_l \} &= \sum_{j \in S} \mathbb{P}_i \{ \tau_k < \tau_l \mid X_1 = j \} \mathbb{P}_i \{ X_1 = j \} \\ &= \sum_{j \in S} P_{i,j} \mathbb{P}_j \{ \tau_k < \tau_l \}, \end{aligned}$$

où la première égalité est donnée par la loi des probabilités totales en conditionnant sur le premier pas, et la seconde égalité tient au fait que, par la propriété de Markov, une fois le premier pas effectué,

$$\mathbb{P}_i \{ \tau_k < \tau_l \mid X_1 = j \} = \mathbb{P}_j \{ \tau_k < \tau_l \},$$

puisque la suite du processus suit exactement la même loi que si on avait simplement commencé en j .

- ii On remarque que pour tout $j \neq l$, $\mathbb{P}_j \{\tau_k < \tau_l^+\} = \mathbb{P}_j \{\tau_k < \tau_l\}$, puisqu'on sait qu'on n'est de toute façon pas en l au temps 0.

On applique donc le même raisonnement que pour la preuve de i.

□

EXEMPLE 1.16. En reprenant le labyrinthe de l'exemple 1.15, on suppose maintenant que le meurtrier sanguinaire est seul dans le labyrinthe, et qu'il part se promener aléatoirement (comme l'aurait fait Alexe si elle était partie du même point de départ).

Quelle est la probabilité que le meurtrier trouve la sortie *avant* d'être revenu à son point d'origine ?

SOLUTION. On a déjà calculé $\mathbb{P}_i \{\tau_9 < \tau_7\}$ pour $i = 1, 2, 3, \dots, 9$; ces valeurs sont données par les u_i trouvés dans la solution de l'exemple 1.15.

Si on suppose que le meurtrier sanguinaire part de la pièce numéro 7, c'est donc dire qu'on cherche $\mathbb{P}_7 \{\tau_9 < \tau_7^+\}$; comme nous venons de le voir avec le Lemme 1.2, on a que

$$\mathbb{P}_7 \{\tau_9 < \tau_7^+\} = \frac{1}{2} \mathbb{P}_4 \{\tau_9 < \tau_7\} + \frac{1}{2} \mathbb{P}_8 \{\tau_9 < \tau_7\} = \frac{1}{2} \left(\frac{3}{10} + \frac{1}{2} \right) = \frac{2}{5}.$$

On va laisser de côté $\mathbb{P}_x \{\tau_x^+ < +\infty\}$ et $\mathbb{P}_x \{\tau_A < +\infty\}$ pour l'instant; bien qu'une méthode similaire soit applicable pour ces événements également, nous verrons plus loin que ce sont des cas de figure un peu particuliers.

On va s'intéresser plutôt à l'espérance des temps d'atteinte avec un exemple fameux :

EXEMPLE 1.17 (La ruine du parieur). Aymeric et Brandon jouent à un périlleux jeu de hasard, au risque de leurs vastes fortunes respectives – en tout, ils possèdent N dollars.

À chaque partie, les deux misent 1\$; le joueur qui l'emporte remporte la mise et sa fortune personnelle s'accroît de 1\$ tandis que celle de l'autre diminue du même montant. Comme Aymeric et Brandon ne se font aucunement confiance, ils arrêtent de jouer lorsque l'un d'entre eux a acquis la fortune entière de l'autre (puisque le plus riche refusera systématiquement de faire crédit au perdant).

Les parties sont indépendantes; chacune est remportée par Aymeric avec probabilité p et par Brandon avec probabilité $q := (1 - p)$.

- Décrire le processus stochastique d'intérêt; donner les états et la matrice de transition.
- En supposant qu'Aymeric commence avec a dollars (et Brandon avec $b = N - a$ dollars), donner la probabilité que celui-ci remporte toute la fortune de Brandon.
- Trouver l'espérance du nombre de parties qui seront jouées dans le cas où $p = q$.

SOLUTION. (a) On considère X_t la fortune d'Aymeric après t parties jouées; les états possibles sont $S = \{0, 1, 2, \dots, N\}$; la matrice de transition (dont les indices commencent à 0) est donnée par les équations suivantes :

$$P_{0,0} = P_{N,N} = 1; \quad P_{i,i+1} = p, P_{i,i-1} = q \quad \forall i : 1 \leq i \leq N-1.$$

Tous les autres éléments de la matrice sont nuls.

- On cherche $\mathbb{P}_a \{\tau_N < \tau_0\}$; on va définir $u_i = \mathbb{P}_i \{\tau_N < \tau_0\}$.

Par le Lemme 1.2, on sait qu'on a le système d'équations suivant :

$$\begin{cases} u_i = qu_{i-1} + pu_{i+1} \quad \forall i : 1 \leq i \leq N-1 \\ u_0 = 0 \\ u_N = 1 \end{cases}$$

Il s'agit d'un système d'équations linéaires avec $N+1$ équations indépendantes et $N+1$ inconnues ; on peut donc trouver une solution. Évidemment, le cas général est un peu dur à traiter ; pour en venir à bout, on va se servir d'équations de récurrence (on dit parfois de différences finies).

En l'occurrence, on remarque que $p+q=1$. Dès lors, pour tous $1 \leq i \leq N-1$,

$$\begin{aligned} & (p+q)u_i = pu_{i+1} + qu_{i-1} \\ \Rightarrow & p(u_{i+1} - u_i) = q(u_i - u_{i-1}) \\ \Rightarrow & u_{i+1} - u_i = \frac{q}{p}(u_i - u_{i-1}). \end{aligned}$$

C'est à ce stade qu'on introduit $\rho = q/p$ et $\delta_i = u_i - u_{i-1}$ pour tous $1 \leq i \leq N$. On a donc :

$$\delta_{i+1} = \rho\delta_i = \rho^2\delta_{i-1} = \dots = \rho^i\delta_1;$$

ou, reformulé :

$$\delta_i = \rho^{i-1}\delta_1.$$

On termine l'élucidation du mystère en remarquant que $\delta_1 = u_1 - u_0 = u_1$ car $u_0 = 0$. De plus, on doit avoir que $u_N - u_0 = u_N = \sum_{i=1}^N \delta_i = 1$. Ainsi, on trouve que :

$$\begin{aligned} 1 &= \sum_{i=1}^N \rho^{i-1}u_1 \\ \Rightarrow & u_1 = \left(\sum_{i=1}^N \rho^{i-1} \right)^{-1}. \end{aligned}$$

Si on suppose que $\rho = 1$, alors $\sum_{i=1}^N \rho^{i-1} = N$, et $u_1 = \frac{1}{N}$. Dès lors, $u_a = \frac{1}{N} \sum_{i=1}^a \rho^{i-1} = \frac{a}{N}$, et la probabilité qu'Aymeric remporte toute la fortune est

$$\mathbb{P}_a \{ \tau_N < \tau_0 \} = u_a = \frac{a}{N}.$$

Si on suppose que $\rho \neq 1$, alors

$$\sum_{i=1}^N \rho^{i-1} = \frac{1 - \rho^N}{1 - \rho}.$$

Il suit que

$$u_1 = \frac{1 - \rho}{1 - \rho^N},$$

et

$$\mathbb{P}_a \{ \tau_N < \tau_0 \} = u_a = \frac{1 - \rho}{1 - \rho^N} \sum_{i=1}^a \rho^{i-1} = \frac{1 - \rho^a}{1 - \rho^N}.$$

On peut vérifier rapidement l'intuition : lorsque $a = N$, cette probabilité est 1. Lorsque $a = 0$, cette probabilité est bien Nulle. Entre les deux, elle est croissante en a .

- (c) On cherche maintenant $\mathbb{E}_a [\tau_{\{0,N\}}]$, le temps d'atteinte de la région comprenant les états 0 et N ; c'est le temps où on a joué la dernière partie, puisqu'après ce temps, on arrête de jouer.

On définit $\mu_i = \mathbb{E}_i [\tau_{\{0,N\}}]$; Peut-on utiliser un conditionnement par le premier pas pour obtenir un système d'équations pour μ_i ? Oui.

En effet,

$$\begin{aligned}\mu_i &= \mathbb{E}_i [\tau_{\{0,N\}}] \\ &= \mathbb{E}_i [\mathbb{E}_i [\tau_{\{0,N\}} \mid X_1]] \\ &= \sum_{j \in S} P_{i,j} \mathbb{E}_i [\tau_{\{0,N\}} \mid X_1 = j].\end{aligned}$$

Or, encore une fois, par la propriété de Markov, une fois qu'on sait qu'on a atteint l'état j , la loi du processus est la même que si on venait tout juste de commencer! Autrement dit, la loi de $\tau_{\{0,N\}}$ sachant que $X_t = j$ est exactement la loi de $t + \tau_{\{0,N\}}$ en ne sachant rien (on ajoute t parce qu'on sait qu'on a au moins fait t pas avant d'arriver).

Donc,

$$\mathbb{E}_i [\tau_{\{0,N\}} \mid X_1 = j] = 1 + \mathbb{E}_j [\tau_{\{0,N\}}].$$

En remplaçant ceci dans notre équation précédente, on obtient :

$$\begin{aligned}\mu_1 &= \sum_{j \in S} P_{i,j} (1 + \mu_j) \\ &= 1 + \sum_{j \in S} P_{i,j} \mu_j.\end{aligned}$$

Spécifiquement, pour nous, cela correspond donc au système d'équations suivantes :

$$\begin{cases} \mu_i = 1 + p\mu_{i+1} + q\mu_{i-1} \quad \forall i : 1 \leq i \leq N-1 \\ \mu_0 = \mu_N = 0 \end{cases}$$

Encore une fois, pour résoudre ce système d'équations, on peut faire appel aux techniques de résolution d'équations aux différences finies. En manipulant un peu les équations et en notant $\delta_i = \mu_i - \mu_{i-1}$, on trouve que pour $1 \leq i \leq N-1$, on trouve que dans ces cas :

$$\delta_{i+1} = p\delta_i - \frac{1}{p}.$$

Résoudre cette récurrence est très compliqué; néanmoins, dans le cas où $p = q = \frac{1}{2}$, on trouve simplement que

$$\delta_{i+1} = \delta_i - 2 = \delta_{i-1} - 4 = \dots = \delta_1 - 2i.$$

Bien sûr, $\delta_1 = \mu_1$ puisque $\mu_0 = 0$; on a donc

$$\delta_i = \mu_1 - 2(i-1),$$

et

$$\begin{aligned}
0 &= \mu_N \\
&= \sum_{i=1}^N \delta_i \\
&= \sum_{i=1}^N (\mu_1 - 2(i-1)) \\
&= N\mu_1 - 2 \sum_{i=1}^{N-1} i \\
&= N\mu_1 - N(N-1),
\end{aligned}$$

d'où $\mu_1 = N - 1$ et

$$\mu_a = \sum_{i=1}^a \delta_i = a(N-1) - a(a-1) = a(N-a).$$

On a donc que $\mathbb{E}_a [\tau_{\{0,N\}}] = \mu_a = a(N-a)$.

On peut donner un résultat similaire au Lemme 1.2 pour l'espérance du temps d'atteinte et des temps de retour à l'état k à partir de n'importe où ; cette fois, on trouve les solutions à l'aide d'un système d'équations inhomogènes (c'est à dire qu'elles ont des termes constants non-nuls) :

LEMME 1.3. *Soit $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ une chaîne de Markov homogène à temps discret sur l'espace d'états dénombrable $S = \{1, 2, 3, \dots\}$; soit $k \in S$ un états distincts de S . Alors,*

i. *on a le système d'équations suivant :*

$$(1.6.5) \quad \begin{cases} \mathbb{E}_i [\tau_k] = 1 + \sum_{j \in S} P_{i,j} \mathbb{E}_j [\tau_k] & \forall i \neq k \\ \mathbb{E}_k [\tau_k] = 0. \end{cases}$$

ii. *De plus, $\mathbb{E}_k [\tau_k^+]$ est donné par*

$$(1.6.6) \quad \mathbb{E}_k [\tau_k^+] = 1 + \sum_{j \in S} P_{k,j} \mathbb{E}_j [\tau_k].$$

DÉMONSTRATION. La démonstration de ce lemme suit les mêmes principes que ceux illustrés dans l'exemple 1.17. \square

1.6.2. Récurrence, transience. C'est maintenant le temps de se pencher sur les quantités qu'on a ignorées plus haut : $\mathbb{P}_i \left\{ \tau_j^+ < +\infty \right\}$, la probabilité d'atteindre j éventuellement (ou de revenir en i si $j = i$).

Cette quantité est d'un grand intérêt pour nous puisqu'elle sera l'une des valeurs clés pour nous aider à décrire la dynamique de notre processus stochastique. On introduit le vocabulaire suivant :

DÉFINITION 1.11 (Récurrence, transience). Soit $\mathbf{X} = (X_t)_{t \geq 0}$ une chaîne de Markov homogène sur l'espace d'états S quelconque, et soit un état $x \in S$.

L'état x est **récurrent** si $\mathbb{P}_x \{\tau_x^+ < +\infty\} = 1$; sinon, il est **transient**.

REMARQUE. Remarquer qu'on s'est replacés dans le contexte des chaînes de Markov homogènes, pour simplifier les choses.

Dans une chaîne de Markov homogène, un état est donc récurrent si on est (presque) sûr qu'on va y revenir lorsqu'on en part ; dans ce cas, par la propriété de Markov, une fois qu'on y est revenu, le processus recommence – et on va donc revenir (presque sûrement) au point de départ une nouvelle fois, et ainsi de suite.

Autrement dit, si la chaîne commence dans un état récurrent, cet état sera visité un nombre infini de fois ; à l'inverse, si un état est transient, on le visite seulement un nombre fini de fois, mais éventuellement, on cesse complètement d'y revenir.

EXEMPLE 1.18. Dans l'exemple 1.10, tous les états sont *récurrents* ; en effet, on remarque qu'on finit toujours par revenir éventuellement à l'état 1. On peut se convaincre assez facilement que les états 2 ou 3 sont récurrents aussi ; Si on commence en 3, par exemple, on atteint éventuellement 1, et de 1, on atteindra éventuellement 3 de nouveau. Même raisonnement pour l'état 2.

Dans l'exemple 1.12, tous les états sont transients ; En effet, il est manifeste qu'une fois qu'on a quitté un état pour un autre, on ne peut plus y revenir. La seule exception est l'état 4, qui est récurrent.

Dans l'exemple 1.13, l'état 1 est transient ; il est évident qu'on ne peut jamais y revenir depuis aucun autre état. Les autres états sont récurrents. En effet, si par exemple on commence à l'état 2 (resp. 3, 4), on y revient forcément après avoir fait un petit tour du cycle 2, 3, 4. Même chose pour les états 5, 6 et 7.

Dans l'exemple 1.15, tous les états sont transients sauf les états 7 et 9, qui sont récurrents. Dans l'exemple 1.17, les états 0 et N sont récurrents tandis que tous les autres sont transients.

Le nombre de visites. Comme nous l'avons déjà mentionné, une façon de distinguer un état récurrent d'un état transient, c'est de compter le nombre de fois qu'on y revient lorsqu'on laisse le processus évoluer indéfiniment. Nous avons déjà identifié que, pour un état transient, le nombre de visites serait forcément infini. Nous allons maintenant utiliser cette observation pour obtenir la proposition suivante (très utile) :

PROPOSITION 1.8. Soit $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ une chaîne de Markov homogène à temps discret sur l'espace d'états discret $S = \{1, 2, 3, \dots\}$ (fini ou dénombrable).

L'état i est récurrent si et seulement si la série $\sum_{t \in \mathbb{N}} P_{i,i}^{(t)}$ diverge.

DÉMONSTRATION. Comme annoncé plus tôt, nous allons utiliser le nombre de visites en i pour déterminer si un état est transient est récurrent. Soit

$$N_i^{(u)} = \sum_{t > u} \mathbb{1}_{\{X_t = i\}}$$

le nombre de visites en i après le temps u . On note aussi $N_i = N_i^{(0)}$ le nombre total de visites en i .

D'une part, on a que :

$$\begin{aligned} \sum_{t \in \mathbb{N}} P_{i,i}^{(t)} &= \sum_{t=1}^{\infty} \mathbb{P}_i \{X_t = i\} \\ &= \sum_{t=1}^{\infty} \mathbb{E}_i [\mathbb{1}_{\{X_t=i\}}] \\ &= \mathbb{E}_i [N_i]. \end{aligned}$$

D'autre part, on peut écrire

$$\mathbb{E}_i [N_i] = \sum_{k=1}^{\infty} \mathbb{P}_i \{N_i \geq k\}.$$

Or, pour tout k , on écrit

$$\begin{aligned} \mathbb{P}_i \{N_i \geq k\} &= \mathbb{E}_i [\mathbb{P}_i \{N_i \geq k \mid \tau_i^+\}] \\ &= \sum_{u=1}^{\infty} \mathbb{P}_i \{N_i \geq k \mid \tau_i^+ = u\} \mathbb{P}_i \{\tau_i^+ = u\} \\ &= \sum_{u=1}^{\infty} \mathbb{P}_i \{N_i^{(u)} \geq k-1 \mid X_u = i\} \mathbb{P}_i \{\tau_i^+ = u\} \\ &= \sum_{u=1}^{\infty} \mathbb{P}_i \{N_i \geq k-1\} \mathbb{P}_i \{\tau_i^+ = u\} \\ &= \mathbb{P}_i \{N_i \geq k-1\} \sum_{u=1}^{\infty} \mathbb{P}_i \{\tau_i^+ = u\} \\ &= \mathbb{P}_i \{N_i \geq k-1\} \mathbb{P}_i \{\tau_i^+ < +\infty\}. \end{aligned}$$

Les deux premières égalités sont simplement une présentation de la loi des probabilités totales. La troisième est due au fait que l'événement $\{\tau_i^+ = u\}$ est identiquement égal à l'événement $X_1, X_2, X_3, \dots, X_{u-1} \neq i; X_u = i$, et qu'on peut donc y appliquer la propriété de Markov et ne garder que la condition que $X_u = i$.

L'égalité suivante est due au fait que le nombre de visites après u si on sait qu'on est en i au temps u , c'est (par homogénéité du processus et par la propriété de Markov) une variable qui a la même loi que le nombre total de retours en i si on a commencé à i . Finalement, on peut mettre le tout en évidence. La dernière égalité vient du fait que

$$\{\tau_i^+ < +\infty\} = \bigcup_{u=1}^{\infty} \{\tau_i^+ = u\},$$

et donc, par un axiome des probabilités, on obtient que

$$\mathbb{P}_i \{\tau_i^+ < +\infty\} = \sum_{u=1}^{\infty} \mathbb{P}_i \{\tau_i^+ = u\}.$$

Bien sûr, cette récurrence linéaire se résout facilement et on trouve :

$$\mathbb{P}_i \{N_i \geq k\} = \mathbb{P}_i \{N_i \geq 1\} \mathbb{P}_i \{\tau_i^+ < +\infty\}^{k-1}.$$

Or, $\mathbb{P}_i \{N_i \geq 1\} = \mathbb{P}_i \{\tau_i^+ < +\infty\}$, puisque c'est simplement la probabilité qu'on revienne en i au moins une fois ! Donc, finalement, on a que :

$$\mathbb{P}_i \{N_i \geq k\} = \mathbb{P}_i \{\tau_i^+ < +\infty\}^k.$$

On a donc que

$$\sum_{t=1}^{\infty} P_{i,i}^{(t)} = \mathbb{E}_i [N_i] = \sum_{k=1}^{\infty} \mathbb{P}_i \{\tau_i^+ < +\infty\}^k.$$

Cette série (l'espérance du nombre de visites) diverge si et seulement si $\mathbb{P}_i \{\tau_i^+ < +\infty\} = 1$; autrement, elle converge.

Donc, par définition, la série $\sum_{t=1}^{\infty} P_{i,i}^{(t)}$ diverge si et seulement si l'état i est récurrent. \square

EXEMPLE 1.19. Dans le processus de l'exemple 1.13 (dont le graphe est présenté à la figure 1.5), on voit clairement que pour tout $k \geq 1$, on a que

$$P_{2,2}^{(3k)} = P_{3,3}^{(3k)} = P_{4,4}^{(3k)} = 1;$$

ce que cela signifie, c'est que forcément, $\sum_{t=1}^{\infty} P_{2,2}^{(t)} \geq \sum_{k=1}^{\infty} P_{2,2}^{(3k)} = +\infty$ et que la série diverge ; donc, les états 2, 3 et 4 sont récurrents. On peut voir que ça fonctionne aussi pour les états 4, 5 et 6.

On remarque une chose importante : la récurrence et la transience des états semble souvent pouvoir être déduites à partir d'un examen sommaire du graphe des transitions possibles, sans avoir vraiment besoin de faire de calculs – du moins, pour les espaces d'états finis.

Récurrence et classes d'équivalence. Ce constat nous mène à la proposition suivante :

PROPOSITION 1.9. Soit $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ une chaîne de Markov homogène à temps discret sur l'espace d'états fini $S = \{1, 2, 3, \dots, n\}$, et soit $\Gamma = (S, E)$ son graphe des transitions possibles. Supposons de plus $i, j \in S$ deux états connectés dans Γ (c'est-à-dire que $i \longleftrightarrow j$).

Alors, i est récurrent si et seulement si j est récurrent.

DÉMONSTRATION. D'abord, si i et j sont connectés, cela signifie que $\mathbb{P}_i \{\tau_j < \tau_i^+\} > 0$ et $\mathbb{P}_j \{\tau_i < \tau_j^+\} > 0$.

Avec les mêmes définitions que dans la preuve de la proposition 1.8 pour N_i et N_j , on va raisonner par « nombres de visites » encore une fois, et obtenir une inégalité utile. En effet, d'une part, à chaque fois qu'on visite i , on atteint visite j au moins une fois avec probabilité $\mathbb{P}_i \{\tau_j < \tau_i^+\} > 0$. Donc, sachant $N_i = k$ pour un certain k , on sait que $N_j \geq \text{Binom}(k, \mathbb{P}_i \{\tau_j < \tau_i^+\})$. Il suit que

$$\mathbb{E}_i [N_j] \geq \mathbb{P}_i \{\tau_j < \tau_i^+\} \mathbb{E}_i [N_i].$$

Bien sûr, on sait aussi que l'espérance du nombre de visites en j partant de j est au moins le nombre de visites en j après avoir touché i ; donc :

$$\begin{aligned} \mathbb{E}_j [N_j] &\geq \mathbb{P}_j \{\tau_i < \tau_j^+\} \mathbb{E}_i [N_j] \\ &\geq \mathbb{P}_j \{\tau_i < \tau_j^+\} \mathbb{P}_i \{\tau_j < \tau_i^+\} \mathbb{E}_i [N_i]. \end{aligned}$$

Et si i est récurrent, alors $\mathbb{E}_i[N_i]$ diverge, et $\mathbb{E}_j[N_j]$ diverge, donc j est forcément récurrent aussi.

Puisque les états sont connectés, on peut inverser i et j dans le raisonnement précédent pour obtenir une preuve similaire dans l'autre sens : si j est récurrent, alors i est récurrent aussi. \square

Et vlan ! On vient de montrer que la récurrence et la transience sont des propriétés partagées par tous les états d'une même classe d'équivalence ! On peut pousser plus loin, et utiliser les structures des classes d'équivalence pour déduire directement des choses sur leur transience et leur récurrence :

PROPOSITION 1.10. *Soit $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ une chaîne de Markov homogène à temps discret sur l'espace d'états dénombrable $S = \{1, 2, 3, \dots\}$, soit $\Gamma = (S, E)$ le graphe des transitions possibles, et soit $C \in S / \longleftrightarrow$ une classe d'équivalence.*

- i. *Si C est une classe ouverte, alors ses états sont transients.*
- ii. *Si C est une classe fermée et finie (c'est-à-dire que $|C| < +\infty$), alors C est une classe récurrente.*

DÉMONSTRATION. i Puisque C est une classe ouverte, il existe $i \in C$ et $j \notin C$ tels que $i \rightarrow j$ – c'est-à-dire que $P_{i,j} > 0$.

Or, une fois en j , il est impossible de revenir dans C – autrement on pourrait revenir en i , et j serait dans C . Donc, pour tout $k \in C$, $\mathbb{P}_j\{\tau_k < +\infty\} = 0$

Donc, la probabilité de ne jamais revenir en i est bornée comme suit :

$$\mathbb{P}_i\{\tau_i^+ = +\infty\} \geq P_{i,j} > 0,$$

et l'état i est transient.

Puisque i est transient, tous les autres états de C sont aussi transients.

- ii À l'inverse, si C est une classe fermée comprenant un nombre fini d'états, soit tous ses états sont transients, soit ils sont tous récurrents.

On note $N_C = \sum_{j \in C} N_j$, où les N_i sont défini comme dans les preuves des propositions 1.8 et 1.9. On a que pour un certain $i \in C$, $\mathbb{E}_i[N_C] = \sum_{j \in C} \mathbb{E}_i[N_j]$. Si on suppose que tous les états de C sont transients, par un raisonnement comme dans la preuve de la proposition 1.9, on doit avoir que $\mathbb{E}_i[N_j] < +\infty$ pour tous $i, j \in C$, et, par conséquent, pour tout $i \in C$,

$$\mathbb{E}_i[N_C] < +\infty,$$

puisqu'on a seulement une somme finie de termes de la forme $\mathbb{E}_i[N_j]$.

Mais par ailleurs, cela signifie que $\mathbb{P}_i\{N_C < +\infty\} = 1$, ce qui est impossible ; en effet, C état une classe fermée, une fois l'état i atteint, il est impossible de sortir de la classe C ; donc $\mathbb{P}_i\{N_C = +\infty\} = 1$. C'est une contradiction.

On conclue que les états de la classe C doivent être récurrents. \square

Au vu des propositions 1.9, 1.10 et 1.12, il vaut la peine de faire la définition suivante :

DÉFINITION 1.12 (irréductibilité). Soit $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ une chaîne de Markov homogène à temps discret sur l'espace des états dénombrable $S = \{1, 2, 3, \dots\}$ (possiblement fini), et soit $\Gamma = (S, E)$ son graphe des transitions possibles.

On dit que \mathbf{X} est **irréductible** si le graphe Γ n'a qu'une seule classe d'équivalence (c'est-à-dire que $\Gamma / \longleftrightarrow = \{S\}$).

COROLLAIRE 1.3. *Soit $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ une chaîne de Markov homogène à temps discret irréductible sur l'espace des états fini $S = \{1, 2, 3, \dots, n\}$.*

Alors tous les états sont récurrents.

On remarque aussi le résultat suivant (présenté sans preuve) :

LEMME 1.4. *Soit $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ une chaîne de Markov homogène à temps discret sur l'espace des états dénombrable $S = \{1, 2, 3, \dots\}$. Soit $\Gamma = (S, E)$ le graphe des transitions possibles pour \mathbf{X} , et soit $C \in \Gamma / \longleftrightarrow$ une classe d'équivalence finie récurrente.*

Pour tout $i \in C$, sachant $X_0 = i$, le processus \mathbf{X} a exactement la loi d'une chaîne de Markov homogène à temps discret irréductible sur l'espace des états C commençant en i .

Autrement dit, une fois qu'on se trouve dans une classe récurrente, on peut considérer notre chaîne de Markov comme une chaîne irréductible sur seulement les états de cette classe.

La probabilité d'atteindre éventuellement un état. Vous souvenez-vous des quantités de la forme $\mathbb{P}_i \{ \tau_j^+ < +\infty \}$ mentionnées plus tôt ? C'est maintenant le temps de voir comment on se débrouille avec.

PROPOSITION 1.11. *Soit $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ une chaîne de Markov homogène à temps discret sur l'espace des états dénombrable $S = \{1, 2, 3, \dots\}$ (possiblement fini).*

i. *Si i est un état récurrent, alors on a, pour tout $j \neq i$:*

$$(1.6.7) \quad \mathbb{P}_i \{ \tau_j < +\infty \} = \begin{cases} 1 & \text{si } i \longleftrightarrow j ; \\ 0 & \text{sinon.} \end{cases}$$

ii. *Si i est un état transient, et j et j' sont deux états récurrents avec $j \longleftrightarrow j'$, alors*

$$(1.6.8) \quad \mathbb{P}_i \{ \tau_j < +\infty \} = \mathbb{P}_i \{ \tau_{j'} < +\infty \}.$$

iii. *Si $R \subset S$ est l'ensemble des états récurrents, et $C \subseteq R$ est une classe d'équivalence récurrente, alors, pour tout $i \in S$, $k \in C$, on a que $\mathbb{P}_i \{ \tau_k < +\infty \} = \mathbb{P}_i \{ \tau_C < +\infty \}$. De plus, on a le système d'équations suivant :*

$$(1.6.9) \quad \begin{cases} \mathbb{P}_i \{ \tau_C < +\infty \} = \sum_{j \in S} P_{i,j} \mathbb{P}_j \{ \tau_C < +\infty \} & \forall i \notin R; \\ \mathbb{P}_i \{ \tau_C < +\infty \} = 0 & \forall i \in R \setminus C; \\ \mathbb{P}_i \{ \tau_C < +\infty \} = 1 & \forall i \in C. \end{cases}$$

iv. *Avec les mêmes notations, si $k \notin R$ est un état transient, on a le système d'équations suivant :*

$$(1.6.10) \quad \begin{cases} \mathbb{P}_i \{ \tau_k < +\infty \} = \sum_{j \in S} P_{i,j} \mathbb{P}_j \{ \tau_k < +\infty \} & \forall i \notin R, i \neq k; \\ \mathbb{P}_i \{ \tau_k < +\infty \} = 0 & \forall i \in R; \\ \mathbb{P}_k \{ \tau_k < +\infty \} = 1. \end{cases}$$

DÉMONSTRATION. **i** D'abord, puisque $i \longleftrightarrow j$, il existe $m, n \in \mathbb{N}$ tels que $P_{i,j}^{(m)} P_{j,i}^{(n)} > 0$.

On va montrer que

$$\mathbb{P}_i \{ \tau_j < +\infty \} \mathbb{P}_j \{ \tau_i = +\infty \} = 0;$$

il s'agit de la probabilité d'atteindre j partant de i , puis de subséquemment ne plus revenir en i ; Cela doit être borné par la probabilité de ne jamais revenir en i :

$$\mathbb{P}_i \{ \tau_j < +\infty \} \mathbb{P}_j \{ \tau_i = +\infty \} \leq \mathbb{P}_i \{ \tau_i^+ = +\infty \}.$$

Or, $\mathbb{P}_i \{ \tau_i^+ = +\infty \} = 0$, puisque i est récurrent. Donc, on doit avoir le produit :

$$\mathbb{P}_i \{ \tau_j < +\infty \} \mathbb{P}_j \{ \tau_i = +\infty \} = 0,$$

et puisque $\mathbb{P}_i \{ \tau_j < +\infty \} \geq P_{i,j}^{(m)} > 0$, la seule possibilité est que $\mathbb{P}_j \{ \tau_i = +\infty \} = 0$; autrement dit, si $i \longleftrightarrow j$, $\mathbb{P}_j \{ \tau_i < +\infty \} = 1$.

En faisant le raisonnement inverse, on obtient le résultat.

ii Pour j, j' deux états récurrents qui communiquent, on doit avoir :

$$\mathbb{P}_i \{ \tau_{j'} < +\infty \} \geq \mathbb{P}_i \{ \tau_j < +\infty \} \mathbb{P}_j \{ \tau_{j'} < +\infty \};$$

La probabilité d'atteindre j' à partir de i est *au moins* la probabilité d'atteindre d'abord j , puis ensuite j' . Or, on vient de voir que $\mathbb{P}_j \{ \tau_{j'} < +\infty \} = 1$.

En renversant le raisonnement, on trouve donc

$$\mathbb{P}_i \{ \tau_{j'} < +\infty \} = \mathbb{P}_i \{ \tau_j < +\infty \},$$

ce qu'il fallait démontrer.

iii, iv La preuve de ces deux systèmes d'équations est entièrement analogue aux preuves du lemme 1.2; il s'agit simplement d'un conditionnement par le premier pas. \square

1.6.3. Périodicité. Une autre propriété importante des états est leur *périodicité*; des états dits « périodiques » sont des états qui, même à long terme, sont « périodiquement accessibles »; c'est-à-dire qu'on ne peut y revenir qu'après des multiples d'un certain intervalle de temps. On fait la définition suivante :

DÉFINITION 1.13 (Période, périodicité). Soit $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ une chaîne de Markov homogène à temps discret sur l'espace d'états dénombrable $S = \{1, 2, 3, \dots\}$ (possiblement fini).

On dit que la **période** de l'état i – notée $d(i)$ – est donnée par

$$(1.6.11) \quad d(i) = \text{pgcd} \left\{ n \in \mathbb{N} : P_{i,i}^{(n)} > 0 \right\}.$$

Les états dont la période est 1 et les états pour lesquels $P_{i,i}^{(n)} = 0$ pour tout $n \in \mathbb{N}$ sont dits **apériodiques**; les autres sont **périodiques**.

Si $\Gamma = (S, E)$ est le graphe des transitions possibles pour i , et \mathcal{C}_i l'ensemble des cycles du graphe Γ qui passent par i ; si de plus on note $|C|$ la longueur d'un cycle pour tout $C \in \mathcal{C}_i$, alors la période de i est égale au PGCD des longueurs de tous les cycles de \mathcal{C}_i :

$$(1.6.12) \quad d(i) = \text{pgcd} \{ |C| : C \in \mathcal{C}_i \}$$

EXEMPLE 1.20. Dans l'exemple 1.10, l'état 1 est apériodique ; en effet, $P_{1,1} > 0$, ce qui signifie que $d(1) \leq 1$. Mais comme $d(1) \geq 1$, $d(1) = 1$. De même pour l'état 2. L'état 3 est également apériodique en effet, selon le graphe des transitions possibles (figure 1.3), on voit que $P_{3,3}^{(2)} > 0$ (il est possible de revenir à 3 en deux pas), mais $P_{3,3}^{(3)} > 0$ aussi (il est possible de revenir à 3 en 3 pas, par exemple par le chemin $3 \rightarrow 1 \rightarrow 1 \rightarrow 3$).

Dans l'exemple 1.12, tous les états sont apériodiques, car pour tout i , $P_{i,i} > 0$.

Dans l'exemple 1.13, l'état 1 est apériodique puisque $P_{1,1}^{(n)} = 0$ pour tout $n \geq 1$; une fois partis de 1, on n'y retourne jamais. Les états 2, 3 et 4 sont périodiques de période 3. En effet, la seule façon de revenir en 2 est par le cycle $2 \rightarrow 3 \rightarrow 4 \rightarrow 2$ (et resp. pour 3 et 4).

De même, dans cet exemple, les états 4, 5 et 6 aussi périodiques, de période 3.

Dans l'exemple 1.15, (figure 1.7), les états 7 et 9 sont apériodiques. Les autres sont de période 2 ; en effet, on remarque que le graphe des transitions possible est *biparti* ; tous les cycles qui reviennent à leur point de départ (autre que 7 et 9) sont de longueur paire (donc le PGCD est 2).

Même chose pour l'exemple 1.17 : tous les états sauf 0 et N sont périodiques de période 2. Les états 0 et N sont apériodiques.

La périodicité est une autre propriété que partagent les états de même classe d'équivalence :

PROPOSITION 1.12. Soit $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ une chaîne de Markov homogène à temps discret sur l'espace d'états dénombrable $S = \{1, 2, 3, \dots\}$ (possiblement fini). Soit $\Gamma = (S, E)$ son graphe des transitions possibles, et soient $i, j \in S$ deux états connectés (c'est-à-dire que $i \longleftrightarrow j$).

Alors, $d(i) = d(j)$.

DÉMONSTRATION. Puisque i et j sont connectés, il existe C un cycle qui passe par i et par j .

Puisque le cycle C passe par i et par j , on a que $|C|$ est un multiple de $d(i)$ et de $d(j)$.

Supposons C_i un cycle quelconque qui passe par i . La concaténation¹² de C avec C_i (notée $C \rightarrow C_i$) est elle-même un cycle qui passe aussi par i et j . Donc, on doit aussi avoir que $|C \rightarrow C_i|$ est un multiple de $d(i)$ et $d(j)$. Mais alors, $|C_i|$, la longueur du cycle C_i lui-même, donnée par $|C_i| = |C \rightarrow C_i| - |C|$ doit aussi être un multiple de $d(i)$ et $d(j)$.

Donc, pour tout cycle C_i passant par i , on a que $|C|$ est un multiple de $d(j)$ et de $d(i)$ – c'est à dire que $d(j)$ est un commun diviseur des longueurs de tous les cycles qui passent par i . On a donc que $d(j) \leq d(i)$; en effet, $d(i)$ est le plus grand commun diviseur.

Par le raisonnement inverse, on voit que $d(i) \leq d(j)$. Donc, $d(i) = d(j)$. \square

1.7. Le théorème ergodique et la distribution stationnaire.

On arrive maintenant à l'un de nos objectifs énoncés à la section 1.2 – identifier le comportement à long terme de notre chaîne de Markov. Pour rappel : on s'intéresse à la limite de la distribution de X_t quand t tend vers l'infini. Comme X_t est une variable discrète (dans le contexte d'une chaîne sur un espace d'états discret), cela correspond à identifier un

12. La concaténation de deux chemins est le chemin qui résulte quand on met ces deux chemins « bout-à-bout ». Voir la définition A.4 v

comportement à la limite pour la *fonction de masse* de X_t (étant donné une distribution initiale) ; donc, trouver un comportement-limite pour les quantités

$$\mathbb{P}_i \{X_t = j\} =: P_{i,j}^{(t)}.$$

1.7.1. La distribution stationnaire. Notre première étape, c'est de faire un petit peu d'analyse ; et si il y avait une solution simple ?

On se penche sur l'équation (1.4.7) – la revoici, pour rappel, dans le contexte où $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ est une chaîne de Markov homogène à temps discret sur l'espace discret d'états S , avec les matrices de transition usuelles :

$$(1.4.7) \quad P_{x,y}^{(s+t)} = \sum_{z \in S} P_{x,z}^{(s)} P_{z,y}^{(t)}.$$

En particulier, un examen sommaire de cette équation permet de déduire presque immédiatement que :

$$(\dagger) \quad P_{x,y}^{(t+1)} = \sum_{z \in S} P_{x,z}^{(t)} P_{z,y}.$$

Si on suppose que la limite de $P_{x,y}^{(t)}$ existe pour toute paire x, y lorsque t tend vers l'infini, et que, de plus, on suppose que cette limite est indépendante de la distribution initiale, et que

$$\lim_{t \rightarrow \infty} P_{x,y}^{(t)} = \pi_y,$$

alors en employant le théorème de convergence dominée¹³ avec (\dagger) , en plus de la condition de normalisation (l'équation (1.4.6)), alors on doit avoir le système d'équations suivant :

$$\begin{cases} \pi_y = \sum_{x \in S} \pi_x P_{x,y} & \forall y \in S; \\ 1 = \sum_{x \in S} \pi_x \end{cases}$$

Il s'avère que les solutions à ce système d'équations sont d'un intérêt tout particulier pour nous ; c'est pour cela que nous leur donnons un nom : les *distributions stationnaires* :

DÉFINITION 1.14 (Distribution stationnaire). Soit $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ une chaîne de Markov homogène à temps discret sur l'espace d'états discret (dénombrable) S , avec sa famille des probabilités de transition en un pas $P = (P_{x,y})_{(x,y) \in S^2}$.

Une **distribution stationnaire** pour \mathbf{X} est une famille $\pi = (\pi_x)_{x \in S}$ indicée par les états S qui satisfait le système d'équations suivant :

$$(\dagger) \quad \begin{cases} \pi_y = \sum_{x \in S} \pi_x P_{x,y} & \forall y \in S; \\ 1 = \sum_{x \in S} \pi_x \end{cases}$$

De façon plus générale, on dira que π est un **vecteur stationnaire** si il satisfait toutes les équations dans (\dagger) sauf la dernière (la condition de normalisation).

13. Il s'agit d'un résultat très utile de la théorie de la mesure et de l'intégration de Lebesgue ; consulter l'annexe la section B.1 de l'annexe B pour en apprendre plus à ce sujet.

REMARQUE. Une chaîne de Markov \mathbf{X} peut possiblement avoir plus d'une distribution stationnaire, selon que le système d'équations (†) admet une ou plusieurs solutions. Elle pourrait également n'avoir aucune solution stationnaire. Nous verrons un peu plus loin à quelles conditions une distribution stationnaire.

La distribution stationnaire est appelée ainsi parce que si, au temps t , la distribution de la variable X_t est stationnaire (le vecteur de masses $\mathbf{x}^{(t)}$ est une distribution stationnaire), alors elle le sera aussi pour tous les temps $u \geq t$; une fois atteinte, elle ne change plus.

Dans le cas spécifique où on considère un espace d'états fini $S = \{1, 2, 3, \dots, n\}$, la famille des probabilités de transition en un pas $P = (P_{x,y})_{x,y \in S}$ est en fait une matrice $n \times n$.

Si on voit une distribution stationnaire $\pi = (\pi_x)_{x \in S}$ comme un vecteur-ligne (dans $\mathbb{R}^{1 \times n}$), alors le système d'équations (†) se réduit à :

$$(\dagger) \quad \begin{cases} \pi = \pi P \\ \pi \cdot \mathbf{1}_n = 1. \end{cases}$$

ATTENTION!. Le vecteur π est un vecteur-ligne, pas un vecteur-colonne. La multiplication par la matrice est *par la droite*, puisque l'expression $P\pi$ n'aurait *pas de sens* – en effet, les formats des matrices P et π sont incompatibles.

1.7.2. Théorème ergodique. Alors ? Peut-on dire qu'on a trouvé la distribution-limite pour X_t ? Non.

Souvenez-vous : on a commencé la section 1.7.1 avec l'importante supposition que la limite de $P_{x,y}^{(t)}$ existe pour toute paire $x, y \in S$; est-ce vrai ? On a également assumé que la distribution-limite était indépendante de l'état de départ. Ça aussi, c'est une hypothèse importante qui, *a priori*, n'est pas évidente à démontrer. Et pour cause : parfois, c'est faux !

EXEMPLE 1.21. On considère la chaîne de Markov $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ à temps discret et homogène, sur l'espace des états $S = \{1, 2, 3\}$, avec la matrice des probabilités de transition en un pas :

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

- (a) Calculer $P^{(2)}$ et $P^{(3)}$.
- (b) Trouver une distribution stationnaire. Est-elle unique ?
- (c) Est-ce que $\lim_{t \rightarrow \infty} P_{i,j}^{(t)}$ existe pour de quelconques i et j dans S ?
- (d) Qu'est-ce qui fait que ça ne fonctionne pas ?

SOLUTION. (a) On a

$$P^{(2)} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \quad P^{(3)} = \mathbb{I}_3.$$

- (b) On doit avoir

$$\pi = \pi P,$$

, c'est à dire que

$$\begin{cases} \pi_1 = \pi_2; \\ \pi_2 = \pi_3; \\ \pi_3 = \pi_1, \end{cases}$$

ce qui signifie que $\pi_1 = \pi_2 = \pi_3$. Avec en plus la condition de normalisation, on doit donc avoir

$$\pi_1 = \pi_2 = \pi_3 = \frac{1}{3},$$

et c'est la seule solution possible.

- (c) Non. En effet, les états sont périodiques; pour tout i , on a que $P_{i,i}^{(3k)} = 1$ tandis que $P_{i,i}^{(3k+1)} = P_{i,i}^{(3k+2)} = 0$; cette limite ne converge donc pas vers $\frac{1}{3}$, malgré qu'il s'agit de la distribution stationnaire.

EXEMPLE 1.22. Soit $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ une chaîne de Markov homogène à temps discret sur l'espace des états $S = \{1, 2, 3, 4, 5, 6, 7\}$ avec la matrice de transition en un pas

$$P = \begin{pmatrix} 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/4 & 3/4 \\ 0 & 0 & 0 & 0 & 0 & 3/4 & 1/4 \end{pmatrix}.$$

- Tracer le graphe des transitions possibles.
- Identifier les classes d'états et leurs propriétés (récurrence/transience, période).
- Y a-t-il plus d'une distribution stationnaire possibles?
- Est-ce que la distribution limite dépend de l'état initial ici?

SOLUTION. (a) Le graphe est reproduit à la figure 1.8.

- (b) Il y a trois classes d'états :

- Les états 1 et 2 sont transients car ils forment une classe ouverte. Ils sont apériodiques (puisque $P_{1,1} > 0$).
- Les états 3, 4 et 5 sont récurrents puisqu'ils forment une classe fermée. Ils sont périodiques avec période 3 (la seule chaîne possible pour revenir en 3 est $3 \rightarrow 4 \rightarrow 5 \rightarrow 3$).
- Les états 6 et 7 sont récurrents puisqu'ils forment une classe fermée. Ils sont apériodiques puisque $P_{6,6} > 0$.

- (c) Oui. On remarque que $\pi^{(1)} = (0, 0, 1/3, 1/3, 1/3, 0, 0)$ et $\pi^{(2)} = (0, 0, 0, 0, 0, 1/2, 1/2)$ sont tous les deux des distributions stationnaires; puisqu'on a affaire à un système d'équations linéaires, toutes les combinaisons linéaires de ces deux solutions sont également des solutions stationnaires. En particulier, si $p \in (0, 1)$,

$$\pi = p\pi^{(1)} + (1 - p)\pi^{(2)}$$

est aussi forcément une distribution stationnaire.

Il existe donc une infinité de distributions stationnaires correspondant aux différentes valeurs du paramètre p .

- (d) Bien sûr. Si on commence en 3, 4 ou 5, la distribution-limite n'existe pas ; si on commence en 6 ou en 7, elle existe peut-être mais il est impossible qu'elle donne du poids aux états 3, 4 ou 5. Si on commence en 1 ou en 2, c'est encore différent, puisqu'on pourrait finir soit dans la classe de l'état 3, ou dans celle de l'état 6.

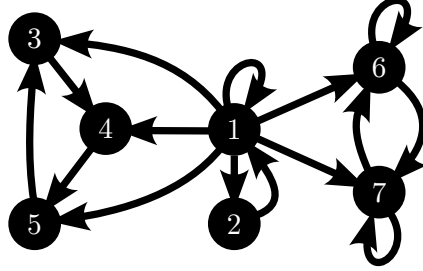


FIGURE 1.8 – Le graphe des transitions possibles pour la chaîne de Markov de l'exemple 1.22

On va maintenant tenter de répondre à trois questions :

- i. Pour quels états x, y la limite $\lim_{t \rightarrow \infty} P_{x,y}^{(t)}$ existe-t-elle ?
- ii. Ces limites (ou leur existence) dépendent-elles de x ? Dans quels cas ?
- iii. Dans le cas où la limite existe, à quoi cette quantité correspond-t-elle ?

On peut déjà répondre à la première question partiellement :

PROPOSITION 1.13. Soit $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ une chaîne de Markov homogène à temps discret sur l'espace des états discret S (dénombrable, possiblement infini). Soit $y \in S$ un état transient, et $x \in S$.

Alors,

$$(1.7.1) \quad \lim_{t \rightarrow \infty} P_{x,y}^{(t)} = 0.$$

DÉMONSTRATION. Dans le cas où $x = y$, on a que, puisque y est transient, on doit avoir que

$$\sum_{u=1}^{\infty} P_{y,y}^{(u)} < +\infty;$$

ceci implique directement que $\lim_{t \rightarrow \infty} P_{y,y}^{(t)} = 0$.

Dans le cas où $x \neq y$, avec $x \rightarrow y$ (c'est-à-dire que y est accessible depuis x), on considère

$$\sum_{u=1}^{\infty} P_{x,y}^{(u)} = \mathbb{E}_x [N_y],$$

le nombre de visites à y si on part de x . On voit que, à condition d'atteindre y (avec probabilité $\mathbb{P}_x \{ \tau_y^+ < +\infty \}$, ça doit être la même espérance que si on part de y directement ; autrement dit :

$$\sum_{u=1}^{\infty} P_{x,y}^{(u)} = \mathbb{P}_x \{ \tau_y^+ < +\infty \} \mathbb{E}_y [N_y] = \mathbb{P}_x \{ \tau_y^+ < +\infty \} \sum_{u=1}^{\infty} P_{y,y}^{(u)},$$

et donc, il faut forcément que $P_{x,y}^{(t)}$ tende aussi vers 0, puisque cette série doit aussi converger.

Dans le cas où $x \not\rightarrow y$, naturellement, $P_{x,y}^{(t)} = 0$ pour tout t .

Donc, dans tous les cas, on a $\lim_{t \rightarrow \infty} P_{x,y}^{(t)} = 0$. \square

Donc, on a déjà un élément de réponse ; pour tous les états transients, la limite existe – elle est nulle et ne dépend pas de la condition initiale.

Reste les états récurrents ; là, ça va être un peu plus difficile. On va revenir à notre première question à la fin ; pour l’instant, on peut répondre à la deuxième :

PROPOSITION 1.14. *Soit $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ une chaîne de Markov homogène à temps discret sur l’espace des états des états discret S (dénombrable, possiblement infini). Soient $y \in S$ un état récurrent et apériodique.*

Alors, on a que pour tout $x \in S$ connecté à y dans le graphe des transitions possibles (c’est-à-dire que $x \longleftrightarrow y$), on a que

$$(1.7.2) \quad \lim_{t \rightarrow \infty} |P_{x,y}^{(t)} - P_{y,y}^{(t)}| = 0.$$

En particulier, si $\lim_{t \rightarrow \infty} P_{y,y}^{(t)}$ existe, alors $\lim_{t \rightarrow \infty} P_{x,y}^{(t)}$ existe pour tout x connecté à y dans le graphe des transitions possibles ; de plus, ces limites sont égales.

DÉMONSTRATION. Sans perdre de généralité, nous allons assumer que la chaîne \mathbf{X} est irréductible.

Pour faire cette preuve, nous allons procéder par un argument de *couplage* – plutôt que de travailler directement sur la chaîne \mathbf{X} , nous allons créer une *autre* chaîne de Markov, dont \mathbf{X} sera « une composante ».

Partie 1 : Le couplage. On introduit d’abord la chaîne $\mathbf{Y} = (Y_t)_{t \in \mathbb{Z}^+}$, indépendante de \mathbf{X} , et ayant exactement la même loi (c’est à dire les mêmes probabilités de transition en un pas). Intuitivement, c’est comme si on avait deux « copies indépendantes » de notre processus, et qu’on suivait simultanément l’évolution des deux.

En soi, cela est un nouveau processus – appelons-le $\mathbf{Z} = (Z_t = (X_t, Y_t))_{t \in \mathbb{Z}^+}$. Que peut-on dire sur \mathbf{Z} ?

Pour commencer, l’espace des états de \mathbf{Z} est S^2 – l’ensemble des couples d’états (x, y) où x et y sont des états de S .

Quelles sont les de \mathbf{Z} ? Pour éviter la confusion, nous allons noter la famille des probabilités de transition après t unités de temps $Q^{(t)} = (Q_{\mathbf{x}, \mathbf{y}}^{(t)})_{\mathbf{x}, \mathbf{y} \in S^2}$, où $\mathbf{x} = (x_1, x_2)$ et $\mathbf{y} = (y_1, y_2)$ sont des états possibles de \mathbf{Z} (dans S^2) ; c’est à dire que x_1, x_2, y_1, y_2 sont dans S .

Puisque \mathbf{X} et \mathbf{Y} sont indépendantes, on doit avoir que pour tous $t \geq 0$ et pour tous $\mathbf{x} = (x_1, x_2), \mathbf{y} = (y_1, y_2) \in S^2$,

$$\begin{aligned} Q_{\mathbf{x}, \mathbf{y}}^{(t)} &= \mathbb{P} \{X_t = y_1, Y_t = y_2 \mid X_0 = x_1, Y_0 = x_2\} \\ &= \mathbb{P} \{X_t = y_1 \mid X_0 = x_1\} \mathbb{P} \{Y_t = y_2 \mid Y_0 = x_2\} \\ &= P_{x_1, y_1}^{(t)} P_{x_2, y_2}^{(t)}. \end{aligned}$$

On cherche maintenant à déterminer deux choses importantes.

- Est-ce que la chaîne \mathbf{Z} est irréductible ? Ça serait utile.
- Est-ce que la chaîne \mathbf{Z} est apériodique ? Ça serait aussi utile.

Pour que la chaîne \mathbf{Z} soit irréductible, il suffit qu'il existe, pour tous $\mathbf{x} = (x_1, x_2)$ et $\mathbf{y} = (y_1, y_2)$ dans S^2 un certain $t \in \mathbb{N}$ tel que $Q_{\mathbf{x}, \mathbf{y}}^{(t)} > 0$; ça voudrait dire que le graphe des transitions possibles admet un chemin de tout état dans S^2 à tout autre état dans S^2 . Pour que \mathbf{Z} soit apériodique en plus, il suffira de trouver pour tous $\mathbf{x}, \mathbf{y} \in S^2$ un $N(\mathbf{x}, \mathbf{y})$ tel que si $t > N$, alors $Q_{\mathbf{x}, \mathbf{y}}^{(t)} > 0$.

Puisque \mathbf{X} (et donc \mathbf{Y} aussi) est *apériodique*, pour tous $x, y \in S$ (souvenez-vous qu'on assume que \mathbf{X} est irréductible – sinon, c'est pour tous x connectés à y), il existe un $N(x, y)$ tel que pour tous $t > N(x, y)$, on a que

$$P_{x,y}^{(t)} > 0.$$

En particulier, si $\mathbf{x} = (x_1, x_2)$ et $\mathbf{y} = (y_1, y_2)$ sont deux états dans S^2 , on peut prendre

$$N(\mathbf{x}, \mathbf{y}) = \max \{N(x_1, y_1), N(x_2, y_2)\}.$$

Dès lors, si $t > N(\mathbf{x}, \mathbf{y})$, on a que $t > N(x_1, y_1)$ et $t > N(x_2, y_2)$, ce qui implique immédiatement :

$$Q_{\mathbf{x}, \mathbf{y}}^{(t)} = P_{x_1, y_1}^{(t)} P_{x_2, y_2}^{(t)} > 0,$$

et on a donc que \mathbf{Z} est bel et bien irréductible et apériodique.

Partie 2 : L'argument. On fixe $x \in S$ et $y \in S$. À partir d'ici, on va noter $\mathbf{x} = (x, y)$ et $\mathbf{y} = (y, y)$; ce sont deux états de S^2 qui sont particulièrement intéressants pour nous.

Il va y avoir deux cas à considérer :

- Soit l'état \mathbf{y} est *récurrent* (et donc tous les états de S^2 sont récurrents) ;
- soit l'état \mathbf{y} est *transient*.

Si \mathbf{y} est *transient*, alors pour tous $x, y \in S$, avec $\mathbf{x} = (x, y)$, on a que

$$Q_{\mathbf{x}, \mathbf{x}}^{(t)} = P_{x,x}^{(t)} P_{y,y}^{(t)} \rightarrow 0,$$

et en particulier, on devra donc avoir $(P_{x,x}^{(t)})^2 \rightarrow 0$ pour tout $x \in S$ lorsque t tend vers l'infini; tant que $P_{x,y}^{(t)}$ converge aussi vers 0, ça va.

Or, $Q_{(x,x),(y,y)}^{(t)} = (P_{x,y}^{(t)})^2$; supposons que $Q_{(x,x),(y,y)}^{(t)}$ ne converge pas vers 0. Alors, il existe $\epsilon > 0$ et une suite strictement croissante de temps t_k telle que $Q_{(x,x),(y,y)}^{(t_k)} > 0$;

Si $N_{(y,y)} = \sum_{u=1}^{\infty} \mathbb{1}_{\{Z_t=(y,y)\}}$ le nombre de visites en (y, y) par \mathbb{Z} , alors

$$\mathbb{E}_{(x,x)} [N_{(y,y)}] = \sum_{u=1}^{\infty} Q_{(x,x),(y,y)}^{(u)} > \sum_{k=1}^{\infty} \epsilon = +\infty,$$

et l'état (y, y) est récurrent, ce qui est contradictoire.

Donc, on doit avoir que $Q_{(x,x),(y,y)}^{(t)} = (P_{x,y}^{(t)})^2$ tend vers 0, et donc il faut aussi que $P_{x,y}^{(t)}$ tende vers 0 pour tous $x \in S$.

Si \mathbf{y} est *récurrent*, alors il y a plus de travail.

On considère maintenant

$$\tau_{\mathbf{y}}^+ = \inf \{t \in \mathbb{N} : Z_t = (X_t, Y_t) = (y, y) = \mathbf{y}\}.$$

Il s'agit du temps du premier retour simultané en y pour \mathbf{X} et \mathbf{Y} . Puisque \mathbf{y} est récurrent (et donc \mathbf{x} aussi, vu qu'ils communiquent), on a bien sûr que $\mathbb{P}_{\mathbf{x}} \{\tau_{\mathbf{y}}^+ < +\infty\} = 1$.

Une première remarque à faire, c'est que l'on a :

$$\{\tau_{\mathbf{y}}^+ = u\} = \{Z_0, Z_1, Z_2, \dots, Z_{u-1} \neq \mathbf{y}; Z_u = \mathbf{y}\};$$

donc, par la propriété de Markov et l'homogénéité de \mathbf{X} et \mathbf{Z} , et l'indépendance de \mathbf{X} et \mathbf{Y} , on a que pour tout $t, u \in \mathbb{Z}^+$ avec $t \geq u$, on a :

$$\mathbb{P}_{\mathbf{x}} \{X_t = y \mid \tau_{\mathbf{y}}^+ = u\} = \mathbb{P}_{\mathbf{x}} \{X_t = y \mid X_u = y\} = P_{y,y}^{(t-u)} = \mathbb{P}_{\mathbf{x}} \{Y_t = y \mid \tau_{\mathbf{y}}^+ = u\}.$$

On remarque d'abord que

$$\begin{aligned} P_{x,y}^{(t)} &= \mathbb{P}_{\mathbf{x}} \{X_t = y\} \\ &= \sum_{u \geq 1} \mathbb{P}_{\mathbf{x}} \{X_t = y, \tau_{\mathbf{y}}^+ = u\}. \end{aligned}$$

en vertu de la loi des probabilités totales.

On va séparer la sommation en deux portions : les termes où $u \leq t$, et les termes où $u > t$:

$$P_{x,y}^{(t)} = \sum_{u=1}^t \mathbb{P}_{\mathbf{x}} \{X_t = y, \tau_{\mathbf{y}}^+ = u\} + \sum_{u=t+1}^{\infty} \mathbb{P}_{\mathbf{x}} \{X_t = y, \tau_{\mathbf{y}}^+ = u\}.$$

Nous allons borner ces termes séparément.

Tout d'abord, on a que

$$\begin{aligned} \sum_{u=1}^t \mathbb{P}_{\mathbf{x}} \{X_t = y, \tau_{\mathbf{y}}^+ = u\} &= \sum_{u=1}^t \mathbb{P}_{\mathbf{x}} \{X_t = y \mid \tau_{\mathbf{y}}^+ = u\} \mathbb{P}_{\mathbf{x}} \{\tau_{\mathbf{y}}^+ = u\} \\ &\leq \sum_{u=1}^t \mathbb{P}_{\mathbf{x}} \{Y_t = y \mid \tau_{\mathbf{y}}^+ = u\} \mathbb{P}_{\mathbf{x}} \{\tau_{\mathbf{y}}^+ = u\} \\ &= \sum_{u=1}^t \mathbb{P}_{\mathbf{x}} \{Y_t = y, \tau_{\mathbf{y}}^+ = u\} \\ &= \mathbb{P}_{\mathbf{x}} \{Y_t = y\} = P_{y,y}^{(t)}. \end{aligned}$$

D'autre part, on a que

$$\sum_{u=t+1}^{\infty} \mathbb{P}_{\mathbf{x}} \{X_t = y, \tau_{\mathbf{y}}^+ = u\} \leq \sum_{u=t+1}^{\infty} \mathbb{P}_{\mathbf{x}} \{\tau_{\mathbf{y}}^+ = u\} = \mathbb{P}_{\mathbf{x}} \{\tau_{\mathbf{y}}^+ > t\}.$$

Finalement, tout ensemble, on obtient donc :

$$P_{x,y}^{(t)} \leq P_{y,y}^{(t)} + \mathbb{P}_{\mathbf{x}} \{\tau_{\mathbf{y}}^+ > t\},$$

et en faisant un argument analogue pour $P_{y,y}^{(t)}$ avec la chaîne \mathbf{Y} , on obtient également

$$P_{y,y}^{(t)} \leq P_{x,y}^{(t)} + \mathbb{P}_{\mathbf{x}} \{\tau_{\mathbf{y}}^+ > t\}.$$

En mettant ces deux inéquations ensemble, on trouve que

$$\left| P_{x,y}^{(t)} - P_{y,y}^{(t)} \right| \leq \mathbb{P}_{\mathbf{x}} \{\tau_{\mathbf{y}}^+ > t\},$$

et, en prenant la limite lorsque t tend vers l'infini, on trouve :

$$\lim_{t \rightarrow \infty} \left| P_{x,y}^{(t)} - P_{y,y}^{(t)} \right| \leq \mathbb{P}_{\mathbf{x}} \{\tau_{\mathbf{y}}^+ = +\infty\} = 0;$$

on complète la preuve par le théorème de la sandwich. \square

REMARQUE. On avait déjà réglé la question pour les états y transients ; la limite existe aussi, et elle ne dépend pas non plus de la condition initiale.

On s'attaque maintenant au plus gros morceau : que peut-on dire sur les états récurrents ?

PROPOSITION 1.15. *Soit $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ une chaîne de Markov homogène à temps discret sur l'espace des états discret S (dénombrable, fini ou non). Soit $y \in S$ un état récurrent apériodique.*

Alors,

- i. si $\mathbb{E}_y [\tau_y^+] < +\infty$, on a que

$$(1.7.3) \quad \lim_{t \rightarrow \infty} P_{y,y}^{(t)} = \frac{1}{\mathbb{E}_y [\tau_y^+]};$$

- ii. si $\mathbb{E}_y [\tau_y^+]$ diverge, on a que

$$(1.7.4) \quad \lim_{t \rightarrow \infty} P_{y,y}^{(t)} = 0.$$

DÉMONSTRATION. Prologue : Un peu d'intuition. La preuve qui suit est longue et assez subtile. Avant de se lancer dedans, je voudrais donner au moins un peu d'intuition pour justifier ce résultat possiblement surprenant.

La quantité $\mathbb{E}_y [\tau_y^+]$ représente le temps qu'il faut, en moyenne, pour revenir à y lorsqu'on en part. Vu sur un long intervalle de temps, si y est un état récurrent, on va y revenir tout le temps ; c'est donc (modulo des variations aléatoires) « un peu comme un processus dynamique périodique » – et $\mathbb{E}_y [\tau_y^+]$, avec l'homogénéité du processus, et la propriété de Markov, ça nous donne la « période », c'est à dire le temps qu'il faut pour revenir, ou le « nombre d'unités de temps par retour en y »

Quelle relation avec $P_{y,y}^{(t)}$? Eh bien, on a déjà vu que

$$\frac{1}{t} \sum_{u=1}^t P_{y,y}^{(t)} = \mathbb{E}_y \left[\frac{1}{t} N_y^{(t)} \right],$$

où $N_y^{(t)}$ est le nombre de retours en y jusqu'au temps t ; à la limite, $N_y^{(t)}$ représente la « fréquence » à laquelle on revient, ou le « nombre de retours par unités de temps ». Il est donc naturel d'imaginer qu'il y aurait une relation de proportionnalité inverse entre les probabilités $P_{y,y}^{(t)}$ à la limite, et $\mathbb{E}_y [\tau_y^+]$ – puisque dans tout système périodique, la période est l'inverse de la fréquence.

En théorie des processus stochastiques, des événements qui reviennent « stocha-périodiquement » ¹⁴, on appelle ça des *renouvellements*.

Étape 1 : L'équation du renouvellement. Notre première étape, ça va être d'obtenir une équation « simple » dont on pourra prendre une limite, au moment opportun (et en faisant très très attention).

L'équation du renouvellement est l'affirmation mathématique simple que « la probabilité qu'il y ait un renouvellement après le temps t est de 1. » Puisque l'état y est récurrent, c'est trivialement vérifié, bien sûr.

Ce qu'il y a de particulier, c'est la façon dont on va l'écrire, cette probabilité. On va la disséquer selon la valeur d'un temps aléatoire un peu particulier : *le temps du dernier renouvellement avant t .*

On définit

$$T_t = \sup \{ u \in \mathbb{Z}^+ : 0 \leq u < t, X_u = y \},$$

14. C'est un mot de mon invention, allez pas dire des affaires comme ça dans des colloques ; ils vont rire de vous !

le temps du dernier retour en y strictement avant t . Puis, on définit

$$\tau_y^{(t)} = \inf \{u \in \mathbb{Z}^+ : u \geq t, X_u = y\}.$$

On reconnaîtra $\tau_y^{(0)} = \tau_y$ et $\tau_y^{(1)} = \tau_y^+$. De façon générale, $\tau_y^{(t)}$ est le temps du *premier retour en y après t* . Si vous avez suivi : ça veut dire que T_t et $\tau_y^{(t)}$ sont des renouvellements consécutifs ; l'un avant t , l'autre après t (inclus).

Comme mentionné précédemment, l'équation du renouvellement, c'est simplement ça :

$$\mathbb{P}_y \left\{ \tau_y^{(t)} < +\infty \right\} = 1.$$

Mais on développe en conditionnant par rapport à la valeur de T_t :

$$\begin{aligned} 1 &= \mathbb{P}_y \left\{ \tau_y^{(t)} < +\infty \right\} \\ &= \sum_{u=1}^t \mathbb{P}_y \left\{ T_t = t - u, \tau_y^{(t)} < +\infty \right\}. \end{aligned}$$

Et c'est ici qu'on fait le tour de magie : observez que

$$\left\{ T_t = t - u, \tau_y^{(t)} < +\infty \right\} = \left\{ X_{t-u} = y, t \leq \tau_y^{(t-u+1)} < +\infty \right\};$$

en effet, si on sait qu'on était en y au temps $t - u$, et que le premier renouvellement après $t - u + 1$ a eu lieu après le temps t , alors forcément

- d'une part, $t - u$ était, effectivement le dernier renouvellement ;
- d'autre part, le premier renouvellement après t aura lieu éventuellement.

On a donc finalement

$$\begin{aligned} 1 &= \sum_{u=1}^t \mathbb{P}_y \left\{ T_t = t - u, \tau_y^{(t)} < +\infty \right\} \\ &= \sum_{u=1}^t \mathbb{P}_y \left\{ X_{t-u} = y, t \leq \tau_y^{(t-u+1)} < +\infty \right\} \\ &= \sum_{u=1}^t \mathbb{P}_y \left\{ \tau^{(t-u+1)} \geq t \mid X_{t-u} = y \right\} \mathbb{P}_y \{X_{t-u} = y\} \end{aligned}$$

Par la propriété de Markov, on a que, sachant que $X_{t-u} = y$, la loi de $\tau_y^{(t-u+1)}$ est la même que celle de $t - u + \tau_y^+$; en fait, on a

$$\mathbb{P}_y \left\{ \tau^{(t-u+1)} \geq t \mid X_{t-u} = y \right\} = \mathbb{P}_y \left\{ \tau_y^+ \geq u \right\}.$$

D'autre part, $\mathbb{P}_y \{X_{t-u} = y\} = P_{y,y}^{(t-u)}$, évidemment.

Donc, finalement, on se retrouve avec *l'équation du renouvellement*, sous sa forme finale :

$$\sum_{u=1}^t P_{y,y}^{(t-u)} \mathbb{P}_y \left\{ \tau_y^+ \geq u \right\} = 1 \dots$$

Enfin, presque. On va utiliser une dernière petite astuce, qui nous aidera plus tard : on va transformer la sommation (dont une borne dépend de t) en une série (dont les bornes sont constantes). Comment ?

On rajoute simplement un facteur $\mathbb{1}_{\{u \leq t\}}$, qui vaut 1 si $u \leq t$, et 0 sinon :

$$(\dagger) \quad \sum_{u=1}^t P_{y,y}^{(t-u)} \mathbb{P}_y \{ \tau_y^+ \geq u \} = \sum_{u=1}^{\infty} \mathbb{1}_{\{u \leq t\}} P_{y,y}^{(t-u)} \mathbb{P}_y \{ \tau_y^+ \geq u \} = 1.$$

Et voilà ! L'équation du renouvellement.

Étape 2 : Trouver la limite pour une sous-suite convergente. Ça pourrait sembler anodin, comme ça. Mais le défi, c'est qu'à priori, on *ne sait pas* que la limite existe. Il faut faire la preuve nous-même !

Par contre, on sait que la suite $P_{y,y}^{(t)}$ est bornée – forcément, puisque ce sont des probabilités. On a donc immédiatement le résultat suivant : cette suite admet une *limite supérieure* et une *limite inférieure*.¹⁵

On va donc définir $\lambda = \limsup_{t \rightarrow \infty} P_{y,y}^{(t)}$.

Bien sûr, il existe forcément une sous-suite $P_{y,y}^{(t_k)}$ (où les t_k sont une suite strictement croissante d'entiers positifs) telle que

$$\lim_{k \rightarrow \infty} P_{y,y}^{(t_k)} = \lambda.$$

Comme y est apériodique, on peut sans perdre de généralité supposer que $P_{y,y} = \mathbb{P}_y \{ \tau_y^+ = 1 \} > 0$.

On va montrer que $\lim_{k \rightarrow \infty} P_{y,y}^{(t_k-1)} = \lambda$ aussi ; en effet, on considère l'équation suivante :

$$\begin{aligned} P_{y,y}^{(t_k)} &= \sum_{u=1}^{t_k} \mathbb{P}_y \{ X_{t_k} = y, \tau_y^+ = u \} \\ &= \sum_{u=1}^{t_k} \mathbb{P}_y \{ X_{t_k} = y \mid \tau_y^+ = u \} \mathbb{P}_y \{ \tau_y^+ = u \}. \end{aligned}$$

On a que

$$\{ \tau_y^+ = u \} = \{ X_1, X_2, X_3, \dots, X_{u-1} \neq y; X_u = y \},$$

et, par homogénéité et par la propriété de Markov, on obtient que

$$\mathbb{P}_y \{ X_{t_k} = y \mid \tau_y^+ = u \} = \mathbb{P}_y \{ X_{t_k} = y \mid X_u = y \} = P_{y,y}^{(t_k-u)}.$$

On obtient donc

$$(\ddagger) \quad P_{y,y}^{(t_k)} = \sum_{u=1}^{t_k} \mathbb{P}_y \{ \tau_y^+ = u \} P_{y,y}^{(t_k-u)}.$$

En prenant la limite inférieure de part et d'autre, on trouve :

$$\lambda \leq \mathbb{P}_y \{ \tau_y^+ = 1 \} \liminf_{t \rightarrow \infty} P_{y,y}^{(t_k-1)} + \sum_{u=2}^{\infty} \mathbb{P}_y \{ \tau_y^+ = u \} \limsup_{t \rightarrow \infty} P_{y,y}^{(t_k-u)},$$

mais bien sûr, on a que $\limsup_{t \rightarrow \infty} P_{y,y}^{(t_k-u)} \leq \lambda$, puisque $P_{y,y}^{(t_k-u)}$ est une sous-suite de $P_{y,y}^{(t)}$ pour tout u .

Finalement, on trouve donc :

$$\lambda \leq \mathbb{P}_y \{ \tau_y^+ = 1 \} \liminf_{k \rightarrow \infty} P_{y,y}^{(t_k-1)} + \lambda \mathbb{P}_y \{ \tau_y^+ > 1 \},$$

¹⁵. Consulter la section B.2 de l'annexe B au besoin.

ou, en réarrangeant :

$$\limsup_{k \rightarrow \infty} P_{y,y}^{(t_k-1)} \leq \lambda \leq \liminf_{k \rightarrow \infty} P_{y,y}^{(t_k-1)},$$

et on obtient immédiatement

$$\lim_{k \rightarrow \infty} P_{y,y}^{(t_k-1)} = \lambda.$$

Par induction, on déduit que

$$\lim_{k \rightarrow \infty} P(t_k - u)_{y,y} = \lambda,$$

pour tout $u \geq 1$.

On utilise maintenant notre équation (†) – l'équation du renouvellement.

On distingue maintenant les deux cas :

i $\mathbb{E}_y [\tau_y^+] < +\infty$. Dans ce cas, on a bien sûr que

$$\mathbb{E}_y [\tau_y^+] = \sum_{u=1}^{\infty} \mathbb{P}_y \{ \tau_y^+ \geq u \}.$$

On peut utiliser le théorème de convergence dominée pour prendre la limite ; on trouve que

$$\begin{aligned} 1 &= \lim_{k \rightarrow \infty} \sum_{u=1}^{\infty} \mathbb{1}_{\{u \leq t_k\}} P_{y,y}^{(t_k-u)} \mathbb{P}_y \{ \tau_y^+ \geq u \} \\ &= \lambda \sum_{u=1}^{\infty} \mathbb{P}_y \{ \tau_y^+ \geq u \} \\ &= \lambda \mathbb{E}_y [\tau_y^+], \end{aligned}$$

d'où on tire que

$$\lim_{k \rightarrow \infty} P_{y,y}^{(t_k)} = \frac{1}{\mathbb{E}_y [\tau_y^+]}. \quad \text{ii}$$

ii $\mathbb{E}_y [\tau_y^+]$ diverge. Dans ce cas, on ne peut pas utiliser directement le théorème de convergence dominée ; il faut user d'un peu de ruse.

On commence par fixer un $n \in \mathbb{N}$; pour $t_k > n$, on a bien sûr que

$$\sum_{u=0}^{t_k} \mathbb{P}_y \{ \tau_y^+ \geq u \} P_{y,y}^{(t_k-u)} \geq \sum_{u=1}^n \mathbb{P}_y \{ \tau_y^+ \geq u \} P_{y,y}^{(t_k-u)};$$

Bien sûr, en utilisant l'équation du renouvellement (†), on trouve que

$$\begin{aligned} 1 &\geq \liminf_{k \rightarrow \infty} \sum_{u=1}^n \mathbb{P}_y \{ \tau_y^+ \geq u \} P_{y,y}^{(t_k-u)} \\ &\geq \lambda \sum_{u=1}^n \mathbb{P}_y \{ \tau_y^+ \geq u \}. \end{aligned}$$

Mais évidemment, puisque $\mathbb{E}_y [\tau_y^+]$ diverge, on peut obtenir une somme $\sum_{u=1}^n \mathbb{P}_y \{ \tau_y^+ \geq u \}$ arbitrairement grande en choisissant n comme on veut ; la seule façon de s'en tirer, c'est qu'on doit avoir $\lambda = 0$; on en tire donc

$$\lim_{k \rightarrow \infty} P_{y,y}^{(t_k)} = 0.$$

Étape 3 : La conclusion. On y est presque !

Ce qu'on a montré, c'est que : si $\lambda = \limsup_{t \rightarrow \infty} P_{y,y}^{(t)}$, alors

$$\lambda = \begin{cases} \frac{1}{\mathbb{E}_y [\tau_y^+]} & \text{si } \mathbb{E}_y [\tau_y^+] \text{ converge;} \\ 0 & \text{sinon.} \end{cases}$$

On peut également refaire une preuve en tout point similaire pour montrer la même chose pour la limite inférieure ; il suit donc que

$$\lim_{t \rightarrow \infty} P_{y,y}^{(t)} = \begin{cases} \frac{1}{\mathbb{E}_y [\tau_y^+]} & \text{si } \mathbb{E}_y [\tau_y^+] \text{ converge;} \\ 0 & \text{sinon.} \end{cases}.$$

□

COROLLAIRE 1.4. Soit $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ une chaîne de Markov homogène à temps discret sur l'espace des états discret S , et soit $y \in S$ un état récurrent apériodique.

Alors, pour tout $x \in S$ connecté à y , on a que

$$(1.7.5) \quad \lim_{t \rightarrow \infty} P_{x,y}^{(t)} = \begin{cases} \frac{1}{\mathbb{E}_y [\tau_y^+]} & \text{si } \mathbb{E}_y [\tau_y^+] < +\infty; \\ 0 & \text{sinon.} \end{cases}$$

DÉMONSTRATION. Il suffit d'appliquer la propositions 1.15 et 1.14. □

OUF ! Grosse preuve, technique, subtile, désagréable... Souvenons-nous de ce qu'on était en train de faire : on était en train d'essayer d'identifier les cas où la distribution-limite existe ; à date, ce qu'on a montré, c'est que :

- Pour les états *transients*, la limite existe tout le temps, indépendamment de la condition initiale (Proposition 1.13) ;
- Si la limite existe pour l'état y partant de y , elle existe aussi pour toutes les conditions initiales au départ des états dans la même classe que y , et toutes ces limites-là sont égales entre elles (Proposition 1.14) ;
- La limite existe pour les états *récurrents apériodiques* (Proposition 1.15)

À date, ça couvre tout, sauf les états récurrents périodiques. On fait la proposition suivante :

PROPOSITION 1.16. Soit $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ une chaîne de Markov homogène à temps discret sur l'espace des états discret S (fini ou dénombrable), et soit $y \in S$ un état récurrent périodique de période $d > 1$.

- i. Si $\mathbb{E}_y [\tau_y^+]$ diverge, on a que

$$(1.7.6) \quad \lim_{t \rightarrow \infty} P_{y,y}^{(t)} = 0.$$

ii. Si $\mathbb{E}_y [\tau_y^+] < +\infty$, on a que

$$(1.7.7) \quad \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{u=1}^t P_{y,y}^{(u)} = \lim_{t \rightarrow \infty} \mathbb{E}_y \left[\frac{1}{t} N_y^{(t)} \right] = \frac{1}{\mathbb{E}_y [\tau_y^+]},$$

où $N_y^{(t)} = \sum_{u=1}^t \mathbb{1}_{\{X_u=y\}}$ est le nombre de retours en y jusqu'au temps t inclus.

DÉMONSTRATION. Étape 1 : la chaîne « dé-périodisée ». On considère $\mathbf{Y} = (Y_t = X_{dt})_{t \in \mathbb{Z}^+}$; il s'agit d'une chaîne de Markov homogène à temps discret sur l'espace des états S . Intuitivement, \mathbf{Y} est la même chaîne que \mathbf{X} , mais si on ne l'observe qu'après toutes les d unités de temps. On va appeler $Q^{(t)}$ la famille des probabilités de transition en t pas de temps ; on a évidemment que pour tous $x, y \in S$,

$$Q_{x,y}^{(t)} = P_{x,y}^{(dt)}.$$

Puisque l'état y est de période $d > 1$ pour la chaîne \mathbf{X} , cela signifie que si $P_{y,y}^{(t)} > 0$, alors il existe $k \in \mathbb{Z}^+$ tel que $t = dk$. En particulier,

$$\sum_{t=1}^{\infty} P_{y,y}^{(t)} = \sum_{k=1}^{\infty} P_{y,y}^{(dk)} = \sum_{t=1}^{\infty} Q_{y,y}^{(t)}.$$

Puisque y est récurrent pour \mathbf{X} , alors $\sum_{t \geq 1} P_{y,y}^{(t)}$ diverge. Donc, $\sum_{t \geq 1} Q_{y,y}^{(t)}$ diverge aussi, et y est récurrent également pour \mathbf{Y} . De plus, si on suppose que y est de période d_Y pour la chaîne \mathbf{Y} , alors y serait de période dd_Y pour la chaîne \mathbf{X} – on doit donc avoir que $d_Y = 1$; y est *apériodique* pour la chaîne \mathbf{Y} .

Il est clair que τ_y^+ est un multiple de d également. En effet, pour tout t non-multiple de d ,

$$\mathbb{P}_y \{ \tau_y^+ = t \} \leq P_{y,y}^{(t)} = 0.$$

On définit

$$\tau_y^{+,(\mathbf{Y})} = \inf \{ t \in \mathbb{N} : Y_t = X_{dt} = y \}$$

le temps du premier retour en y dans la chaîne \mathbf{Y} . On a directement

$$\tau_y^{+,(\mathbf{Y})} = \frac{1}{d} \tau_y^+;$$

la conséquence, c'est bien sûr que $\mathbb{E}_y [\tau_y^{+,(\mathbf{Y})}] < +\infty$ si et seulement si $\mathbb{E}_y [\tau_y^+] < +\infty$.

i Si $\mathbb{E}_y [\tau_y^+]$ diverge, la preuve est relativement simple ; pour tout t non-multiple de d , on a que $P_{y,y}^{(t)} = 0$. Donc,

$$\liminf_{t \rightarrow \infty} P_{y,y}^{(t)} = 0.$$

On a également que, puisque les seuls termes non-nuls de la suite $(P_{y,y}^{(t)})_{t \in \mathbb{Z}^+}$ sont ceux où t est multiple de d , alors

$$\limsup_{t \rightarrow \infty} P_{y,y}^{(t)} = \limsup_{t \rightarrow \infty} P_{y,y}^{(dt)}.$$

Par ailleurs, on a que

$$\begin{aligned} \limsup_{t \rightarrow \infty} P_{y,y}^{(dt)} &= \lim_{t \rightarrow \infty} Q_{y,y}^{(t)} \\ &= 0, \end{aligned}$$

en vertu de la Proposition 1.15 puisque $\mathbb{E}_y [\tau_y^{+,(\mathbf{Y})}]$ diverge également.

Donc, $\lim_{t \rightarrow \infty} P_{y,y}^{(t)} = 0$.

ii Si $\mathbb{E}_y [\tau_y^+] < +\infty$, alors on a également que $\mathbb{E}_y [\tau_y^{+,(\mathbf{Y})}] = \frac{1}{d} \mathbb{E}_y [\tau_y^+] < +\infty$.

Comme on l'a déjà vu, la suite des $P_{y,y}^{(t)}$ n'admet pas forcément de limite (voir l'exemple 1.21).

On définit

$$q_t = \lfloor t/d \rfloor =: \sup \{n \in \mathbb{Z} : n \leq t/d\}, \quad r_t = t - dn_t,$$

soit n_t le quotient (entier) de t par d , et r_t le reste de la division. Il est clair que $0 \leq r_t \leq d-1$. On a que

$$\begin{aligned} \frac{1}{t} \sum_{u=1}^t P_{y,y}^{(u)} &= \frac{1}{t} \sum_{u=1}^{dn_t+r_t} P_{y,y}^{(u)} \\ &= \frac{1}{t} \sum_{u=1}^{dn_t} P_{y,y}^{(u)} + \frac{1}{t} \sum_{u=dn_t+1}^{dn_t+r_t} P_{y,y}^{(u)}, \end{aligned}$$

mais pour u entre dn_t+1 et dn_t+r_t inclusivement, u est un non-multiple de d , et $P_{y,y}^{(u)} = 0$. Donc, on a

$$\begin{aligned} \frac{1}{t} \sum_{u=1}^t P_{y,y}^{(u)} &= \frac{1}{t} \sum_{u=1}^{dn_t} P_{y,y}^{(u)} \\ &= \frac{1}{t} \sum_{k=0}^{n_t-1} \sum_{r=1}^d P_{y,y}^{(dk+r)}, \end{aligned}$$

mais, similairement, $P_{y,y}^{(dk+r)}$ est non-nul seulement si $r = d$ (puisque $1 \leq r \leq d$). Donc,

$$\begin{aligned} \frac{1}{t} \sum_{u=1}^t P_{y,y}^{(u)} &= \frac{1}{t} \sum_{k=0}^{n_t-1} P_{y,y}^{(dk+d)} \\ &= \frac{1}{t} \sum_{k=1}^{n_t} P_{y,y}^{(dk)} \\ &= \frac{1}{t} \sum_{u=1}^{n_t} Q_{y,y}^{(u)}. \end{aligned}$$

On passe à la limite, en multipliant par n_t/n_t :

$$\lim_{t \rightarrow \infty} \frac{1}{t} \sum_{u=1}^t P_{y,y}^{(t)} = \lim_{t \rightarrow \infty} \frac{n_t}{t} \cdot \lim_{t \rightarrow \infty} \frac{1}{n_t} \sum_{u=1}^{n_t} Q_{y,y}^{(u)}.$$

D'une part, $(t-1)/d \leq n_t/d \leq t/d$, et par le théorème de la Sandwich,

$$\lim_{t \rightarrow \infty} \frac{n_t}{t} = \frac{1}{d}.$$

D'autre part, on sait que

$$\lim_{t \rightarrow \infty} Q_{y,y}^{(t)} = \frac{1}{\mathbb{E}_y [\tau_y^{+,(\mathbf{Y})}]} = \frac{d}{\mathbb{E}_y [\tau_y^+]},$$

en vertu de la Proposition 1.15 puisque y est un état récurrent apériodique pour la chaîne \mathbf{Y} .

Or, par un résultat d'analyse¹⁶, puisque $\lim_{t \rightarrow \infty} Q_{y,y}^{(t)}$ existe, il suit que

$$\lim_{t \rightarrow \infty} \frac{1}{t} \sum_{u=1}^t Q_{y,y}^{(u)} = \lim_{t \rightarrow \infty} Q_{y,y}^{(t)} = \frac{d}{\mathbb{E}_y [\tau_y^+]};$$

en particulier, la sous-suite $\frac{1}{n_t} \sum_{u=1}^{n_t} Q_{y,y}^{(u)}$ doit admettre la même limite; on obtient donc finalement

$$\lim_{t \rightarrow \infty} \frac{1}{t} \sum_{u=1}^t P_{y,y}^{(t)} = \frac{1}{d} \cdot \frac{d}{\mathbb{E}_y [\tau_y^+]} = \frac{1}{\mathbb{E}_y [\tau_y^+]}.$$

□

Et ça y est enfin! On a fait le tour de tous les types d'états qui existent, et identifié les moments où une limite existe, et laquelle! Pour célébrer, nous allons regrouper tous ces résultats dans un seul gros théorème!

THÉORÈME 1.1 (Théorème ergodique – temps discret, espace d'états discret). *Soit $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ une chaîne de Markov homogène à temps discret sur l'espace des états discret S (fini ou dénombrable), avec probabilités de transition en t pas $P^{(t)} = (P_{x,y}^{(t)})_{x,y \in S}$.*

- i. *Pour tout état $y \in S$ exceptés les états récurrents périodiques où $\mathbb{E}_y [\tau_y^+] < +\infty$, on a que pour tout $x \in S$,*

$$(1.7.8) \quad \lim_{t \rightarrow \infty} P_{x,y}^{(t)} = \begin{cases} \mathbb{P}_x \{ \tau_y < +\infty \} \cdot \frac{1}{\mathbb{E}_y [\tau_y^+]} & \text{si } \mathbb{E}_y [\tau_y^+] < +\infty; \\ 0 & \text{sinon.} \end{cases}$$

- ii. *Soit $N_y^{(t)} = \sum_{u=1}^t \mathbb{1}_{\{X_u=y\}}$ le nombre de visites/retours en y entre les temps 1 et t inclus. Alors, pour tous $x, y \in S$, sans exception,*

$$(1.7.9) \quad \lim_{t \rightarrow \infty} \mathbb{E}_y \left[\frac{1}{t} N_y^{(t)} \right] = \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{u=1}^t P_{x,y}^{(u)} = \begin{cases} \mathbb{P}_x \{ \tau_y < +\infty \} \cdot \frac{1}{\mathbb{E}_y [\tau_y^+]} & \text{si } \mathbb{E}_y [\tau_y^+] < +\infty; \\ 0 & \text{sinon.} \end{cases}$$

DÉMONSTRATION. Les propositions 1.13, 1.15 et 1.16 se chargent déjà de la preuve lorsque $x = y$; dans ce qui suit, on assume que $x \neq y$.

16. Voir la section B.2 de l'annexe B.

i Si on assume que $\lim_{t \rightarrow \infty} P_{y,y}^{(t)} = \lambda$, alors on a que

$$\begin{aligned}
 P_{x,y}^{(t)} &= \sum_{u=1}^t \mathbb{P}_x \{X_t = y, \tau_y^+ = u\} \\
 &= \sum_{u=1}^t \mathbb{P}_x \{X_t = y \mid \tau_y = u\} \mathbb{P}_x \{\tau_y = u\} \\
 &= \sum_{u=1}^t \mathbb{P}_x \{X_t = y \mid X_u = y\} \mathbb{P}_x \{\tau_y = u\} \\
 &= \sum_{u=1}^t P_{y,y}^{(t-u)} \mathbb{P}_x \{\tau_y = u\} \\
 (\dagger) \quad &= \sum_{u=1}^{\infty} \mathbb{1}_{\{u \leq t\}} P_{y,y}^{(t-u)} \mathbb{P}_x \{\tau_y = u\}.
 \end{aligned}$$

Par le théorème de convergence dominée, puisque

$$\lim_{t \rightarrow \infty} \mathbb{1}_{\{u \leq t\}} P_{y,y}^{(t-u)} \mathbb{P}_x \{\tau_y = u\} = \lambda \mathbb{P}_x \{\tau_y = u\},$$

on obtient donc que :

$$\lim_{t \rightarrow \infty} P_{x,y}^{(t)} = \lambda \sum_{u=1}^{\infty} \mathbb{P}_x \{\tau_y = u\} = \lambda \mathbb{P}_x \{\tau_y < +\infty\}.$$

On complète la preuve en appliquant simplement les propositions 1.13, 1.15 et 1.16 ; le seul cas non-couvert par ces propositions est le cas où y serait un état récurrent périodique mais que $\mathbb{E}_y [\tau_y^+] < +\infty$.

ii Si on sait que $\lim_{t \rightarrow \infty} P_{x,y}^{(t)}$ existe, alors par un résultat d'analyse, on a que

$$\lim_{t \rightarrow \infty} \frac{1}{t} \sum_{u=1}^t P_{x,y}^{(u)} = \lim_{t \rightarrow \infty} P_{x,y}^{(t)};$$

il suffit alors d'appliquer la partie **i** du présent théorème pour obtenir le résultat.

Dans le cas où un état y est récurrent périodique et que $\mathbb{E}_y [\tau_y^+] < +\infty$, on utilise l'équation (†) pour obtenir

$$\begin{aligned}
 \frac{1}{t} \sum_{u=1}^t P_{x,y}^{(t)} &= \frac{1}{t} \sum_{u=1}^t \sum_{k=1}^u P_{y,y}^{(u-k)} \mathbb{P}_x \{\tau_y = k\} = \frac{1}{t} \sum_{k=1}^t \sum_{u=k}^t P_{y,y}^{(u-k)} \mathbb{P}_x \{\tau_y = k\} \\
 &= \sum_{k=1}^{\infty} \mathbb{1}_{\{k \leq t\}} \left(\frac{1}{t} \sum_{u=1}^{t-k} P_{y,y}^{(u)} \right) \mathbb{P}_x \{\tau_y = k\}.
 \end{aligned}$$

Supposons que $\lim_{t \rightarrow \infty} \frac{1}{t} \sum_{u=1}^t P_{y,y}^{(u)} = \lambda$.

Pour tout $k \in \mathbb{N}$, on a bien sûr que

$$\begin{aligned}
 \lim_{t \rightarrow \infty} \mathbb{1}_{\{k \leq t\}} \left(\frac{1}{t} \sum_{u=1}^{t-k} P_{y,y}^{(u)} \right) &= \lim_{t \rightarrow \infty} \mathbb{1}_{\{k \leq t\}} \cdot \lim_{t \rightarrow \infty} \frac{t-k}{t} \cdot \lim_{t \rightarrow \infty} \frac{1}{t-k} \sum_{u=1}^{t-k} P_{y,y}^{(u)} \\
 &= \lambda.
 \end{aligned}$$

Par le théorème de la convergence dominée, on obtient donc :

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{u=1}^t P_{x,y}^{(u)} &= \lim_{t \rightarrow \infty} \sum_{k=1}^{\infty} \mathbb{1}_{\{k \leq t\}} \left(\frac{1}{t} \sum_{u=k}^t P_{y,y}^{(u)} \right) \mathbb{P}_x \{ \tau_y = k \} \\ &= \lambda \sum_{k=1}^{\infty} \mathbb{P}_x \{ \tau_y = k \} = \lambda \mathbb{P}_x \{ \tau_y < +\infty \}. \end{aligned}$$

Il suffit alors d'appliquer la partie **ii** de la proposition 1.16 pour compléter la preuve. □

Et voilà ! On a un théorème qui encapsule complètement le comportement de la suite $P_{x,y}^{(t)}$, pour tous x et y ; on sait

- quand la limite existe : elle existe pour tous les états $y \in S$ *apériodiques* ou *récurrents* avec $\mathbb{E}_y [\tau_y^+] = +\infty$, ou *transients* ;
- la valeur de cette limite si elle existe.

1.7.3. De retour aux distributions stationnaires. Pour obtenir cette dernière information, on a utilisé des méthodes analytiques, et ç'a été très ardu. Mais souvenez-vous, au début de la section 1.7.1, quand on parlait de distributions stationnaires... Le motif n'était-il pas de trouver une solution simple au problème de la distribution limite ?

Voici venu le temps de boucler la boucle, en rendant plus rigoureux le résultat suggéré au début du chapitre :

THÉORÈME 1.2. *Soit $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ une chaîne de Markov homogène à temps discret sur l'espace d'états discret $S = \{1, 2, 3, \dots\}$ (fini ou dénombrable) irréductible et récurrente, avec $\mathbb{E}_j [\tau_j^+] < +\infty$ pour tout j .*

- i. *La distribution stationnaire $\pi = (\pi_i)_{i \in S}$ existe ; elle est unique.*
- ii. *Pour tous $j \in S$, on a que*

$$(1.7.10) \quad \pi_j = \frac{1}{\mathbb{E}_j [\tau_j^+]}. \quad \square$$

DÉMONSTRATION. i *L'argument est présenté dans le cas d'un nombre d'états fini, parce que c'est plus simple. Toutefois, il existe une version de cet argument qui fonctionne pour des espaces d'états infinis dénombrables aussi.*

Dans le cas où S a un nombre d'états finis, on peut considérer P comme une matrice $n \times n$, où $n = |S|$.

La matrice $M = P - \mathbb{I}_n$ n'est pas inversible ; en effet, on remarque que, si on écrit $\mathbf{v}_j = (P_{i,j} - \delta_{i,j})_{i \in S}$ le vecteur correspondant à la j ème colonne de la matrice $P - \mathbb{I}_n$, alors on a que :

$$\sum_{j \in S} \mathbf{v}_j = \mathbf{0};$$

vu que pour tout i , on a que

$$\sum_{j \in S} [v_j]_i = \sum_{j \in S} (P_{i,j} - \delta_{i,j}) = \sum_{j \in S} P_{i,j} - 1 = 0.$$

Donc, les colonnes sont linéairement dépendantes.

Les solutions stationnaires sont les éléments de $\ker M$, le noyau de M (si on applique la multiplication par la gauche – attention !)

Sans perdre de généralité, on assume que $P_{i,j} > 0$ pour tous $i, j \in S$; donc, $P_{i,j} \in (0, 1)$ pour tous $i, j \in S$. Si ce n'est pas le cas, puisque \mathbf{X} est une chaîne de Markov irréductible, il faut nécessairement que ce soit le cas pour tout $P^{(t_N)}$, pour un certain $t_N \geq N$ pour tous N , mais l'argument est le même.

Donc, $M_{i,i} < 0$ pour tout $i \in S$ et $M_{i,j} > 0$ pour tous $i \neq j$ dans S . Les $\{\mathbf{v}_i : 1 \leq i \leq n\}$ engendrent le même espace que les vecteurs $\{\mathbf{e}_i : 1 \leq i \leq n-1\}$; on conclue que $\dim \ker P = n - (n-1) = 1$; un espace de dimension $n-1$.

Donc, toutes les solutions stationnaires sont proportionnelles.

Supposons donc que $\lambda' = (\lambda'_j)_{j \in S}$ est aussi une distribution stationnaire. Alors, on doit avoir que $\lambda' = \alpha \lambda$ pour un certain $\alpha \in \mathbb{R}$. Mais on doit aussi avoir $\alpha \sum_{j \in S} \lambda'_j = 1$, donc $\alpha = 1$. Donc $\lambda' = \lambda$ et la distribution stationnaire est unique.

- ii Notre objectif maintenant, c'est de montrer que $\pi_x = \frac{1}{\mathbb{E}_x[\tau_x^+]}$ pour tout x ; on commence par écrire

$$\lambda_y = \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{u=1}^t P_{x,y}^{(u)};$$

on peut se permettre d'ignorer la dépendance en x puisque comme notre chaîne est irréductible et récurrente, on a forcément que $\mathbb{P}_x\{\tau_y < +\infty\} = 1$ pour tous x, y , et la limite ne dépend pas de x .

Bien sûr, par le théorème ergodique, on a que

$$\lambda_y = \frac{1}{\mathbb{E}_y[\tau_y^+]};$$

Ce qu'il faut faire maintenant, c'est montrer que λ_y est une distribution stationnaire. On a d'une part que :

$$\begin{aligned}
\sum_{z \in S} \lambda_z P_{z,y} &= \sum_{z \in S} \left(\lim_{t \rightarrow \infty} \frac{1}{t} \sum_{u=1}^t P_{x,z}^{(u)} \right) P_{z,y} \\
&= \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{u=1}^t \sum_{z \in S} P_{x,z}^{(u)} P_{z,y} \\
&= \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{u=1}^t P_{x,y}^{(u+1)} \\
&= \lim_{t \rightarrow \infty} \frac{t+1}{t} \frac{1}{t+1} \left(\sum_{u=1}^{t+1} P_{x,y}^{(u)} - P_{x,y} \right) \\
&= \lim_{t \rightarrow \infty} \frac{1}{t+1} \sum_{u=1}^{t+1} P_{x,y}^{(u)} \\
&= \lambda_y.
\end{aligned}$$

(Pour obtenir la seconde égalité, il faut sortir la limite de la sommation, possiblement à l'aide du théorème de convergence dominée. Le reste est le résultat des équations de Chapman-Kolmogorov, et de réarrangement de termes et de changements d'indices.)

On a donc que les λ_x forment une solution stationnaire ; sont-ils une distributions stationnaire ? Eh bien oui ! En effet,

$$\begin{aligned}
\sum_{y \in S} \lambda_y &= \sum_{x \in S} \left(\lim_{t \rightarrow \infty} \frac{1}{t} \sum_{u=1}^t P_{x,y}^{(u)} \right) \\
&= \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{u=1}^t \sum_{y \in S} P_{x,y}^{(u)} \\
&= \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{u=1}^t 1 = 1.
\end{aligned}$$

Par unicité de la distribution stationnaire, on doit donc conclure que

$$\pi_x = \lambda_x.$$

□

EXEMPLE 1.23. Alexandre a eu un désastreux accident de voiture sur l'autoroute 15. Heureusement, pour Marianne, sa passagère, il n'y a pas de blessé·e·s. Pour Alexandre, ce qui le soulage surtout, c'est qu'il est assuré chez la *Métropo'n'cenne*.

Il y a 5 classes d'assuré·e·s – la classe dans laquelle un·e assuré·e se trouve détermine la valeur de la prime que ce·tte dernier/ère devra payer en fonction d'une prime maximale : 100%, 90%, 85%, 80%, et 75%.

À chaque renouvellement (annuel) de son contrat, Alexandre se voit assigné à une classe selon le montant qui lui a été accordé suite aux réclamations qu'il a soumises dans la dernière

année : il est déclassé une fois pour chaque tranche complète de 100\$ qui lui ont été versées, tandis qu'il est surclassé si il n'a effectué aucune réclamation.

La *Métropo'n'cenne* estime que, chaque année,

- la moitié de ses client·e·s reçoivent moins de 100\$.
- 35% de ses client·e·s reçoivent entre 100 et 200\$.
- 12% de ses client·e·s reçoivent entre 200 et 300\$.
- 2,8% de ses client·e·s reçoivent entre 300 et 400\$.
- 0,2% de ses client·e·s reçoivent plus 400\$.

On considère X_t la classe d'assuré·e·s à laquelle Alexandre appartient pour l'année t , de 1 à 5 – 1 étant la classe la plus désavantagée (qui paye 100% de la prime maximale) et 5 étant la classe la plus avantagée (qui ne paye que 75% de la prime maximale).

- (a) Donner la matrice des probabilités de transition en un pas pour $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$.
- (b) Tracer le graphe des transitions possibles, et identifier la/les classe·s d'équivalence. Pour chaque classe, indiquer le type (récurrent/transient) et donner la période.
- (c) Est-ce que cette chaîne admet une distribution stationnaire ? Si oui, trouver la distribution stationnaire unique, ou donner une expression générale pour toutes les distributions stationnaires possibles si il y en a plus d'une.
- (d) À long terme, quelle proportion des assuré·e·s paye plus de 80% de la prime maximale en moyenne chaque année ?
- (e) Avant son accident, cela faisait déjà longtemps qu'Alexandre était assuré chez la *Métropo'n'cenne* pour son auto. Si la prime maximale est de 1000\$ par année, en moyenne combien Alexandre payait-il par année au cours des dernières années ?
- (f) Si Alex est à la classe 5 au départ, en moyenne combien de temps met-il avant de revenir à la classe 5 ?

SOLUTION. (a) La matrice des probabilités de transition en un pas est la suivante :

$$P = \begin{pmatrix} 50\% & 50\% & 0 & 0 & 0 \\ 50\% & 0 & 50\% & 0 & 0 \\ 15\% & 35\% & 0 & 50\% & 0 \\ 3\% & 12\% & 35\% & 0 & 50\% \\ 0,2\% & 2,8\% & 12\% & 35\% & 50\% \end{pmatrix}.$$

- (b) Le graphe des transitions possibles est représenté à la figure 1.9.

La chaîne ne comporte qu'une seule classe – donc elle est irréductible. Puisqu'elle est finie, on détermine donc immédiatement qu'elle est récurrente. Puisque $P_{1,1} > 0$, elle est apériodique.

- (c) En vertu du théorème 1.2, il y a une distribution stationnaire unique. Pour le trouver, on résout :

$$\left\{ \begin{array}{l} \pi_1 = \frac{500}{1000}\pi_1 + \frac{500}{1000}\pi_2 + \frac{150}{1000}\pi_3 + \frac{30}{1000}\pi_4 + \frac{2}{1000}\pi_5 \\ \pi_2 = \frac{500}{1000}\pi_1 + \frac{350}{1000}\pi_3 + \frac{120}{1000}\pi_4 + \frac{28}{1000}\pi_5 \\ \pi_3 = \frac{500}{1000}\pi_2 + \frac{350}{1000}\pi_4 + \frac{120}{1000}\pi_5 \\ \pi_4 = \frac{500}{1000}\pi_3 + \frac{350}{1000}\pi_5 \\ \pi_5 = \frac{500}{1000}\pi_4 + \frac{500}{1000}\pi_5 \\ 1 = \pi_1 + \pi_2 + \pi_3 + \pi_4 + \pi_5; \end{array} \right.$$

on trouve que la solution doit être :

$$\pi = \frac{1}{3537}(1057, 830, 650, 500, 500).$$

Ou, en approximations décimales,

$$\pi \approx (30\%, 24\%, 18\%, 14\%, 14\%).$$

- (d) On demande à long terme, quelle est la probabilité qu'Alexandre paye plus de 80% de la prime pour une année donnée. Il s'agit de la distribution-limite pour les gens qui sont dans les classes 1, 2 ou 3, soit

$$\pi_1 + \pi_2 + \pi_3 = \frac{2537}{3537} \approx 72\%.$$

- (e) À la limite où Alexandre souscrit depuis longtemps, on assume que la fonction de masse pour sa classe est donnée par la distribution stationnaire. Alors, on a que

$$\begin{aligned} \mathbb{P}\{N = 1000\$ \times 100\%\} &= \pi_1 \\ \mathbb{P}\{N = 1000\$ \times 90\%\} &= \pi_2 \\ \mathbb{P}\{N = 1000\$ \times 85\%\} &= \pi_3 \\ \mathbb{P}\{N = 1000\$ \times 80\%\} &= \pi_4 \\ \mathbb{P}\{N = 1000\$ \times 75\%\} &= \pi_5. \end{aligned}$$

On cherche $\mathbb{E}[N]$ – c'est :

$$\begin{aligned} \mathbb{E}[N] &= 1000\$ \times (100\%\pi_1 + 90\%\pi_2 + 85\%\pi_3 + 80\%\pi_4 + 75\%\pi_5) \\ &= 1000\$ \times \left(\frac{1057}{3537} + \frac{9}{10} \cdot \frac{830}{3537} + \frac{85}{100} \cdot \frac{650}{3537} + \frac{8}{10} \cdot \frac{500}{3537} + \frac{75}{100} \cdot \frac{500}{3537} \right) \\ &= 1000\$ \times \frac{6263}{7074} \\ &\approx 885\$. \end{aligned}$$

- (f) On cherche

$$\mathbb{E}_5[\tau_5^+] = \frac{1}{\pi_5} = \frac{3537}{500} \approx 7,1 \text{ ans.}$$

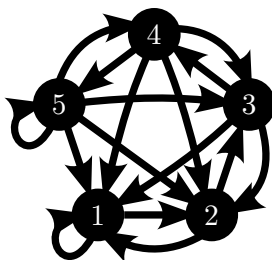


FIGURE 1.9 – Le graphe des transitions possibles pour la chaîne en question dans l'exercice 1.23.

Ceci complète notre étude de la théorie des chaînes de Markov homogènes à temps discret sur les espaces d'états discrets.

Dans le chapitre 2, nous allons nous intéresser de plus près à deux exemples de chaînes de Markov à temps discret sur des espaces d'états infinis. Nous verrons comment utiliser certains autres outils analytiques pour nous tirer d'affaire et résoudre des problèmes intéressants.

Au chapitre 3, nous allons nous intéresser aux chaînes de Markov à temps continu, d'abord en restant toujours sur les espaces d'états discrets. Au chapitre 4, nous nous pencherons de façon similaire sur différentes applications de cette théorie.

Au chapitre 5, nous nous intéresserons à la théorie des renouvellements, et à certains résultats utiles. Nous verrons même comment on peut arriver à les utiliser pour traiter les cas de certains processus qui n'ont pas la propriété de Markov !

Au chapitre 6, nous allons nous pencher sur une classe de processus stochastiques très intéressants : les *martingales*.

Finalement, au chapitre 7, nous ferons une brève incursion dans le monde des processus stochastiques à temps continus sur des espaces d'états continus.

Bref, il y a de quoi faire !

Et la fin du chapitre ? La section 1.8 est un supplément qui discute brièvement des chaînes de Markov d'un point de vue plus abstrait, en utilisant des notions très générales d'analyse fonctionnelle et d'algèbre linéaire.

La section 1.9 regroupe des exercices à faire pour se pratiquer à manier la matière. Lorsque ceux-ci sont tirés d'une source précise, cette source sera citée.

1.8. Bonus : Du point de vue de l'algèbre et de l'analyse.

Dans la section 1.4, nous avons introduit les probabilités de transition d'un état vers un autre, sur un intervalle de temps. Rapidement, nous avons organisé ces probabilités en matrices. Et, *a fortiori*, il a semblé que ce fût un judicieux éclat de génie – en effet, ces matrices, presque par coïncidence, semblent avoir des propriétés intéressantes. Étrange, non ?

Dans cette petite section, pour celles et ceux qui sont curieux/ses, je propose d'étaler un peu de vocabulaire et de théorie, d'algèbre et d'analyse, afin de donner peut-être un peu plus de profondeur aux notions présentées dans ce chapitre.

1.8.1. Distributions, mesures, vecteurs. Dans le contexte des probabilités, on appelle **distribution** d'une variable aléatoire X la **mesure** induite par la variable X et la mesure de probabilité \mathbb{P} . Plus précisément, si \mathbb{P} est une mesure de probabilités sur Ω , et

$X : \Omega \rightarrow S$ est une variable aléatoire à valeurs dans S , alors la mesure μ_X définie par

$$\mu_X(A) = \mathbb{P}\{X \in A\}$$

pour tout $A \subseteq S$ (ou au moins tout A dans une tribu de parties mesurables de S), correspond à la distribution de X .¹⁷

Comme mentionné en début de chapitre, donc : ce qui nous intéresse, c'est la distribution de notre variable aléatoire X_t – et plus particulièrement, on voudrait savoir comment cette distribution évoluera dans le temps.

Dans tout le chapitre 1, on s'attaque à cette question d'un point de vue calculatoire presque entièrement calculatoire. Dans ce qui suit, on va plutôt se pencher sur les structures abstraites. Plutôt que de détailler ce qu'est la distribution de X_t , on va se contenter de la manipuler comme un objet abstrait – comme un *vecteur*, en fait.

Les mesures signées. Pour faire ça, il faut prendre un petit pas de recul. En effet, si on se limite seulement aux distributions sur S , on n'a pas un espace vectoriel. Même si on se restreint seulement aux mesures positives, on n'a toujours pas un espace vectoriel. Pour ça, il faut considérer l'espace des mesures signées. Une **mesure signée** est une mesure (au sens de la définition B.4) mais qui ne respecte pas le premier axiome – c'est-à-dire qu'on peut assigner à une partie mesurable une mesure négative.

Si μ est une mesure signée sur S , le théorème de décomposition de Hahn¹⁸ garantit qu'il existe une partition de S en deux parties P et N , et deux mesures positives μ^+ et μ^- , supportées respectivement sur P et N , et telles que

$$\mu = \mu^+ - \mu^-.$$

On peut définir $|\mu|$, la valeur absolue de la mesure μ , par

$$|\mu| = \mu^+ + \mu^-.$$

Attention ! En général, $|\mu|(A) \neq |\mu(A)|$.

On dit qu'une mesure signée μ sur S est finie lorsque $|\mu|(S) < +\infty$.

On introduit $\mathcal{M}(S)$ l'ensemble des mesures signées finies sur S . On fait les définitions évidentes pour l'addition de deux mesures, et la multiplication d'une mesure par un scalaire : pour tout A mesurable dans S , et pour tout $k \in \mathbb{R}$,

$$(\mu_1 + \mu_2)(A) = \mu_1(A) + \mu_2(A); \quad (k\mu)(A) = k(\mu(A)).$$

Avec ces définitions, \mathcal{M} vérifie les axiomes d'un espace vectoriel !

Le cône positif. Pour retrouver nos distributions, il faut rajouter les axiomes des mesures de probabilités qui manquent – d'abord, la positivité. On notera $\mathcal{M}^+(S)$ l'ensemble des mesures positives sur S – c'est à dire des mesures de $\mathcal{M}(S)$ qui prennent seulement des valeurs non-négatives. Tous les éléments de $\mathcal{M}^+(S)$ sont des mesures au sens exact de la définition B.4. Nos distributions y sont, mais pas que ! Toutes les mesures positives finies y sont ! On appelle $\mathcal{M}^+(S)$ le **cône positif** de $\mathcal{M}(S)$.

La norme. Pour achever de repérer nos distributions dans l'espace $\mathcal{M}(S)$, on va définir une norme sur $\mathcal{M}(S)$. On notera :

$$\|\mu\| = |\mu|(S).$$

17. On consultera l'annexe B, ou les notes d'André Giroux sur la théorie de la mesure, pour davantage de détails sur la théorie de la mesure.

18. Consulter l'article Wikipédia en anglais.

Armés de cette norme, c'est évident : les distributions sont précisément *les éléments unitaires du cône positif* – c'est à dire les mesures positives dont la mesure totale est de 1.

1.8.2. Les opérateurs stochastiques. Bon, d'accord – on a bel et bien placé nos distributions – toutes les distributions possibles sur S – dans le contexte d'un espace vectoriel. Mais qu'est-ce que ça peut bien nous faire ? Eh bien, pour tout temps t , on peut dénoter par $\mu_t \in \mathcal{M}^+(S)$ la distribution de X_t . Ce qu'on cherche maintenant, bien entendu, c'est une façon de « faire évoluer » μ_t dans le temps – on voudrait bien savoir μ_{t+1} , par exemple, ou plus généralement, μ_{t+u} , pour un quelconque $u \geq 0$.

Cet objet, qu'on cherche, c'est une application $P^{(u)} : \mathcal{M}(S) \rightarrow \mathcal{M}(S)$, qui a pour argument une mesure, et qui nous redonne la mesure correspondante après que se soit écoulé un intervalle de temps de longueur u (on assume que le processus stochastique étudié est homogène dans le temps).

Immédiatement, on observe que de telles applications doivent respecter certains axiomes :

- i. L'opérateur $P^{(0)}$ est l'opérateur identité ;
- ii. Pour tout $u \geq 0$, l'opérateur $P^{(u)}$ est *linéaire* : ¹⁹

$$P^{(u)}(k\mu_1 + \mu_2) = kP^{(u)}\mu_1 + P^{(u)}\mu_2;$$

- iii. Pour tout $u \geq 0$ et toute mesure μ l'opérateur $P^{(u)}$ préserve la masse totale :

$$P^{(u)}\mu(S) = \mu(S),$$

et en particulier, pour les mesures positives, $P^{(u)}$ préserve la norme ;

- iv. Pour tous $u, v \geq 0$, on a l'équation de Chapman-Kolmogorov :

$$P^{(u+v)} = P^{(u)}P^{(v)}.$$

Les plus fûté-e-s d'entre vous aurez reconnus les propriétés détaillées pour les probabilités de transition à plusieurs reprises, plus haut.

La famille d'opérateurs $(P^{(u)})_{u \geq 0}$ constitue un *semi-groupe stochastique*. Algébriquement, elle a la structure d'un *semi-groupe* parce qu'elle est algébriquement fermée pour la composition, qu'elle admet un élément neutre, et que le produit y est associatif ; notons toutefois que les éléments n'y admettent en général pas d'inverse. L'appellation de semi-groupe *stochastique* renvoie à l'équation de Chapman-Kolmogorov spécifiquement, et à la condition que les opérateurs doivent préserver la norme.

1.8.3. L'espace des opérateurs. Si $\mathcal{M}(S)$ est un espace vectoriel, on peut introduire l'espace vectoriel $\mathcal{L}(\mathcal{M}(S))$:

$$\mathcal{L}(\mathcal{M}(S)) = \{T : \mathcal{M}(S) \rightarrow \mathcal{M}(S) : T \text{ est linéaire et continue.}\}$$

$\mathcal{L}(\mathcal{M}(S))$ est l'espace des transformations linéaires continues de notre espace des mesures signées sur S . Les éléments du semi-groupe stochastique – les $P^{(u)}$ – sont des éléments de $\mathcal{L}(\mathcal{M}(S))$.

On peut introduire une norme aussi sur $\mathcal{L}(\mathcal{M}(S))$: pour tout opérateur linéaire $T : \mathcal{M}(S) \rightarrow \mathcal{M}(S)$, la **norme** de T dans $\mathcal{L}(\mathcal{M}(S))$ est donnée par :

$$\|T\| := \sup \{\|T(\mu)\| : \mu \in \mathcal{M}(S), \|\mu\| \leq 1.\}$$

19. Intuitivement, ceci s'explique par un principe de « superposition » ; si on s'imagine que l'opérateur agit sur la somme de deux distributions, on veut que chacune de ces deux distributions soit propagée dans le temps indépendamment l'une de l'autre, un peu comme le sont les ondes ou la chaleur (ce n'est pas une coïncidence non plus !)

On peut même se simplifier la vie et remarquer que cela est équivalent à

$$\|T\| = \sup \{ \|T(\mu)\| : \mu \in \mathcal{M}(S), \|\mu\| = 1 \}.$$

La norme de $P^{(u)}$. On peut obtenir rapidement un résultat sur la norme de l'opérateur stochastique $P^{(u)}$; en effet, on a exigé que $P^{(u)}$ préserve la norme de toute mesure positive. Dès lors, pour toute mesure quelconque $\mu \in \mathcal{M}(S)$:

$$\begin{aligned} \|P^{(u)}\mu\| &= \|P^{(u)}(\mu^+ - \mu^-)\| \\ &\leq \|P^{(u)}\mu^+\| + \|P^{(u)}\mu^-\| \\ &= \|\mu^+\| + \|\mu^-\| \\ &= \|\mu\| ; \end{aligned}$$

Il suit donc immédiatement que $\|P^{(u)}\| \leq 1$. Or, puisqu'il existe au moins une distribution $\mu \in \mathcal{M}^+(S)$ avec $\|\mu\| = 1$ (donc telle que $\|P^{(u)}\mu\| = 1$), on doit aussi avoir que $\|P^{(u)}\| \geq 1$, et finalement, pour tout $u \geq 0$, l'opérateur stochastique $P^{(u)}$ a la propriété suivante :

$$\|P^{(u)}\| = 1.$$

Intuitivement, la norme d'une transformation linéaire indique « le facteur maximal par lequel la transformation peut allonger un vecteur. » En l'occurrence, ici, ce facteur est de 1 pour les opérateurs stochastiques.

La distance entre deux distributions. C'est bien beau tout ça, mais à quoi ça sert ? Pour le découvrir, on peut s'intéresser à la *distance* entre deux distributions μ_1 et μ_2 :

$$\|\mu_2 - \mu_1\|.$$

Cette distance donne une façon de déterminer si deux distributions sont semblables ou pas. On s'intéresse à ce qui arrive à la distance entre deux distributions lorsqu'on les fait « avancer dans le temps » en appliquant l'opérateur $P^{(u)}$:

$$\begin{aligned} \|P^{(u)}\mu_1 - P^{(u)}\mu_2\| &= \|P^{(u)}(\mu_1 - \mu_2)\| \\ &\leq \|\mu_1 - \mu_2\|. \end{aligned}$$

Ce que ce résultat nous indique, c'est que, en avançant dans le temps, la distance entre deux distributions **ne peut pas augmenter**.

opérateurs adjoints, valeurs propres et points fixes. En analyse fonctionnelle, l'**espace dual** de l'espace vectoriel $\mathcal{M}(S)$, noté $\mathcal{M}(S)'$, est l'espace vectoriel constitué de toutes les applications linéaires de $\mathcal{M}(S)$ dans le corps \mathbb{C} des nombres complexes.

Par exemple, l'application $f : \mathcal{M}(S) \rightarrow \mathbb{R}$ donnée par

$$f(\mu) = \mu(S)$$

pour tout $\mu \in \mathcal{M}(S)$

est une application linéaire – c'est donc un élément de l'espace dual de $\mathcal{M}(S)$.

Les éléments de l'espace dual sont appelés des *covecteurs*. Donc, ici, f est un covecteur.

Dans l'espace $\mathcal{L}(\mathcal{M}(S))$ des opérateurs linéaires continus sur $\mathcal{M}(S)$, on introduit pour tout opérateur T la notion **d'opérateur adjoint** noté T^* . L'opérateur adjoint T^* est un opérateur linéaire sur l'espace dual $\mathcal{M}(S)'$, qui est défini pour tout covecteur g par :

$$T^*(g) = g \circ T.$$

Autrement dit, pour tout $\mu \in \mathcal{M}(S)$, l'opérateur adjoint T^* doit satisfaire :

$$T^*g(\mu) = g(T\mu).$$

Nous allons nous servir des opérateurs adjoints pour montrer une propriété intéressante de l'opérateur stochastique $P^{(u)}$, en utilisant la propriété suivante : *pour des espaces vectoriels de dimension finie, les valeurs propres de l'opérateur adjoint T^* sont les conjugués complexes des valeurs propres de l'opérateur T .*

On se penche maintenant sur $(P^{(u)})^*$, l'opérateur adjoint de $P^{(u)}$. On remarque que

$$(P^{(u)})^*f = f \circ P^{(u)};$$

en particulier, pour tout $\mu \in \mathcal{M}(S)$, on a donc :

$$\begin{aligned} (P^{(u)})^*f(\mu) &= f(P^{(u)}\mu) \\ &= P^{(u)}\mu(S) \\ &= \mu(S) = f(\mu), \end{aligned}$$

et on conclue donc que

$$(P^{(u)})^*f = f.$$

Résultat : f est un *vecteur propre* de $(P^{(u)})^*$, donc 1 est une valeur propre de $(P^{(u)})^*$. Et lorsque (S) est un espace vectoriel de dimension finie (c'est-à-dire lorsque $|S| < +\infty$), on peut conclure immédiatement que 1 est aussi une valeur propre de $P^{(u)}$.

Conclusion : il existe donc une mesure signée $\pi \in (S)$ non-triviale telle que $P^{(u)}\pi = \pi$. C'est ce qu'on appelle un **point fixe** pour l'opérateur $P^{(u)}$. Les distributions propres associées à cette valeur propre sont les *distributions stationnaires* !

1.8.4. Les parallèles avec la théorie des sections précédentes. Ce qu'on vient de faire, de façon très abstraite, c'est de décrire très généralement la façon dont les opérateurs stochastiques doivent agir sur la distribution de notre processus aléatoire dans le temps. En autant qu'on respecte les principes que l'on vient d'énoncer, on peut, en principe, utiliser n'importe quel espace vectoriel pour représenter les mesures et les distributions, et tout le reste suivra.

Ce qu'on va faire dans les prochains paragraphes, c'est montrer comment la théorie qu'on a présenté dans le reste du chapitre 1 s'insère dans le cadre très abstrait de la présente section.

Les mesures sur S : les vecteurs-ligne. S est l'espace des états ; ici, nous allons assumer que $|S| < +\infty$. Pour nous simplifier la vie, on assumera que $S = \{1, 2, 3, \dots, n\}$.

Puisque S est fini, toutes ses parties sont mesurables, et en particulier, pour décrire une mesure sur S , il suffit de spécifier sa *fonction de masse*, qui indique, pour chaque élément individuel $i \in S$, la mesure du singleton $\{i\}$. Soit $m : S \rightarrow \mathbb{R}$ la fonction de masse de la mesure μ ; alors, pour tout $E \subseteq S$,

$$\mu(E) = \sum_{i \in E} m(i).$$

L'espace des fonctions de S dans \mathbb{R} serait un bon candidat pour représenter nos distributions, donc. Mais on peut faire un peu plus joli – en prenant les vecteurs-ligne !

On note $\mathbf{m} = (m(1), m(2), m(3), \dots, m(n)) \in \mathbb{R}^{1 \times n}$; il s'agit d'un vecteur-ligne qui représente la mesure μ dont la fonction de masse est m . On peut noter les éléments $m_i := m(i)$, par exemple.

Inversement, si $\mathbf{x} = (x_1, x_2, x_3, \dots, x_n) \in \mathbb{R}^{1 \times n}$ est n'importe quel vecteur-ligne, celui-ci correspond immédiatement à une fonction de masse, et donc à une mesure sur S .

Les opérateurs stochastiques : les matrices ! Notre opérateur stochastique, c'est simplement la matrice des probabilités de transition – mais qui agit par la *droite* sur le vecteur ligne ! En effet, si $p_{i,j}^{(u)}$ est la probabilité de passer de l'état i à l'état j dans un intervalle de temps t , et qu'on note $\mathbf{x}^{(t)}$ la distribution du processus au temps u , alors il suit que pour tout $j \in S$:

$$x_j^{(t+u)} = \sum_{i \in S} x_i^{(t)} p_{i,j}^{(u)}.$$

Les covecteurs et opérateurs adjoints : les vecteurs-colonnes ! Toutes les fonctionnelles linéaires de l'espace des vecteurs lignes $\mathbb{R}^{1 \times n}$ vers \mathbb{C} peuvent être représentés par des vecteurs colonnes (dans $\mathbb{C}^{n \times 1}$ dont on prend le produit par la droite avec nos vecteurs-ligne.

Par exemple, la fonctionnelle f , définie plus haut par $f(\mu) = \mu(S)$, est représentée par le vecteur colonne

$$\mathbf{1} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix};$$

en effet, pour tout vecteur-ligne $\mathbf{x} = (x_1, x_2, x_3, \dots, x_n)$, on a que

$$f(\mathbf{x}) = x_1 + x_2 + x_3 + \dots + x_n = \mathbf{x} \cdot \mathbf{1}.$$

L'opérateur adjoint à une matrice M est alors représenté par ... *la même matrice !*, mais cette fois, on la multiplie **par la gauche** au vecteur colonne.

Et effectivement, si $P^{(u)}$ est la matrice de transition sur un intervalle de temps de longueur u , on a bien que $P^{(u)}\mathbf{1} = \mathbf{1}$, ce qui correspond à la notion qu'on avait vu plus haut, que l'opérateur adjoint $(P^{(u)})^*$ appliqué sur f donne f .

D'autres covecteurs intéressants ? Les covecteurs sont des fonctionnelles linéaires qui, pour n'importe quelle mesure, rendent une valeur réelle associée à cette mesure.

Voici quelques exemples de telles fonctionnelles :

- Si E est un quelconque événement, alors on peut écrire $\mathbb{P}_i\{E\}$ la probabilité que l'événement E se réalise sachant que $X_0 = i$. Étant donné la distribution initiale $\mathbf{x}^{(0)}$ pour le processus stochastique (avec $x_i^{(0)} = \mathbb{P}\{X_0 = i\}$), on a alors que

$$\mathbb{P}\{E\} = \mathbf{x}^{(0)} \cdot \mathbf{u}_E,$$

où

$$\mathbf{u}_E = \begin{pmatrix} \mathbb{P}_1\{E\} \\ \mathbb{P}_2\{E\} \\ \mathbb{P}_3\{E\} \\ \vdots \\ \mathbb{P}_n\{E\} \end{pmatrix}.$$

Nous avons vu ce vecteur à la section 1.6, lorsque nous appliquions la méthode du conditionnement par le premier pas.

- Si Y est une variable aléatoire quelconque, alors on peut écrire $\mathbb{E}_i[Y]$ l'espérance de Y sachant que $X_0 = i$. Étant donné la distribution initiale $\mathbf{x}^{(0)}$ pour le processus stochastique, on a alors que

$$\mathbb{E}[Y] = \mathbf{x}^{(0)} \cdot \mathbf{v}_Y,$$

où

$$\mathbf{v}_Y = \begin{pmatrix} \mathbb{E}_1[Y] \\ \mathbb{E}_2[Y] \\ \mathbb{E}_3[Y] \\ \vdots \\ \mathbb{E}_n[Y] \end{pmatrix}.$$

Quelques commentaires sur les dimensions. Pour terminer cette section, nous allons faire quelques remarques amusantes sur les dimensions de certains espaces et sous-espaces intéressants.

Tout d'abord, il convient de remarquer que $\dim \mathcal{M}(S) = |S|$ – cela a été abordé dans les paragraphes précédents, mais ça vaut la peine de le souligner.

Ce qui intrigue plus spécialement, c'est plutôt l'espace propre associé à la valeur propre 1. En effet, lorsque $|S|$ est fini, du moins, on est assurés que l'opérateur stochastique $P^{(u)}$ a 1 pour valeur propre – et donc qu'il existe au moins une mesure qui soit un point fixe non-trivial pour $P^{(u)}$. Sans entrer dans les détails, on pourra assumer qu'ici plus particulièrement qu'il existe une distribution stationnaire – une distribution $\mu \in \mathcal{M}^+(S)$ avec $\mu(S) = 1$ et telle que $P^{(u)}\mu = \mu$.

On peut donc noter V_1 le sous-espace de $\mathcal{M}(S)$ engendré par les vecteurs propres associés à la valeur propre 1 – quelle est la dimension de cet espace ?

Pour répondre à cette question, on peut retourner plus haut dans le chapitre, à la section 1.7.2 ; on y montre que lorsque la chaîne de Markov est irréductible, la distribution stationnaire est unique – ça correspond, on le sait maintenant, à dire que $\dim V_1 = 1$.

Mais que se passe-t-il lorsque la chaîne de Markov n'est PAS irréductible ? Supposons qu'il y ait $C_1, C_2 \subseteq S$, deux classes d'équivalence (disjointes) fermées (donc récurrentes).

En nous restreignant à seulement les états de C_1 , on a une chaîne irréductible – donc il existe une unique distribution stationnaire s_1 dans $\mathcal{M}(C_1)$, et on remarque que, nécessairement, la distribution π_1 dans $\mathcal{M}(S)$ qui est donnée par $\pi_1(E) = s_1(E \cap C_1)$ pour tout $E \subseteq S$ est une distribution stationnaire. Mais en faisant le même raisonnement pour C_2 , on trouve aussi une autre distribution stationnaire π_2 . Les deux sont linéairement indépendantes, puisque C_1 et C_2 sont deux classes d'équivalence disjointes.

Résultat : on peut trouver exactement autant de distributions stationnaires linéairement indépendantes les unes des autres qu'il y a de classes d'équivalence fermées. Autrement dit, lorsque $|S| < +\infty$, on a que

$$\dim V_1 = k,$$

où k est le nombre de classes d'équivalence fermées de S !

1.9. Exercices

EXERCICE 1.1. Connaissez-vous des phénomènes de la vie courante (au sens large) que vous pourriez modéliser par des processus stochastiques ? Quel serait l'espace des états ? Ces processus seraient-ils à temps discret ou continu ?

EXERCICE 1.2. Pour les processus suivants, dites si oui ou non ils ont la propriété de Markov ; si oui, sont-ils homogènes ? Si c'est bien le cas, donner une matrice des probabilités de transition en un pas.

- (a) On lance un dé à six faces façon répétée. Soit S_n le total des points observés après n lancers ; on définit alors $X_n \equiv S_n \pmod{6}$ le reste de la division de S_n par 6.
- (b) On pige un dé d'une boîte qui contient des dés à 6, 8, 10, 12 et 20 faces, en proportions égales. Puis, sans regarder ce qu'on a pigé, on le lance de façon répétée, et un ami nous annonce le résultat X_n du n ème lancer.
- (c) À chaque fois qu'il neige, le nombre de jours avant la prochaine averse de neige suit une loi de Poisson de paramètre $\lambda > 0$. On note X_t le nombre de jours où il a neigé t jours après la toute première averse de neige.
- (d) À chaque fois qu'elle va courir, Clémence note le temps qu'elle met pour parcourir les 5,3 kilomètres de sa boucle habituelle. En particulier, elle prend note du compteur des secondes sur son chronomètre – qui affiche un nombre entre 00 et 59. À chaque fois que le nombre qu'elle observe apparaît pour la première fois, elle le raye d'une liste. X_t est le nombre de nombres (entre 0 et 59 inclusivement) que Clémence a rayé de sa liste.

EXERCICE 1.3 (LESSARD, ex. 1.1). Supposons que le temps qu'il fait d'une journée à la suivante est décrit par une chaîne de Markov sur les états 1, 2, 3 (1 pour *ensoleillé*, 2 pour *nuageux*, 3 pour *pluvieux*), dont la matrice de transition en un pas est donnée par

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

On est jeudi et c'est nuageux. Déterminer :

- (a) la probabilité que les trois prochains jours soient ensoleillés ;
- (b) la probabilité que dimanche prochain soit ensoleillé.

EXERCICE 1.4 (LESSARD, ex. 1.2). En utilisant les propriétés des matrices de transition pour les chaînes de Markov homogènes à temps discret, montrer que si $P^{(t)}$ et P sont les matrices de probabilités de transition en t unités de temps et en un pas respectivement, pour une chaîne de Markov à deux états $\{1, 2\}$, on a toujours

$$P_{1,1}^{(2)} \geq P_{2,1}^{(2)}.$$

EXERCICE 1.5 (LESSARD, ex. 1.3). Une chaîne de production comprend deux machines-outils qui fonctionnent indépendamment l'une de l'autre. Chaque machine-outil a une fiabilité de 90% au cours d'une journée, ce qui signifie que sa probabilité de tomber en panne pendant cette période est de 10%. Il faut une nuit pour réparer une machine-outil qui tombe en panne, mais une seule à la fois peut être réparée.

- (a) Quelle est la matrice de transition pour le nombre de machines-outils en état de fonctionnement au début d'une journée (les indices pourront commencer en 0).
- (b) S'il faut deux nuits pour réparer une machine-outil, quels états doit-on considérer pour que notre chaîne ait la propriété de Markov ? Quelle est la matrice de transition ?

EXERCICE 1.6. Soit P la matrice des probabilités de transition en un pas pour une chaîne de Markov à temps discret homogène sur un espace d'états $S = \{1, 2, 3, \dots, n\}$.

- (a) Montrer que $\lambda = 1$ est une valeur propre de P .
- (b) Montrer que le vecteur-colonne $\mathbf{1}_n = (1, 1, 1, \dots, 1)$ est un vecteur propre de P associé à la valeur propre $\lambda = 1$.
- (c) Montrer que $\dim \ker(P - \mathbb{I}_n) < n$.

EXERCICE 1.7 (LESSARD, ex. 1.4). On considère un pentagone régulier dont les sommets sont numérotés de 1 à 5 dans le sens horaire. Initialement, deux coccinelles sont placées aux sommets 1 et 3. À chaque instant suivant, chacune des coccinelles se déplace indépendamment de l'autre, vers l'un des deux sommets adjacents, chacun ayant probabilité $1/2$.

Combien de temps faudra-t-il en moyenne pour que les deux coccinelles se rencontrent au même sommet ?

Indice : Considérer la distance en nombre d'arêtes entre les deux coccinelles...

EXERCICE 1.8 (LESSARD, ex. 1.5). Anastasia et Barrah jouent au tennis de table ; toutes les deux sont excellentes et elles sont rendues à 21 points chacune – à partir de maintenant, l'une d'entre elles doit avoir au moins deux points d'avance pour remporter la victoire. Une joueuse avec un point d'avance est dite avoir « l'avantage ».

En supposant qu'Anastasia remporte chaque point indépendamment avec probabilité p , et Barrah avec probabilité $q = 1 - p$,

- (a) déterminer la probabilité qu'Anastasia remporte la partie, et
- (b) donner l'espérance du nombre de fois où Anastasia sera en avantage avant la fin du jeu.

EXERCICE 1.9 (LESSARD, ex. 1.6). On lance une pièce de monnaie équilibrée plusieurs fois indépendamment, jusqu'à obtenir trois « Face » de suite. Les résultats des trois premiers jets sont « Face, Pile, Face ».

En incluant ces trois premiers jets, déterminer la probabilité d'obtenir trois « Pile » consécutifs avant la fin du jeu.

EXERCICE 1.10 (LESSARD, ex. 1.7). Une espèce de fleurs peut se trouver dans trois états :

- (1) viable ;
- (2) en voie de disparition ;
- (3) éteinte.

Les probabilités des transitions d'une année à l'autre sont données par les matrices suivantes :

$$\begin{aligned}
 P(0,1) &= \begin{pmatrix} 85\% & 15\% & 0 \\ 0 & 70\% & 30\% \\ 0 & 0 & 100\% \end{pmatrix} & P(1,2) &= \begin{pmatrix} 90\% & 10\% & 0 \\ 10\% & 70\% & 20\% \\ 0 & 0 & 100\% \end{pmatrix} \\
 P(2,3) &= \begin{pmatrix} 95\% & 5\% & 0 \\ 20\% & 70\% & 10\% \\ 0 & 0 & 100\% \end{pmatrix} & P(t-1,t) &= \begin{pmatrix} 95\% & 5\% & 0 \\ 50\% & 50\% & 0 \\ 0 & 0 & 100\% \end{pmatrix} \quad t > 3.
 \end{aligned}$$

Calculer la probabilité que l'espèce de fleurs s'éteigne ultimement, étant donné qu'elle est initialement en voie de disparition.

EXERCICE 1.11 (LESSARD, ex. 1.14). Tracer les graphes des transitions possibles, identifier les classes d'équivalence et donner les types d'états pour les chaînes de Markov dont les matrices de transition en un pas sont les suivantes :

$$\begin{aligned}
 \text{(a)} \quad & \begin{pmatrix} 0 & 1/3 & 1/3 & 0 & 1/3 \\ 1/2 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1/2 & 1/2 & 0 \end{pmatrix} & \text{(b)} \quad & \begin{pmatrix} 0 & 1/4 & 0 & 0 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 1/3 & 0 & 0 & 1/3 \\ 1/2 & 0 & 0 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 0 & 1/2 & 1/2 \end{pmatrix}
 \end{aligned}$$

EXERCICE 1.12. Pour chacune des matrices de probabilités de transition en un pas suivantes, dessiner les graphes des transitions possibles, identifier les classes d'équivalence et dire si elles sont ouvertes ou fermées.

(a)

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

(b)

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

(c)

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

EXERCICE 1.13. Cinq boules blanches et cinq boules noires sont distribuées dans deux urnes A et B ; chacune contient cinq boules (pas forcément de la même couleur). On va procéder à des échanges successifs entre deux boules choisies au hasard dans les urnes. On note X_t le nombre de boules blanches dans l'urne A.

- (a) Expliquer pourquoi X_t est une chaîne de Markov homogène.
- (b) Quel est l'espace des états possibles pour $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$.
- (c) Donner les probabilités de transition.
- (d) Dessiner le graphe des transitions possibles. Combien y a-t-il de classes d'équivalence ?

EXERCICE 1.14. Soit

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

la matrice des transitions en un pas d'une chaîne de Markov homogène à temps discret.

- (a) Trouver les classes d'équivalence d'états. Dire si elles sont ouvertes ou fermées.
- (b) Trouver $P^{(t)}$, la matrice de transition en t pas.
- (c) Quelle est la limite lorsque t tend vers l'infini ?
- (d) Est-ce qu'on peut dire qu'à long terme, le processus est indépendant de sa condition initiale ?

EXERCICE 1.15. On lance un dé à six faces de façon répétée, et on note X_t le *minimum* des résultats observés après t lancers (pour $t \geq 1$).

- (a) Donner la matrice des probabilités de transition en un pas P pour cette chaîne de Markov.
- (b) Combien y a-t-il de classes d'équivalence ? Lesquelles sont ouvertes ? Fermées ?
- (c) Donner $\lim_{t \rightarrow \infty} P^{(t)}$.

EXERCICE 1.16. Soit $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ une chaîne de Markov homogène à temps discret sur l'espace des états $S = \{1, 2, 3, \dots, n\}$, avec matrice de probabilités de transition en un pas P .

On suppose que P est une matrice triangulaire supérieure.

- (a) Montrer que chaque état est seul dans sa classe d'équivalence.
Indice : deux états $i, j \in S$ communiquent si et seulement si il existe un cycle qui passe par i et par j .
- (b) Montrer que $P_{n,n} = 1$.
- (c) On suppose que $P_{i,i} < 1$ pour tout $i \neq n$. Expliquer pourquoi, dans ce cas, on a que $\lim_{t \rightarrow \infty} P_{i,n}^{(t)} = 1$.

EXERCICE 1.17. Soit $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ une chaîne de Markov sur l'espace des états $S = \mathbb{Z}^+ = \{0, 1, 2, \dots\}$, avec

$$P_{0,1} = 1; P_{k,0} = \frac{1}{k+1}, P_{k,k+1} = \frac{k}{k+1} \quad \forall k \geq 1.$$

- (a) Esquisser le graphe des transitions possibles pour cette chaîne de Markov.
- (b) Combien y a-t-il de classes d'équivalence? Pour chacune, dire si elles sont fermées ou ouvertes.
- (c) Calculer $P_{0,t}^{(t)}$ pour $t \geq 1$.
- (d) En partant de 0, quelle est la probabilité qu'on revienne en 0 pour la première fois au temps t ?
- (e) Quelle est la probabilité de ne jamais revenir à l'état 0 sachant qu'on part de 0?
Indice : prendre la limite de la partie (d) lorsque t tend vers l'infini.

EXERCICE 1.18. Pour chaque état de chacune des matrices à l'exercice 1.12,

- (a) déterminer s'il est récurrent ou transient;
- (b) déterminer sa période.

EXERCICE 1.19. Pour le processus décrit à l'exercice 1.13,

- (a) dire, pour chaque état, s'il est récurrent ou transient.
- (b) En supposant qu'on commence avec toutes les boules blanches dans l'urne A et toutes les noires dans l'urne B , en moyenne combien d'échanges faut-il faire pour que toutes les boules blanches aient toutes été remplacées par des boules noires et vice-versa?

EXERCICE 1.20. Pour le processus avec la matrice comme à l'exercice 1.14, calculer $\mathbb{E}_1[\tau_2]$ et $\mathbb{E}_3[\tau_2]$.

EXERCICE 1.21. Pour le processus de l'exercice 1.17,

- (a) Calculer $\mathbb{P}_0\{\tau_0^+ = t\}$.
- (b) L'état 0 est-il transient ou récurrent? Qu'en est-il de l'état n , pour $n \in \mathbb{N}$?

EXERCICE 1.22. Soit $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ une chaîne de Markov homogène à temps discret sur l'espace des états $S = \mathbb{Z}$, avec les probabilités de transition en un pas

$$P_{i,i+1} = p > q = 1 - p = P_{i,i-1}.$$

- (a) On définit la famille d'événements $(E_n = \{\tau_n < \tau_{-1}\})_{n \in \mathbb{N}}$. Calculer $\mathbb{P}_0 \{E_n\}$.
- (b) Montrer que si $X_0 = 0$, alors $E_n \supseteq E_{n+1}$ pour tout n ; déduire que

$$\lim_{n \rightarrow \infty} \mathbb{P}_0 \{E_n\} = \mathbb{P}_0 \left\{ \bigcap_{k \in \mathbb{N}} E_k \right\}.$$

- (c) Expliquer pourquoi $\bigcap_{k \in \mathbb{N}} E_k = \{\tau_{-1} = +\infty\}$. Calculer $\mathbb{P}_0 \{\tau_{-1} = +\infty\}$.
- (d) La chaîne est-elle récurrente ou transiente?

EXERCICE 1.23 (LESSARD, ex. 1.34). Kamila et Mahily jouent à roche-papier-ciseaux jusqu'à ce que l'une des deux mène par deux parties sur son adversaire. Pour rappel : Le papier enveloppe la roche, la roche casse les ciseaux et les ciseaux coupent le papier ; le choix de deux stratégies identiques résulte en une partie nulle.

Soit X_t la différence entre le nombre de victoires de Kamila et le nombre de victoires de Mahily.

- (a) Quels sont les états pour cette chaîne ? Donner la matrice de transition en un pas.
- (b) Donner l'espérance du nombre de parties avant la fin du jeu.

EXERCICE 1.24 (LESSARD, ex 1.16). On considère la chaîne de Markov sur \mathbb{Z} , avec les probabilités de transition en un pas suivantes pour tout $i \in \mathbb{Z}$:

$$P_{i,i-1} = 1 - p, \quad P_{i,i+1} = p$$

- (a) Déterminer $P_{0,0}^{(t)}$.
- (b) Déterminer pour quelles valeurs de p les états sont récurrents, et pour quelles valeurs de p ils sont transients.

EXERCICE 1.25. Sur une autoroute, il y a des camions et des autos. Le $3/4$ des autos sont suivies par une autre auto, et $1/5$ des camions sont suivis par un autre camion. On note X_t le type du t ème véhicule sur la route – 1 pour une auto, et 2 pour un camion.

- (a) Donner la matrice des probabilités de transition en un pas pour ce processus.
- (b) En moyenne sur un long tronçon de cette route, quel pourcentage des véhicules sont des camions ?

EXERCICE 1.26 (LESSARD, ex. 1.18). Supposons que, d'une génération à la suivante, les familles changent de groupe de revenu (bas, moyen, élevé) selon une chaîne de Markov dont la matrice des probabilités de transition en un pas est donnée par

$$P = \begin{pmatrix} 80\% & 15\% & 5\% \\ 30\% & 60\% & 10\% \\ 0\% & 20\% & 80\% \end{pmatrix}.$$

Quel sera la répartition à long terme des familles dans ces trois classes ?

EXERCICE 1.27 (LESSARD, ex. 1.19). Les résultats successifs de parties d'échecs d'un joueur contre un logiciel d'échecs suivent une chaîne de Markov sur les états $S = \{1, 2, 3\}$, où 1 correspond à une victoire, 2 correspond à une défaite, et 3 correspond à une partie nulle. La matrice des probabilités de transition en un pas est donnée par :

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

- Déterminer la proportion moyenne p de victoires à long terme pour ce joueur.
- Donner l'espérance du nombre de parties entre deux victoires.
- En supposant plutôt que le joueur remporte chaque partie de façon indépendante avec probabilité p , déterminer maintenant l'espérance du nombre de parties entre deux victoires. Comparer avec le résultat précédent, et commenter.
- Chaque victoire compte pour trois points, et chaque partie nulle compte pour un point. Les défaites ne comptent pour aucun point. À long terme, en moyenne combien de points le joueur accumule-t-il par partie ?

EXERCICE 1.28 (LESSARD, ex. 1.20). Soit S_n la somme du nombre de points observés après n lancers d'un dé équilibré à six faces. Est-ce que $\mathbb{P}\{S_n \equiv 0 \pmod{6}\}$ converge lorsque n tend vers l'infini ?

EXERCICE 1.29 (LESSARD, ex. 1.24). Une chaîne de Markov sur un nombre fini d'états avec matrice de probabilités de transition en un pas P est dite **régulière** s'il existe un certain N tel que la matrice des probabilités de transition en N pas P^N est strictement positive (c'est-à-dire que toutes les entrées sont strictement positives).

- (a) Donner un exemple d'une chaîne de Markov irréductible qui n'est *pas* régulière.
- (b) Montrer que la chaîne de Markov sur $S = \{1, 2, 3, 4, 5\}$ avec la matrice des probabilités de transition en un pas

$$P = \begin{pmatrix} 3/4 & 1/4 & 0 & 0 & 0 \\ 3/4 & 0 & 1/4 & 0 & 0 \\ 3/4 & 0 & 0 & 1/4 & 0 \\ 3/4 & 0 & 0 & 0 & 1/4 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

est régulière.

- (c) A-t-elle une distribution stationnaire unique? Si oui, quelle est-elle?
- (d) Donner une condition suffisante pour qu'une chaîne de Markov soit régulière.

EXERCICE 1.30 (LESSARD, ex. 1.25). Une sauterelle se déplace sur les sites 1, 2 et 3 disposés sur un cercle en allant à chaque saut au site adjacent (dans le sens horaire) avec probabilité p , et dans le sens anti-horaire avec probabilité $1 - p$.

- (a) Déterminer les valeurs de p pour lesquelles $P_{i,j}^{(t)}$ converge lorsque t tend vers l'infini, pour tous i, j . Dans ces cas, trouver les limites.
- (b) Trouver les valeurs de p pour lesquelles $\frac{1}{t} \sum_{u=1}^t P_{i,j}^{(u)}$ converge lorsque t tend vers l'infini. Trouver les limites.
- (c) Répondre aux mêmes questions si on a plutôt N sites, pour $N \in \mathbb{N}$.

EXERCICE 1.31 (LESSARD, ex. 1.29). Il vous en coûte 2\$ pour jouer dans une machine à sous, qui vous redonne 3\$ lorsque vous gagnez la partie. Pour vous encourager à y jouer, le casino annonce en grosses lettres qu'il y a toujours au moins 50% de chances de gagner. En effet, le fonctionnement de la machine est de telle sorte que, pour chaque partie, la probabilité de gagner est égale à $(k+1)/(k+2)$, où k est le nombre de défaites aux deux parties précédentes.

On souhaite déterminer si cette machine à sous est profitable pour le casino en utilisant une chaîne de Markov.

- (a) Quelle chaîne de Markov peut-on utiliser pour résoudre ce problème? Donner les états et la matrice des probabilités de transition en un pas.
- (b) Trouver la distribution stationnaire pour cette chaîne de Markov.
- (c) En vous servant de la distribution stationnaire, déterminer si l'espérance de l'argent gagné par le casino en moyenne en une partie est positive ou négative.

Chaînes de Markov à temps discret : études.

Au chapitre 1, nous avons introduit les notions de base de la théorie des chaînes de Markov, et nous avons développé des outils analytiques pratiques pour l'étude de chaînes homogènes à temps discret sur des espaces d'états discrets; notamment, nous avons fourni une caractérisation assez complète du comportement de chaînes sur un nombre fini d'états.

Les chaînes sur des nombres infinis d'états peuvent être plus complexes à étudier, et même si les fondements théoriques sont les mêmes, l'étude de chaînes particulières requièrera l'emploi de techniques et de méthodes particulières adaptées aux sujets à l'étude.

Dans ce chapitre, nous allons nous intéresser à deux types de chaînes de Markov spécifique; pour chacune d'entre elles, nous explorerons les principaux résultats qu'il est possible d'obtenir, et explorerons les méthodes employées pour y parvenir.

2.1. Le processus de Galton-Watson.

Le *processus de Galton-Watson* est un processus stochastique aussi appelé *un processus branchant*. Il a été introduit à l'origine vers la fin du XIXe siècle pour étudier les arbres de descendance, et plus particulièrement la question de déterminer à quelles conditions les noms de famille (passés à l'époque de père en fils) sont assurés de perdurer dans le temps.

De façon plus générale, on peut utiliser un processus de Galton-Watson pour modéliser l'évolution d'une *population* qui se reproduit « par clonage », indépendamment à chaque génération. La façon typique de représenter l'évolution d'un processus de Galton-Watson est de tracer un *arbre*, soit un graphe où chaque sommet représente un individu, relié par une arête à son *parent* (figure 2.1).

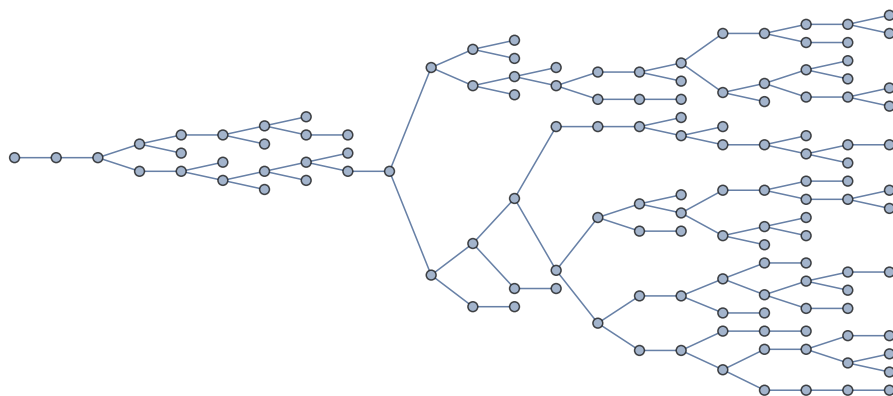


FIGURE 2.1 – Exemple des 20 premières générations d'un arbre de Galton-Watson.

2.1.1. Définition du modèle. Nous allons faire les hypothèses suivantes :

- i. Chaque individu ne produit d'enfants qu'une seule fois.
- ii. Chaque individu produit tous ses enfants en même temps. Ses enfants font partie de la génération suivante.
- iii. Les nombres d'enfants produits par chaque individus respectivement sont des variables aléatoires entières non-négatives, indépendantes et identiquement distribuées ; on notera $p : \mathbb{Z}^+ \rightarrow [0, 1]$ la fonction de masse associée à cette distribution (c'est à dire que $p(k)$ est la probabilité pour un individu d'avoir k descendants directs).

La quantité principale qui nous intéresse, ce sera le *nombre d'individus nés à la t -ième génération*, que l'on notera simplement Z_t . Pour faire simple, on notera également Z_0 le nombre d'individus présents au départ – on fixe donc $Z_0 = 1$.

On notera $\xi_{t,n}$ le nombre d'enfants produits par le n -ième individu né à la t -ième génération. On a immédiatement que

$$Z_1 = \xi_{0,1}.$$

Par nos hypothèses, on sait que les $\xi_{t,i}$ sont indépendants et identiquement distribués, et leur fonction de masse est donnée par p . De plus, pour n'importe que t fixé, on a :

$$(2.1.1) \quad Z_t = \sum_{n=1}^{Z_{t-1}} \xi_{t-1,n}.$$

Il suit que Z_t ne dépend que de $(\xi_{u,i})_{u < t, i \in \mathbb{N}}$; en particulier, pour tout t , Z_t est indépendant de $\xi_{t,n}$ pour tout $n \in \mathbb{N}$.

LEMME 2.1. *Le processus $\mathbf{Z} = (Z_t)_{t \in \mathbb{Z}^+}$ est une chaîne de Markov homogène à temps discret sur l'espace des états $S = \mathbb{Z}^+$.*

DÉMONSTRATION. En effet, sachant que $Z_t = k$, la loi de Z_{t+1} est tout simplement la loi d'une somme de k variables aléatoires i.i.d. avec fonction de masse p ; cette loi ne dépend pas de Z_u pour $u < t$ – donc c'est une chaîne de Markov. Mais cette loi ne dépend pas non plus de t ; donc le processus est homogène. \square

La loi du processus \mathbf{Z} est entièrement caractérisée par la fonction de masse $p : \mathbb{Z}^+ \rightarrow [0, 1]$ des $\xi_{t,n}$:

$$p(k) = \mathbb{P} \{ \xi_{t,n} = k \},$$

qui constitue en quelque sorte un ensemble de *paramètres* pour notre modèle. On introduit également *l'espérance du nombre d'enfants par individu* :

$$(2.1.2) \quad \mu = \mathbb{E} [Z_1] = \sum_{k=0}^{\infty} kp(k)$$

EXEMPLE 2.1. Supposons que $\mathbf{Z} = (Z_t)_{t \in \mathbb{Z}^+}$ soit un processus de Galton-Watson où le nombre d'enfants de chaque individu est de loi géométrique avec paramètre q : pour tout $k \geq 1$,

$$p(k) = (1 - q)^{k-1}q.$$

Soit $P_{i,j}$ la probabilité de transition en un pas de i à j . Donner $P_{i,j}$.

SOLUTION. On cherche $\mathbb{P}\{Z_{t+1} = j \mid Z_t = i\}$. Comme mentionné plus haut, sachant que $Z_t = i$, la loi de Z_{t+1} est simplement la loi d'une somme de i variables aléatoires géométriques indépendantes de paramètre q ; on sait aussi que cette loi est une loi binomiale négative de paramètres (i, q) ; donc, pour $j \geq i$:

$$P_{i,j} = \binom{j-1}{i-1} (1-q)^{j-i} q^i.$$

EXEMPLE 2.2. Soit $\mathbf{Z} = (Z_t)_{t \in \mathbb{Z}^+}$ un processus de Galton-Watson.

- (a) Donner $P_{0,0}$ et expliquer pourquoi l'état 0 est toujours seul dans sa classe d'équivalence, récurrent et apériodique.
- (b) Expliquer pourquoi $P_{k,0}^{(t)} = (P_{1,0}^{(t)})^k$.

SOLUTION. (a) L'état 0 est l'état où il n'y a aucun individu né à la génération courante. La population ne peut plus se reproduire; il n'y aura aucun individu né dans la génération subséquente, et on doit donc avoir $P_{0,0} = 1$. L'état est donc récurrent, apériodique, et seul dans sa classe d'équivalence puisqu'il s'agit d'un état absorbant.

- (b) On a que $P_{k,0}^{(t)}$ est la probabilité qu'en t générations, on passe de k individus à l'extinction complète du processus.

Si on se penche sur la descendance de chacun de ces individus, on voit que le processus est éteint si et seulement si la descendance respective de chacun des k individus est éteinte au temps t . Comme chacune de ces branches sont indépendantes, et identiquement distribuées, et que la probabilité qu'elles soient éteintes au temps t est $P_{1,0}^{(t)}$, on a donc :

$$P_{k,0}^{(t)} = (P_{1,0}^{(t)})^k.$$

Dans l'exemple 2.2, ce que nous venons de voir, c'est que l'état 0 joue un rôle un peu particulier dans un processus de Galton-Watson : il signifie la fin de la reproduction pour les individus de la population. Si on assume que chaque individu meure après s'être reproduit, l'état 0 représente la fin du processus, l'extinction de la population.

Questions intéressantes. De façon générale, les quantités qui vont nous intéresser sont :

- $P_{1,k}^{(t)} = \mathbb{P}\{Z_t = k \mid Z_0 = 1\}$, la distribution du nombre d'individus de la t -ième génération, et en particulier,
- $P_{1,0}^{(t)}$, la probabilité qu'il n'y ait aucun individu né à la t -ième génération – c'est-à-dire que la population s'est éteinte à un moment avant la t -ième génération inclusivement.

On note $\tau_0 = \inf \{t \in \mathbb{Z}^+ : Z_t = 0\}$ le temps d'atteinte de l'état 0. La question primordiale est de connaître la probabilité de toucher 0 éventuellement; on recherche

$$\mathbb{P}_1 \{ \tau_0 < +\infty \},$$

ce que l'on appellera la **probabilité d'extinction**.

On peut déjà entrevoir au moins un résultat amusant :

LEMME 2.2. Soit $\mathbf{Z} = (Z_t)_{t \in \mathbb{Z}^+}$ un processus de Galton-Watson où le nombre d'enfants pour chaque individu suit une distribution dont la fonction de masse est p .

Alors :

- i. Si $p(0) = 0$, alors $\mathbb{P}_1 \{\tau_0 < +\infty\} = 0$.
- ii. Si $p(0) > 0$, alors on a $\mathbb{P}_1 \{\tau_0 < +\infty\} > 0$; de plus,

$$\mathbb{E}_1 [\tau_0] \geq \frac{1}{p(0)}.$$

DÉMONSTRATION. i Si $p(0) = 0$, la probabilité de n'avoir aucun enfant est nulle, et forcément, on a $Z_t \geq 1$ pour tout t .

ii Si $p(0) > 0$, alors clairement, $\mathbb{P}_1 \{\tau_0 < +\infty\} \geq \mathbb{P}_1 \{\tau_0 = 1\} = p(0) > 0$.

On a que

$$\mathbb{P}_1 \{\tau_0 > t\} = \mathbb{P}_1 \{Z_t > 0\}.$$

On va noter $q = 1 - p(0)$ la probabilité d'avoir au moins un enfant. Alors, pour tout t , on a que :

$$\mathbb{P} \{Z_{t+1} > 0 \mid Z_t = 1\} = q.$$

De plus, on a montré à l'exemple 2.2 que pour tout $k \geq 1$,

$$P_{k,0} = (P_{1,0})^k;$$

cela signifie que

$$\begin{aligned} \mathbb{P}_1 \{Z_{t+1} > 0 \mid Z_t = k\} &= 1 - P_{k,0} \\ &= 1 - p(0)^k \\ &\geq 1 - p(0) = q, \end{aligned}$$

et il suit que

$$\begin{aligned} \mathbb{P}_1 \{Z_t > 0\} &= \sum_{k=1}^{\infty} \mathbb{P}_1 \{Z_t > 0 \mid Z_{t-1} = k\} \mathbb{P}_1 \{Z_{t-1} = k\} \\ &\geq q \sum_{k=1}^{\infty} \mathbb{P}_1 \{Z_{t-1} = k\} \\ &= q \mathbb{P}_1 \{Z_{t-1} > 0\}. \end{aligned}$$

Par récurrence, on obtient donc que

$$\mathbb{P}_1 \{\tau_0 > t\} = \mathbb{P}_1 \{Z_t > 0\} \geq q^t \mathbb{P}_1 \{Z_0 > 0\} = q^t.$$

Il suit que

$$\begin{aligned} \mathbb{E}_1 [\tau_0] &= \sum_{t=0}^{\infty} \mathbb{P}_1 \{\tau_0 > t\} \\ &\geq \sum_{t=0}^{\infty} q^t \\ &= \frac{1}{1-q} = \frac{1}{p(0)}. \end{aligned}$$

□

2.1.2. Fonction génératrice, loi de la t -ième génération et probabilité d'extinction. On a déjà identifié dans le lemme 2.2 que, dans le cas où la probabilité de n'avoir aucun enfant est nulle, la probabilité d'extinction est aussi nulle – comme le disent les mathématicien-ne-s dans leur jargon technique, *œngu'à'ouerre-on'oué-ben*.

Mais qu'en est-il de façon plus générale ? Peut-on calculer cette probabilité ? Peut-on obtenir la loi de la t -ième génération facilement ?

La réponse à ces questions se trouve dans un outil très puissant : les *fonctions génératrices des probabilités*.

On va définir :

$$(2.1.3) \quad \psi(s) = \sum_{k=0}^{\infty} p(k)s^k = \mathbb{E}[s^{Z_1}] \text{ / ,}$$

où $p(k)$ est la probabilité d'avoir k enfants, et s est une variable libre. ψ est la *fonction génératrice des probabilités de la distribution du nombre d'enfants*.

Rappels sur les fonctions génératrices des probabilités. Nous allons d'abord procéder à quelques rappels concernant les fonctions génératrices des probabilités.

LEMME 2.3. *La série $\sum_{k \geq 0} p(k)s^k$ converge uniformément pour $s \in [-1, 1]$.*

DÉMONSTRATION. Puisque les $p(k)$ sont des probabilités, on a que $|p(k)| \leq 1$ pour tout k , et il suit par comparaison que la série

$$\sum_{k=0}^{\infty} p(k)s^k$$

converge pour tout $s \in (-1, 1)$. Si $s = -1$, on a que

$$\psi(-1) = \sum_{k=0}^{\infty} (-1)^k p(k).$$

Or, cette série est absolument convergente – donc elle est convergente – puisque $\sum_{k=0}^{\infty} p(k) = 1$; c'est également pour cela qu'on peut affirmer que la série converge pour $s = 1$. Donc, $\psi(s)$ est bien définie. Par ailleurs, si on écrit $f_n(s) = \sum_{k=0}^n p(k)s^k$ pour la n -ième somme partielle de la série $\sum_{k \geq 0} p(k)s^k$, alors on a que

$$\begin{aligned} \sup_{s \in [-1, 1]} |\psi(s) - f_n(s)| &= \sup_{s \in [-1, 1]} \left| \sum_{k > n} p(k)s^k \right| \\ &\leq \sum_{k > n} p(k); \end{aligned}$$

On conclue en remarquant que $\sum_{k > n} p(k)$ tend vers 0 lorsque n tend vers l'infini. □

REMARQUE. Le rayon de convergence n'est pas forcément strictement égal à 1 ; il peut être plus grand. Mais on sait au moins que $\psi(s)$ est définie sur l'intervalle $[-1, 1]$.

On note les égalités suivantes, qui sont importantes :

$$(2.1.4) \quad \psi(0) = p(0); \quad \psi(1) = 1.$$

La conséquence du Lemme 2.3, c'est qu'on peut faire tout un tas de choses amusantes avec ψ – principalement, on peut la dériver en dérivant la série qui la définit terme par terme, etc.

LEMME 2.4. *La fonction $\psi(s)$ est infiniment dérivable (on dit aussi lisse) sur l'intervalle $(-1, 1)$. Sur cet intervalle, la n ième dérivée de ψ , notée $\psi^{(n)}$, correspond à la n ième dérivée, terme par terme, de la série $\sum_{k \geq 0} p(k)s^k$:*

$$(2.1.5) \quad \psi^{(n)}(s) = \sum_{k \geq n} p(k) \frac{k!}{(k-n)!} s^{k-n} = \mathbb{E} \left[\frac{Z_1!}{(Z_1-n)!} s^{Z_1-n} \mathbb{1}_{\{Z_1 \geq n\}} \right];$$

Cette série converge uniformément sur tout intervalle $[-\epsilon, \epsilon]$ pour $0 < \epsilon < 1$. En particulier, pour tout n , on a que :

$$(2.1.6) \quad \psi^{(n)}(0) = n!p(n)$$

Si $\mathbb{E}[Z_1^n] < +\infty$, alors cette série converge uniformément sur $[-1, 1]$ et on peut étendre, par continuité, la fonction $\psi^{(n)}$ à l'intervalle fermé $[-1, 1]$ avec :

$$(2.1.7) \quad \psi^{(n)}(\pm 1) = \lim_{s \rightarrow \pm 1^\mp} \psi^{(n)}(s).$$

En particulier, on a que :

$$(2.1.8) \quad \psi^{(n)}(1) = \mathbb{E} \left[\frac{Z_1!}{(Z_1-n)!} \mathbb{1}_{\{Z_1 \geq n\}} \right] = \mathbb{E}[Z_1(Z_1-1)(Z_1-2) \cdots (Z_1-n+1)]$$

est le n ième moment factoriel de Z_1 .

DÉMONSTRATION. On a que $\frac{k!}{(k-n)!} \leq k^n$. Il suffit donc de montrer que la série $\sum_{k \geq n} p(k)k^n s^{k-n}$ converge uniformément sur l'intervalle $(-\epsilon, \epsilon)$ pour tout $\epsilon \in (0, 1)$. En supposant que $s \in [-\epsilon, \epsilon]$, la valeur absolue du terme général est bornée par $\leq k^n \epsilon^k$; c'est une série qui converge. On prouve alors la convergence uniforme de façon similaire à la preuve du Lemme 2.3. On peut donc dériver terme par terme une infinité de fois

Maintenant, si on suppose que $\mathbb{E}[Z_1^n] < +\infty$, la série $\sum_{k \geq n} k^n p(k)$ converge, et donc $\sum_{k \geq n} k \geq n \frac{k!}{(k-n)!} p(k)$ converge aussi, et on peut prouver la convergence uniforme de la série $\sum_{k \geq n} p(k) \frac{k!}{(k-n)!} s^{k-n}$ sur l'intervalle $[-1, 1]$ de la même façon qu'on l'a fait dans le Lemme 2.3. Il suit que les limites $\lim_{s \rightarrow \pm 1^\mp} \psi^{(n)}(s)$ existent et sont bien égales aux valeurs correspondantes des séries. \square

COROLLAIRE. *La fonction ψ caractérise entièrement la distribution de Z_1 , puisqu'on peut retrouver sa fonction de masse en prenant des dérivées successives en 0 :*

$$p(n) = \frac{\psi^{(n)}(0)}{n!}.$$

LEMME 2.5. *La fonction ψ et ses dérivées sont positives, croissantes et convexes¹ partout où elles sont définies dans l'intervalle $[0, 1]$.*

DÉMONSTRATION. Si $s \geq 0$, alors le terme général de la série qui définit ψ (et de celles qui sont égales à ses dérivées) est positif : $p(k)s^k \geq 0$. Par conséquent, $\psi(s) \geq 0$, et de même, $\psi^{(n)}(s) \geq 0$. En particulier, les premières et secondes dérivées de ψ et de $\psi^{(n)}$ sont positives partout où elles sont définies sur $[0, 1]$, et il suit immédiatement qu'elles sont croissantes et convexes (pas forcément strictement). \square

1. Consulter l'annexe B pour des rappels sur les fonctions convexes.

De retour à notre programmation régulière. On va faire les hypothèses suivantes :

- i. $0 < p(0) < 1$;
- ii. $0 < p(0) + p(1) < 1$;
- iii. $\mu = \sum_{k \geq 1} kp(k) = \mathbb{E}[Z_1] < +\infty$.

Les hypothèses **i** et **ii** nous assurent que, non seulement la fonction ψ est positive, croissante et convexe – elle est *strictement* positive, croissante et convexe pour $s \in (0, 1)$. En effet, si il faut qu'il existe un $k > 1$ tel que $p(k) > 0$, alors les dérivées premières et seconde de ψ seront strictement positives sur l'intervalle $(0, 1)$.

Ces hypothèses sont naturelles à formuler. Si on assume que $p(0) = 1$, le problème est vide ; le processus s'éteint après la génération 0. Si on assume que $p(0) = 0$, il ne s'éteint jamais. Si on assume que $p(0) + p(1) = 1$, alors il faut que $p(1) = 1 - p(0)$. Le processus se résume donc à une suite de tentatives de Bernoulli indépendantes, ce qui n'est que très légèrement moins trivial.

L'hypothèse **iii** nous assure que $\psi'(1) = \lim_{s \rightarrow 1^-} \psi'(s)$ existe et est finie. La figure 2.2 montre l'allure typique de fonctions génératrices qui respectent ces conditions.

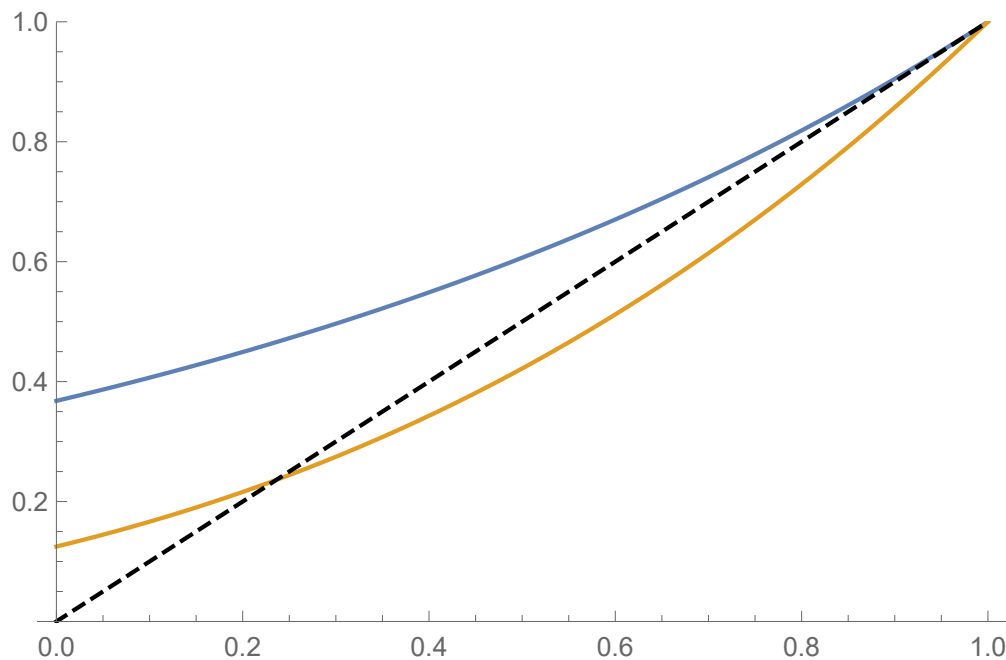


FIGURE 2.2 – Les fonctions génératrices pour les distributions de Poisson (en bleu ; paramètre $\lambda = 1$) et binomiale (en jaune ; $n = 3$ et $p = 1/2$). La diagonale $y = s$ est représentée en pointillés.

2.1.2.1. *La probabilité d'extinction.* On cherche la probabilité d'extinction

$$\mathbb{P}_1 \{ \tau_0 < +\infty \}.$$

On peut se servir du théorème ergodique (théorème 1.1) ; en effet, l'état 0 est récurrent et apériodique. On a évidemment que $\mathbb{E}_0 [\tau_0^+]$. Donc, par le théorème ergodique, on sait

déjà que la limite $\lim_{t \rightarrow \infty} P_{1,0}^{(t)}$ existe, et que :

$$\lim_{t \rightarrow \infty} P_{1,0}^{(t)} = \mathbb{P}_1 \{ \tau_0 < +\infty \}.$$

On a également vu, à la partie (b) de l'exemple 2.2, que $P_{k,0}^{(t)} = (P_{1,0}^{(t)})^k$. On va se servir de ces deux faits pour raisonner, en obtenant une récurrence, par un conditionnement par le premier pas :

$$\begin{aligned} P_{1,0}^{(t)} &= \sum_{k=0}^{\infty} \mathbb{P}_1 \{ Z_t = 0 \mid Z_1 = k \} \mathbb{P}_1 \{ Z_1 = k \} \\ &= \sum_{k=0}^{\infty} p(k) P_{k,0}^{(t-1)}. \end{aligned}$$

La seconde égalité s'obtient en vertu du fait que $\mathbb{P}_1 \{ Z_t = 0 \mid Z_1 = k \}$ est simplement égale à $\mathbb{P}_k Z_{t-1} = 0$ par homogénéité et la propriété de Markov ; quant à $\mathbb{P}_1 \{ Z_1 = k \}$, c'est la probabilité que l'individu de la génération 0 ait eu k enfants, soit $p(k)$.

On utilise maintenant le résultat de l'exemple 2.2 (b), pour écrire :

$$(2.1.9) \quad P_{1,0}^{(t)} = \sum_{k=0}^{\infty} p(k) (P_{1,0}^{(t-1)})^k = \psi(P_{1,0}^{(t-1)}).$$

Cette relation de récurrence s'avère très utile ; en effet, puisque la fonction ψ est continue sur $[0, 1]$, et que la limite $\lim_{t \rightarrow \infty} P_{1,0}^{(t)} = u$ (et $\lim_{t \rightarrow \infty} P_{1,0}^{(t-1)} = u$) existe, on *doit* avoir que cette limite u satisfait :

$$u = \psi(u).$$

La limite u est donc ce que l'on appelle un *point fixe* de la fonction ψ – c'est un point de l'intervalle où ψ est définie, et dont l'image par ψ est simplement le point lui-même.

Sur un graphe de la fonction $\psi(s)$ pour $s \in [0, 1]$, les points fixes sont les endroits où le graphe de la fonction croise la diagonale $y = s$ (représentée en pointillés sur la figure 2.2).

Notre objectif, maintenant, c'est d'identifier les points fixes de la fonction ψ . On peut déjà faire la proposition suivante :

PROPOSITION 2.1. *Soit $\mathbf{Z} = (Z_t)_{t \in \mathbb{Z}^+}$ un processus de Galton-Watson avec fonction de masse du nombre d'enfants p , fonction génératrice des probabilités pour le nombre d'enfants ψ , et $\mu = \mathbb{E}[Z_1] = \psi'(1)$. Alors,*

- i. *si $\mu \leq 1$, alors la probabilité d'extinction est de 1 :*

$$\mathbb{P}_1 \{ \tau_0 < +\infty \} = 1.$$

- ii. *si $\mu > 1$, alors la probabilité $\mathbb{P}_1 \{ \tau_0 < +\infty \} = u$ est l'unique nombre dans l'intervalle $(0, 1)$ qui satisfait $u = \psi(u)$.*

DÉMONSTRATION.

i On sait déjà que la valeur $u = 1$ satisfait $u = \psi(u)$; il faut maintenant montrer que c'est la seule valeur possible qui satisfait cette équation dans l'intervalle $[0, 1]$. Pour ce faire, on utilise des arguments d'analyse simples. On peut comprendre intuitivement ce qui se passe lorsqu'on regarde la courbe bleue de la figure 2.2 – lorsque $\mu = \psi'(1) \leq 1$, le graphe de ψ doit se situer au-dessus de la diagonale pointillée sauf en $s = 1$.

On considère la fonction $f(s) = \psi(s) - s$; on veut montrer qu'elle ne peut pas s'annuler avant $s = 1$. On a que $f(0) = p(0) > 0$, et $f(1) = 0$. Or, la fonction

f a pour dérivée $f'(s) = \psi'(s) - 1$, et bien sûr, la fonction $\psi'(s)$ est strictement croissante en s (puisque ψ est strictement convexe); Puisque $\psi'(1) = \mu \leq 1$, on doit avoir que $\psi'(s) < 1$ pour tout $s < 1$.

Donc, on a que $f'(s) = \psi'(s) - 1 < 0$ pour tout $s < 1$, et il suit que la fonction f est strictement décroissante sur l'intervalle $[0, 1]$; en supposant qu'il existe $u < 1$ tel que $f(u) = 0$, on devrait donc avoir $f(1) < 0$, ce qui est contradictoire puisqu'on sait que $f(1) = 0$.

Donc, $u = 1$ est bel et bien la seule solution de l'équation $u = \psi(u)$ dans ce cas.

Or, la limite $u = \lim_{t \rightarrow \infty} P_{1,0}^{(t)} = \mathbb{P}_1 \{ \tau_0 < +\infty \}$ par le théorème ergodique, et il suit nécessairement que $\mathbb{P}_1 \{ \tau_0 < +\infty \} = 1$.

- ii Si on suppose plutôt que $\mu > 1$, alors, toujours avec notre fonction $f(s) = \psi(s) - s$, on a que $f'(1) > 0$, tandis que $f'(0) = \psi'(0) - 1 = p(1) - 1 < 0$. Par continuité, monotonie de f' et le théorème des valeurs intermédiaires, il existe exactement un unique point $x \in (0, 1)$ tel que $f'(x) = 0$; ce point doit être un minimum absolu pour la fonction f sur l'intervalle $[0, 1]$, puisque cette dernière est strictement convexe (en effet, $f''(s) = \psi''(s) > 0$). On doit donc avoir $f(x) < 0$.

Encore par le théorème des valeurs intermédiaires, et par monotonie de f sur l'intervalle $[0, x]$, on doit donc avoir un unique $u < x$ tel que $f(u) = 0$. Il existe donc un unique $u \in (0, 1)$ qui satisfait $f(u) = 0$.

Par convexité de f , on a que :

- pour tout $s \in [0, u)$, $f(s) > 0$;
- pour tout $s \in (u, 1)$, $f(s) < 0$.

Encore une fois, on comprend bien ce qui arrive en regardant la courbe jaune de la figure 2.2 : la fonction génératrice ψ ne peut croiser la diagonale pointillée qu'une seule fois avant $s = 1$.

Mais on a donc deux valeurs possible dans $[0, 1]$ qui satisfont l'équation $u = \psi(u)$: 1, et une autre strictement plus petite que 1. On va montrer que la suite $P_{1,0}^{(t)}$ ne converge pas vers 1, en procédant par contradiction.

Supposons que $\lim_{t \rightarrow \infty} P_{1,0}^{(t)} = 1$. Dans ce cas, il existe un certain $N \in \mathbb{N}$ tel que pour tout $t > N$, on a que $P_{1,0}^{(t)} > u$, où u est le point fixe non-trivial de ψ (donc $0 < u < 1$).

puisque $P_{1,0}^{(t)} > u$ pour tout $t > N$, alors il suit que, si $t > N$, on a que

$$P_{1,0}^{(t+1)} = \psi(P_{1,0}^{(t)}) < P_{1,0}^{(t)},$$

soit que la suite décroît strictement. Puisque la suite $P_{1,0}^{(t)}$ est strictement décroissante et bornée supérieurement par 1, sa limite doit être strictement inférieure à 1, ce qui contredit l'hypothèse que nous avons introduite; on doit donc la rejeter : la suite $P_{1,0}^{(t)}$ ne peut pas converger vers 1 si $\mu > 1$.

Donc, on est forcé·e·s de conclure que la limite de la suite $P_{1,0}^{(t)}$ est l'unique valeur u contenue dans $(0, 1)$ et qui satisfait $u = \psi(u)$.

□

EXEMPLE 2.3. Soit $\mathbf{Z} = (Z_t)_{t \in \mathbb{Z}^+}$ un processus de Galton-Watson où le nombre d'enfants suit une distribution binomiale avec $n = 3$ et $p = 1/2$:

$$\frac{p(k) = \binom{3}{k}}{8}.$$

- (a) Quelle est la valeur de $\mathbb{E}[Z_1]$?
- (b) Ce processus s'éteindra-t-il forcément ? Si non, quelle est la probabilité de survie ?

SOLUTION. (a) On a que $\mathbb{E}[Z_1] = np = \frac{3}{2} = \mu$.

- (b) Puisque $\mu = \frac{3}{2} > 1$, il suit que le processus a une probabilité d'extinction inférieure à 1 – donc une probabilité de survie positive.

Pour trouver la probabilité de survie, il faut d'abord trouver $\psi(s)$; c'est :

$$\psi(s) = \frac{1}{8} + \frac{3}{8}s + \frac{3}{8}s^2 + \frac{1}{8}s^3.$$

On veut maintenant résoudre :

$$\psi(s) = s.$$

On obtient l'équation :

$$s^3 + 3s^2 - 5s + 1 = 0.$$

Évidemment, $s = 1$ satisfait cette équation ; on peut donc factoriser $(s - 1)$. On trouve que

$$s^3 + 3s^2 - 5s + 1 = (s - 1)(s^2 + 4s - 1),$$

et on peut maintenant résoudre

$$\begin{aligned} 0 &= s^2 + 4s - 1 \\ &= (s^2 + 4s + 4) - 5 \\ &= (s + 2)^2 - 5. \end{aligned}$$

Finalement, on trouve que $(s + 2) = \pm\sqrt{5}$, donc que $s = -2 \pm \sqrt{5}$.

Bien sûr, on veut une solution entre 0 et 1 exclusivement. Il faut donc prendre $s = \sqrt{5} - 2$.

Finalement, on a donc que

$$\mathbb{P}_1 \{ \tau_0 < +\infty \} = \sqrt{5} - 2 \approx 23,6\%.$$

La probabilité de *survie* est donc

$$1 - \mathbb{P}_1 \{ \tau_0 < +\infty \} = 3 - \sqrt{5} \approx 76,4\%.$$

2.1.2.2. *La distribution de Z_t .* Pour connaître la distribution de Z_n , on utilise encore les fonctions génératrices ; on définit :

$$\psi_t(s) = \sum_{k=0}^{\infty} \mathbb{P}_1 \{ Z_t = k \} s^k = \sum_{k=0}^{\infty} P_{1,k}^{(t)} s^k = \mathbb{E} [s^{Z_t}].$$

Comme pour ψ et Z_1 , si on parvient à trouver ψ_t , on aura entièrement caractérisé la distribution de Z_t ; il suffira alors de prendre les dérivées pertinentes pour extraire les informations voulues sur la distribution.

PROPOSITION 2.2. Soit $\mathbf{Z} = (Z_t)_{t \in \mathbb{Z}^+}$ un processus de Galton-Watson avec la fonction de masse du nombre d'enfants donnée par p , la fonction génératrice des probabilités du nombre d'enfants donnée par ψ , et la fonction génératrice des probabilités de Z_t donnée par ψ_t .

Alors, on a que $\psi_1 = \psi$, et

$$(2.1.10) \quad \psi_t = \psi_{t-1} \circ \psi.$$

Plus spécifiquement,

$$(2.1.11) \quad \psi_t = \underbrace{\psi \circ \psi \circ \psi \circ \cdots \circ \psi}_{t \text{ fois}}.$$

DÉMONSTRATION. Pour ce faire, on va utiliser les espérances conditionnelles : en effet, on a que :

$$\psi_t = \mathbb{E} [\mathbb{E} [s^{Z_t} \mid Z_{t-1}]] .$$

On s'intéresse donc à $\mathbb{E} [s^{Z_t} \mid Z_{t-1}]$. Par l'équation 2.1.1, on a que, si $Z_{t-1} = k$,

$$Z_t = \xi_{t-1,1} + \xi_{t-1,2} + \xi_{t-1,3} + \cdots + \xi_{t-1,k}.$$

Donc,

$$\mathbb{E} [s^{Z_t} \mid Z_{t-1} = k] = \mathbb{E} [s^{\xi_{t-1,1} + \xi_{t-1,2} + \xi_{t-1,3} + \cdots + \xi_{t-1,k}} \mid Z_{t-1} = k] .$$

Mais les $\xi_{t-1,i}$ sont indépendants de Z_{t-1} , indépendant entre eux, et identiquement distribués. On a donc que :

$$\mathbb{E} [s^{Z_t} \mid Z_{t-1} = k] = \mathbb{E} [s^{\xi_{0,1}}]^k = \mathbb{E} [s^{Z_1}]^k = \psi(s)^k.$$

Donc,

$$\mathbb{E} [s^{Z_t} \mid Z_{t-1}] = \psi(s)^{Z_{t-1}},$$

et en appliquant l'espérance de part et d'autre, on trouve :

$$\psi_t(s) = \mathbb{E} [\mathbb{E} [s^{Z_t} \mid Z_{t-1}]] = \mathbb{E} [\psi(s)^{Z_{t-1}}] = \psi_{t-1}(\psi(s)).$$

On conclue donc :

$$\psi_t = \psi_{t-1} \circ \psi.$$

Évidemment $\psi_1 = \psi$ par définition de ψ ; on voit facilement que l'équation 2.1.11 est vraie par induction sur t . \square

Les détails de la distribution de Z_t peuvent tout de même rester ardues à obtenir ; il faut d'abord calculer ψ_t , puis la dériver plusieurs fois. Mais c'est théoriquement possible. Par ailleurs, on peut montrer certaines choses assez facilement.

PROPOSITION 2.3. Soit $\mathbf{Z} = (Z_t)_{t \in \mathbb{Z}^+}$ un processus de Galton-Watson avec $\mu = \mathbb{E}_1 [Z_1]$. Alors, on a que

$$(2.1.12) \quad \mathbb{E}_1 [Z_t] = \mu^t.$$

DÉMONSTRATION. Soient ψ et ψ_t les fonctions génératrices des probabilités respectivement pour Z_1 et Z_t ; alors, on a que : $\psi_t(s) = \psi_{t-1}(\psi(s))$, et en dérivant, on trouve :

$$\psi'_t(s) = \psi'_{t-1}(\psi(s))\psi'(s).$$

Donc,

$$\psi'_t(1) = \psi'_{t-1}(\psi(1))\psi'(1) = \psi'_{t-1}(1)\mu.$$

Bien sûr, avec $\psi'_1(1) = \psi'(1) = \mu$, on trouve donc, par récurrence, que

$$\psi'_t(1) = \mathbb{E}_1[Z_t] = \mu^t.$$

□

Si on définit la *population totale* du processus de Galton-Watson comme la somme des nombres d'individus de chaque génération, on a que la population totale (disons T) est donnée par la série

$$T = \sum_{t=0}^{\infty} Z_t,$$

et

$$\mathbb{E}[T] = \sum_{t=0}^{\infty} \mu^t = \begin{cases} \frac{1}{1-\mu} & \text{si } \mu < 1 \\ +\infty & \text{si } \mu \geq 1 \end{cases}.$$

REMARQUE. On remarque que dans le cas dit *critique*, où $\mu = 1$, la probabilité d'extinction est de 1, mais la taille espérée de la population totale est tout de même infinie.

2.2. La marche aléatoire sur \mathbb{Z}^d

De façon très informelle, on appelle *marche aléatoire* tout processus stochastique (généralement à temps discret) que l'on peut visualiser comme le déplacement aléatoire d'un « marcheur » sur un espace mathématique quelconque – un graphe, le plan, une variété Riemannienne...

REMARQUE. Le graphe des transitions possibles peut-être vu comme cet espace mathématique, et donc si on y tient, n'importe quel processus stochastique à temps discret peut être considéré, si l'on veut, comme une marche aléatoire sur son graphe des transitions possibles.

Dans la section qui suit, nous étudions les marches aléatoires à pas entiers en d dimensions ; l'espace des états est \mathbb{Z}^d , soit l'ensemble des points de l'espace euclidien avec coordonnées entières.

2.2.1. Le cas $d = 1$. Le modèle en une dimension nous est déjà quelque peu familier : on suppose que les probabilités de transition sont données par

$$P_{i,i+1} = p \quad P_{i,i-1} = q = 1 - p,$$

où p fait office de paramètre.

La figure 2.3 montre des trajectoires typiques pour différentes valeurs de p .

L'analyse de ce processus nous est déjà quelque peu familière ; nous savons déjà qu'il n'y a qu'une seule classe d'équivalence.

PROPOSITION 2.4. Soit $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ une marche aléatoire sur \mathbb{Z} avec probabilités de transition en un pas $P_{i,i+1} = p$ et $P_{i,i-1} = 1 - p =: q$.

Alors, l'espérance du temps de retour en 0 est infinie :

$$\mathbb{E}_0[\tau_0^+] = +\infty.$$

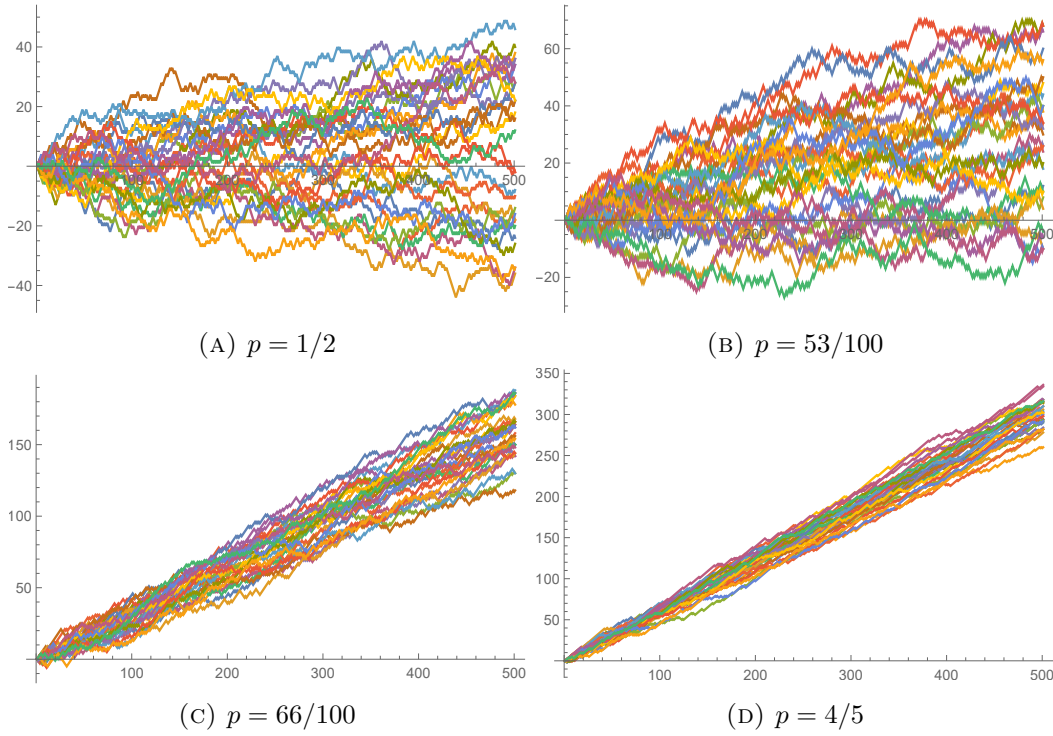


FIGURE 2.3 – Pour chaque valeur de p , 30 réalisations des 500 premiers pas d’une marche aléatoire sur \mathbb{Z} .

DÉMONSTRATION. En supposant que cette espérance est finie, puisque la chaîne est irréductible, on devrait avoir que

$$\lim_{t \rightarrow \infty} P_{0,0}^{(t)} = 1/\mathbb{E}_0 [\tau_0^+],$$

or, puisque la chaîne est irréductible et que l’espérance est finie, il existe une distribution stationnaire. Mais on a

$$\pi_i = \pi_{i-1}p + \pi_{i+1}q;$$

En soustrayant $q\pi_i + p\pi_{i-1}$ de part et d’autre on trouve

$$p(\pi_i - \pi_{i-1}) = q(\pi_{i+1} - \pi_i),$$

ou, réarrangeant :

$$\pi_{i+1} - \pi_i = \frac{p}{q}(\pi_i - \pi_{i-1}).$$

On remarque donc que la différence entre les π_i est au mieux constante lorsque $p = q$, ou exponentielle si $p \neq q$; puisque ces équations sont valides pour tous $i \in \mathbb{Z}$, si $p \neq q$, on peut toujours trouver un i tel que $\pi_i > 1$, peu importe la valeur de π_0 et π_1 . Ceci est une contradiction, et on doit conclure que, bien qu’il puisse exister une solution stationnaire, ça n’est pas une distribution.

Si $p = q$, la différence entre π_i et π_{i-1} est constante – elle doit être nulle, sinon, il y a forcément un i tel que $\pi_i < 0$. Mais alors, les seules solutions stationnaires sont de la forme $\pi_i = C$, et il n’en existe aucune qui satisfait la condition de normalisation. Donc il n’y a toujours pas de distribution stationnaire.

Donc, on doit avoir $\mathbb{E}_0 [\tau_0^+] = +\infty$. \square

La question de déterminer si la chaîne est transiente ou récurrente. Pour ce faire, il existe plusieurs méthodes. Dans le cas qui nous intéresse ici, nous allons faire les preuves en utilisant le théorème de la limite centrale.

PROPOSITION 2.5. *Soit $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ une marche aléatoire sur \mathbb{Z} avec probabilités de transition en un pas $P_{i,i+1} = p$ et $P_{i,i-1} = 1 - p =: q$.*

Alors la marche aléatoire est récurrente si seulement si $p = 1/2$.

DÉMONSTRATION. On peut écrire $X_t = X_0 + \sum_{u=1}^t Y_u$, où les Y_u sont des variables aléatoires indépendantes et identiquement distribuées avec

$$\mathbb{P}\{Y_u = 1\} = p, \quad \mathbb{P}\{Y_u = -1\} = 1 - p.$$

Or, ce faisant, on trouve immédiatement que

$$\mathbb{E}[Y_1] = p - q, \quad \text{Var}[Y_1] = 1 - (p - q)^2 = 4pq$$

Par conséquent, il suit que, lorsque t tend vers l'infini,

$$Z_t = \frac{X_t - (p - q)t}{\sqrt{4pqt}}$$

converge en distribution vers une loi normale standard.

En particulier,

si $p > q$,

$$\begin{aligned} \mathbb{P}_0\{X_t = 0\} &= \mathbb{P}_0\left\{Z_t = -\frac{(p - q)t}{\sqrt{4pqt}}\right\} \\ &\leq \mathbb{P}_0\{Z_t \leq -C\sqrt{t}\} \\ &\leq \frac{1}{2}e^{-C^2t/2}, \end{aligned}$$

où C est une constante ne dépendant pas de t , et où on a utilisé un estimé pratique sur la queue d'une variable normale :

$$\mathbb{P}\{\mathcal{N}(0, 1) \geq u\} \leq \frac{1}{2}e^{-u^2/2}.$$

Donc, par le critère donné par la proposition 1.8, on a que

$$\sum_{t=1}^{\infty} P_{0,0}^{(t)} \leq \frac{1}{2} \sum_{t=1}^{\infty} e^{-C^2t/2}$$

converge, et les états sont transients.

si $p = q = 1/2$, c'est un peu plus difficile, mais on peut tout de même faire le calcul directement. En effet, dans ce cas, $X_t = 0$ si et seulement si $t = 2n$ et on a fait exactement n pas à gauche, et n à droite ; c'est à dire que :

$$\mathbb{P}_0\{X_t = 0\} = \frac{1}{4^n} \binom{2n}{n} = \frac{1}{4^n} \frac{(2n)!}{n! \cdot n!}.$$

Pour nous tirer d'affaire, on utilise l'approximation de Stirling pour les factoriels :

$$n! \sim \sqrt{2\pi n} (n/e)^n.$$

On a donc :

$$\frac{1}{4^n} \binom{2n}{n} \sim \frac{\sqrt{4\pi n} (2n/e)^{2n}}{4^n 2\pi n (n/e)^{2n}} = \frac{1}{\sqrt{\pi n}},$$

et bien sûr, $\sum_{n=1}^{\infty} \frac{1}{\sqrt{\pi n}}$ diverge, donc $\sum_{t=1}^{\infty} \mathbb{P}_0 \{X_t = 0\}$ aussi, et l'état 0 est récurrent. \square

Les limites supérieures et inférieures ; la vitesse de croisière. Par la loi des grands nombres, on a bien sûr que

$$(2.2.1) \quad \lim_{t \rightarrow \infty} \frac{X_t}{t} = p - q \text{ p. s.}$$

Ceci nous donne une indication sur la *vitesse de croisière* de la marche ; elle s'éloigne de son point d'origine à un rythme moyen égal à $(p - q)$ pas par unité de temps.

Bien sûr, lorsque $p = q = \frac{1}{2}$, cette vitesse de croisière est nulle. Néanmoins, on sait que pour tout élément $\pm i$, avec $i \in \mathbb{N}$, on a que, puisque la chaîne est irréductible et que les états sont récurrents,

$$\mathbb{P}_0 \{\tau_{\pm i} < +\infty\} = 1;$$

c'est à dire qu'on atteindra éventuellement *tous* les éléments de \mathbf{Z} . Autrement dit, dans le cas récurrent :

$$(2.2.2) \quad \limsup_{t \rightarrow \infty} \pm X_t = +\infty \text{ p. s.}$$

Si on souhaite être plus précis, on peut utiliser à nouveau le théorème de la limite centrale dans le cas récurrent : dans ce cas, la variable aléatoire $Z_t = X_t/\sqrt{t}$ converge en distribution vers une variable aléatoire de loi normale standard ; conséquemment, $Z_t^2 = X_t^2/t$ converge en distribution vers une variable aléatoire de loi χ^2 ; on a donc que

$$\text{Var}[X_t] \sim t.$$

Excursions. On introduit maintenant la notion d'*excursion*.

Une **excursion** correspond à la portion de temps consécutif avant un retour à l'état 0, partant de 0.

On introduit la *hauteur maximale* d'une excursion

$$H = \max \{|X_t| : t \leq \tau_0^+\}.$$

La figure 2.4 montre les excursions lors des 50 premiers pas d'une marche aléatoire récurrente sur \mathbb{Z} .

On peut se convaincre que :

$$\mathbb{P}_0 \{H \geq n\} = \mathbb{P}_1 \{\tau_n < \tau_0\};$$

en effet, l'excursion sera de hauteur au moins n si et seulement si, partant de 0, selon qu'on est allé.e.s d'abord à ± 1 , on touche ensuite $\pm n$ avant de revenir en 0. Comme nous l'avons vu à l'exemple 1.17, cette probabilité est donc :

$$\mathbb{P}_0 \{H \geq n\} = \frac{1}{n};$$

on peut donc caractériser complètement la fonction de masse de H :

$$\mathbb{P}_0 \{H = n\} = \mathbb{P} \{H \geq n\} - \mathbb{P} \{H \geq n+1\} = \frac{1}{n} - \frac{1}{n+1} = \frac{1}{n^2 + n}.$$

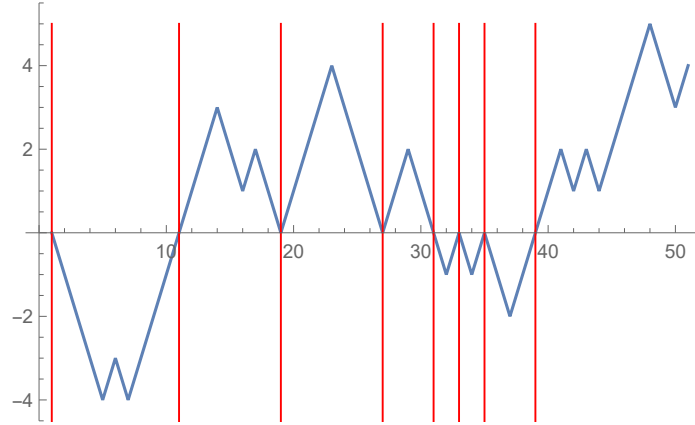


FIGURE 2.4 – Les excursions sont séparées par des lignes verticales rouges.

D'autres problèmes intéressants. Nous avons déjà vu que $\mathbb{E}_0 \{\tau_0^+\} = +\infty$; il pourrait être intéressant d'obtenir plus de détails sur la distribution de τ_0^+ ; dans les faits, on a le résultat suivant :

$$\mathbb{P}_0 \{\tau_0^+ = 2n\} = \frac{1}{(2n-1)4^n} \binom{2n}{n} = \frac{1}{2n-1} \mathbb{P}_0 \{X_{2n} = 0\} \dots$$

2.2.2. Le cas $d = 2$. On passe maintenant à « la vitesse supérieure »; qu'est-ce qui se passe si on regarde plutôt une marche aléatoire, non pas sur \mathbb{Z} , mais sur une grille \mathbb{Z}^2 , où les transitions possibles sont entre les sommets voisins (à distance 1).

Soit $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ une marche aléatoire sur \mathbb{Z}^2 avec les probabilités de transition suivantes :

$$P_{(i,j),(i,j \pm 1)} = P_{(i,j),(i \pm 1,j)} = \frac{1}{4}.$$

La figure 2.5 montre les dix mille premiers pas d'une réalisation de marche aléatoire sur \mathbb{Z}^2 .

Récurrente ou transiente ? D'abord, on cite tout de suite une évidence : si on note $\mathbf{0} = (0,0)$ notre point d'origine,

$$\mathbb{E}_{\mathbf{0}} [\tau_{\mathbf{0}}^+] = +\infty.$$

Si c'était le cas en une dimension, le fait d'avoir rajouté encore plus d'endroits où aller pour éviter de retourner à 0 n'aura pas amélioré les choses.

La question demeure : ce processus est-il récurrent ou transient ? La réponse n'est pas triviale du tout !

Pour la trouver, il faut...

Une analogie avec les circuits électriques. Le problème posé est le suivant : on s'imagine que notre graphe est en fait un vaste circuit électrique; sur chaque arête, on a placé une résistance d'un Ohm (figure 2.6).

A priori, on semble avoir changé complètement de sujet... mais en fait, non ! Il existe un parallèle puissant entre circuits électriques et marches aléatoires ! C'est ce qu'on explore dans les prochains paragraphes.

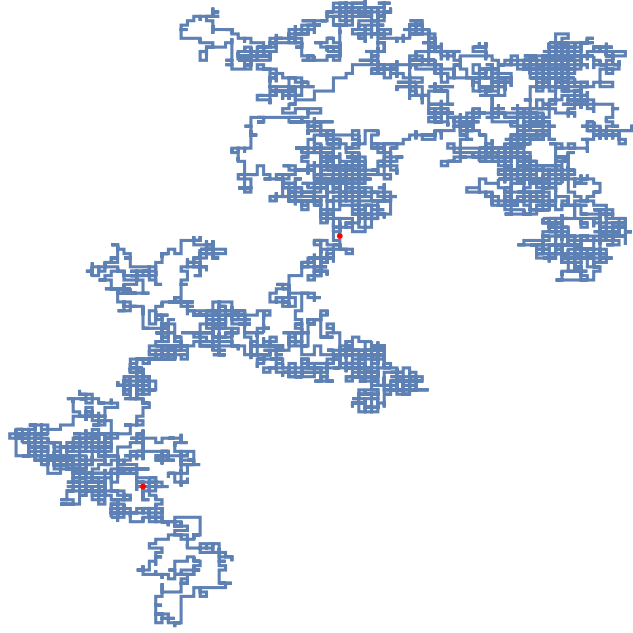


FIGURE 2.5 – Les 10000 premiers pas d’une marche aléatoire dans \mathbb{Z}^2 ; les points de départ et d’arrivée sont marqués en rouge.

Rappel de notions de physique, partie 1 : définitions et postulats. Si on représente un circuit électrique comme un graphe simple $G = (S, E)$, où les arêtes (non-orientées) correspondent aux composantes résistives du circuit, et les sommets sont les points de contact électrique (les *nœuds*) entre ces composantes, la physique nous donne des lois simples pour décrire mathématiquement les déplacements de charges dans le circuit :

DÉFINITION 2.1 (Courant, potentiel, conductance, résistance, sources). *i.* Le **courant électrique** qui circule entre deux nœuds voisins x et y , dans la direction de x vers y , est le débit de charges électriques transférés de x à y par unité de temps. On le note $i(x, y)$. Dans le système international d’unités, il est mesuré en Ampère ; un Ampère correspond à un Coulomb de charge par seconde.

Le courant est positif si le débit de charges net circule vers y , et négatif si il circule plutôt vers x . On a en outre que

$$i(x, y) = -i(y, x).$$

- ii.* Le **potentiel électrique** au nœud x indique la quantité d’énergie disponible par charge lorsqu’elle se trouve en x , et est noté $v(x)$. Dans le système international d’unités, il est mesuré en Volts ; ça peut être une quantité positive ou négative. En fait, le potentiel est toujours déterminé à une constante près, puisque le niveau de référence de potentiel électrique est arbitraire.
- iii.* La **conductance** d’une composante électrique mesure la propension de celle-ci à laisser passer librement le courant électrique. Plus une conductance est élevée, plus les charges qui y circulent doivent dépenser de l’énergie pour la traverser.

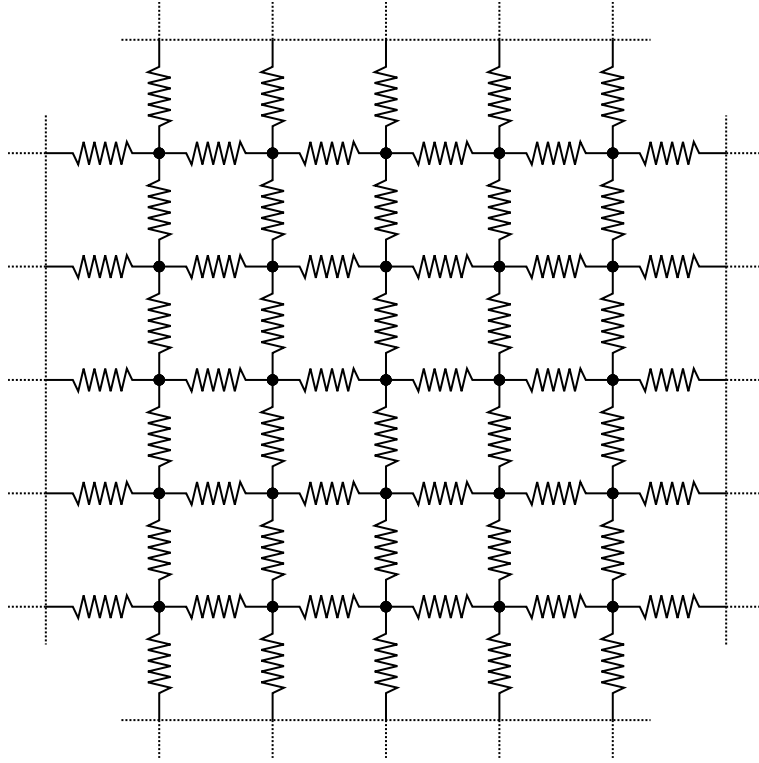


FIGURE 2.6 – Une grille infinie de résistances d'un Ohm.

Dans notre circuit électrique, si les sommets x et y sont adjacents (c'est à dire qu'ils sont reliés par une composante électrique), on note la conductance de cette composante $c(x, y)$. La conductance est une quantité positive (ou nulle si le courant ne peut pas passer), et on a que, bien sûr,

$$c(x, y) = c(y, x).$$

- iv.* La **résistance** d'une composante électrique mesure sa propension à empêcher le passage du courant ; c'est simplement l'inverse de la conductance :

$$r(x, y) = c(x, y)^{-1}.$$

Dans le système international d'unités, les résistances sont mesurées en Ohm (noté Ω), et les conductances en Ohm^{-1} .

- v.* Lorsqu'on applique à un circuit une différence de potentiel entre deux nœuds x et y , ces nœuds sont appelés des **sources** de courant.

Les lois physiques qui régissent les interactions entre ces différentes quantités sont dictées par des principes physiques que nous accepterons ici sans trop de justification, comme des postulats ; ce sont les lois d'Ohm et de Kirchhoff :

POSTULAT 1 (Loi d'Ohm). Le courant qui circule à travers un composant électrique est en proportion de sa conductance et de la différence de potentiel à ses bornes ; en fait :

$$(2.2.3) \quad i(x, y) = c(x, y)(v(x) - v(y)),$$

ou, de façon équivalente,

$$v(x) - v(y) = r(x, y)i(x, y).$$

POSTULAT 2 (Loi des nœuds de Kirchhoff). La conservation des charges impose que le débit de charges net circulant à tout nœud x qui n'est pas une source de courant doit être nul ; autrement dit, pour tout nœud x excepté les sources de courant,

$$(2.2.4) \quad \sum_{y: x \sim y} i(x, y) = 0.$$

Parmi les questions intéressantes à se poser :

- (a) En supposant qu'on applique une différence de potentiel totale $v(x) - v(y) = 1$ entre deux nœuds-source x et y , peut-on savoir le potentiel partout dans le circuit ?
- (b) Quel est le courant total qui circule dans tout le circuit ? Et sur chaque branche individuellement ?

Pour répondre à ces questions, on se sert des lois d'Ohm et de Kirchhoff.

La loi des nœuds et le potentiel. Si on applique une différence de potentiel $v(x) - v(y) = 1$ entre deux sources de courant x et y dans notre circuit électrique, on peut choisir la jauge de telle sorte que $v(y) = 0$ et $v(x) = 1$.

Pour tout autre nœud z , on montre assez facilement que :

$$(2.2.5) \quad v(z) = \sum_{w: z \sim w} \frac{c(z, w)}{\lambda(z)} v(w),$$

où nous avons écrit $\lambda(z) = \sum_{w: z \sim w} c(z, w)$ la somme de toutes les conductances des composantes incidentes au nœud z .

On dit que le potentiel est une fonction *harmonique* pour notre circuit, puisque $v(z)$ est la moyenne du potentiel sur les nœuds voisins, pondérée par les conductances ; en résolvant ce système d'équations linéaire (l'équation (2.2.5) pour tout $z \neq x, y$, avec $v(x) = 1$ et $v(y) = 0$), on peut trouver le potentiel.

La conductance effective. Pour savoir quel est le courant *total* qui circule entre les sources x et y lorsqu'on y applique une différence de potentiel d'un volt, on peut s'imaginer un instant qu'on cherche à remplacer tout le circuit entre ces deux nœuds par *une seule* composante. La quantité d'énergie dissipée par cette composante doit être la même que celle dissipée par tout le circuit.

DÉFINITION 2.2 (Conductance, résistance effectives). La **conductance effective** du circuit entre les sources x et y est la conductance qu'il faut donner à un unique composant de circuit placé entre ces deux sources pour que le courant qui y circule soit égal au courant total qui circulait dans le circuit original. On la note $C_{\text{eff}}(x, y)$.

La **résistance effective** du circuit entre les sources x et y , notée $R_{\text{eff}}(x, y)$, est simplement :

$$R_{\text{eff}}(x, y) = C_{\text{eff}}(x, y)^{-1}.$$

Pour calculer ces quantités, on réduit le circuit de façon itérative en y effectuant des transformations qui réduisent chaque fois le nombre de composantes, tout en préservant la conductance effective, jusqu'à ce qu'il ne reste plus qu'une seule composante de circuit (soit une seule arête dans le graphe). Les transformations possibles sont les suivantes (figure 2.7) :

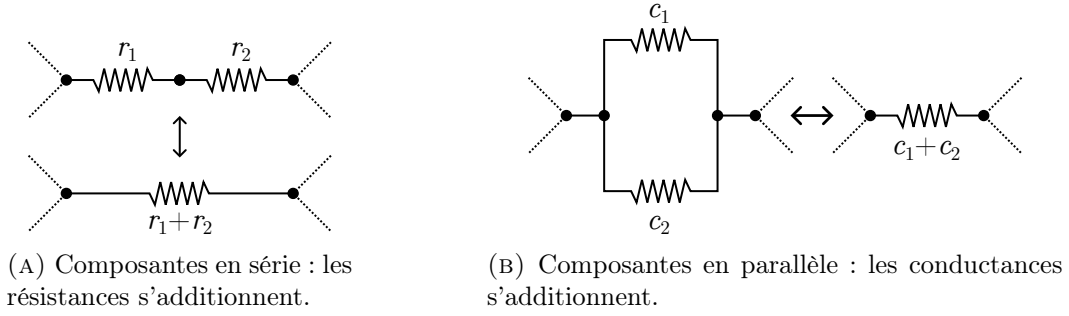


FIGURE 2.7 – Règles élémentaires de réduction des circuits.

POSTULAT 3. On peut toujours remplacer deux composantes placées en série par une seule dont la résistance est la somme des deux composantes remplacées.

POSTULAT 4. On peut toujours remplacer deux composantes placées en parallèle par une seule dont la conductance est la somme des deux composantes remplacées.

Il existe finalement une autre équivalence pratique qui ne réduit pas le nombre d'arêtes, mais qui peut se révéler utile tout de même pour la réduction de circuits – on l'appelle l'équivalence étoile-triangle (figure 2.8) :

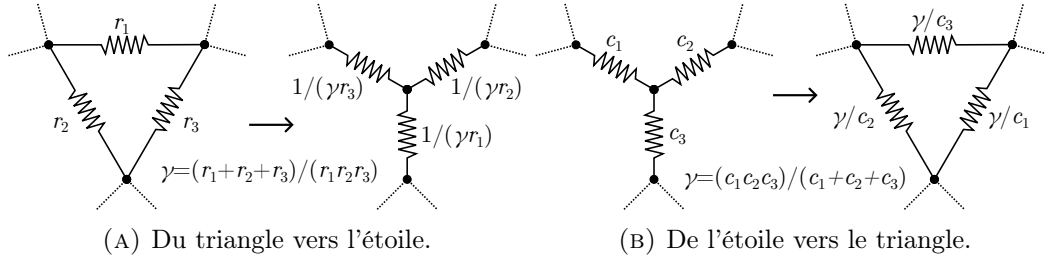


FIGURE 2.8 – L'équivalence étoile-triangle.

POSTULAT 5. En supposant qu'on a trois sommets z_1, z_2, z_3 tels que $z_i \sim z_{i+1}$ (où les indices sont modulo 3), le triangle de composantes est équivalent à une « étoile », où on ajoute un quatrième sommet w ; on retire alors toutes les composantes entre les z_i et $z_{i\pm 1}$, et on ajoute à la place des composantes respectivement entre z_i et w , avec les conductances

$$c'(z_i, w) = \gamma / c(z_{i-1}, z_{i+1}),$$

où

$$\gamma = \frac{r(z_1, z_2) + r(z_2, z_3) + r(z_3, z_1)}{r(z_1, z_2)r(z_2, z_3)r(z_3, z_1)}.$$

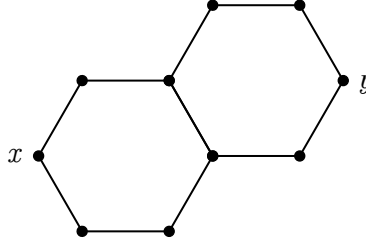
En utilisant les postulats 3, 4 et 5 (en fait ce sont des résultats démontrables mais on les accepte comme tels), on peut toujours réduire n'importe quel circuit jusqu'à n'obtenir qu'une seule composante ; on a alors que, avec un voltage unitaire ($v(x) - v(y) = 1$) :

$$I_{\text{tot}} = \sum_{z: x \sim z} i(x, z) = C_{\text{eff}}(x, y),$$

ou, en développant :

$$(2.2.6) \quad C_{\text{eff}}(x, y) = \lambda(x) \left(1 - \sum_{z: x \sim z} \frac{c(x, z)}{\lambda(x)} v(z) \right).$$

EXEMPLE 2.4. Réduire le circuit électrique suivant (où les conductances sont toutes unitaires) pour trouver $C_{\text{eff}}(x, y)$.



SOLUTION. En appliquant les postulats, on obtient la chaîne de réductions et de transformations présentée à la figure 2.9 (tirée de LYONS, PERES)

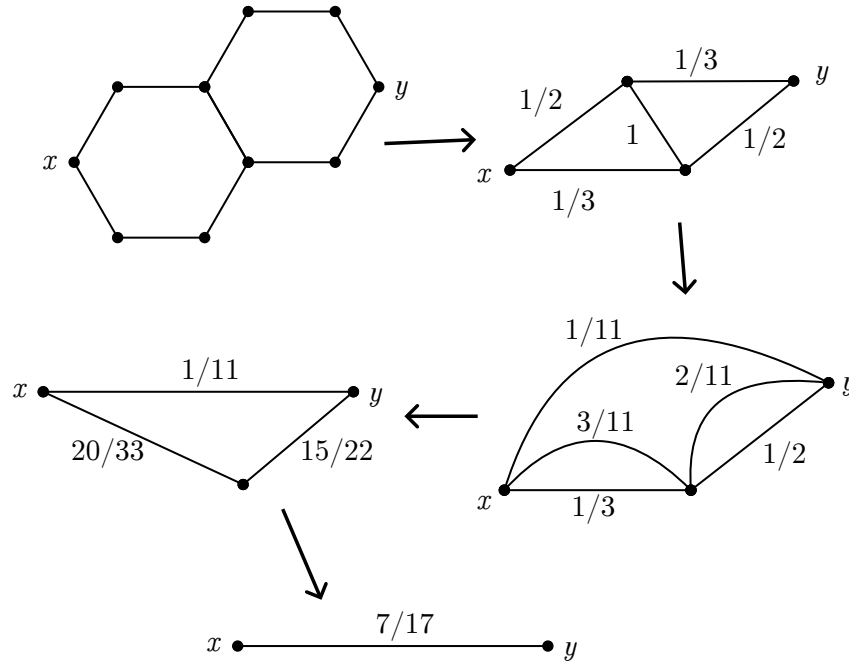


FIGURE 2.9 – Chaîne de réductions du circuit électrique de l'exemple 2.4.

On a donc $C_{\text{eff}}(x, y) = 7/17$.

Et les marches aléatoires ? Revenons-y. Qu'est-ce que tout ça a à voir avec les marches aléatoires, donc ?

Eh bien c'est simple : il y a une connexion directe. Si on définit une marche aléatoire $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ sur les états S avec le graphe $G = (S, E)$ de notre circuit, et $c : S^2 \rightarrow \mathbb{R}^+$ la fonction qui donne les conductances pour chaque composante de circuit, on peut choisir les

probabilités de transition suivante : pour tous $x, y \in S$,

$$(2.2.7) \quad P_{x,y} = \frac{c(x,y)}{\lambda(x)}.$$

Avec ces notations, d'une part, on a déjà vu que le potentiel satisfait le système d'équations

$$(2.2.8) \quad \begin{cases} v(z) = \sum_{w: z \sim w} P_{z,w} v(w) \quad \forall z \neq x, y \\ v(x) = 1 \\ v(y) = 0. \end{cases}$$

Or, d'autre part, si on note $u_z = \mathbb{P}_z \{ \tau_x < \tau_y \}$, on a que les u_z satisfont exactement le même système d'équations (Lemme 1.2). Puisque la solution à ce système est unique, on a la correspondance suivante :

$$(2.2.9) \quad \mathbb{P}_z \{ \tau_x < \tau_y \} = v(z).$$

N'est-ce pas merveilleux ? Mais ce n'est pas tout – on peut utiliser la seconde partie du lemme 1.2 pour montrer que

$$\begin{aligned} \mathbb{P}_x \{ \tau_y < \tau_x^+ \} &= 1 - \mathbb{P}_x \{ \tau_x^+ < \tau_y \} \\ &= 1 - \sum_{z: x \sim z} P_{x,z} v(z) \\ &= \frac{C_{\text{eff}}(x,y)}{\lambda(x)}. \end{aligned}$$

En d'autres termes, la conductance effective du circuit entre x et y correspond (à un facteur $\lambda(x)$ près) à la probabilité, partant de x , de toucher y avant de revenir en x ! N'est-ce pas beau ?

Graphes infinis. Nous revoici maintenant face à notre problème : on souhaite identifier si la marche aléatoire en $d = 2$ dimensions est récurrente ou transiente. Comment faire ?

On vient de voir que lorsqu'on met une tension unitaire en x et en y , $I_{\text{tot}} = C_{\text{eff}}(x,y)$ correspond à la probabilité de toucher l'état y avant de revenir en x lorsqu'on part de x ; ça, ça nous donne une petite idée...

Plutôt que de considérer le graphe \mathbb{Z}^2 au complet, on va d'abord considérer une « boîte » $B_n = (\mathbb{Z} \cap [-n, n])^2$, de $(2n+1) \times (2n+1)$ sommets. On va « contracter » tout le reste du graphe en un point, qu'on va appeler ∞_n (figure 2.10).

Les probabilités de transition sont restées les mêmes ; on garde les mêmes conductances, on fait juste toutes les « brancher » sur le même court-circuit !

Bien sûr, si on met une tension unitaire entre $\mathbf{0}$ et ∞_n , on aura nécessairement que

$$C_{\text{eff}}(\mathbf{0}, \infty_n) = 4 \mathbb{P}_{\mathbf{0}} \{ \tau_{\infty_n} < \tau_{\mathbf{0}}^+ \}.$$

La conductance effective $C_{\text{eff}}(\mathbf{0}, \infty_n)$ est donc proportionnelle à la probabilité de sortir de B_n avant de revenir en $\mathbf{0}$.

On prend maintenant la limite lors que n tend vers l'infini ; on note :

$$\frac{1}{4} C_{\text{eff}}(\mathbf{0}, \infty) = \lim_{n \rightarrow \infty} \frac{1}{4} C_{\text{eff}}(\mathbf{0}, \infty_n) = \lim_{n \rightarrow \infty} \mathbb{P}_{\mathbf{0}} \{ \tau_{\infty_n} < \tau_{\mathbf{0}}^+ \}.$$

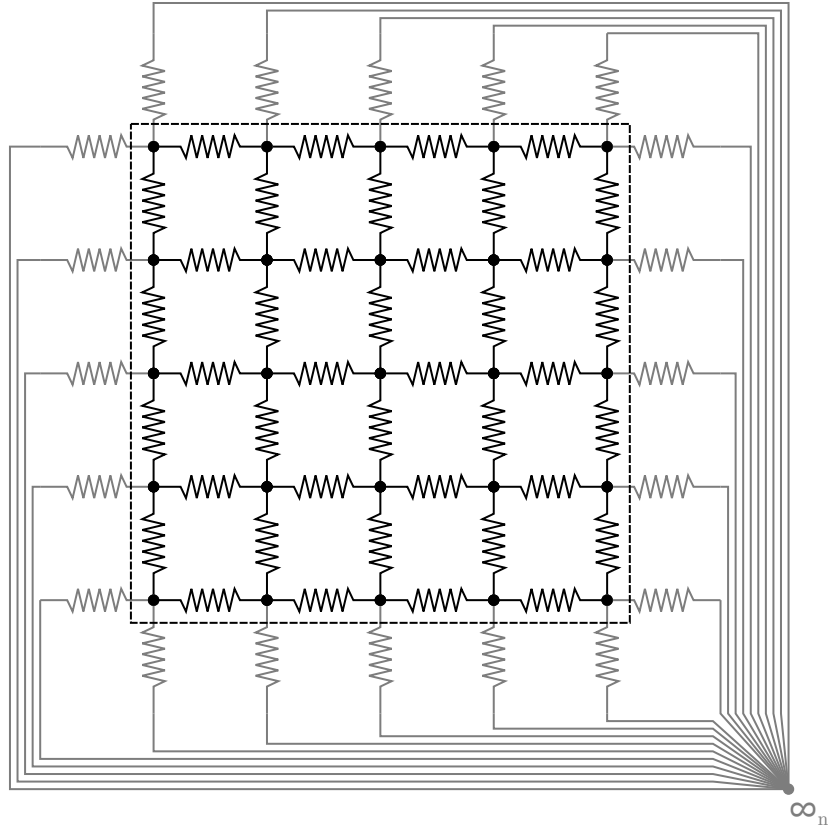


FIGURE 2.10 – La boîte B_n (ici représentée avec $n = 2$). Tous les états à l'extérieur ont été condensés dans l'état ∞_n .

C'est la « conductance effective » entre le nœud $\mathbf{0}$ et l'infini ; ça correspond au courant total qui passe lorsqu'on met une différence de potentiel unitaire entre $\mathbf{0}$ et « l'infini ».

On obtient le critère suivant :

LEMME 2.6. *Dans la marche aléatoire sur un graphe $G = (S, E)$ infini avec probabilités données par les conductances $c : S^2 \rightarrow \mathbb{R}^+$ (symétrique), l'état x est récurrent si et seulement si*

$$C_{\text{eff}}(x, \infty) = 0.$$

De façon équivalente, l'état x est récurrent si et seulement si

$$R_{\text{eff}}(x, \infty) = +\infty.$$

Dans ce cas, le circuit infini est tellement résistant au courant que celui-ci ne peut passer. À l'inverse, si $C_{\text{eff}}(x, \infty) > 0$, le processus est transient !

Rayleigh et Nash-Williams. Pour estimer C_{eff} , on peut utiliser deux outils.

LEMME 2.7 (Monotonie de Rayleigh). *Soit $G = (S, E)$ un circuit électrique avec conductances c . Supposons qu'on a aussi les conductances c' avec $c'(x, y) \geq c(x, y)$ pour tout x, y ; alors, la conductance effective obéit à la même inégalité : pour tous $x, y \in S$,*

$$C'_{\text{eff}}(x, y) \geq C_{\text{eff}}(x, y).$$

À l'inverse, la résistance efficace obéit l'inégalité opposée :

$$R'_{\text{eff}}(x, y) \leq R_{\text{eff}}(x, y).$$

En particulier, si on enlève des arêtes (mettre des conductances à 0), la conductance effective diminue, la résistance effective augmente. Si on rajoute des arêtes, la conductance effective augmente, et la résistance effective diminue.

Pour notre second outil, on a besoin d'une autre définition :

DÉFINITION 2.3 (Coupure). Soit $G = (S, E)$ un graphe infini. Une **coupure** pour $x \in S$ est un ensemble d'arêtes E_0 tel que tout chemin infini partant de x doit traverser au moins une fois une arête de E_0 .

Une coupure s'appelle comme ça parce qu'il suffit d'en couper l'ensemble des arêtes pour déconnecter x de l'infini.

Le prochain lemme est un résultat amusant :

LEMME 2.8. Soit E_0 une coupure de x dans le graphe $G = (S, E)$ avec les conductances $c : S^2 \rightarrow \mathbb{R}^+$ symétriques.

Alors, on a que

$$C_{\text{eff}}(x, \infty) \leq \sum_{e \in E_0} c(e);$$

la conductance effective est au plus égale à la somme des conductances de la coupure E_0 .

DÉMONSTRATION. Intuitivement, la totalité du courant qui va jusqu'à l'infini doit passer par les arêtes de E_0 ; si la coupure est minimale, on a

$$C_{\text{eff}}(x, \infty) = \sum_{e \in E_0} i(e) = \sum_{e \in E_0} c(e)(\delta v(e)),$$

où $\delta v(e)$ est la différence de potentiel aux bornes de e . Mais $\delta v(e) \leq 1$, ce qui complète la « preuve »... \square

En fait, on peut même passer ce résultat plus loin ; c'est le théorème de Nash-Williams :

THÉORÈME 2.1 (Nash-Williams). Soient $E_1, E_2, E_3, \dots, E_n$ une famille de coupures disjointes de x . Alors, on doit avoir que

$$R_{\text{eff}}(x, \infty) \geq \sum_{i=1}^n \left(\sum_{e \in E_i} c(e) \right)^{-1}.$$

DÉMONSTRATION. Encore une fois, très loussement, ça fonctionne parce que si on fait des coupures successives, et que chaque coupure a une « résistance effective » égale à l'inverse de la somme des conductances, alors les coupures successives sont comme « en série », et leurs résistances effectives s'additionnent. \square

La marche sur \mathbb{Z}^2 est récurrente. Enfin, on lève le suspense. La marche aléatoire sur \mathbb{Z}^2 est récurrente. Pour le prouver, on va montrer que la résistance effective entre $\mathbf{0}$ et l'infini est elle-même infinie.

Pour ça, on retourne à nos boîtes B_n ; les arêtes qui en sortent, on les appellera E_n . On a bien sûr que $|E_n| = 4(2n + 1)$; chaque côté a $2n + 1$ sommet sur le bord et une arête qui

sort par ce côté-là pour chacun de ces sommets. Bien sûr, donc :

$$\sum_{e \in E_n} c(e) = |E_n| = 4(2n+1),$$

puisque l'on se souvient que dans notre graphe, les probabilités de transition sont toutes égales – donc les arêtes ont toutes la même conductance : $c(e) = 1$ pour tout e .

Mais, par Nash-Williams, pour tout $n \in \mathbb{N}$,

$$\begin{aligned} R_{\text{eff}}(x, \infty) &\geq \sum_{i=1}^n \left(\sum_{e \in E_i} c(e) \right)^{-1} \\ &= \sum_{i=1}^n \frac{1}{4(2n+1)}. \end{aligned}$$

En prenant une limite lorsque n tend vers l'infini, on trouve finalement que

$$R_{\text{eff}}(x, \infty) = +\infty,$$

ce qui signifie qu'aucun courant ne peut passer, et que la marche aléatoire est récurrente. \square

2.2.3. Le cas $d \geq 3$. De façon générale, on peut définir la marche aléatoire à d dimensions comme étant la marche aléatoire sur le circuit électrique dans \mathbb{Z}^d avec les conductances toutes à 1 ; la probabilité de transitionner vers l'un des nœuds voisins est alors $1/(2d)$.

On peut appliquer les mêmes outils qu'on vient de développer (et il en faut encore plus) pour démontrer le théorème suivant :

THÉORÈME 2.2 (Pólya). *La marche aléatoire dans \mathbb{Z}^d est récurrente pour $d \leq 2$; elle est transiente pour $d \geq 3$.*

Ça va même plus loin ; on fait la définition suivante :

DÉFINITION 2.4 (Dimension de Hausdorff). Soit $G = (S, E)$ un graphe infini. On note $B_n(x)$ la boule centrée en x de rayon n ; $B_n(x) = \{y \in S : d(x, y) \leq n\}$. On dit que le graphe G a la dimension $d \geq 0$ (ça peut être fractionnaire ou pas) si et seulement si pour tout $x \in S$ on a que

$$\lim_{n \rightarrow \infty} \frac{\log |B_n(x)|}{\log n} = d.$$

Par exemple, si $|B_n(x)|$ croît comme n^2 , on est en dimension 2. Si $|B_n(x)|$ croît comme n^3 , on est en dimension 3. Mais la puissance n'est pas obligée d'être entière.

Avec la notion de dimension de Hausdorff, on a le théorème suivant, plus général :

THÉORÈME 2.3 (Pólya). *La marche aléatoire sur un graphe $G = (S, E)$ infini de dimension de Hausdorff d avec conductances unitaires est récurrente si et seulement si $d \leq 2$; si $d > 2$, la marche est transiente.*

On peut avoir une vague idée de comment l'argument fonctionne si on utilise des coupures à la frontières des boules $B_n(x)$, et qu'on remplace l'inégalité dans Nash-Williams par un \sim ; ce n'est pas exact, mais ça donne quand même une idée de ce qui se passe : Si $|B_n(x)| = O(n^d)$, alors $|E_n| = O(n^{d-1})$, on aurait « $R_{\text{eff}}(x, \infty) \sim \sum_{i=1}^n \frac{1}{i^{d-1}}$ » ; ça converge pour $d > 2$, et ça diverge sinon !

2.3. Exercices

2.3.1. Exercices sur les processus de Galton-Watson.

EXERCICE 2.1. Montrer que si $p(0), p(1) > 0$ et $p(0) + p(1) < 1$, on a forcément que :

$$\mathbb{P}_1 \{ \tau_0 < +\infty \} > \frac{p(0)}{1 - p(1)}.$$

EXERCICE 2.2. Donner $\lim_{t \rightarrow \infty} \mathbb{P}_1 \{ Z_t = k \}$.

EXERCICE 2.3 (LESSARD, ex. 1.8). Dans un processus de branchement avec générations séparées, chaque individu de chaque génération, indépendamment de tous les autres, produit k individus de la génération suivante avec probabilité $(1 - p)p^k$, pour $k \geq 0$. Déterminer la probabilité d'extinction :

- (a) si $0 < p < 1/2$;
- (b) si $p = 1/2$;
- (c) si $1/2 < p < 1$.

Dans les cas où l'extinction est certaine, déterminer le nombre total moyen de descendants d'un individu.

EXERCICE 2.4 (LESSARD, ex. 1.9). Supposons que, dans un processus de branchement, le nombre d'individus produits par chaque individu de chaque génération a comme fonction génératrice $\psi(s) = \frac{1}{2} + \frac{3}{2}s^3$. Quelle est la probabilité d'extinction de la population si il y a 2 individus à la première génération ?

EXERCICE 2.5 (LESSARD, ex. 1.10). Supposons que, dans un processus de branchement, le nombre d'individus produits par chaque individu de la génération n a comme fonction génératrice $\psi(s) = \frac{1}{4} + \frac{3}{4}s$ si n est pair, et $\phi(s) = \frac{1}{4} + \frac{3}{4}s^2$ si n est impair. Quelle est la probabilité d'extinction de la population si il y a un seul individu à la génération 0 ?

EXERCICE 2.6 (LESSARD, ex. 1.11). Une plante survit d'un printemps au suivant avec probabilité $3/4$ et, dans ce cas, elle produit le printemps suivant 0, 1 ou 2 autres plantes identiques à elle-même avec la même probabilité $1/3$. Une plante peut donc survivre et se reproduire plusieurs printemps de suite. On suppose une seule plante au départ, et on considère le nombre total de plantes à chaque printemps suivant.

- (a) L'extinction est-elle certaine ?
- (b) Si elle ne l'est pas, quelle est sa probabilité ?

EXERCICE 2.7 (LESSARD, ex. 1.12). On considère un processus de branchement où la distribution du nombre d'enfants suit une loi binomiale de paramètres $p = 1/2$, $n = 4$. À la génération initiale, il y a deux individus.

- (a) A-t-on extinction certaine de la population ?

- (b) Sinon, quelle est la probabilité d'extinction ?

EXERCICE 2.8 (LESSARD, ex. 1.36). Dans un processus de branchement, le nombre d'enfants est de 1, 2 ou 3, chacun avec probabilité p , et de 0 avec probabilité $1 - 3p$.

Déterminer

- (a) la condition sur p pour que la descendance d'un individu ne s'éteigne pas avec probabilité 1 ;
 (b) la probabilité d'extinction de la descendance d'un individu sous cette condition.

2.3.2. Exercices sur les marches aléatoires.

EXERCICE 2.9. Soit $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ la *marche aléatoire paresseuse* sur \mathbb{Z} : il s'agit d'une chaîne de Markov homogène à temps discret sur les états $S = \mathbb{Z}$, avec les probabilités de transition en un pas suivantes :

$$P_{i,i+1} = p, P_{i,i-1} = q, P_{i,i} = r, ,$$

et $p + q + r = 1$. On dit que la marche est « paresseuse » parce qu'elle peut rester au même état avec une probabilité positive.

- (a) Soit $Y_t = X_t - X_{t-1}$. Montrer que les Y_t sont indépendantes et identiquement distribuées pour tout $t \geq 1$.
 (b) Calculer $\mathbb{E}[Y_t]$. Déterminer pour quelles valeurs de p, q et r le processus est récurrent ou transient.

EXERCICE 2.10. Soient $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ et $\mathbf{Y} = (Y_t)_{t \in \mathbb{Z}^+}$ deux marches aléatoires symétriques sur \mathbb{Z} ($p = q = \frac{1}{2}$).

On note

$$U_t = \frac{X_t + Y_t}{2}, \quad V_t = \frac{X_t - Y_t}{2},$$

et finalement

$$\mathbf{Z}_t = (U_t, V_t).$$

- (a) Montrer que $\mathbf{Z} = (Z_t)_{t \in \mathbb{Z}^+}$ est une chaîne de Markov à temps discret sur l'espace des états \mathbb{Z}^2 . Montrer que cette chaîne est irréductible. Donner les probabilités de transition en un pas.
 (b) On écrit

$$N = |\{t \in \mathbb{Z}^+ : X_t = Y_t\}|.$$

Que peut-on dire sur $\mathbb{E}_{(0,0)}[N]$, où $\mathbb{E}_{(i,j)}$ est l'espérance sachant que $X_0 = i, Y_0 = j$?

EXERCICE 2.11. On considère $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ la marche aléatoire simple sur \mathbb{Z}^d – c'est à dire que pour $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \dots, \mathbf{e}_d\}$ la base orthonormale canonique de \mathbb{R}^d , on a les probabilités de transition suivantes pour tout $x \in \mathbb{Z}^d$:

$$P_{x, x \pm \mathbf{e}_i} = \frac{1}{2d}.$$

On note $U_t^{(i)} = X_t \cdot \mathbf{e}_i$, pour $i = 1, 2, 3, \dots, d$; il s'agit de la i ème coordonnée du vecteur X_t :

$$X_t = (U_t^{(1)}, U_t^{(2)}, \dots, U_t^{(d)}).$$

- (a) Montrer que pour tout i , on a que $\mathbf{U}^{(i)} = (U_t^{(i)})_{t \in \mathbb{Z}^+}$ est une marche aléatoire paresseuse sur \mathbb{Z} comme à l'exercice 2.9 avec $p = q = \frac{1}{2d}$ et $r = \frac{d-1}{d}$.
- (b) On suppose $d \geq 3$. Soient i, j fixés et distincts. Est-ce que $\mathbf{U}^{(i)}$ et $\mathbf{U}^{(j)}$ sont indépendants ?
- (c) Est-ce que les processus $(\mathbf{U}^{(i)})_{i=1,2,3,\dots,d}$ forment une famille indépendante ?

EXERCICE 2.12. Soit $G = (S, E)$ où $S = \mathbb{Z}$ et $E = \{\{i, i+1\} : i \in \mathbb{Z}\}$ le graphe habituel pour \mathbb{Z} . On le considère comme un circuit électrique. Montrer que si on pose les conductances $c(i, i+1) = \beta^i$ (pour $\beta > 0$), alors la marche aléatoire sur ce circuit électrique (telle que définie précédemment) avec les probabilités de transition :

$$P_{i,i\pm 1} = \frac{c(i, i \pm 1)}{c(i, i-1) + c(i, i+1)}$$

est une marche aléatoire sur \mathbb{Z} avec $p = \beta/(1+\beta)$ et $q = 1/(1+\beta)$.

EXERCICE 2.13. On considère une marche aléatoire sur l'espace des états $S = \mathbb{Z}^2$, mais avec les probabilités de transition suivantes : pour tout $j \neq 0$:

$$P_{(i,j),(i,j\pm 1)} = \frac{1}{2}.$$

Du reste,

$$P_{(i,0),(i+1,0)} = p, P_{(i,0),(i-1,0)} = q, P_{(i,0),(i,\pm 1)} = \frac{r}{2}, \quad (p+q+r=1)$$

et toutes les autres transitions sont impossibles. La figure 2.11 illustre le graphe des transitions possibles et les probabilités de transition.

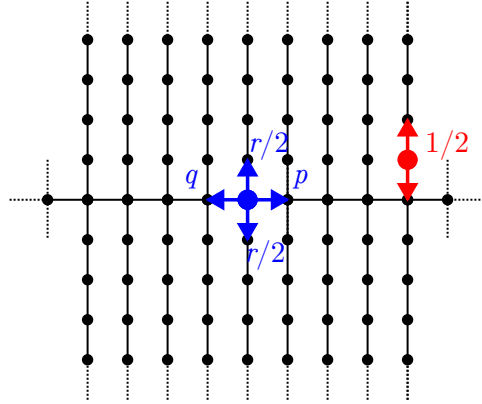


FIGURE 2.11 – Le graphe décrit à l'exercice 2.13.

- (a) Soit $\beta = p/q$. On donne les conductances suivantes ;

$$c((i,0),(i+1,0)) = \beta^i, \quad c((i,j),(i,j\pm 1)) = \frac{\beta^{i-1}(1+\beta)r}{2(1-r)}.$$

Montrer que la marche aléatoire sur le graphe décrit, avec ces conductances, correspond au processus décrit plus haut.

- (b) Dans le cas où $\beta = 1$, utiliser le théorème de Nash-Williams pour montrer que le processus X_t est récurrent.
- (c) Dans le cas où $\beta \neq 1$ (vous pouvez assumer que $\beta > 1$ sans perdre de généralités), le processus est-il récurrent ? Transient ?

Chaînes de Markov à temps continu

Dans les deux premiers chapitres, nous avons discuté des processus stochastiques à temps discret – spécifiquement, le temps était une variable entière non-négative, et on pouvait parler de « pas de temps » entiers. Dans les chapitres qui suivent, nous allons reprendre cette discussion, mais cette fois, nous allons laisser le temps être « continu » au sens topologique – c’est à dire que le temps sera une variable *réelle* non-négative.

Par chance, tout n’est pas perdu ! En effet, une bonne partie de la discussion au début du chapitre 1 ne dépend pas du fait que le temps soit discret et/ou continu. Nous commençons donc ce chapitre en rappelant ces éléments

3.1. Le chapitre 1 à nouveau.

Nous avons commencé notre discussion des processus stochastiques en parlant de l’espace des états S – pour tout de suite, cet espace ne change pas ; il sera encore discret. De plus, nous nous concentrerons encore une fois sur les processus qui ont la propriété de Markov.

3.1.1. Ce qui reste pareil. Dans ce contexte, presque la section 1.4 tient encore la route. Spécifiquement, dans le cas où notre chaîne de Markov est *homogène*, on va récupérer exactement la même notation pour les probabilités de transitions après un intervalle de temps t : si $\mathbf{X} = (X_t)_{t \in \mathbb{R}^+}$ est une chaîne de Markov à temps continu, on a donc :

$$P_{x,y}^{(t)} = \mathbb{P} \{X_t = y \mid X_0 = x\} = \mathbb{P}_0 \{X_t = y\}.$$

La seule différence, c’est qu’ici, le temps t peut être n’importe quel nombre réel.

On rappelle les corollaires 1.1 et 1.2, qui donnent respectivement les propriétés des probabilités de transition et des matrices de probabilités de transition dans le cas où $|S| < +\infty$:

COROLLAIRE (1.1). Soit $\mathbf{X} = (X_t)_{t \geq 0}$ une chaîne de Markov homogène sur l’espace d’états discret S . Les probabilités de transition ont les propriétés suivantes :

i Pour tout $x, y \in S$:

$$(1.4.5) \quad P_{x,y}^{(0)} = \delta_{x,y} = \begin{cases} 1 & \text{si } x = y \\ 0 & \text{si } x \neq y. \end{cases}$$

ii Pour tous $t \geq 0$, $x \in S$:

$$(1.4.6) \quad \sum_{y \in S} P_{x,y}^{(t)} = 1.$$

iii Pour tous $s, t \geq 0$, et tous $x, y \in S$:

$$(1.4.7) \quad P_{x,y}^{(s+t)} = \sum_{z \in S} P_{x,z}^{(s)} P_{z,y}^{(t)}.$$

COROLLAIRE (1.2). Soit $\mathbf{X} = (X_t)_{t \in \mathbb{R}^+}$ une chaîne de Markov homogène sur un espace fini d'états $S = \{1, 2, 3, \dots, n\}$ et $P^{(t)}$ sa matrice de transition en un intervalle de temps t , pour $t \geq 0$. Alors les matrices de probabilités de transition ont les propriétés suivantes :

i

$$(1.4.11) \quad P^{(0)} = \mathbb{I}_n.$$

ii Pour tous $t \geq 0$ et pour tous $1 \leq i \leq n$, on a :

$$(1.4.12) \quad \sum_{j=1}^n P_{i,j}^{(t)} = 1. \quad (\Leftrightarrow P^{(t)} \mathbf{1}_n = \mathbf{1}_n)$$

Autrement dit, pour chaque ligne de $P^{(t)}$, la somme est 1.

iii Soient $s, t \geq 0$. Alors, on a :

$$(1.4.13) \quad P^{(s+t)} = P^{(s)} P^{(t)}.$$

3.1.2. Ce qui change. Jusqu'ici, on voit bien que rien n'a changé – sauf bien sûr que t est une variable réelle plutôt qu'entière. Qu'est-ce qui nous empêche de continuer comme ça ?

Examinons de plus près la section 1.4.1 – on y introduit la matrice de transition « en un pas » :

$$P = P^{(1)}.$$

On y montre également que $P^{(t)} = P^t$ pour tout $t \in \mathbb{Z}^+$. Et... c'est encore vrai !

Mais ce n'est pas très pratique ; en effet, dans le contexte du processus continu, la notion d'un « pas » n'a plus vraiment de signification spéciale. On aimerait avoir une façon plus évidente et intuitive de caractériser les probabilités de transition et la loi de notre chaîne de Markov. Mais pour ça, il faudra travailler un peu plus fort.

3.2. Une description qualitative de X_t .

Soit $\mathbf{X} = (X_t)_{t \in \mathbb{R}^+}$ une chaîne de Markov homogène à temps continu. Pour une réalisation quelconque de ce processus, à quoi ressemblerait le graphe de la courbe (t, X_t) ?

L'espace des états S est toujours *discret*. Puisque le temps t est *continu*, on n'aura pas le choix : le graphe de la courbe (t, X_t) ressemblera à une *fonction en escaliers* ; c'est-à-dire que ce sera une fonction *constante par morceaux*, avec des sauts à des endroits discrets. La figure 3.1 montre un exemple d'un graphe pour un tel processus.

On peut identifier deux « phases » d'évolution du processus (figure 3.2) :

- i.* Le processus passe un certain temps (aléatoire ; disons T) dans l'état x . La loi de T ne peut dépendre de rien d'autre que de l'état x ; en effet, autrement, on contredirait l'homogénéité et/ou la propriété de Markov.
- ii.* Après que ce temps se soit écoulé, le processus transitionne *instantannément* à un état y . Puis on repart à *i.* à l'état y , et ainsi de suite.

L'homogénéité et la propriété de Markov contraignent beaucoup les distributions possibles pour la loi du temps T . On a la proposition suivante :

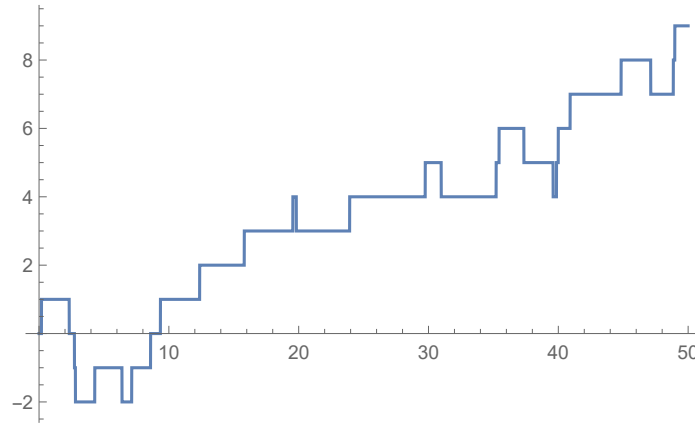


FIGURE 3.1 – Le graphe d’une chaîne de Markov à temps continu. Ici, X_t est un processus de Poisson composé ; nous reparlerons de celui-ci au chapitre 4.

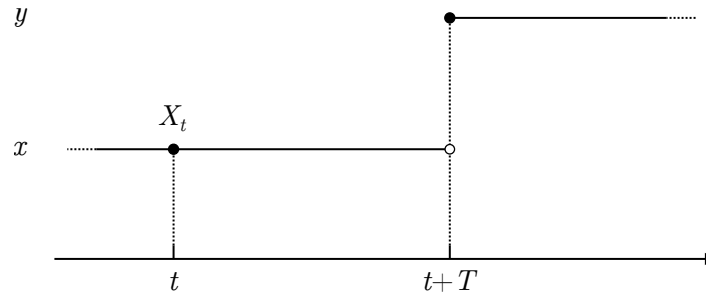


FIGURE 3.2 – Si on est en x au temps t , on y reste pour un temps aléatoire T ; au temps $t + T$, on transitionne instantanément vers un autre état y ; puis, on recommence.

PROPOSITION 3.1. Soit $\mathbf{X} = (X_t)_{t \in \mathbb{R}^+}$ une chaîne de Markov homogène à temps continu. Si on définit

$$T = \inf \{t \geq 0 : X_t \neq x\}$$

le temps passé à l’état x avant de transitionner vers un autre état, alors sous \mathbb{P}_x , T suit une distribution exponentielle : il existe $\lambda \geq 0$ tel que pour tout $u \geq 0$:

$$(3.2.1) \quad \mathbb{P}_x \{T > u\} = e^{-\lambda u}.$$

De plus, λ ne peut dépendre que de x . Si $\lambda = 0$, l’état x est absorbant – on ne transitionnera plus jamais une fois qu’on l’a atteint.

DÉMONSTRATION. On remarque que l’on a l’égalité d’événements suivante :

$$\{T > u\} = \{X_s = x \forall s : 0 \leq s \leq u\};$$

en effet, T est supérieur à u si et seulement si le processus s’est trouvé en x en tout temps entre les temps 0 et u .

Mais alors, on doit avoir que, pour tous $u, h \geq 0$:

$$\begin{aligned}\mathbb{P}_x \{T > u + h \mid T > u\} &= \mathbb{P}_x \{X_s = x \ \forall s : 0 \leq s \leq u + h \mid X_t = x \ \forall t : 0 \leq t \leq u\} \\ &= \mathbb{P}_x \{X_{u+s} = x \ \forall s : 0 \leq s \leq h \mid X_t = x \ \forall t : 0 \leq t \leq u\} \\ &= \mathbb{P}_x \{X_s = x \ \forall s : 0 \leq s \leq h\} \\ &= \mathbb{P}_x \{T > h\};\end{aligned}$$

la seconde égalité est obtenue en remarquant que si on sait déjà que $X_t = x$ pour t entre 0 et u , alors $X_s = x$ pour tous les temps s entre 0 et $u + h$ si et seulement si $X_s = x$ pour tous les temps entre u et $u + h$ – ou, de façon équivalente, que $X_{u+s} = x$ pour s entre 0 et h . La troisième égalité est obtenue en appliquant la propriété de Markov pour retirer les conditions sur le passé et ne garder que la condition sur $X_u = x$, puis l'homogénéité, puisque la loi de X_{u+s} sachant X_u est la même que celle de X_s sachant X_0 .

Mais alors, on doit avoir que pour tous $u, h \geq 0$,

$$\mathbb{P}_x \{T > u + h \mid T > u\} = \mathbb{P}_x \{T > h\},$$

ce qui signifie que T a la propriété d'*absence de mémoire*. Par un théorème élémentaire, on peut donc immédiatement conclure que T suit une loi exponentielle avec un certain paramètre λ .

Par les définitions données, on voit que le paramètre λ ne peut dépendre que de x . \square

Ce que nous venons de montrer, c'est que les temps entre les transitions dans une chaîne de Markov à temps continu sur un espace d'état discret sont de loi exponentielle, et que le taux λ de ces lois dépend potentiellement de l'état x dans lequel on se trouve. On définit formellement les temps de transition et inter-transition, et les taux inter-transition :

DÉFINITION 3.1 (Temps de transition, temps inter-transition, taux inter-transition). Soit $\mathbf{X} = (X_t)_{t \in \mathbb{R}^+}$ une chaîne de Markov homogène à temps continu sur l'espace des états discret S (fini ou dénombrable).

Avec $T_0 = 0$, on va noter le **temps de la k ième transition** :

$$(3.2.2) \quad T_k = \inf \{t \geq T_{k-1} : X_t \neq X_{T_{k-1}}\};$$

On peut également noter le **temps inter-transition** :

$$(3.2.3) \quad U_k = T_k - T_{k-1}.$$

La famille $U = (U_k)_{k \in \mathbb{N}}$ est une famille indépendante. Sachant que $X_{T_{k-1}} = x$, on a que U_k suit une loi exponentielle de paramètre $\lambda_x \geq 0$;

Les $(\lambda_x)_{x \in S}$ sont appelés les **taux inter-transition**.

Si on note $p_{x,y}$ la probabilité de transitionner de l'état x vers l'état y au moment où on effectue notre transition, alors on peut préciser un peu notre description de nos deux « phases »

- i. Au temps T_{k-1} , on arrive à l'état x . On y reste un temps $U_k = T_k - T_{k-1}$. Puisqu'on sait qu'on se trouve en x pendant ce temps, on sait que U_k suit une loi exponentielle de paramètre λ_x .
- ii. À la fin de cet intervalle de temps, on transitionne vers un état y avec probabilité $p_{x,y}$; On arrive donc en y au temps T_k ; on recommence alors.

REMARQUE. On remarque qu'avec ces définitions, le processus $\mathbf{Y} = (X_{T_k})_{k \in \mathbb{Z}^+}$ est une chaîne de Markov à temps discret. Ses probabilités de transition en un pas sont simplement $P_{x,y} = p_{x,y}$.

Le graphe des transitions possibles, les classes d'états et leurs types. De façon analogue aux chaînes de Markov à temps discret, on peut tracer le graphe des transitions possibles pour le processus $\mathbf{X} = (X_t)_{t \in \mathbb{R}^+}$ en plaçant une arête de x vers y si et seulement si $p_{x,y} > 0$.

À noter qu'il ne sert à rien d'inclure une probabilité de transitionner de x vers x – en effet, ça signifierait seulement rester à l'état x encore plus longtemps ; ça n'est aucunement nécessaire – voir l'exercice 3.2. Donc, on pourra toujours assumer que $p_{x,x} = 0$.

Tout le reste des raisonnements développés par rapport aux classes d'équivalence sont les mêmes. On a les mêmes définitions de classes ouvertes et fermées, etc. On peut montrer que si on définit $Y_k = X_{T_k}$, alors la chaîne de Markov homogène à temps discret $(Y_k)_{k \in \mathbb{Z}^+}$ a les mêmes classes d'états et des mêmes types que la chaîne $\mathbf{X} = (X_t)_{t \in \mathbb{R}^+}$. En somme : les problèmes de récurrence et de transience dans le cas continu se réduisent facilement à des problèmes de récurrence et de transience dans le cas discret ; nous n'en parlerons pas beaucoup plus.

3.3. Probabilités de transitions sur un intervalle de temps.

Comme dans le cas des probabilités de transition en temps discret, on va chercher à déterminer les probabilités de transition sur n'importe quel intervalle de temps t ; c'est à dire qu'on aimerait bien connaître la famille $P^{(t)} = (P_{x,y}^{(t)})_{x,y \in S}$.

Pour ce faire, nous allons commencer par nous concentrer sur ce qui se produit dans un petit intervalle de temps $[t, t + h]$.

3.3.1. Les transitions dans un intervalle de temps très petit. La première chose que nous allons montrer, c'est que, en général (sous quelques conditions peu encombrantes), la probabilité qu'il survienne plus d'une transition dans un intervalle de temps de longueur h est une valeur négligeable devant h lorsque h est très petit.

PROPOSITION 3.2. *Soit $\mathbf{X} = (X_t)_{t \in \mathbb{R}^+}$ une chaîne de Markov à temps continu sur l'espace des états discret S (fini ou dénombrable).*

*De plus, on fait l'hypothèse que pour tout état x , les taux des temps inter-transition $(\lambda_y)_{y: x \rightarrow y}$ pour états y adjacents à x sont bornés par $\lambda < \infty$; cela est vérifié immédiatement si $|S| < +\infty$, ou si tout x a un nombre fini d'états adjacents.*¹

Soit T_k le temps de la k ème transition.

Alors, pour tout $x \in S$, on a que

$$\lim_{h \rightarrow 0^+} \frac{\mathbb{P}_x \{T_2 \leq h\}}{h} = 0.$$

DÉMONSTRATION. Pour démontrer cela, nous allons conditionner par l'état X_{T_1} auquel nous avons transitionné lors de la première transition. On sait que $T_1 = U_1$ est de loi exponentielle avec paramètre λ_x ; on a bien sûr que $T_2 = U_1 + U_2$; sachant que $X_{T_1} = y$, alors U_2 est de loi exponentielle avec paramètre λ_y , nous allons temporairement écrire :

$$\mathbb{P}_{x \rightarrow y} \{\cdot\} = \mathbb{P}_x \{\cdot \mid X_{T_1} = y\}.$$

1. Cette condition est très permissive ; tous les cas traités dans ces notes la satisfont sans problème.

Sous la mesure $\mathbb{P}_{x \rightarrow y}$, U_1 et U_2 sont deux variables aléatoires indépendantes de lois exponentielles, respectivement de taux λ_x et λ_y . avec x fixé, on note :

$$\lambda = \sup \{ \lambda_x, \lambda_y : x \rightarrow y \};$$

c'est possible puisque les λ_y tels que y est adjacent à x sont bornés (par hypothèse).

Alors, on a que :

$$\begin{aligned} \mathbb{P}_{x \rightarrow y} \{U_1 + U_2 \leq h\} &= \int_0^h \lambda_x e^{-\lambda_x u} \mathbb{P}_{x \rightarrow y} \{U_2 \leq h - u\} du. \\ &= (1 - e^{-\lambda_x h}) - \lambda_x \int_0^h e^{-\lambda_x u} e^{-\lambda_y (h-u)} du. \\ &= (1 - e^{-\lambda_x h}) - \lambda_x e^{-\lambda_y h} \frac{(1 - e^{-(\lambda_x - \lambda_y)h})}{\lambda_x - \lambda_y}. \end{aligned}$$

Avec le développement en série limitée pour $e^z = 1 + z + O(z^2)$, on trouve finalement que pour $h \rightarrow 0^+$,

$$\begin{aligned} \mathbb{P}_{x \rightarrow y} \{T_2 \leq h\} &= \lambda_x h + O(h^2) - \lambda_x (1 - \lambda_y h - O(h^2))(h + O(h^2)) \\ &= \lambda_x h - \lambda_x h + \lambda_x \lambda_y h^2 + O(h^2) + O(h^3) + O(h^4) \\ &= \lambda_x \lambda_y h^2 + O(h^2). \end{aligned}$$

Pour x fixé, si on prend $\lambda = \sup \{ \lambda_x, \lambda_y : x \rightarrow y \}$, on sait qu'il existe $\epsilon > 0$ et C_ϵ tels que si $h < \epsilon$, on a

$$\mathbb{P}_{x \rightarrow y} \{T_2 \leq h\} \leq (C_\epsilon + \lambda^2)h^2.$$

Donc, lorsque $h < \epsilon$, on a que

$$\begin{aligned} \mathbb{P}_x \{T_2 \leq h\} &= \sum_{y \in S} \mathbb{P}_{x \rightarrow y} \{T_2 \leq h\} p_{x,y} \\ &\leq (C_\epsilon + \lambda^2)h^2 \left(\sum_{y \in S} p_{x,y} \right) \\ &= (C_\epsilon + \lambda^2)h^2. \end{aligned}$$

On peut donc immédiatement conclure que

$$\lim_{h \rightarrow 0^+} \frac{\mathbb{P}_x \{T_2 \leq h\}}{h} = 0.$$

□

COROLLAIRE 3.1. *Soit $\mathbf{X} = (X_t)_{t \in \mathbb{R}^+}$ une chaîne de Markov à temps continu sur l'espace des états discret S (fini ou dénombrable). Alors, si T_k est le temps de la k ième transition, on a que*

$$\lim_{h \rightarrow 0^+} \frac{\mathbb{P} \{ |\{k \in \mathbb{N} : t \leq T_k < t + h\}| \}}{h} = 0.$$

Autrement dit, pour tout intervalle de temps $[t, t + h]$, la probabilité qu'il survienne plus de deux transitions dans cet intervalle est négligeable devant h lorsque h est très petit.

Ce qu'on vient de montrer, c'est que dans un tout petit intervalle de largeur h , la probabilité de transitionner deux fois est négligeable – plus précisément, c'est $o(h)$.

On raisonne donc comme suit : si h est très petit,

$$P_{x,y}^{(h)} = \mathbb{P}_x \{X_h = y\} = \mathbb{P}_x \{T_1 < h \text{ et on transitionne vers } y \text{ et } T_2 > h\} + \mathbb{P}_x \{T_2 \geq h \text{ et } X_h = y\}.$$

Comme on peut approximer qu'il n'y a qu'une transition entre les temps 0 et h (vu que h est petit), le premier terme est simplement :

$$\mathbb{P}_x \{T_1 < h \text{ et on transitionne vers } y\};$$

le second terme est bien sûr borné par $\mathbb{P}_x \{T_2 \leq h\} = o(h)$. Finalement, on a donc que, lorsque h est très petit,

$$P_{x,y}^{(h)} = (1 - e^{-\lambda_x h})p_{x,y} + o(h),$$

et en faisant encore une fois un développement en série limitée pour l'exponentielle, on trouve que, si $y \neq x$:

$$(3.3.1) \quad P_{x,y}^{(h)} = \lambda_x p_{x,y} h + o(h).$$

Bien sûr, si au plus une transition peut se produire dans l'intervalle de temps de longueur h , on doit aussi avoir

$$(3.3.2) \quad P_{x,x}^{(h)} = \mathbb{P}_x \{T_1 > h\} + o(h) = 1 - \lambda_x h + o(h).$$

3.3.2. Les équations de Kolmogorov et le générateur. On va immédiatement utiliser les équations (3.3.1) et (3.3.2) pour obtenir le lemme suivant :

LEMME 3.1. *Soit $\mathbf{X} = (X_t)_{t \in \mathbb{R}^+}$ une chaîne de Markov homogène à temps continu sur l'espace des états discret S (fini ou dénombrable), avec les taux inter-transition $(\lambda_x)_{x \in S}$ et les probabilités de transition $(p_{x,y})_{x,y \in S}$.*

Alors, on a le résultat suivant :

$$(3.3.3) \quad \left. \frac{d}{dt} P_{x,y}^{(t)} \right|_{t=0} = \begin{cases} \lambda_x p_{x,y} & \text{si } x \neq y \\ -\lambda_x & \text{si } x = y. \end{cases}$$

DÉMONSTRATION. Ce résultat suit directement des équations (3.3.1) et (3.3.2) en se souvenant que l'on a toujours, pour toute fonction f dérivable en u

$$\left. \frac{df}{dt} \right|_{t=u} = \lim_{t \rightarrow u} \frac{f(t) - f(u)}{t - u}.$$

□

Cela motive l'introduction de valeurs que nous appellerons les *taux générateurs* :

DÉFINITION 3.2 (Taux générateurs). Soit $\mathbf{X} = (X_t)_{t \in \mathbb{R}^+}$ une chaîne de Markov homogène à temps continu sur l'espace des états discrets S (fini ou dénombrable), avec les taux inter-transition $(\lambda_x)_{x \in S}$ et les probabilités de transition $(p_{x,y})_{x,y \in S}$.

Alors, on définit le **taux générateur** des transitions de x vers y , noté $G_{x,y}$, par :

$$(3.3.4) \quad G_{x,y} = \left. \frac{d}{dt} P_{x,y}^{(t)} \right|_{t=0} = \begin{cases} \lambda_x p_{x,y} & \text{si } x \neq y \\ -\lambda_x & \text{si } x = y. \end{cases}$$

Lorsque $|S| = n < +\infty$, G est une matrice $n \times n$, que l'on appelle le **générateur** de \mathbf{X} .

Les taux générateurs ont la propriété suivante :

PROPOSITION 3.3. Soit $\mathbf{X} = (X_t)_{t \in \mathbb{R}^+}$ une chaîne de Markov à temps continu sur l'espace des états discrets S (fini ou dénombrable), avec les taux générateurs $G = (G_{x,y})_{x,y \in S}$. Alors, pour tout $x \in S$, on a ;

$$(3.3.5) \quad \sum_{y \in S} G_{x,y} = 0.$$

DÉMONSTRATION. En supposant que l'on a les taux inter-transition $(\lambda_x)_{x \in S}$ et les probabilités de transition $(p_{x,y})_{x,y \in S}$, on a forcément que :

$$\begin{aligned} \sum_{y \in S} G_{x,y} &= -\lambda_x + \sum_{y: y \neq x} \lambda_x p_{x,y} \\ &= -\lambda_x + \lambda_x \left(\sum_{y: x \neq y} p_{x,y} \right). \end{aligned}$$

Or, tel que précédemment discuté, on doit avoir $p_{x,x} = 0$; dans ce cas, par la condition de normalisation, on a bien

$$\sum_{y: x \neq y} p_{x,y} = 1,$$

et il suit que

$$\sum_{y \in S} G_{x,y} = -\lambda_x + \lambda_x = 0.$$

□

COROLLAIRE 3.2. Lorsque $|S| = n < +\infty$, G est une matrice $n \times n$ de nombres réels ; la somme de chaque ligne est égale à 0 ; on a donc l'équation matricielle :

$$(3.3.6) \quad G \mathbf{1}_n = \mathbf{0};$$

Autrement dit, $\mathbf{1}_n$ est toujours un élément du noyau de G .

On peut maintenant utiliser la propriété de Markov pour montrer les *équations progressives et régressives de Kolmogorov* :

PROPOSITION 3.4. Soit $\mathbf{X} = (X_t)_{t \in \mathbb{R}^+}$ une chaîne de Markov homogène à temps continu sur l'espace des états discret S (fini ou dénombrable). Soit $G = (G_{x,y})_{x,y \in S}$ la famille des taux générateurs pour les transitions de \mathbf{X} .

i. *L'équation progressive de Kolmogorov est vérifiée pour tous $x, y \in S$:*

$$(3.3.7) \quad \frac{d}{dt} P_{x,y}^{(t)} = \sum_{z \in S} P_{x,z}^{(t)} G_{z,y}.$$

ii. *L'équation régressive de Kolmogorov est vérifiée pour tous $x, y \in S$:*

$$(3.3.8) \quad \frac{d}{dt} P_{x,y}^{(t)} = \sum_{z \in S} G_{x,z} P_{z,y}^{(t)}.$$

DÉMONSTRATION. Par les équations (1.4.5) et (1.4.7) du corollaire 1.1, on a que pour tous $t, h \geq 0$:

$$\begin{aligned} \frac{P_{x,y}^{(t+h)} - P_{x,y}^{(t)}}{h} &= \frac{1}{h} \left(\sum_{z \in S} P_{x,z}^{(t)} P_{z,y}^{(h)} - P_{x,y}^{(t)} \right) \\ &= \sum_{z \in S} \frac{P_{x,z}^{(t)} (P_{z,y}^{(h)} - \delta_{z,y})}{h} \\ &= \sum_{z \in S} P_{x,z}^{(t)} \frac{P_{z,y}^{(h)} - P_{z,y}^{(0)}}{h}; \end{aligned}$$

Par le théorème de convergence dominée, on peut passer la limite à l'intérieur de la somme ; par définition des taux générateurs, on trouve

$$\begin{aligned} \frac{d}{dt} P_{x,y}^{(t)} &= \lim_{h \rightarrow 0^+} \frac{P_{x,y}^{(t+h)} - P_{x,y}^{(t)}}{h} \\ &= \sum_{z \in S} P_{x,z}^{(t)} \left(\lim_{h \rightarrow 0^+} \frac{P_{z,y}^{(h)} - P_{z,y}^{(0)}}{h} \right) \\ &= \sum_{z \in S} P_{x,z}^{(t)} G_{z,y}. \end{aligned}$$

La preuve de l'équation régressive est entièrement analogue. \square

COROLLAIRE 3.3. Si $|S| = n < +\infty$, G est une matrice $n \times n$, ainsi que $P^{(t)}$; on a alors que :

i. l'équation progressive de Kolmogorov devient, sous forme matricielle :

$$(3.3.9) \quad \frac{d}{dt} P^{(t)} = P^{(t)} G$$

ii. l'équation régressive de Kolmogorov devient, sous forme matricielle :

$$(3.3.10) \quad \frac{d}{dt} P^{(t)} = G P^{(t)}$$

3.3.3. Solutions aux équations de Kolmogorov dans le cas fini. Spécifiquement dans le cas où on a un nombre fini d'états, on se souvient que dans le cas des chaînes de Markov à temps discret, on avait réussi à trouver, tout simplement, la solution générale pour la matrice des probabilités de transition en t pas ; il s'agissait de $P^{(t)} = P^t$, où P était la matrice des probabilités de transition en un pas.

En temps continu, c'est un peu plus compliqué ; Pour y arriver, il faut introduire l'exponentielle de matrices.

L'exponentielle de matrices. Soit $A \in \mathbb{R}^{n \times n}$. On introduit l'exponentielle :

$$(3.3.11) \quad \exp A := \mathbb{I}_n + \sum_{k=1}^{\infty} \frac{1}{k!} A^k.$$

Si A est une matrice diagonale avec éléments $a_{1,1}, a_{2,2}, a_{3,3}, \dots, a_{n,n}$, alors $\exp(A)$ est aussi une matrice diagonale, et ses éléments sont $e^{a_{1,1}}, e^{a_{2,2}}, e^{a_{3,3}}, \dots, e^{a_{n,n}}$; il suit que si A

est diagonalisable et que

$$A = QDQ^{-1}$$

avec D la matrice dont la diagonale est constituée des valeurs propres $\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_n$ de A , alors

$$\exp(A) = Q \exp(D) Q^{-1};$$

La matrice Q diagonalise aussi $\exp(A)$, et les valeurs propres de $\exp(A)$ sont $e^{\lambda_1}, e^{\lambda_2}, e^{\lambda_3}, \dots, e^{\lambda_n}$.

PROPOSITION 3.5. Soit $\mathbf{X} = (X_t)_{t \in \mathbb{R}^+}$ une chaîne de Markov homogène à temps continu sur l'espace des états $S = \{1, 2, 3, \dots, n\}$ fini, et soit G le générateur des transitions de \mathbf{X} .

Alors, on a que

$$(3.3.12) \quad P^{(t)} = \exp(tG).$$

DÉMONSTRATION. Dans cette preuve, nous assumerons que la solution aux équations de Kolmogorov est unique ; démontrer cela est difficile et hors du cadre du cours. On assume aussi qu'on peut dériver la série définissant $\exp(tG)$ terme par terme, ce qui n'est pas trivial.

Nous allons simplement vérifier que la solution proposée satisfait les équations de Kolmogorov :

$$\begin{aligned} \frac{d}{dt} \exp tG &= \frac{d}{dt} \sum_{k=1}^{\infty} \frac{1}{k!} t^k G^k \\ &= \sum_{k=1}^{\infty} \frac{k}{k!} t^{k-1} G^k \\ &= \sum_{k=0}^{\infty} \frac{1}{k!} t^k G^{k+1} \\ &= \exp(tG)G, \end{aligned}$$

et de même pour l'équation régressive. Puisque $\exp(0) = \mathbb{I}_n = P^{(0)}$, la solution

$$P^{(t)} = \exp(tG)$$

est l'unique solution à notre système d'équations différentielles. \square

EXEMPLE 3.1. On va faire comme pour les marches aléatoires à temps discret, et on va régler tout de suite la question du cas où $|S| = n = 2$; dans ce cas, le générateur de notre chaîne de Markov prendra toujours la forme :

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

Pour calculer $P^{(t)} = \exp(tG)$, on va commencer par diagonaliser G .

Étape 1 : Trouver les valeurs propres de G . Pour trouver les valeurs propres de G , on calcule :

$$\begin{aligned} 0 &= \det(G - \lambda \mathbb{I}) \\ &= (-\alpha - \lambda)(-\beta - \lambda) - \alpha\beta \\ &= (\alpha + \lambda)(\beta + \lambda) - \alpha\beta \\ &= \lambda^2 + (\alpha + \beta)\lambda \\ &= \lambda(\lambda + \alpha + \beta). \end{aligned}$$

Les valeurs propres sont donc $\lambda_1 = 0$ et $\lambda_2 = -\alpha - \beta$.

Étape 2 : Trouver des vecteurs propres de G . Comme vecteur propre associé à la valeur propre $\lambda_1 = 0$ on peut prendre $\mathbf{v}_1 = \mathbf{1}_2 = (1, 1)$.

Comme vecteur propre associé à la valeur propre $\lambda_2 = -\alpha - \beta$, on doit solutionner :

$$\begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix} \begin{pmatrix} v_{2,1} \\ v_{2,2} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Si on pose $v_{2,1} = 1$, on trouve $v_{2,2} = -\beta/\alpha$; en multipliant par α , on voit donc qu'on peut prendre $\mathbf{v}_2 = (\alpha, -\beta)$.

Étape 3 : Trouver Q et D . On a bien sûr :

$$Q = \begin{pmatrix} 1 & \alpha \\ 1 & -\beta \end{pmatrix}, \quad D = \begin{pmatrix} 0 & 0 \\ 0 & -\alpha - \beta \end{pmatrix},$$

Et on peut vérifier que $QD = GQ$.

Étape 4 : Trouver Q^{-1} . Ici, c'est relativement simple; on se sert (encore une fois) de la formule d'inversion : si

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

alors

$$A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

En l'occurrence, on trouve :

$$Q^{-1} = -\frac{1}{\alpha + \beta} \begin{pmatrix} -\beta & -\alpha \\ -1 & 1 \end{pmatrix} = \frac{1}{\alpha + \beta} \begin{pmatrix} \beta & \alpha \\ 1 & -1 \end{pmatrix}.$$

Étape 5 : Calculer $\exp(tG)$. On a maintenant

$$\begin{aligned} P^{(t)} &= \exp(tG) \\ &= Q \exp(tD) Q^{-1} \\ &= \frac{1}{\alpha + \beta} \begin{pmatrix} 1 & \alpha \\ 1 & -\beta \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & e^{-(\alpha+\beta)t} \end{pmatrix} \begin{pmatrix} \beta & \alpha \\ 1 & -1 \end{pmatrix} \\ &= \frac{1}{\alpha + \beta} \begin{pmatrix} 1 & \alpha \\ 1 & -\beta \end{pmatrix} \begin{pmatrix} \beta & \alpha \\ e^{-(\alpha+\beta)t} & -e^{-(\alpha+\beta)t} \end{pmatrix} \\ &= \frac{1}{\alpha + \beta} \begin{pmatrix} \beta + \alpha e^{-(\alpha+\beta)t} & \alpha - \alpha e^{-(\alpha+\beta)t} \\ \beta - \beta e^{-(\alpha+\beta)t} & \alpha + \beta e^{-(\alpha+\beta)t} \end{pmatrix} \end{aligned}$$

En particulier, on voit que, encore une fois, puisque $\alpha, \beta > 0$ (sinon c'est ennuyeux), on a

$$\lim_{t \rightarrow \infty} P^{(t)} = \frac{1}{\alpha + \beta} \begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix}.$$

3.4. Taux générateurs et temps d'arrivée exponentiels.

À date, nous avons vu les taux générateurs uniquement comme des symboles, définis à partir des taux inter-transition $(\lambda_x)_{x \in S}$ et des probabilités de transition $(p_{x,y})_{x,y \in S}$. Dans cette section, notre objectif sera de montrer que les taux générateurs des transitions ont en fait une interprétation plus naturelle que ça.

3.4.1. Des événements imprévisibles. La façon la plus naturelle de comprendre les taux générateurs est de s'imaginer que les transitions sont causées par l'arrivée imprévisible d'*événements* (au sens colloquial du terme). Des événements de « types » différents peuvent engendrer des transitions vers différents états ; le temps inter-transition correspond alors au temps écoulé lorsque le premier événement imprévisible s'est produit ; l'état vers lequel on transitionne est déterminé par le type d'événement qui s'est produit.

Dans ce contexte, les $G_{x,y}$ représentent les taux moyens auxquels se produisent des « événements » qui occasionnent une transition vers y lorsque l'on est dans l'état x . L'élément $-G_{x,x}$ représente le taux total auquel on quitte l'état x ; on a donc naturellement

$$\sum_{y:x \neq y} G_{x,y} = -G_{x,x}$$

pour tout $x \in S$.

Supposons que dans l'état x , n types d'événements imprévisibles ; chacun de ces événements engendre respectivement une transition vers les états $y_1, y_2, y_3, \dots, y_n$.

On notera T_i le temps avant que ne se produise un événement du type i ; puisque l'événement doit être complètement imprévisible, T_i doit être sans-mémoire – donc être de loi exponentielle, disons avec taux μ_i . Ce taux reflète le nombre moyen d'événements qui se produiront par unité de temps.

La transition surviendra lorsqu'arrivera le premier événement de tous types confondus ; le temps inter-transition T est donc :

$$T = \min \{T_1, T_2, T_3, \dots, T_n\}.$$

On fait l'hypothèse que les événements de différent types surviennent indépendamment les uns des autres ; avec cette hypothèse, le temps inter-transition T a aussi une loi exponentielle :

$$\begin{aligned} \mathbb{P}_x \{T > u\} &= \mathbb{P}_x \{\min \{T_1, T_2, T_3, \dots, T_n\} > u\} \\ &= \mathbb{P}_x \{T_1 > u, T_2 > u, T_3 > u, \dots, T_n > u\}. \\ &= \mathbb{P}_x \{T_1 > u\} \mathbb{P}_x \{T_2 > u\} \mathbb{P}_x \{T_3 > u\} \cdots \mathbb{P}_x \{T_n > u\} \\ &= e^{-\mu_1 u} e^{-\mu_2 u} e^{-\mu_3 u} \dots e^{-\mu_n u} \\ &= e^{-\mu u}, \end{aligned}$$

où $\mu = \sum_{k=1}^n \mu_k$ est le taux inter-transition depuis l'état x ; on a donc $\lambda_x = \mu$.

On calcule la probabilité que le minimum soit T_1 ; si $T' = \min \{T_2, T_3, \dots, T_n\}$, alors $T = \min \{T_1, T'\}$ et $\mu' = \sum_{k=2}^n \mu_k$ est le taux de la loi de T' ; alors $T = T_1$ si et seulement si $T_1 < T'$; par un résultat de probabilité, il s'agit de :

$$PR_x \{T_1 \leq T'\} = \frac{\mu_1}{\mu_1 + \mu'} = \frac{\mu_1}{\mu};$$

on a donc que la probabilité de transition vers y_1 est $p_{x,y_1} = \mu_1/\mu$.

De façon similaire, pour $i = 1, 2, 3, \dots, n$, on a que

$$p_{x,y_i} = \frac{\mu_i}{\mu}.$$

Les éléments générateurs sont donc :

$$G_{x,x} = -\mu; \quad G_{x,y_i} = \mu \cdot \frac{\mu_i}{\mu} = \mu_i.$$

EXEMPLE 3.2. Deux rampes mobiles qui amènent les étudiant-e-s au pavillon Roger-Gaudry de l'Université de Montréal (avant 2018, mettons) fonctionnent indépendamment l'une de l'autre. Chacune d'entre elle se brise à un taux $\beta = 1$ fois par heure.

Cependant, comme il y a un seul technicien, il ne peut réparer qu'une seule rampe à la fois ; lorsqu'au moins une rampe est brisée, le technicien peut réparer $\rho = 2$ rampes par heure en moyenne ; tous les temps de réparation et de fonctionnement sont indépendants.

- Quel est le générateur de la chaîne $\mathbf{X} = (X_t)_{t \in \mathbb{R}^+}$, où X_t est le nombre de rampes en fonctionnement au temps t ?
- Trouver $P^{(t)}$ et donner la limite lorsque t tend vers 0.
- À long terme, quelle proportion de temps les deux rampes fonctionnent-elles ?

SOLUTION. (a) Pour trouver le générateur, nous devons d'abord identifier, pour chaque état,

- Quels événements peuvent survenir.
- À quel taux chacun de ces événements survient-il ?
- Pour chaque type d'événement qui peut survenir : quelle transition cet événement engendre-t-il ?

Alors allons-y :

État 0 : Lorsqu'on est dans l'état 0, aucune des rampes ne fonctionne. Le seul événement qui peut survenir est que le réparateur a fini de réparer l'une des deux rampes ; Il effectue cette réparation à un taux moyen de $\rho = 2$ réparations par heure, et lorsque cet événement survient, on transitionne dans l'état 1.

On a donc $G_{0,1} = \rho$ et $G_{0,0} = -\rho$. Les autres taux générateurs de cette ligne sont nuls.

État 1 : Lorsqu'on est dans l'état 1, il peut survenir deux types d'événements :

- Le réparateur finit de réparer la rampe brisée à un taux de $\rho = 2$ réparation par heure ; dans ce cas, on transitionne vers l'état 2. Donc, $G_{1,2} = \rho$.
- La rampe fonctionnelle se brise à un taux de $\beta = 1$ bris par heure ; dans ce cas, on transitionne vers l'état 0. Donc, $G_{1,0} = \beta$.

Finalement, on doit donc avoir $G_{1,1} = -\rho - \beta$. Les autres taux générateurs de cette ligne sont nuls.

État 2 : Lorsqu'on est dans l'état 2, il peut survenir deux événements :

- La rampe n° 1 se brise avec un taux de $\beta = 1$ bris par heure, ou
- La rampe n° 2 se brise, aussi avec un taux de $\beta = 1$ bris par heure.

Ces deux événements causent tous les deux des transitions vers l'état 1 ; on a donc $G_{2,1} = \beta + \beta = 2\beta$.

Finalement, on doit donc avoir $G_{2,2} = -2\beta$. Les autres taux générateurs de cette ligne sont nuls.

En mettant tout cela ensemble, on trouve que :

$$G = \begin{pmatrix} -2 & 2 & 0 \\ 1 & -3 & 2 \\ 0 & 2 & -2 \end{pmatrix}.$$

(b) On diagonalise G ;

Étape 1 : Trouver les valeurs propres. On a $\lambda_1 = 0$, $\lambda_2 = -2$ et $\lambda_3 = -5$.

Étape 2 : Trouver des vecteurs propres. On peut prendre

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad \mathbf{v}_2 = \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}, \quad \mathbf{v}_3 = \begin{pmatrix} -2 \\ 3 \\ -2 \end{pmatrix}.$$

Étape 3 : Trouver les matrices Q et D . On a :

$$Q = \begin{pmatrix} 1 & -2 & -2 \\ 1 & 0 & 3 \\ 1 & 1 & -2 \end{pmatrix}, \quad D = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -5 \end{pmatrix}.$$

Étape 4 : Trouver Q^{-1} . On trouve :

$$Q^{-1} = \frac{1}{15} \begin{pmatrix} -3 & -6 & -6 \\ 5 & 0 & -5 \\ 1 & -3 & 2 \end{pmatrix}.$$

Étape 5 : Calculer $\exp(tG)$. On a :

$$P^{(t)} = \exp(tG) = \frac{1}{15} \left(\begin{pmatrix} 3 & 6 & 6 \\ 3 & 6 & 6 \\ 3 & 6 & 6 \end{pmatrix} + e^{-2t} \begin{pmatrix} 10 & 0 & -10 \\ 0 & 0 & 0 \\ -5 & 0 & 5 \end{pmatrix} + e^{-5t} \begin{pmatrix} 2 & -6 & 4 \\ 3 & 9 & -6 \\ 2 & -6 & 4 \end{pmatrix} \right)$$

(c) On voit que

$$\lim_{t \rightarrow \infty} P_{x,y}^{(t)} = \pi_y,$$

où

$$\pi_0 = \frac{1}{5}, \quad \pi_1 = \frac{2}{5}, \quad \pi_3 = \frac{2}{5}.$$

On a donc que, à long terme, la proportion de temps passée avec les deux rampes fonctionnelles est de $2/5 = 40\%$ du temps.

3.5. Espérance de temps d'atteinte.

Comme nous l'avons déjà mentionné, nous ne nous intéresserons pas nécessairement à la récurrence et/ou à la transience ; on peut répondre à ces questions pour des chaînes à temps continu en utilisant des chaînes à temps discret tout à fait analogues ; ce n'est donc pas un enjeu particulièrement intéressant.

Toutefois, il peut arriver qu'on veuille obtenir des résultats (dans le cas où une chaîne est récurrente et irréductible, par exemple) sur la distribution des temps d'atteinte ou de retour à un état ; dans ce cas.

Pour rappel, on définit le temps d'atteinte de l'état y :

$$(3.5.1) \quad \tau_y = \inf \{ t \in \mathbb{R}^+ : X_t = y \} ;$$

il s'agit du premier temps où on atteint l'état y . Encore une fois, on peut utiliser le conditionnement par la première transition pour arriver à nos fins. On a :

LEMME 3.2. Soit $\mathbf{X} = (X_t)_{t \in \mathbb{R}^+}$ une chaîne de Markov homogène à temps continu sur l'espace des états S avec taux générateurs $G = (G_{x,y})_{x,y \in S}$; on note $\lambda_x = -G_{x,x}$ le taux inter-transition de l'état x et $p_{x,y} = G_{x,y}/\lambda_x$ la probabilité de transition de l'état x à l'état y .

Pour $x, y \in S$ fixés, on a le système d'équations suivant :

$$(3.5.2) \quad \begin{cases} \mathbb{P}_z \{ \tau_x < \tau_y \} = \sum_{w \in S} p_{z,w} \mathbb{P}_w \{ \tau_x < \tau_y \} \quad \forall z \neq x, y \\ \mathbb{P}_x \{ \tau_x < \tau_y \} = 1 \\ \mathbb{P}_y \{ \tau_x < \tau_y \} = 0. \end{cases}$$

REMARQUE. On constate que les probabilités $\mathbb{P}_z \{ \tau_x < \tau_y \}$ satisfont exactement le même système d'équations que pour le processus $Y_k = X_{T_k}$, à temps discret. Ce n'est pas un hasard; le temps passé à état (donc les taux inter-transitions) n'affectent aucunement la probabilité d'atteindre un état avant un autre.

C'est – encore une fois – la raison pour laquelle nous ne nous éterniserons pas sur ces résultats.

La vraie différence importante interviendra lorsqu'on voudra calculer les *espérance* des temps d'atteinte; en effet, à ce moment-là, on aura besoin d'information sur les temps inter-transitions.

On a le lemme suivant :

LEMME 3.3. Soit $\mathbf{X} = (X_t)_{t \in \mathbb{R}^+}$ une chaîne de Markov homogène à temps continu sur l'espace des états S avec taux générateurs $G = (G_{x,y})_{x,y \in S}$; on note $\lambda_x = -G_{x,x}$ le taux inter-transition de l'état x , et $p_{x,y} = G_{x,y}/\lambda_x$ la probabilité de transition de l'état x à l'état y .

Pour $y \in S$ fixé, on a le système d'équations suivant :

$$(3.5.3) \quad \begin{cases} \mathbb{E}_x [\tau_y] = \frac{1}{\lambda_x} + \sum_{z \in S} p_{x,z} \mathbb{E}_z [\tau_y] \quad \forall x \neq y \\ \mathbb{E}_y [\tau_y] = 0. \end{cases}$$

De plus, si $T_1 = \inf \{ t \in \mathbb{R}^+ : X_t \neq X_0 \}$ est le temps de la première transition, et on définit

$$\tau_x^+ = \inf \{ t > T_1 : X_t = x \}$$

le temps d'atteinte après la première transition de x , alors

$$(3.5.4) \quad \mathbb{E}_x [\tau_x^+] = \frac{1}{\lambda_x} + \sum_{z \in S} p_{x,z} \mathbb{E}_z [\tau_x].$$

DÉMONSTRATION. Pour obtenir ce résultat, on conditionne par la première transition; c'est à dire par X_{T_1} .

On a :

$$\begin{aligned} \mathbb{E}_x [\tau_y] &= \mathbb{E}_x [\mathbb{E}_x [\tau_y \mid X_{T_1}]] \\ &= \sum_{z \in S} p_{x,z} \mathbb{E}_x [\tau_y \mid X_{T_1} = z] \end{aligned}$$

Et bien sûr, si on sait qu'en T_1 , on est au point z , alors

$$\mathbb{E}_x [\tau_y \mid X_{T_1} = z] = \mathbb{E}_x [T_1] + \mathbb{E}_z [\tau_y],$$

par la propriété de Markov et l'homogénéité.²

Or, on a que $\mathbb{E}_x [T_1] = \frac{1}{\lambda_x}$, et donc

$$\begin{aligned} \mathbb{E}_x [\tau_y] &= \sum_{z \in S} p_{x,z} \left(\frac{1}{\lambda_x} + \mathbb{E}_z [\tau_y] \right) \\ &= \frac{1}{\lambda_x} + \sum_{z \in S} p_{x,z} \mathbb{E}_z [\tau_y]. \end{aligned}$$

Bien sûr, si $x = y$, $\mathbb{E}_y [\tau_y] = 0$, puisque $X_0 = y$. On a donc montré le système d'équations (3.5.3). Le raisonnement est similaire pour montrer (3.5.4). \square

EXEMPLE 3.3. En reprenant l'exemple 3.2, calculer l'espérance du temps avant d'avoir un arrêt complet du service des rampes si on commence la journée avec deux rampes fonctionnelles.

SOLUTION. Le générateur pour ce processus est

$$G = \begin{pmatrix} -2 & 2 & 0 \\ 1 & -3 & 2 \\ 0 & 2 & -2 \end{pmatrix}.$$

Si on pose $u_i = \mathbb{E}_i [\tau_0]$, on a le système d'équations suivant :

$$\begin{cases} u_0 = 0 \\ u_1 = \frac{1}{3} + \frac{1}{3}u_0 + \frac{2}{3}u_2 \\ u_2 = \frac{1}{2} + u_1. \end{cases}$$

On trouve donc :

$$u_0 = 0, u_1 = 2, u_2 = \frac{5}{2}.$$

Donc, $\mathbb{E}_2 [\tau_0] = \frac{5}{2}$, et il faut en moyenne 2h 30 minutes avant de se trouver dans une situation de panne complète lorsque les deux rampes fonctionnent.

3.6. Distribution stationnaire et théorème ergodique.

Ah tiens ! On a déjà vu une section avec presque le même titre...

Ça vaut la peine, encore une fois, de commencer avec un petit rappel de ce qu'on avait vu à la section 1.7.2.

2. En fait, ici, T_1 est un temps aléatoire ; *a priori*, la propriété de Markov ne s'applique que si le temps est constant. Toutefois, il s'avère qu'on a aussi que la propriété de Markov fonctionne pour T_1 ; c'est ce qu'on appelle la propriété de Markov *forte*.

3.6.1. Rappel de la section 1.7.2. On suppose un instant que $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ soit une chaîne de Markov homogène à temps discret sur l'espace des états discret S (fini ou dénombrable), et que $x, y \in S$ sont deux états.

D'une part, le théorème ergodique (Théorème 1.1) nous garantit que si l'état y satisfait au moins l'une des conditions suivantes :

- l'état y est transient ;
- l'espérance du temps de retour à y , $\mathbb{E}_y [\tau_y^+]$ est infinie ;
- l'état y est récurrent et apériodique ;

alors, la limite de la suite $P_{x,y}^{(t)}$ existe ; dans ce cas, elle est donnée par :

$$(1.7.8) \quad \lim_{t \rightarrow \infty} P_{x,y}^{(t)} = \begin{cases} \mathbb{P}_x \{ \tau_y < +\infty \} \cdot \frac{1}{\mathbb{E}_y [\tau_y^+]} & \text{si } \mathbb{E}_y [\tau_y^+] < +\infty \\ 0 & \text{sinon ;} \end{cases}$$

nous avons aussi vu que dans tous les cas, la fréquence moyenne des visites à y convergeait vers cette même limite :

$$(1.7.9) \quad \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{u=1}^t P_{x,y}^{(u)} = \begin{cases} \mathbb{P}_x \{ \tau_y < +\infty \} \cdot \frac{1}{\mathbb{E}_y [\tau_y^+]} & \text{si } \mathbb{E}_y [\tau_y^+] < +\infty \\ 0 & \text{sinon.} \end{cases}$$

Nous avons aussi trouvé une astuce algébrique pour trouver facilement ces limites dans certains cas ; en effet, nous avons établi que le vecteur-ligne $\pi = (\pi_x)_{x \in S}$ constituait une *distribution stationnaire* (définition 1.14) si et seulement si elle satisfaisait le système d'équations suivantes :

$$(1) \quad \begin{cases} \pi_y = \sum_{x \in S} \pi_x P_{x,y} & \forall x, y \in S; \\ 1 = \sum_{x \in S} \pi_x; \end{cases}$$

ce même vecteur est une *solution stationnaire* s'il satisfait toutes ces équations sauf la dernière (la condition de normalisation).

Finalement, le théorème 1.2 nous garantissait que dans le cas où \mathbf{X} est une chaîne irréductible et récurrente, et que les espérances des temps de retour sont finies pour tout état y ($\mathbb{E}_y [\tau_y^+] < +\infty$ pour tout $y \in S$), alors *il existe une distribution stationnaire unique*. De plus, nous avons vu que $\pi_x = \mathbb{E}_x [\tau_x^+]^{-1}$, de telle sorte que lorsque la chaîne est apériodique et que la limite existe, on doit forcément avoir $\lim_{t \rightarrow \infty} P_{x,y}^{(t)} = \pi_y$.

Ce que nous allons voir maintenant, c'est comment on obtient des résultats analogues dans le cas continu.

3.6.2. Les distributions stationnaires, version continue. Ce qu'on commence par faire, c'est donner une idée de ce que sont les distributions stationnaires dans le cas où on a une chaîne de Markov continue. Pour ce faire, on remarque d'abord une chose :

Dans le cas discret, si $\pi = (\pi_x)_{x \in S}$ est une solution stationnaire, alors

$$\pi = \pi P,$$

ce qui signifie, par récurrence, que $\pi = \pi P^n$ pour tout n ; en particulier, π est une solution stationnaire si et seulement si $\pi = \pi P^{(n)}$ pour tout $n \in \mathbb{N}$; Ça nous donne donc l'idée de généraliser la définition comme suit :

DÉFINITION 3.3. Soit $\mathbf{X} = (X_t)_{t \in \mathbb{R}^+}$ une chaîne de Markov homogène à temps continu sur l'espace des états discret S (fini ou infini dénombrable).

$\pi = (\pi_x)_{x \in S}$ est une **solution stationnaire** pour la chaîne \mathbf{X} si et seulement si elle satisfait les équations

$$(3.6.1) \quad \pi_y = \sum_{x \in S} \pi_x P_{x,y}^{(t)} \quad \forall y \in S, t \in \mathbb{R}^+.$$

C'est une **distribution stationnaire** si elle satisfait en plus la *condition de normalisation* :

$$(3.6.2) \quad \sum_{y \in S} \pi_y = 1.$$

Bien sûr, lorsque $S = \{1, 2, 3, \dots, n\}$ et qu'on a les matrices de transition sur un intervalle de temps t données par $P^{(t)}$, les équations de stationnarité se réduisent à

$$\pi = \pi P^{(t)} \quad \forall t \in \mathbb{R}^+.$$

Cette définition est une généralisation naturelle de la définition de la stationnarité pour les chaînes à temps discret ; toutefois, elle n'est pas très pratique à vérifier. On remarque tout de suite la proposition suivante :

PROPOSITION 3.6. Soit $\mathbf{X} = (X_t)_{t \in \mathbb{R}^+}$ une chaîne de Markov homogène à temps continu sur l'espace des états discret S (fini ou infini dénombrable). Soit $G = (G_{x,y})_{x,y \in S}$ la famille des taux générateurs pour \mathbf{X} .

Alors, $\pi = (\pi_x)_{x \in S}$ est une solution stationnaire pour \mathbf{X} si et seulement si elle satisfait les équations

$$(3.6.3) \quad \sum_{x \in S} \pi_x G_{x,y} = 0 \quad \forall y \in S.$$

REMARQUE. Encore une fois, dans le cas où $S = \{1, 2, 3, \dots, n\}$ est fini, on peut noter ces équations tout simplement comme l'équation matricielle :

$$\pi G = \mathbf{0},$$

où π est un vecteur-ligne.

DÉMONSTRATION. Nous allons faire la preuve dans le cas où S est fini ; ce sera plus simple, mais l'idée reste la même dans le cas infini.

Dans un sens, la preuve est immédiate ; en effet, si $\pi G = \mathbf{0}$, alors $\pi G^n = 0$ pour tout $n \in \mathbb{N}$, et

$$\begin{aligned} \pi P^{(t)} &= \pi \exp(tG) \\ &= \pi \left(\mathbb{I}_n + \sum_{k=1}^{\infty} \frac{1}{k!} G^k \right) \\ &= \pi + \sum_{k=1}^{\infty} \frac{1}{k!} (\pi G^k) \\ &= \pi, \end{aligned}$$

et on respecte les équations de stationnarité ; π est une solution stationnaire.

Dans l'autre sens, c'est légèrement plus angoissant. On y arrive avec les équations de Kolmogorov ; on remarque, bien sûr, que le vecteur π est constant ; donc :

$$\frac{d}{dt}\pi = \mathbf{0}.$$

Or, par les équations de stationnarité, si π est une solution stationnaire, on doit avoir que

$$\mathbf{0} = \frac{d}{dt}(\pi P^{(t)}) = \pi G P^{(t)},$$

par l'équation régressive de Kolmogorov. Par associativité, on a donc

$$\mathbf{0} = (\pi G) P^{(t)},$$

ce qui doit être vrai pour tout t ; Bien sûr, en particulier, ça doit être vrai pour $t = 0$, et on conclue finalement que $\mathbf{0} = \pi G$. \square

3.6.3. Le théorème ergodique, version continue. Maintenant, nous allons voir un premier résultat assez surprenant : contrairement au cas discret, où il pouvait arriver qu'une distribution limite n'existe pas, dans le cas des chaînes de Markov homogènes à temps continu, la distribution-limite existera *toujours*.

THÉORÈME 3.1. *Soit $\mathbf{X} = (X_t)_{t \in \mathbb{R}^+}$ une chaîne de Markov homogène à temps continu sur l'espace des états discret S (fini ou infini dénombrable).*

i. *Pour tout point de départ $x \in S$, la distribution-limite existe ; on peut noter*

$$\lambda_{x,y} = \lim_{t \rightarrow \infty} P_{x,y}^{(t)};$$

ii. *La famille $\lambda_x = (\lambda_{x,y})_{y \in S}$ est une solution stationnaire (potentiellement triviale) pour tout $x \in S$.*

DÉMONSTRATION. ii Ce sera plus simple de commencer par montrer que si on suppose que les limites existent pour tous $x, y \in S$, alors elles sont toutes des solutions stationnaires.

Si on suppose que $\lambda_{x,y} = \lim_{t \rightarrow \infty} P_{x,y}^{(t)}$, alors bien sûr puisque $\lambda_{x,y}$ est une constante, sa dérivée par rapport à t est nulle, et on doit avoir :

$$\lim_{t \rightarrow \infty} \frac{d}{dt} P_{x,y}^{(t)} = 0.$$

Mais par l'équation progressive de Kolmogorov, on obtient :

$$\begin{aligned} 0 &= \lim_{t \rightarrow \infty} \sum_{z \in S} P_{x,z}^{(t)} G_{z,y} \\ &= \sum_{z \in S} \left(\lim_{t \rightarrow \infty} P_{x,z}^{(t)} \right) G_{z,y} \\ &= \sum_{z \in S} \lambda_{x,z} G_{z,y}, \end{aligned}$$

En particulier, donc, on trouve que la distribution $\lambda_x = (\lambda_{x,y})_{y \in S}$ est stationnaire pour tout $x \in S$.

i Il ne reste plus qu'à montrer que les limites existent. Pour ce faire, nous allons considérer des chaînes discrétisées.

Notons $\mathbf{Y}^{(\delta)} = (Y_n^{(\delta)} = X_{n+\delta})_{n \in \mathbb{Z}^+}$; pour tout $\delta \in \mathbb{R}^+$, $\mathbf{Y}^{(\delta)}$ est une chaîne de Markov homogène à temps discret dont les probabilités de transition en un pas est $P_{x,y} = P_{x,y}^{(1)}$ correspondent aux probabilités de transition sur un intervalle de longueur 1 pour \mathbf{X} .

En l'occurrence, pour tout $x \in S$, $P_{x,x} > e^{G_{x,x}} > 0$, et tous les états dans S sont apériodiques pour la chaîne $\mathbf{Y}^{(\delta)}$; par le théorème ergodique, on a donc que pour tout état de départ $x \in S$, la distribution limite existe; notons

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

Par un argument similaire à ce que l'on a fait plus haut, on constate que $\lambda_{x,y}(0)$ doit être une solution stationnaire; on doit avoir

$$\sum_{z \in S} \lambda_{x,z}(0) G_{z,y} = 0$$

pour tout $x, y \in S$.

Par les équations de Chapman-Kolmogorov, on trouve que

$$P_{x,y}^{(n+\delta)} = \sum_{z \in S} P_{x,z}^{(n)} P_{z,y}^{(\delta)},$$

et en prenant la limite de part et d'autre, on voit que

$$\lambda_{x,y}(\delta) = \lim_{n \rightarrow \infty} P_{x,y}^{(n+\delta)} = \sum_{z \in S} \lambda_{x,z}(0) P_{z,y}^{(\delta)}.$$

Or en dérivant par δ , on trouve que

$$\begin{aligned} \frac{d}{d\delta} \lambda_{x,y}(\delta) &= \sum_{z \in S} \lambda_{x,z}(0) \left(\frac{d}{d\delta} P_{z,y}^{(\delta)} \right) \\ &= \sum_{z \in S} \sum_{w \in S} \lambda_{x,z}(0) G_{z,w} P_{w,y}^{(\delta)} \\ &= \sum_{w \in S} \left(\sum_{z \in S} \lambda_{x,z}(0) G_{z,w} \right) P_{w,y}^{(\delta)} \\ &= 0. \end{aligned}$$

On conclue donc que $\lambda_{x,y}(\delta)$ est constante en δ ; toutes les limites sont égales, et on peut conclure que la limite $\lim_{t \rightarrow \infty} P_{x,y}^{(t)} = \lambda_{x,y}$ existe bel et bien. □

Ce théorème n'est *a priori* pas d'une très grande aide dans le cas où de multiples solutions stationnaires linéairement indépendantes existent; toutefois, on peut formuler un certain corollaire :

COROLLAIRE 3.4. *Soit $\mathbf{X} = (X_t)_{t \in \mathbb{R}^+}$ une chaîne de Markov homogène à temps continu sur l'espace des états discret S (fini ou dénombrable); on suppose que \mathbf{X} est irréductible.*

Alors, toutes les solutions stationnaires non-triviales sont linéairement dépendantes (c'est-à-dire qu'elles sont des multiples les unes des autres). En particulier, on distingue deux cas :

- i. Soit il existe une unique distribution stationnaire; dans ce cas, soit $\pi = (\pi_x)_{x \in S}$ cette distribution stationnaire (c'est-à-dire la solution stationnaire qui satisfait aussi $\sum_{x \in S} \pi_x = 1$). Alors, pour tout $x, y \in S$,

$$(3.6.4) \quad \lim_{t \rightarrow \infty} P_{x,y}^{(t)} = \pi_y.$$

- ii. Soit il n'existe aucune distribution stationnaire parce que pour toute solution stationnaire $\pi = (\pi_x)_{x \in S}$, on a que $\sum_{x \in S} \pi_x = +\infty$; dans ce cas, pour tout $x, y \in S$,

$$(3.6.5) \quad \lim_{t \rightarrow \infty} P_{x,y}^{(t)} = 0.$$

DÉMONSTRATION. La preuve assumera que $S = \{1, 2, 3, \dots, n\}$ est fini; encore une fois, c'est pour nous simplifier un peu la tâche, mais l'idée resterait la même, moyennant des embêtements techniques qui nous dépassent un peu.

Pour montrer qu'il existe une unique solution stationnaire (à constante multiplicative près), on commence par remarquer que si $\pi = (\pi_x)_{x \in S}$ est une solution stationnaire, alors en particulier, $\pi = \pi P$ où $P = \exp(G)$; or, on peut immédiatement voir que, puisque \mathbf{X} est irréductible, on doit avoir que tous les éléments de la matrice P sont non-nuls.

À partir de là, la preuve qu'il existe une unique solution stationnaire à constante près est identique à la preuve faite dans le théorème 1.2.

Soit donc $\pi = (\pi_x)_{x \in S}$ une solution stationnaire.

- i Soit $\sum_{x \in S} \pi_x = M < +\infty$; dans ce cas, sans perdre de généralité, on peut simplement assumer que $M = 1 - \pi$ est donc l'unique distribution stationnaire.

Parallèlement, on pose $\lambda_{x,y} = \lim_{t \rightarrow \infty} P_{x,y}^{(t)}$; on sait que cette limite existe. On sait aussi que c'est une solution stationnaire. Puisqu'on doit avoir $\sum_{y \in S} P_{x,y}^{(t)} = 1$, en prenant la limite, avec le théorème de convergence dominée, on doit conclure que $\lambda_{x,y} = \pi_y$.

- ii Si $\sum_{x \in S} \pi_x = +\infty$, alors la seule solution stationnaire qui peut encore fonctionner est $\lambda_{x,y} = \lim_{t \rightarrow \infty} P_{x,y}^{(t)} = 0$; dans ce cas, vous aurez compris qu'on ne peut pas appliquer le théorème de convergence dominée aussi facilement...

□

3.6.4. Un exemple. Pour illustrer l'application de ces résultats, nous allons voir quelques exemples.³

EXEMPLE 3.4. Pour aller faire sa carte OPUS donnant droit au tarif réduit, il faut d'abord attendre en file pour faire valider l'attestation qui donne bel et bien le droit à ce tarif. Puis, une fois l'attestation validée, on doit attendre ensuite dans une autre file, afin d'être photographié-e; on reçoit ensuite sa carte identifiée avec sa photo imprimée dessus, et on est libre de partir.

On fait les hypothèses suivantes :

- Les usagers arrivent pour faire valider leur attestation à un taux moyen de 2 par minute, à intervalles aléatoires de loi exponentielle.
- Le processus de validation est d'une durée aléatoire de loi exponentielle, qui dure en moyenne 1/4 de minutes (15 secondes).

3. Ces exemples sont tirés et/ou adaptés de l'ouvrage de Sabin Lessard.

- Le temps requis pour prendre photographier un usager et lui remettre sa carte est un temps aléatoire de loi exponentielle, qui dure en moyenne 1/2 minutes (30 secondes).
- Un usager qui se présente à un comptoir déjà occupé s'impatiente immédiatement et décide qu'il reviendra plus tard.

On considère l'espace des états $S = \{1, 2, 3, 4\}$, selon la convention suivante :

- 1 : Il n'y a aucun usager à la validation ni à la prise de photos ;
- 2 : Il y a un usager à la validation, mais pas à la prise de photos ;
- 3 : Il y a un usager à la prise de photos, mais pas à la validation ;
- 4 : Il y a un usager respectivement à la validation et à la prise de photos.

On note X_t l'état du processus (dans S) au temps t .

- (a) Donner le générateur G pour le processus $\mathbf{X} = (X_t)_{t \in \mathbb{R}^+}$.
- (b) Donner la distribution stationnaire.
- (c) Quelle est la proportion moyenne de temps où le photographe est occupé ?
- (d) Quel est le nombre moyen de comptoirs occupés à long terme ?
- (e) Parmi tous les usagers qui se présentent, quelle proportion d'entre eux recevront leur carte avec photo ?
- (f) Quel est le temps moyen que passera un usager dans ce processus ?

SOLUTION. (a) On raisonne état par état :

- Si on est dans l'état 1, on passe à l'état 2 au taux de 2 par minute lorsqu'un usager arrive.
- Si on est dans l'état 2, il y a un usager au comptoir de validation ; on passe dans l'état 3 à un taux moyen de 4 par minutes puisqu'il n'y a toujours personne à la prise de photos.
- Si on est dans l'état 3, il y a un usager à la prise de photos ; on passe à l'état 0 à un taux de 2 par minutes lorsque la prise de photos se termine avant l'arrivée d'un nouvel usager ; on passe à l'état 4 à un taux de 2 par minutes si il arrive un nouvel usager pendant qu'on est encore à la prise de photos.
- Si on est dans l'état 4, il y a un usager à la prise de photos et un autre à la validation. On passe à l'état 2 à un taux de 2 par minutes en moyenne si la prise de photo se termine en premier ; on passe à l'état 3 à un taux de 4 par minutes en moyenne si la validation se termine en premier.

En tout, voici donc le générateur qu'on obtient.

$$G = \begin{pmatrix} -2 & 2 & 0 & 0 \\ 0 & -4 & 4 & 0 \\ 2 & 0 & -4 & 2 \\ 0 & 2 & 4 & -6 \end{pmatrix}$$

- (b) On résout le système d'équations $\pi G = 0$. Attention ; π est un vecteur ligne qu'on multiplie à G par la gauche ! Les équations sont :

$$\begin{cases} 0 = -2\pi_1 & + 2\pi_3 \\ 0 = +2\pi_1 & - 4\pi_2 & + 2\pi_4 \\ 0 = & + 4\pi_2 & - 4\pi_3 & + 4\pi_4 \\ 0 = & & + 2\pi_3 & - 6\pi_4. \end{cases}$$

En résolvant, on trouve :

$$(\pi_1, \pi_2, \pi_3, \pi_4) = \frac{1}{9}(3, 2, 3, 1).$$

- (c) La proportion moyenne du temps où le photographe est occupé est la proportion moyenne du temps passée dans les états 3 ou 4 ; c'est :

$$\pi_3 + \pi_4 = \frac{4}{9}.$$

- (d) Soit $N(x)$ le nombre de comptoirs qui sont occupés à l'état x ; on a que $N(1) = 0$, $N(2) = N(3) = 1$ et $N(4) = 2$. On cherche $\lim_{t \rightarrow \infty} \mathbb{E}[N(X_t)]$; c'est :

$$N(1)\pi_1 + N(2)\pi_2 + N(3)\pi_3 + N(4)\pi_4 = 0\pi_1 + 1\pi_2 + 1\pi_3 + 2\pi_4 = \frac{7}{9}$$

- (e) On distingue les cas suivants en conditionnant selon l'état dans lequel on se trouve lorsque l'utilisateur arrive.

- Sachant que l'utilisateur est arrivé alors qu'on était à l'état 1 (il n'y a personne), la probabilité que celui-ci reçoive sa carte avec photo est de 1.
- Sachant que l'utilisateur est arrivé pendant qu'on était dans l'état 3 (pendant qu'il y a quelqu'un à la prise de photos), la probabilité que celui-ci reçoive sa carte avec photo correspond à la probabilité qu'on transitionne de 4 (où on est maintenant que l'utilisateur est arrivé à la validation et qu'il y a encore quelqu'un à la prise de photos) vers l'état 2 (la prise de photos s'est libérée avant que notre utilisateur ait fini sa validation) – c'est une probabilité de $2/(2+4) = 1/3$.
- Sachant que l'utilisateur est arrivé pendant que le comptoir de validation était occupé (états 2 ou 4), la probabilité qu'il reçoive sa carte avec photo est 0.
- La probabilité que l'utilisateur arrive pendant qu'on est dans l'état i est π_i .

Donc, la probabilité totale qu'un utilisateur reçoive sa carte avec photo est :

$$1\pi_1 + 0\pi_2 + \frac{1}{3}\pi_3 + 0\pi_4 = \frac{4}{9}.$$

- (f) Encore une fois, on va prendre l'espérance conditionnellement à l'état dans lequel on se trouve quand l'utilisateur arrive.

- Sachant que l'utilisateur arrive alors qu'on est dans l'état 1, celui-ci passe 15 secondes en moyenne à la validation, puis 30 secondes en moyenne à la prise de photos ; c'est donc 45 secondes, ou 3/4 de minutes en moyenne.
- Sachant que l'utilisateur arrive alors qu'on est dans l'état 3, le client va faire tout le processus de validation – donc il va passer au moins 15 secondes en moyenne au comptoir. Puis, il y a deux possibilités :
 - soit la prise de photos s'est libérée entre temps (probabilité 1/3), auquel cas il va passer en moyenne 30 secondes (1/2 minutes) à la prise de photos ;
 - soit la prise de photos ne s'est PAS libérée entre temps (probabilité 2/3), auquel cas il va passer s'en aller tout de suite.

Donc, sachant qu'on est arrivé dans l'état 3, le temps moyen passé dans le système est : $\frac{1}{4} + (\frac{1}{3} \times \frac{1}{2} + \frac{2}{3} \times 0) = \frac{5}{12}$ minutes (25 secondes).

- Sachant que l'utilisateur arrive alors qu'on est dans les états 2 ou 4, il s'en va tout de suite et passe 0 secondes dans le système.

Donc, notre espérance conditionnelle est :

$$\frac{3}{4}\pi_1 + \frac{5}{12}\pi_3 = \frac{7}{18}.$$

3.7. Exercices

EXERCICE 3.1. Soit $\mathbf{X} = (X_t)_{t \in \mathbb{R}^+}$ une chaîne de Markov à temps continu sur l'espace des états discret S (fini ou dénombrable), et soient $(\lambda_x)_{x \in S}$ les taux respectifs des temps inter-transition pour chaque état dans S .

On considère maintenant la restriction de cette chaîne aux temps entiers ; il s'agit d'une chaîne de Markov à temps discret.

Montrer que les probabilités de transition en un pas pour cette nouvelle chaîne satisfont :

$$P_{x,x} \geq e^{-\lambda_x} > 0.$$

Déduire que tous les états d'une telle chaîne sont forcément apériodiques.

EXERCICE 3.2. Soit $\mathbf{X} = (X_t)_{t \in \mathbb{R}^+}$ une chaîne de Markov homogène à temps continu sur l'espace des états discret S (fini ou dénombrable). Soient $(p_{x,y})_{x,y \in S}$ les probabilités de transitionner de l'état x à y pour tous $x, y \in S$ au moment d'une transition.

Supposons que $p_{x,x} > 0$; c'est à dire qu'on a une probabilité non-nulle de transitionner de x vers x au moment de la transition.

- (a) On suppose que N est le nombre de transitions qu'il faut faire pour finalement quitter l'état x . Quelle est la loi de N ? Donner sa fonction de masse et sa fonction génératrice des probabilités $\psi_N(s) = \mathbb{E}[s^N]$.
- (b) Soit T_i le temps passé en x avant la i ème transition (on suppose qu'on est revenus en x à la $(i-1)$ ème transition). Expliquer pourquoi les T_i sont indépendants et identiquement distribués, de loi exponentielle de paramètre λ_x . Quelle est la fonction génératrice des moments $M(s)$ de la distribution des T_i ?
- (c) On note T le temps passé à l'état x avant de transitionner vers un état différent. On a donc

$$T = \sum_{i=1}^N T_i.$$

Montrer que

$$\mathbb{E}[\exp(sT) \mid N = k] = M(s)^k.$$

- (d) On note $M_T(s)$ la fonction génératrice des moments de T . Montrer que

$$M_T(s) = \psi_N(M(s));$$

déduire que T suit aussi une loi exponentielle avec taux

$$\lambda = (1 - p_{x,x})\lambda_x.$$

REMARQUE. Cet exercice illustre la raison pour laquelle il est inutile d'inclure la possibilité que $p_{x,x} > 0$.

EXERCICE 3.3. Soit G le générateur des transitions pour une chaîne de Markov homogène à temps continu sur un espace d'états fini de n états.

Est-ce que $\det(G - \lambda \mathbb{I})$ est toujours factorisable par λ ? Expliquer ?

EXERCICE 3.4. Soit $\mathbf{X} = (X_t)_{t \in \mathbb{R}^+}$ une chaîne de Markov homogène à temps continu avec $G = (G_{x,y})_{x,y \in S}$ sa famille de taux générateurs. Montrer qu'on a, pour tout $y \in S$, le système d'équations :

$$\begin{cases} -\sum_{z \in S} G_{x,z} \mathbb{E}_z[\tau_y] = 1 & \forall x \neq y \\ \mathbb{E}_y[\tau_y] = 0. \end{cases}$$

EXERCICE 3.5 (LESSARD, ex. 2.1). Deux employées d'une maison de courtage reçoivent des appels de clients au sujet d'achat ou de vente de fonds mutuels. Lorsque leurs lignes téléphoniques étaient indépendantes, chacune d'elles était occupé un temps de loi exponentielle avec une espérance de 15 minutes (1/4 d'heure) avec chaque client ; pendant ce temps, tout nouvel appel reçu sur la ligne était rejeté. Il passait en moyenne 60 minutes (1 heure) entre chaque nouvel appel sur la ligne téléphonique.

Depuis une réorganisation du service, lorsqu'un appel est reçu à n'importe lequel des deux numéros de téléphone, il est transféré à l'autre courtière si la première destinataire de l'appel est occupée ; si aucune des deux n'est libre, l'appel est rejeté.

Déterminer le générateur pour le nombre de lignes occupées

- (a) avant la réorganisation du service, et
- (b) après.

EXERCICE 3.6 (LESSARD, ex. 2.2). On suppose qu'un appareil tombe en panne suite à un k ième choc avec probabilité $k^2/9$ pour $k = 1, 2, 3$; dans le cas où il est en panne, il est remplacé par un appareil neuf. Si les chocs surviennent à des intervalles de temps de loi exponentielle à un taux de 9 chocs par semaine, quel est le générateur pour le nombre de chocs subis par l'appareil ?

EXERCICE 3.7 (LESSARD, ex. 2.3). Une chaîne de Markov homogène à temps continu sur les états 0, 1 et 2 a comme générateur :

$$\begin{pmatrix} -3 & 2 & 1 \\ 2 & -4 & 2 \\ 0 & 1 & -1 \end{pmatrix}.$$

Déterminer le temps moyen pour atteindre l'état 2 à partir de l'état 0.

EXERCICE 3.8 (LESSARD, ex. 2.4). Un sous-marin dispose au départ de trois systèmes de navigation et il reste en mer tant qu'au moins deux de ces systèmes fonctionnent. Les temps de fonctionnement de ces systèmes sont indépendants et de loi exponentielle d'espérance 1 an, 1, 5 ans et 3 ans respectivement. Quel est le temps moyen que le sous-marin reste en mer après son départ ?

EXERCICE 3.9 (LESSARD, ex. 2.28). Une grenouille sur le bord d'un étang (état 1) saute dans l'étang (état 2) et vice versa, selon une chaîne de Markov à temps continu dont le générateur est

$$G = \begin{pmatrix} -3 & 3 \\ 2 & -2 \end{pmatrix}.$$

- (a) Déterminer la probabilité à long terme de trouver la grenouille hors de l'étang.
- (b) Déterminer le générateur pour le nombre de grenouilles dans l'étang si 3 grenouilles sautent indépendamment dans et hors de l'étang comme notre première grenouille.

EXERCICE 3.10 (LESSARD, ex. 2.29). Un médecin de garde reçoit des appels à un taux d'environ 1 toutes les 12 heures ; chaque appel nécessite un temps exponentiel de 3 heures en moyenne, indépendant de tout le reste.

Si le médecin commence sa garde le vendredi à 18h, quelle est la probabilité qu'il puisse passer sa soirée de samedi (de 18h à minuit) avec sa famille sans être dérangé ?

EXERCICE 3.11. Soit $\mathbf{X} = (X_t)_{t \in \mathbb{R}^+}$ une chaîne de Markov homogène à temps continu sur l'espace des états discret S , avec des taux inter-transition $(\lambda_x)_{x \in S}$, et soit $\pi = (\pi_x)_{x \in S}$ une distribution stationnaire pour \mathbf{X} .

On note $\mathbf{Y} = (Y_k = X_{T_k})_{k \in \mathbb{Z}^+}$, où Y_k est la position immédiatement après la k ème transition. Les éléments de la matrice des probabilités de transition en un pas de \mathbf{Y} est donnée par $p_{x,y}$.

- (a) Montrer que π satisfait les équations

$$0 = -\pi_y \lambda_y + \sum_{x \in S: x \neq y} \pi_x \lambda_x p_{x,y}.$$

- (b) Dédurre que le vecteur $\tilde{\pi} = (\tilde{\pi}_x)_{x \in S}$, défini par

$$\tilde{\pi}_x = \pi_x \lambda_x,$$

est une solution stationnaire pour la chaîne à temps discret \mathbf{Y} .

EXERCICE 3.12. (*Difficile ; requiert de l'analyse*)

Soit $\mathbf{X} = (X_t)_{t \in \mathbb{R}^+}$ une chaîne de Markov homogène à temps continu sur l'espace des états discret S (fini ou infini dénombrable).

On définit

$$V_y(t) = \frac{1}{t} \int_{u=0}^t \mathbb{1}_{\{X_u=y\}} du.$$

- (a) Montrer que

$$\mathbb{E}_x [V_y(t)] = \frac{1}{t} \int_{u=0}^t P_{x,y}^{(u)} du.$$

- (b) Montrer que si la chaîne \mathbf{X} est irréductible, alors

$$\nu_{x,y} = \lim_{t \rightarrow \infty} \mathbb{E}_x [V_y(t)] = \pi_y,$$

où $\pi = (\pi_y)_{y \in S}$ est l'unique distribution stationnaire.

EXERCICE 3.13 (LESSARD, ex. 2.13). Une chaîne de Markov à temps continu sur les états 0, 1 et 2 possède comme générateur la matrice

$$G = \begin{pmatrix} -2 & 1 & 1 \\ 2 & -4 & 2 \\ 0 & 1 & -1 \end{pmatrix}.$$

Déterminer la fraction moyenne de temps, à long terme, que la chaîne passe à l'état 0.

EXERCICE 3.14 (LESSARD, ex. 2.14). Un courtier d'assurances reçoit des appels de clients à un taux moyen de 4 par heure, à intervalles de longueur aléatoire de loi exponentielle. Chaque conversation téléphonique est d'une durée aléatoire de loi exponentielle et dure en moyenne $1/4$ d'heure. Si un autre appel arrive durant cette période, il est mis en attente jusqu'à la fin de la conversation en cours, mais alors la ligne téléphonique est inaccessible, et tous les appels subséquents sont rejetés jusqu'à ce que la ligne se libère de nouveau.

Un client en attente s'impatience au bout d'un intervalle de temps aléatoire de loi exponentielle d'une durée moyenne d'un quart ($1/4$) d'heure.

Déterminer

- (a) La proportion moyenne de temps où le courtier est au téléphone ;
- (b) la proportion moyenne de clients auxquels le courtier finira par parler ;
- (c) le temps moyen qu'un client passe au téléphone.

EXERCICE 3.15 (LESSARD, ex. 2.17). Des automobilistes se présentent à une station-service au taux de 20 par heure, à des intervalles aléatoires de loi exponentielle, mais n'entrent pas si il y a au moins quatre voitures dans la station. Il y a deux pompes disponibles dans la station, et le temps de service à chacune d'elle est de loi exponentielle d'espérance 6 minutes.

- (a) Quelle est la distribution stationnaire pour le nombre de voitures dans la station ?
- (b) En moyenne, à long terme, combien d'automobilistes sont servis par heure ?

EXERCICE 3.16 (LESSARD, ex. 2.18). Une boutique qui vend des ordinateurs en garde au plus 3 en stock. Les acheteurs se présentent selon à des intervalles aléatoires indépendants de loi exponentielle à un taux de 2 par semaine en moyenne, et ils repartent avec un ordinateur s'il y en a au moins un en stock. Lorsqu'il reste un seul ordinateur, la boutique en commande 2 autres ; elle les reçoit après un temps aléatoire exponentiel d'une durée moyenne d'une semaine.

Déterminer :

- (a) Le nombre moyen d'ordinateurs en stock, à long terme.
- (b) Le nombre moyen d'ordinateurs vendus chaque semaine.

EXERCICE 3.17 (LESSARD, ex. 2.19). Un coiffeur a de la place pour 2 clients seulement dans son salon. Les clients arrivent à des intervalles aléatoires exponentiels indépendants, au taux moyen de 2 par heure, mais ils vont ailleurs si il y a déjà deux clients dans le salon. Si le coiffeur met un temps aléatoire de loi exponentielle pour couper les cheveux d'un client, combien de temps en moyenne doit-il passer par client afin de ne perdre pas plus de $1/7$ de sa clientèle à long terme ?

Processus de naissances et de mort

Au chapitre 3, nous avons fourni le cadre d'analyse général pour les chaînes de Markov homogènes à temps continu sur des espaces d'états discrets ; nous avons également obtenu des solutions complètes pour la distribution marginale du processus dans le cas des chaînes sur des nombres d'états finis.

Dans ce qui suit, nous allons nous pencher sur une certaine catégorie de processus à temps continu simples sur un nombre infini d'états, et nous verrons comment il est possible d'analyser ces problèmes également.

4.1. Le processus de Poisson.

Le plus simple des processus à temps continu est le *processus de Poisson*. Il existe plusieurs façons de décrire ce processus – comme un objet géométrique, ou comme une mesure aléatoire sur un espace...

Dans ce qui suit, nous allons étudier le processus de Poisson en tant que chaîne de Markov à temps continu, évidemment. La prémisse est la suivante :

4.1.1. Une séquence d'événements imprévisibles. On imagine qu'une séquence d'événements d'un certain type se produisent à des intervalles aléatoires indépendants et identiquement distribués, *sans mémoire* (donc de loi exponentielle). L'exemple-type est celui des « clics » d'un compteur Geiger.¹ La figure 4.1a montre un enregistrement des cliquetis d'un compteur Geiger.

Un *processus de Poisson* est un processus qui « compte » simplement le nombre total de ces événements qui sont survenus. Typiquement, on emploiera N_t pour noter le nombre d'événements survenus jusqu'au temps t inclusivement. La figure 4.1b montre le processus N_t correspondant à l'enregistrement illustré au-dessus.

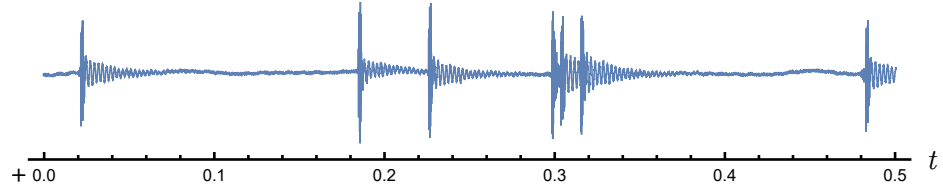
Évidemment, N_t est un nombre entier non-négatif (possiblement nul). Si on note $T_0 = 0$ et

$$T_k = \inf \{t > T_{k-1} : N_t \neq N_{T_{k-1}}\}$$

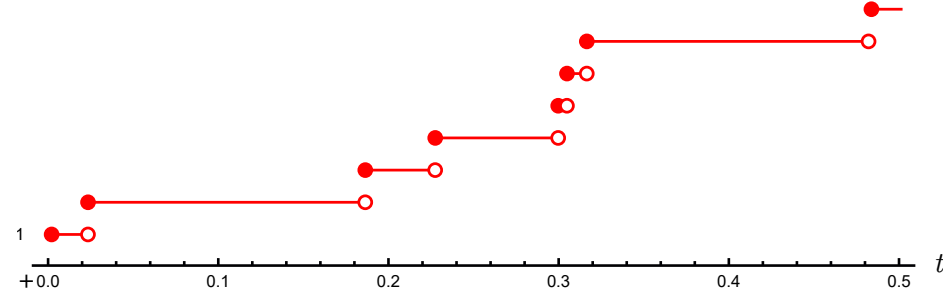
le temps de la k ième transition, alors forcément, on doit avoir que $N_{T_k} = k$ – en effet, les transitions se produisent lorsqu'un événement survient, et quand un événement survient, on ne fait qu'augmenter le compteur.

Puisqu'on fait l'hypothèse que les intervalles $U_k = T_k - T_{k-1}$ sont identiquement distribués et *sans-mémoire* (c'est important), on conclue donc qu'il existe un paramètre $\lambda > 0$ tel que pour les $(U_k)_{k \in \mathbb{N}}$ forment une suite de variables aléatoires indépendantes et identiquement distribuées, de loi exponentielle de taux λ .

1. Un compteur Geiger est un dispositif constitué d'un tube rempli d'hélium aux extrémités duquel est appliqué une tension électrique à l'aide d'électrodes ; lorsqu'une particule radioactive ionise le gaz, la présence momentanée de charges électriques libres provoque une décharge de courant, qui actionne un petit haut-parleur, produisant un déclic. Ces compteurs servent à détecter la présence et l'intensité de rayonnement ionisant.



(A) La forme d'onde de quelques secondes d'un enregistrement d'un compteur Geiger. On peut voir les « clics » très nettement.



(B) Le processus de Poisson N_t comptant le nombre de clics survenus jusqu'au temps t .

FIGURE 4.1 – L'exemple d'un compteur Geiger.

4.1.2. Définitions. On voit que le paramètre λ correspond au taux inter-transition λ_k pour tout $k \in \mathbb{Z}^+$; de plus, les probabilités de transitions sont évidemment données par $p_{k,k+1} = 1$. Le processus N_t peut donc être décrit comme une chaîne de Markov homogène à temps continu. On fait la définition suivante :

DÉFINITION 4.1. Un **processus de Poisson** d'intensité $\lambda > 0$ est une chaîne de Markov homogène à temps discret $\mathbf{N} = (N_t)_{t \in \mathbb{R}^+}$ sur l'espace des états $S = \mathbb{Z}^+$ dont les taux générateurs sont donnés par :

$$(4.1.1) \quad G_{k,k} = -\lambda; \quad G_{k,k+1} = \lambda.$$

4.1.3. Propriétés. Les processus de Poisson sont nommés ainsi en raison de leur plus remarquable propriété :

PROPOSITION 4.1. Soit $\mathbf{N} = (N_t)_{t \in \mathbb{R}^+}$ un processus de Poisson d'intensité $\lambda > 0$. Alors, pour tout $t \in \mathbb{R}^+$, N_t suit une loi de Poisson de paramètre λt .

DÉMONSTRATION. On remarque qu'on a l'égalité entre les événements suivants :

$$\{N_t \leq k\} = \{T_{k+1} > t\};$$

en effet, N_t est inférieur ou égal à k si et seulement si la $k + 1$ ème transition a eu lieu strictement après le temps t .

Or, on a que

$$T_{k+1} = \sum_{i=1}^{k+1} U_i,$$

où les U_i sont des variables aléatoires de loi exponentielle de taux λ indépendantes. Il suit donc que T_{k+1} suit une loi Gamma de paramètres $k+1, \lambda$, donc on a que

$$\begin{aligned}
\mathbb{P}_0 \{N_t \leq k\} &= \mathbb{P}_0 \{T_{k+1} \geq t\} \\
&= \int_{u=t}^{\infty} \frac{\lambda^{k+1}}{k!} u^k e^{-\lambda u} du \\
&= e^{-\lambda t} \int_{v=0}^{\infty} \frac{\lambda^{k+1}}{k!} (v+t)^k e^{-\lambda v} dv \\
&= e^{-\lambda t} \frac{\lambda^{k+1}}{k!} \int_{v=0}^{\infty} \sum_{i=0}^k \frac{k!}{i!(k-i)!} t^i v^{k-i} e^{-\lambda v} dv \\
&= \sum_{i=0}^k e^{-\lambda t} \frac{(\lambda t)^i}{i!} \int_{v=0}^{\infty} \frac{\lambda^{k+1-i}}{(k-i)!} v^{k-i} e^{-\lambda v} dv \\
&= \sum_{i=0}^k e^{-\lambda t} \frac{(\lambda t)^i}{i!}.
\end{aligned}$$

Ici, la dernière égalité est obtenue en vertu de la condition de normalisation pour la densité d'une variable aléatoire de loi Gamma de paramètres $(k+1-i), \lambda$:

$$\int_{v=0}^{\infty} \frac{\lambda^{k+1-i}}{(k-i)!} v^{k-i} e^{-\lambda v} dv = 1.$$

En prenant la différence entre $\mathbb{P}_0 \{N_t \leq k\}$ et $\mathbb{P}_0 \{N_t \leq k-1\}$, on trouve que :

$$\mathbb{P}_0 \{N_t = k\} = e^{-\lambda t} \frac{(\lambda t)^k}{k!};$$

la fonction de masse de N_t est donc celle d'une variable aléatoire de Poisson de paramètre λt . \square

Homogénéité spatiale. Le processus de Poisson possède une propriété particulière que l'on appellera l'*homogénéité spatiale*. Voici une définition générale de cette propriété :

DÉFINITION 4.2. Soit $\mathbf{X} = (X_t)_{t \geq 0}$ un processus stochastique (temps discret ou continu) sur un espace d'états discret S additif – c'est à dire un espace d'états muni d'une opération d'addition. Soient, pour tous $0 \leq s \leq t$, les probabilités de transition

$$P_{x,y}(s, t) = \mathbb{P} \{X_t = y \mid X_s = x\}.$$

On dit que le processus \mathbf{X} est **spatialement homogène** si et seulement si $P_{x,y}(s, t)$ ne dépend que de $y - x$ – plus spécifiquement, on doit avoir :

$$(4.1.2) \quad P_{x,y}(s, t) = P_{0, (y-x)}(s, t).$$

REMARQUE. On remarque l'analogie avec l'*homogénéité temporelle*, qui requiert plutôt que $P_{x,y}(s, t) = P_{x,y}(0, t-s)$.

On a le lemme suivant :

LEMME 4.1. Soit $\mathbf{X} = (X_t)_{t \geq 0}$ une chaîne de Markov homogène et spatialement homogène sur l'espace des états discret S additif.

Alors, le processus $\mathbf{Y}^{(s)} = (Y_t^{(s)} = X_{s+t} - X_s)_{t \geq 0}$ suit, pour tout $s \geq 0$, exactement la même loi que le processus \mathbf{X} . En particulier, $X_{s+t} - X_s \sim X_t - X_0$, où l'équivalence est en distribution.

DÉMONSTRATION. On fixe un $s \geq 0$ arbitraire. Soient $P_{x,y}(t, u)$ et $Q_{x,y}(t, u)$ les probabilités de transition respectivement pour les chaînes \mathbf{X} et $\mathbf{Y}^{(s)}$; il suffit de montrer que ces quantités sont égales.

On a :

$$\begin{aligned}
 Q_{x,y}(t, u) &= \mathbb{P} \left\{ Y_u^{(s)} = y \mid Y_t^{(s)} = x \right\} \\
 &= \mathbb{P} \{ X_{s+u} - X_s = y \mid X_{s+t} - X_s = x \} \\
 &= \sum_{z \in S} \mathbb{P} \{ X_{s+u} - X_s = y \mid X_{s+t} - X_s = x, X_s = z \} \mathbb{P} \{ X_s = z \} \\
 &= \sum_{z \in S} \mathbb{P} \{ X_{s+u} = z + y \mid X_{s+t} = z + y, X_s = z \} \mathbb{P} \{ X_s = z \} \\
 &= \sum_{z \in S} \mathbb{P} \{ X_{s+u} = z + y \mid X_{s+t} = z + y \} \mathbb{P} \{ X_s = z \} \\
 &= \sum_{z \in S} P_{z+x, z+y}(s+t, s+u) \mathbb{P} \{ X_s = z \} \\
 &= \sum_{z \in S} P_{x,y}(t, u) \mathbb{P} \{ X_s = z \} \\
 &= P_{x,y}(t, u).
 \end{aligned}$$

Dans l'ordre on applique successivement les définitions des probabilités de transition et du processus $\mathbf{Y}^{(s)}$; ensuite, on conditionne par l'état X_s , puis on applique la propriété de Markov, l'homogénéité temporelle et spatiale, et enfin la condition de normalisation qui garantit que $\sum_{z \in S} \mathbb{P} \{ X_s = z \} = 1$. \square

PROPOSITION 4.2. Soit $\mathbf{N} = (N_t)_{t \in \mathbb{R}^+}$ un processus de Poisson d'intensité $\lambda > 0$. Alors, ce processus a la propriété d'homogénéité spatiale.

DÉMONSTRATION. Pour montrer cette propriété, il suffit de vérifier que les taux générateurs satisfont la propriété que $G_{x,y}$ ne dépend que de $y - x$; c'est vrai. En effet, d'après la définition, on a que

$$G_{x,y} = \begin{cases} -\lambda & \text{si } y - x = 0 \\ \lambda & \text{si } y - x = 1. \end{cases}$$

\square

COROLLAIRE 4.1. Soit $\mathbf{N} = (N_t)_{t \in \mathbb{R}^+}$ un processus de Poisson d'intensité $\lambda > 0$. Alors, puisqu'il est spatialement et temporellement homogène, on a que pour tous $0 \leq s \leq t$,

$$N_t - N_s \sim N_{t-s};$$

en particulier, $N_t - N_s$ suit donc une loi de Poisson de paramètre $\lambda(t-s)$.

Incréments indépendants. On fait la définition suivante :

DÉFINITION 4.3. Soit $\mathbf{X} = (X_t)_{t \geq 0}$ un processus stochastique sur l'espace des états discret additif S (c'est à dire que S est muni d'une opération d'addition.)

L'incrément de \mathbf{X} entre les temps s et t est donné par $X_t - X_s$.

LEMME 4.2. Soit $\mathbf{X} = (X_t)_{t \geq 0}$ un processus de Markov homogène (temps continu ou discret) spatialement homogène sur l'espace des états discret additif S . Soient $0 \leq s \leq t$.

Alors, l'incrément $X_t - X_s$ est indépendant de X_s .

DÉMONSTRATION. Il suffit de calculer :

$$\begin{aligned} \mathbb{P}\{X_s = x, X_t - X_s = y\} &= \mathbb{P}\{X_s = x, X_t = x + y\} \\ &= \mathbb{P}\{X_t = x + y \mid X_s = x\} \mathbb{P}\{X_s = x\} \\ &= P_{0,y}(s, t) \mathbb{P}\{X_s = x\}. \end{aligned}$$

On a que la fonction de masse jointe de $X_t - X_s$ et X_s se factorise en un facteur qui ne dépend que de x , et un facteur qui ne dépend que de y . On a donc que

$$\mathbb{P}\{X_t - X_s = y\} = P_{0,y}(s, t).$$

□

COROLLAIRE 4.2. Soit $\mathbf{N} = (N_t)_{t \in \mathbb{R}^+}$ un processus de Poisson d'intensité $\lambda > 0$.

Alors, les incréments de \mathbf{N} sur des intervalles disjoints sont indépendants. En particulier, si $(s_i)_{i \in \mathbb{N}}$ et $(t_i)_{i \in \mathbb{N}}$ sont des suites croissantes de temps, avec $0 \leq s_i \leq t_i \leq s_{i+1}$, alors $(N_{t_i} - N_{s_i})_{i \in \mathbb{N}}$ est une suite indépendante de variables aléatoires, chacune suivant respectivement une loi de Poisson de paramètre $\lambda(t_i - s_i)$.

EXEMPLE 4.1. Un radar photo placé sur le bord de l'autoroute enregistre un grand excès de vitesse chaque fois qu'un-e automobiliste passe devant celui-ci à une vitesse supérieure à 130 km/h.² Selon les données du radar, le jour, ces excès surviennent au taux de 3 par heure en moyenne.

- (a) Quelle est la probabilité que l'on n'ait observé aucun excès de vitesse pendant une période d'une heure consécutive ?
- (b) Sachant qu'on a observé cinq excès de vitesse entre 13h et 15h, quelle est la probabilité que ceux-ci se soient tous produits avant 14h ?

SOLUTION. (a) On cherche $\mathbb{P}\{N_1 = 0\} = e^{-\lambda} = e^{-3}$.

- (b) Si t est le temps écoulé depuis 13h, alors on cherche :

$$\mathbb{P}\{N_1 = 5 \mid N_2 = 5\} = \frac{\mathbb{P}\{N_1 = 5, N_2 - N_1 = 0\}}{\mathbb{P}\{N_2 = 5\}}.$$

On peut utiliser l'indépendance des incréments pour obtenir que

$$\mathbb{P}\{N_1 = 5 \mid N_2 = 5\} = \frac{e^{-\lambda} \frac{\lambda^5}{5!} \cdot e^{-\lambda}}{e^{-2\lambda} \frac{(2\lambda)^5}{5!}} = \frac{1}{32}.$$

2. Cette mise en situation pourrait ne pas refléter la réalité – je ne sais pas trop comment ça marche ces machins-là. Conduisez prudemment, bande de *dare-devils* délinquants !

La distribution sachant N_T . Ce qu'on va montrer maintenant, c'est que si on sait que le nombre total d'événements qui sont survenus sur l'intervalle de temps T est égal à $N_T = n$, alors la distribution de ceux-ci à l'intérieur de l'intervalle $[0, T]$ est simplement la distribution de n variables aléatoires uniformes indépendantes.

PROPOSITION 4.3. *Soit $\mathbf{N} = (N_t)_{t \in \mathbb{R}^+}$ un processus de Poisson d'intensité $\lambda > 0$, et soit $T > 0$ une constante fixe quelconque.*

Alors,

$$(4.1.3) \quad \mathbb{P}\{N_t = k \mid N_T = n\} = \binom{n}{k} \frac{t^k (T-t)^{n-k}}{T^n}.$$

DÉMONSTRATION. Comme dans l'exemple précédent, on utilise l'indépendance des incréments :

$$\begin{aligned} \mathbb{P}\{N_t = k \mid N_T = n\} &= \frac{\mathbb{P}\{N_t = k, N_T - N_t = n - k\}}{N_T = n} \\ &= \frac{e^{-\lambda T} \frac{(\lambda t)^k}{k!} \cdot \frac{(\lambda(T-t))^{n-k}}{(n-k)!}}{e^{-\lambda T} \frac{(\lambda T)^n}{n!}} \\ &= \frac{n!}{k!(n-k)!} \frac{t^k (T-t)^{n-k}}{T^n}. \end{aligned}$$

□

REMARQUE. On remarque que si X_1, \dots, X_n sont des variables aléatoires uniformes sur l'intervalle $[0, T]$ indépendantes, et que $U_t = |\{i : X_i \leq t\}|$ est le nombre de variables X_i qui sont inférieures à t , alors U_t est une binomiale de paramètres $n, p = t/T$, qui est exactement la distribution de N_t sachant que $N_T = n$. On conclue donc que sachant $N_T = n$, les temps des n premières transitions T_1, T_2, \dots, T_n suivent la distribution des statistiques d'ordre $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ de nos variables uniformes X_i ...

EXEMPLE 4.2. Au Québec, parmi les jeunes filles nées entre le 1er janvier 1999 et le 31 décembre 2001, 1483 ont reçu le prénom « Camille ».³

Quelle est la probabilité qu'au moins une jeune femme prénommée Camille soit née le 29 février 2000 ? Quelles hypothèses fait-on pour répondre à cette question ?

SOLUTION. Pour répondre à la question, on fait l'hypothèse que les intervalles successifs entre les naissances de Camilles dans ces années-là sont sans-mémoire, de taux identique et indépendants ; Si N_t est le nombre de Camilles nées à la fin du t ième jour depuis le 1er janvier 1999, alors on sait que $N_{1096} = 1483$.

La probabilité qu'il n'y ait eu aucune Camille née le 29 février 2000 correspond à la probabilité que les 1483 soient toutes nées les autres jours ; C'est :

$$\mathbb{P}\{N_{425} - N_{424} = 0 \mid N_{1096} = 1483\} = \left(\frac{1095}{1096}\right)^{1483} \approx 25,8\%.$$

Il est donc très probable qu'une Camille soit née le 29 février 2000.

3. D'après la liste des prénoms de Retraite-Québec des enfants admissibles aux prestations familiales, nés ou ayant immigré au Québec depuis 1980.

4.1.4. Superpositions, conditions. On peut imaginer un processus qui compte le nombre d'événements de « deux types différents », indépendants les uns des autres. Si les événements de chaque type surviennent selon un processus de Poisson, alors le processus qui dénombre les événements des deux types en même temps est *aussi* un processus de Poisson.

PROPOSITION 4.4. Soient $\mathbf{N}^{(1)} = (N_t^{(1)})_{t \in \mathbb{R}^+}$ et $\mathbf{N}^{(2)} = (N_t^{(2)})_{t \in \mathbb{R}^+}$ deux processus de Poisson d'intensités respectives $\lambda > 0$ et $\mu > 0$.

Alors, le processus $\mathbf{N} = (N_t = N_t^{(1)} + N_t^{(2)})_{t \in \mathbb{R}^+}$ est un processus de Poisson d'intensité $\lambda + \mu$.

DÉMONSTRATION. On a que N_t est le nombre total d'événements qui se sont produits, soit pour le processus $\mathbf{N}^{(1)}$ ou le processus $\mathbf{N}^{(2)}$; le temps inter-transition dans l'état k est $U = \min\{U_1, U_2\}$, où U_1 et U_2 sont des variables aléatoires exponentielles indépendantes, respectivement de taux λ et μ ; donc, U est une loi exponentielle de taux $\lambda + \mu$; c'est donc dire que le taux générateur $G_{k,k} = -\lambda - \mu$; évidemment, la seule transition possible est vers l'état $k + 1$ (on ignore la possibilité que deux transitions surviennent en même temps puisque la probabilité de cela est nulle).

Donc, on a finalement que $G_{k,k} = -(\lambda + \mu)$ et $G_{k,k+1} = \lambda + \mu$, pour tout k ; donc \mathbf{N} est un processus de Poisson d'intensité $\lambda + \mu$. \square

À l'inverse, supposons qu'on considère que pour chacun des événements dénombrés par le processus \mathbf{N} , celui-ci peut être d'un certain type « A » avec probabilité p , indépendamment pour chacun des événements, alors le processus $\mathbf{N}^{(A)}$ qui dénombre uniquement les événements de type A est *aussi* un processus de Poisson.

PROPOSITION 4.5. Soit $\mathbf{N} = (N_t)_{t \in \mathbb{R}^+}$ un processus de Poisson d'intensité λ , et soient $(A_k)_{k \in \mathbb{N}}$ une suite indépendante et identiquement distribuée de variables aléatoires de Bernoulli avec probabilité p (c'est-à-dire que $\mathbb{P}\{A_k = 1\} = p$ et $\mathbb{P}\{A_k = 0\} = 1 - p$).

Alors, si on note $N_t^{(A)} = \sum_{k=1}^{N_t} A_k$, le processus $\mathbf{N}^{(A)} = (N_t^{(A)})_{t \in \mathbb{R}^+}$ est un processus de Poisson d'intensité λp .

DÉMONSTRATION. Soit T_k le temps de la k ième transition du processus \mathbf{N} , et $U_k = T_k - T_{k-1}$ les temps inter-transition de ce processus. Si on note $T_1^{(A)}$ le temps de la première transition pour $\mathbf{N}^{(A)}$, et $K = \min\{k : A_k = 1\}$ la première transition de \mathbf{N} qui entraîne une transition pour $\mathbf{N}^{(A)}$, alors on a que, d'une part :

$$T_1^{(A)} = \sum_{k=1}^K U_k,$$

où K suit une loi géométrique de paramètre p (puisque c'est le premier indice où on a un succès), et les U_k sont des variables aléatoires exponentielles indépendantes de paramètre λ .

Par un résultat du cours de probabilités, on voit que la fonction génératrice des moments $M_A(t) = \mathbb{E}[\exp(tT_1^{(A)})]$ est donnée par :

$$M_A(t) = \psi_K(M_U(t)),$$

où ψ_K est la fonction génératrice des probabilités pour la distribution de K , et $M_U(t)$ est la fonction génératrice des moments pour la distribution des U_k . En l'occurrence, on a que :

$$\psi_K(s) = \frac{ps}{1 - (1-p)s}, \quad M_U(t) = \frac{\lambda}{\lambda - t},$$

et il suit que

$$M_A(t) = \frac{p\lambda}{(\lambda - t - (1-p)\lambda)} = \frac{\lambda p}{\lambda p - t},$$

et $T_1^{(A)}$ suit une loi exponentielle de paramètre λp ; donc, le taux inter-transition pour l'état 0 est de λp , et bien sûr la seule transition possible est vers l'état 1.

En faisant le même raisonnement pour les états k et $k+1$, on trouve que les taux générateurs sont $G_{k,k+1} = \lambda p$ et $G_{k,k} = -\lambda p$; donc, le processus $\mathbf{N}^{(A)}$ est un processus de Poisson d'intensité λp . \square

4.2. Les processus de naissances

Les *processus de naissance et de mort* sont une catégorie assez large de processus stochastiques à temps continu sur l'espace des états $S = \mathbb{Z}^+$, qui ont la propriété spéciale que les transitions possibles depuis l'état k sont limitées seulement à des *naissances* (une transition vers $k+1$) et des *morts* (une transition vers $k-1$); On conçoit X_t comme le « nombre d'individus en vie au temps t »; l'analogie n'est pas très difficile.

DÉFINITION 4.4. Soit $\mathbf{X} = (X_t)_{t \in \mathbb{R}^+}$ une chaîne de Markov homogène à temps continu sur l'espace des états $S = \mathbb{Z}^+$.

\mathbf{X} est un **processus de naissance et de mort** lorsque $G_{i,j} = 0$ pour tous $i, j \in \mathbb{Z}^+$ tels que $|i - j| > 1$.

On parle d'un *processus de naissances* tout-court lorsque $G_{i,i-1} = 0$ pour tout i – c'est à dire qu'il n'y a « jamais de morts. »

Le processus de Poisson, que nous venons de voir, est un processus de naissance. C'est le plus simple, puisque c'est un processus de naissance spatialement homogène; c'est à dire que $G_{k,k}$ et $G_{k,k+1}$ ne dépendent pas de k dutout.

Une toute petite coche de complexité au-dessus, on trouve...

Le cas le plus général d'un processus de naissance est celui où le taux générateur $G_{k,k} = -\nu_k$ dépend de k . Puisqu'il s'agit d'un processus de naissance, on sait que le seul autre taux générateur non-nul est $G_{k,k+1} = \nu_k$.

Dans ce cas très général, il est difficile d'obtenir une idée aussi précise des distributions impliquées; cela dépendra bien sûr des valeurs exactes des taux de naissances $(\nu_k)_{k \in \mathbb{Z}^+}$

On observe tout de même le phénomène suivant :

PROPOSITION 4.6. Soit $\mathbf{X} = (X_t)_{t \in \mathbb{R}^+}$ un processus de naissances avec taux de naissances $(\nu_k)_{k \in \mathbb{Z}^+}$. On définit $T_0 = 0$ et

$$T_k = \inf \{t > T_{k-1} : X_t \neq X_{T_{k-1}}\}$$

le temps de la k ième transition (de sorte que $X_{T_k} = X_0 + k$ pour tout $k \in \mathbb{Z}^+$.)

Alors, on a que

$$(4.2.1) \quad \mathbb{E}_0 [T_n] = \sum_{k=1}^n \frac{1}{\nu_k}.$$

DÉMONSTRATION. On sait d'emblée que les variables $(U_k = T_k - T_{k-1})_{k \in \mathbb{N}}$ (les temps inter-transition) sont des variables aléatoires indépendantes. De plus, sachant que $X_{T_{k-1}} = x$, on sait que U_k est une variable aléatoire exponentielle de paramètre ν_x .

Or, bien sûr, puisque \mathbf{X} est un processus de naissance, la suite $(X_{T_n})_{n \in \mathbb{Z}^+}$ est entièrement déterminée – en effet, on a $X_{T_n} = n$. Donc, on sait immédiatement que U_k suit une loi exponentielle de paramètre ν_k . Donc,

$$\mathbb{E}_0[U_k] = \frac{1}{\nu_k}.$$

On conclue en utilisant la linéarité de l'espérance et une somme télescopique :

$$\mathbb{E}_0[T_n] = \sum_{k=1}^n \mathbb{E}_0[U_k] = \sum_{k=1}^n \frac{1}{\nu_k}.$$

□

EXEMPLE 4.3. Dans un processus de Poisson d'intensité $\lambda > 0$, l'espérance du temps de la n ième transition (le temps d'atteinte de l'état n) est donnée par

$$\mathbb{E}_0[\tau_n] = \frac{n}{\lambda}.$$

C'est un fait qu'on peut vérifier puisqu'on sait que T_n suit une loi Gamma de paramètres n, λ dans ce cas.

4.2.1. Les explosions. On vient de voir que dans un processus de naissances avec les taux de naissances $(\lambda_k)_{k \in \mathbb{Z}^+}$, l'espérance du temps requis pour atteindre l'état n est donnée par :

$$(4.2.1) \quad \mathbb{E}_0[T_n] = \sum_{k=1}^n \frac{1}{\nu_k}.$$

Clairement, cette suite est croissante. Mais que se passe-t-il si elle est *bornée* ? En effet, *a priori*, il n'est pas exclu, par exemple, d'avoir

$$\lim_{n \rightarrow \infty} \mathbb{E}_0[T_n] = t_\infty < +\infty;$$

il suffit que la série $\sum_{k \geq 1} \nu_k^{-1}$ converge !

Cela voudrait dire que, *peu importe la valeur de n qu'on choisit*, on aurait que $\mathbb{E}_0[T_n] < t_\infty$. Autrement dit, « en moyenne », le processus pourrait partir à $+\infty$ en un temps *fini* ! C'est ce qu'on appelle gentiment une *explosion*.

En fait, on va faire la définition suivante, très générale :

DÉFINITION 4.5. Soit $\mathbf{X} = (X_t)_{t \in \mathbb{R}^+}$ un processus stochastique à temps continu sur l'espace des états discret S .

On dit que \mathbf{X} **explose** si il existe un intervalle fini de temps durant lequel se produisent une infinité de transitions.

À noter que, *a priori*, le fait que $t_\infty < +\infty$ existe ne signifie pas forcément que notre processus explose à chaque réalisation – mais on sait que les temps successifs des transitions ont une espérance bornée – mais ce qu'on va voir, c'est que dans le cas où cette limite existe, la probabilité que le processus explose est de 1. Pour ce faire, on introduit le *temps de vie* :

DÉFINITION 4.6. Soit $\mathbf{X} = (X_t)_{t \in \mathbb{R}^+}$ un processus stochastique à temps continu sur l'espace des états discret S . On définit $T_0 = 0$ et le temps de la n ième transition

$$T_n = \inf \{t > T_{n-1} : X_t \neq X_{T_{n-1}}\}.$$

Le **temps de vie** de \mathbf{X} , noté T_∞ , est donné par :

$$T_\infty = \sup \{T_n : n \in \mathbb{Z}^+\} = \lim_{n \rightarrow \infty} T_n.$$

À noter que le temps de vie peut être infini, dans le cas où la suite $(T_n)_{n \in \mathbb{Z}^+}$ ne serait pas bornée supérieurement.

Le temps de vie du processus est le temps requis pour effectuer toutes les transitions – comme on vient de le voir, dans le cas d'une explosion, ça sera un temps fini !

PROPOSITION 4.7. Soit $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ un processus de naissances avec taux de naissances $(\nu_k)_{k \in \mathbb{Z}^+}$. On note $T_0 = 0$, T_n le temps de la n ième transition et $T_\infty = \sup \{T_n : n \in \mathbb{Z}^+\}$ le temps de vie de \mathbf{X} .

Alors, si $\sum_{k \geq 1} \nu_k^{-1}$ converge, $\mathbb{P}\{T_\infty < +\infty\} = 1$. Si $\sum_{k \geq 1} \nu_k^{-1}$ diverge, alors $\mathbb{P}_0\{T_\infty < +\infty\} = 0$.

DÉMONSTRATION. En reprenant la notation de la proposition précédente, on a que $T_n = \sum_{k=1}^n U_k$, et

$$T_\infty = \lim_{n \rightarrow \infty} T_n.$$

Lorsque $\mathbb{E}_0[T_\infty] < +\infty$, (soit quand la série $\sum_{k \geq 1} \nu_k^{-1}$ converge, on doit forcément avoir que la probabilité $\mathbb{P}\{T_\infty = +\infty\} = 0$.

Supposons qu'on fixe un temps t ; on a alors que $T_\infty > t$ si et seulement si, au temps t , on se trouve quelque part passé l'état k , sur \mathbb{Z}^+ ; donc, on a en somme :

$$\mathbb{P}_k\{T_\infty > t\} = \sum_{i=k}^{\infty} P_{k,i}^{(t)}.$$

Si on calcule maintenant $\mathbb{P}_0\{T_\infty > mt\}$, pour $m \in \mathbb{N}, m > 1$, on trouve que

$$\mathbb{P}_k\{T_\infty > mt\} = \sum_{i=k}^{\infty} \mathbb{P}_k\{T_\infty > mt \mid X_t = i\} P_{k,i}^{(t)}.$$

Or, par la propriété d'homogénéité et la propriété de Markov,

$$\mathbb{P}_k\{T_\infty > mt \mid X_t = i\} = \mathbb{P}_i\{T_\infty > (m-1)t\},$$

et on trouve :

$$\mathbb{P}_k\{T_\infty > mt\} = \sum_{i=k}^{\infty} \mathbb{P}_i\{T_\infty > (m-1)t\} P_{k,i}^{(t)}.$$

Bien sûr,

$$\mathbb{P}_i\{T_\infty > (m-1)t\} = \mathbb{P}_k\{T_\infty - T_{i-k} > (m-1)t\} \leq \mathbb{P}_k\{T_\infty > (m-1)t\}$$

, puisque $\mathbb{P}_0\{T_k < +\infty\} = 1$ (vu que partant de 0 on est sûr.e.s d'atteindre k).

Donc,

$$\mathbb{P}_k\{T_\infty > mt\} \leq \mathbb{P}_k\{T_\infty > (m-1)t\} \sum_{i=k}^{\infty} P_{k,i}^{(t)} = \mathbb{P}_k\{T_\infty > (m-1)t\} \mathbb{P}_k\{T_\infty > t\},$$

pour tous k (en particulier pour $k = 0$). Donc,

$$\mathbb{P}_0 \{T_\infty > mt\} \leq \mathbb{P}_0 \{T_\infty > t\}^m,$$

En particulier, si on suppose que pour un certain t , $\mathbb{P}_0 \{T_\infty > t\} < 1$, on trouve que

$$\mathbb{E}_0 [T_\infty] = \int_{u=0}^{\infty} \mathbb{P}_0 \{T_\infty > u\} du \leq \sum_{m=1}^{\infty} p^m = \frac{p}{1-p},$$

où $p = \mathbb{P}_0 \{T_\infty > t\}$. Or, cela est une contradiction, puisqu'on suppose que $\mathbb{E}_0 [T_\infty] = +\infty$. Donc, la seule possibilité est qu'on doit avoir que pour tout t , $\mathbb{P}_0 \{T_\infty > t\} = 1$; en prenant la limite lorsque t tend vers 0, on trouve donc que $\mathbb{P}_0 \{T_\infty = +\infty\} = 1$. \square

EXEMPLE 4.4. On suppose une colonie de bactéries, dans laquelle chaque bactérie se clone à un taux de ν fois par minutes.

- (a) Quels-sont les taux générateurs du processus $(X_t)_{t \in \mathbb{R}^+}$, où X_t compte le nombre de bactéries au temps t ?
- (b) Combien de temps cela prend-t-il en moyenne avant que la colonie atteigne n bactéries?
- (c) Y a-t-il explosion?

SOLUTION. (a) Lorsqu'il y a k bactéries dans la colonie, chacune d'entre elle se clone avec un taux λ . Donc, il y aura une transition au taux de $\lambda_k = k\lambda$, aussitôt que l'une des bactéries se sera clonée! On a donc $G_{k,k} = -k\lambda$ et $G_{k,k+1} = k\lambda$.

- (b) On cherche

$$\mathbb{E}_0 [T_n] = \sum_{k=1}^n \frac{1}{k\lambda} \approx \frac{\log(k)}{\lambda}.$$

- (c) Non, il n'y a pas d'explosions, puisque cette suite diverge. Donc, le temps de vie est infini. La croissance de la colonie est exponentielle de taux λ .

4.3. Les processus de naissances et de morts généraux.

La suite naturelle à notre étude, c'est bien sûr de considérer le cas le plus général des processus de naissances et de mort :

DÉFINITION 4.7. Soit $\mathbf{X} = (X_t)_{t \in \mathbb{R}^+}$ une chaîne de Markov homogène à temps continu sur l'espace des états $S = \mathbb{Z}^+$.

On dit que \mathbf{X} est un **processus de naissances et de morts** avec taux de naissance $(\nu_i)_{i \in \mathbb{Z}^+}$ et taux de mort $(\mu_i)_{i \in \mathbb{N}}$ si ses taux générateurs sont donnés par

$$(4.3.1) \quad G_{0,0} = -\nu_0, G_{0,1} = \nu_0; \quad G_{k,k-1} = \mu_k, G_{k,k} = -\mu_k - \nu_k, G_{k,k+1} = \nu_k \quad \forall k \in \mathbb{N}.$$

REMARQUE. Le taux de mort μ_0 n'est pas défini parce qu'il n'est pas nécessaire; intuitivement, si la population est nulle (vide) personne ne peut « mourir »!

Ces processus sont assez complexes pour modéliser plusieurs types de phénomènes intéressants. Les processus de naissances (tout court) sont simplement un cas particulier des processus de naissances et de mort qui survient lorsque $\mu_k = 0$ pour tout $k \in \mathbb{N}$.

4.3.1. Les explosions. Les processus de naissances et de morts sont plus compliqués à gérer que les simples processus de naissances ; en effet, on ne peut pas estimer aussi facilement l'espérance du temps de vie, puisque'il est impossible de l'écrire comme une somme de variables aléatoires indépendantes dont la distribution est connue d'avance.

On peut toutefois remarquer une borne intéressante :

LEMME 4.3. *Soit $\mathbf{X} = (X_t)_{t \in \mathbb{R}^+}$ une chaîne de Markov homogène à temps continu sur l'espace des états discret S .*

Si le processus \mathbf{X} explose avec une probabilité positive, alors l'ensemble $\{X_t : t \geq 0\} \subseteq S$ des états visités est infini.

DÉMONSTRATION. *Rappel de notation : on va utiliser $T_0 = 0, T_k = \inf \{t > T_{k-1} : X_t \neq X_{T_{k-1}}\}$, $T_\infty = \lim_{n \rightarrow \infty} T_n$ et $U_k = T_k - T_{k-1}$.*

Si on note $A = \{T_\infty < +\infty\}$ l'événement où \mathbf{X} explose, et $B = \{|\{X_t : t \geq 0\}| < +\infty\}$ l'événement où le nombre d'états visités est fini. Ce qu'on va montrer, c'est que $\mathbb{P}_0 \{A \mid B\} = 0$; sachant qu'on a visité un nombre fini d'états, la probabilité d'exploser doit être nulle.

Supposons le contraire ; alors, on doit avoir que $\mathbb{P} \{A, B\} > 0$. Alors, cela signifie qu'on doit avoir une explosion en visitant un nombre fini d'états. Pour faire simple, supposons sans perdre de généralité que $S = \mathbb{N}$; on peut supposer que sur l'événement B , il existe une variable aléatoire N telle que $X_t \leq N$ pour tout t .

Ce qu'on peut faire, c'est partitionner B selon la valeur de N ; sachant que $N = n$, on va noter $m_n = \max \{\lambda_i : 1 \leq i \leq n\}$, où λ_i est le taux inter-transiiton de l'état i ($\lambda_i = -G_{i,i}$).

Pour les besoins de cette preuve, on notera

$$\bar{\mathbb{P}}^{(n)} \{\cdot\} = \mathbb{P} \{\cdot \mid N = n\}.$$

C'est la mesure de probabilités sachant qu'on se restreint seulement aux états entre 1 et n . Alors,

$$\begin{aligned} \bar{\mathbb{P}}^{(n)} \{U_k > \delta\} &= \sum_{i=1}^n \bar{\mathbb{P}}^{(n)} \{U_k > \delta \mid X_{T_{k-1}} = i\} \bar{\mathbb{P}}^{(n)} \{X_{T_{k-1}} = i\} \\ &= \sum_{i=1}^n \mathbb{P} \{U_k > \delta \mid X_{T_{k-1}} = i\} \bar{\mathbb{P}}^{(n)} \{X_{T_{k-1}} = i\} \\ &= \sum_{i=1}^n e^{-\lambda_i \delta} \bar{\mathbb{P}}^{(n)} \{X_{T_{k-1}} = i\} \\ &\geq e^{-m_n \delta}. \end{aligned}$$

On a donc que

$$\sum_{k=1}^{\infty} \bar{\mathbb{P}}^{(n)} \{U_k > \delta\} = +\infty;$$

par le Lemme de Borel-Cantelli et la loi zéro-un de Kolmogorov⁴, on peut conclure que puisque la série diverge, avec probabilité 1, une infinité de U_k sont supérieurs à δ , et $T_\infty = +\infty$. Donc, sur $N = n$, on a que

$$\bar{\mathbb{P}}^{(n)} \{A\} = 0.$$

4. Il s'agit de résultats importants en théorie des probabilités. Consulter l'annexe C pour plus de détails.

Donc,

$$\mathbb{P}\{A \mid B\} = \sum_{n=1}^{\infty} \overline{\mathbb{P}}^{(n)}\{A\} \mathbb{P}\{N = n\} = 0,$$

et on a montré que la probabilité d'explosion sachant qu'on n'atteint qu'un nombre fini d'états est nulle. \square

Le résultat des courses : si on veut une explosion dans une chaîne de Markov homogène à temps continu, il faut visiter un nombre infini d'états. Et si on visite un nombre infini d'états, il faut les quitter un par un. On a donc la proposition suivante :

PROPOSITION 4.8. *Soit $\mathbf{X} = (X_t)_{t \in \mathbb{R}^+}$ un processus de naissances et de morts avec taux de naissances $(\nu_i)_{i \in \mathbb{Z}^+}$ et taux de morts $(\mu_i)_{i \in \mathbb{N}}$.*

Alors, si

$$\sum_{i=1}^n (\nu_i + \mu_i)^{-1} = \infty,$$

on a que $\mathbb{P}\{T_\infty < +\infty\} = 0$; le processus n'explose presque sûrement pas.

DÉMONSTRATION. *Cette preuve n'est pas entièrement rigoureuse mais elle donne une idée du raisonnement.*

Ce qu'il suffit de faire ici, c'est de remarquer que par le Lemme 4.3, pour qu'il y ait une explosion, il faut visiter un nombre infini d'états – et pour faire ça, il faut absolument juste visiter tous les états (à partir de celui où on a commencé), vu qu'on ne peut que passer de k à $k \pm 1$.

Si on note V_i le temps passé à l'état i la première fois qu'on y est arrivés, V_i suit une loi exponentielle de taux $\lambda_i = \mu_i + \nu_i$. Or, $T_\infty > \sum_{i=1}^{\infty} V_i$ si on commence en 1 ; si la série $\sum_{i=1}^{\infty} (\mu_i + \nu_i)^{-1}$ diverge, on doit alors avoir que $\mathbb{E}[T_\infty]$ diverge, et donc on doit avoir une explosion. \square

Autrement dit, donc, il suffit de vérifier que $\sum_{i=1}^{\infty} (\nu_i + \mu_i)^{-1}$ diverge pour s'assurer qu'il n'y aura pas d'explosion.

4.3.2. Quelques exemples.

EXEMPLE 4.5. Dans une colonie de bactéries, chaque bactérie se divise en deux bactéries identiques après un temps sans mémoire de taux ν , indépendamment de toutes les autres bactéries. De plus, chaque bactérie meure après un temps sans mémoire de taux μ , indépendamment de toutes les autres.

- Soit X_t la taille de la population de bactéries (en nombre de bactéries) après un temps $t \in \mathbb{R}^+$. Quels sont les taux générateurs pour $\mathbf{X} = (X_t)_{t \in \mathbb{R}^+}$?
- Si on commence avec une seule bactérie, est-ce que la population finit par s'éteindre ?
- Quelle condition garantit l'extinction éventuelle de la population ?
- Dans le cas où l'extinction est garantie, quelle est l'espérance du temps avant l'extinction ?

SOLUTION. (a) Lorsqu'on est dans l'état $k \in \mathbb{N}$, il y a k bactéries dans la colonie. Deux types d'événements peuvent survenir :

- Une des k bactérie se reproduit. Cela survient à un taux $k\nu$, et on transitionne vers l'état $k + 1$.

- Une des k bactéries meure. Cela survient à un taux $k\mu$, et on transitionne vers l'état $k - 1$.

On a donc les taux de transition suivants :

$$G_{k,k-1} = k\mu, G_{k,k} = -k(\mu + \nu), G_{k,k+1} = k\nu.$$

Bien sûr, dans l'état 0, aucune transition n'est possible, donc les taux générateurs des transitions sont tous nuls dans cet état ; l'état 0 est *absorbant* – une fois atteint, on y reste indéfiniment.

- (b) Si on commence avec une seule bactérie, on peut considérer le processus discret $\mathbf{Y} = (Y_n = X_{T_n})_{n \in \mathbb{Z}^+}$ où T_n est le temps de la n ième transition (et $T_0 = 0$). Les probabilités de transition en un pas pour \mathbf{Y} sont données par :

$$P_{0,0} = 1; P_{k,k-1} = \frac{k\mu}{k(\mu + \nu)} = \frac{\mu}{\mu + \nu}, P_{k,k+1} = \frac{\nu}{\mu + \nu}.$$

Pour se simplifier la vie, on va noter :

$$p = \frac{\nu}{\mu + \nu}, q = 1 - p.$$

Donc, $P_{0,0} = 1$ et $P_{k,k+1} = p$ et $P_{k,k-1} = q$.

La chaîne \mathbf{Y} a deux classes d'équivalence :

- $\{0\}$: l'état 0 est tout seul dans sa classe ; il est récurrent car c'est une classe fermée sur un nombre fini d'états. Il est apériodique car $P_{0,0} = 1$.
- \mathbb{N} : tous les autres états communiquent entre eux ; ils sont transients car $P_{1,0} > 0$ donc \mathbb{N} est une classe ouverte. Ils sont de période 2, puisqu'on ne peut transitionner que d'états impairs à pairs et vice-versa, et que les plus petits cycles sont de longueur 2.

Si τ_0 et τ_N sont les temps d'atteinte respectifs des états 0 et N pour la chaîne \mathbf{Y} , alors on a, par l'argument pour la ruine du parieur (exemple 1.17), que si $\mu \neq \nu$,

$$\mathbb{P}_1 \{ \tau_0 < \tau_N \} = \frac{\rho - \rho^N}{1 - \rho^N},$$

où $\rho = q/p = \mu/\nu$.

Si $\mu = \nu$, on a plutôt que

$$\mathbb{P}_1 \{ \tau_0 < \tau_N \} = \frac{N-1}{N}.$$

En prenant la limite lorsque N tend vers l'infini on trouve que

- la probabilité que la population s'éteigne est ρ si $\rho < 1$;
- c'est 1 si $\rho \geq 1$.

- (c) Finalement, on trouve donc que pour garantir l'extinction éventuelle de la population, il faut que $\rho \geq 1$ – soit que $\mu \geq \nu$. C'est très logique ; ça veut dire qu'il faut que les bactéries se reproduisent plus vite qu'elles ne meurent.

REMARQUE. Une autre façon amusante de répondre aux questions est de raisonner par arbres de Galton-Watson. La descendance immédiate de chaque bactérie correspond au nombre de fois qu'elle se divisera avant de mourir ; c'est une variable aléatoire géométrique (commençant en 0) avec fonction de masse

$$p(k) = p^k q \quad (k \geq 0).$$

La fonction génératrice des probabilités pour la descendance immédiate d'une bactérie est donc :

$$\psi(s) = \frac{q}{1 - ps};$$

on trouve

$$\psi'(s) = \frac{pq}{(1 - ps)^2},$$

d'où $\psi'(1) = p/q = 1/\rho$. L'extinction est assurée si $\psi'(1) \leq 1$, d'où $\rho \geq 1$, ce qui correspond exactement à ce que nous avons trouvé.

(d) Ce problème est un peu plus épineux.

Rappel de notation : on notera $\tau_i = \inf \{t > 0 : X_t = i\}$ le premier temps auquel on atteint l'état i . Ce que l'on cherche est donc $\mathbb{E}_1[\tau_0]$.

On note $u_i = \mathbb{E}_i[\tau_0]$. Par le Lemme 3.3, on a immédiatement que $u_0 = 0$ et que pour tous $i \in \mathbb{N}$:

$$u_i = \frac{1}{i(\nu + \mu)} + pu_{i+1} + qu_{i-1}.$$

En réarrangeant ces termes, on trouve que

$$\begin{aligned} u_{i+1} - u_i &= \frac{q}{p}(u_i - u_{i-1}) - \frac{1}{p} \frac{1}{i(\nu + \mu)} \\ &= \frac{\mu}{\nu}(u_i - u_{i-1}) - \frac{1}{i\nu}. \end{aligned}$$

Si on note $\delta_i = u_i - u_{i-1}$, on trouve donc

$$\delta_{i+1} = -\frac{1}{i\nu} + \frac{\mu}{\nu}\delta_i,$$

et en itérant on vérifie (exercice!) que

$$\delta_i = \left(\frac{\mu}{\nu}\right)^i u_1 - \sum_{j=1}^i \frac{1}{j\nu} \left(\frac{\mu}{\nu}\right)^{i-j}.$$

Clairement, δ_i est positif puisque u_i est une suite croissante.

En particulier, lorsque $\mu = \nu$,

$$u_1 = \delta_{i+1} + \sum_{j=1}^i \frac{1}{j\nu} \geq \sum_{j=1}^i \frac{1}{j\nu}.$$

Comme c'est vrai pour tout i , et que la série $\sum_{j=1}^{\infty} \frac{1}{j\nu}$ diverge, on doit avoir que u_1 est infini ; donc lorsque $\mu = \nu$, le temps d'atteinte de 0 est infini.

Il reste le cas où $\nu < \mu$. Dans ce cas, on commence par remarquer que pour tout i , on a que pour tout i , $\delta_i \leq \delta_1 = u_1$.

En effet

$$\begin{aligned} u_i &= \mathbb{E}_i [\tau_0] = \mathbb{E}_i [\tau_{i-1}] + \mathbb{E}_{i-1} [\tau_0] \\ &= \mathbb{E}_i [\tau_{i-1}] + u_{i-1}. \end{aligned}$$

Intuitivement, on voit qu'on doit avoir $\mathbb{E}_i [\tau_{i-1}] \leq \mathbb{E}_1 [\tau_0]$ parce que toutes les transitions surviennent plus rapidement lorsqu'on passe de i à $i-1$ (vu qu'il y a plus d'individus).⁵

On rappelle que

$$\delta_i = \rho^i u_1 - \sum_{j=1}^i \frac{1}{j\nu} \rho^{i-j}.$$

En divisant par ρ^i de part et d'autre et en réarrangeant, on trouve que

$$u_1 = \sum_{j=1}^i \frac{1}{j\nu} \rho^{-j} + \rho^{-i} \delta_i.$$

D'une part, $\delta_i \leq u_1$, comme nous avons argué. D'autre part,

$$u_1 \geq \sum_{j=1}^i \frac{1}{j\nu} \rho^{-j}$$

On obtient donc que

$$\sum_{j=1}^i \frac{1}{j\nu} \rho^{-j} \leq u_1 \leq \sum_{j=1}^i \frac{1}{j\nu} \rho^{-j} + \rho^{-i} u_1.$$

Par le théorème de la sandwich, cela donne :

$$u_1 = \lim_{n \rightarrow \infty} \sum_{i=1}^n \frac{1}{i\nu} \rho^{-i} = \frac{1}{\nu} \sum_{i=1}^{\infty} \frac{(\rho^{-1})^i}{i}.$$

On rappelle la série de Taylor pour la fonction $-\log(1-x)$:

$$-\log(1-x) = \sum_{i=1}^{\infty} \frac{x^i}{i}.$$

Par comparaison, on trouve donc que

$$u_1 = -\frac{1}{\nu} \log(1 - \rho^{-1}).$$

Un processus de naissances et de morts comme celui de l'exemple 4.5 est appelé un processus **linéaire** parce que les taux de naissance et de mort augmentent linéairement en fonction de l'état. On peut rajouter une petite modification simple :

5. L'argument rigoureux ici est un peu compliqué et pas super important.

EXEMPLE 4.6. Soit $\mathbf{X} = (X_t)_{t \in \mathbb{R}^+}$ un processus de naissances et de morts avec taux de naissances $(\nu_i)_{i \in \mathbb{Z}^+}$ et taux de mort $(\mu_i)_{i \in \mathbb{N}}$. On dit que \mathbf{X} est un processus linéaire **avec immigration** si les taux de naissance et de mort sont donnés par :

$$\nu_k = \lambda + k\nu, \quad \mu_k = k\mu.$$

À long terme, peut-on estimer la taille de la population si on assume que $\mu \neq \nu$?

SOLUTION. Pour faire cet estimé, nous allons simplement calculer $\mathbb{E}_0[X_t]$, l'espérance du nombre d'individus au temps t . On a :

$$\mathbb{E}_i[X_t] = \sum_{j=1}^{\infty} j P_{i,j}^{(t)}.$$

On va maintenant utiliser l'équation progressive de Kolmogorov. On a que

$$\begin{aligned} \frac{d}{dt} P_{i,j}^{(t)} &= \sum_{k=0}^{\infty} P_{i,k}^{(t)} G_{k,j} = \nu_{j-1} P_{i,j-1}^{(t)} + \mu_{j+1} P_{i,j+1}^{(t)} - (\nu_j + \mu_j) P_{i,j}^{(t)} \\ &= (\lambda + (j-1)\nu) P_{i,j-1}^{(t)} + (j+1)\mu P_{i,j+1}^{(t)} - (\lambda + j\nu + j\mu) P_{i,j}^{(t)}. \end{aligned}$$

Donc, on a que

$$\begin{aligned} \frac{d}{dt} \mathbb{E}_i[X_t] &= \sum_{j=1}^{\infty} j \frac{d}{dt} P_{i,j}^{(t)} \\ &= \sum_{j=1}^{\infty} j \left((\lambda + (j-1)\nu) P_{i,j-1}^{(t)} + (j+1)\mu P_{i,j+1}^{(t)} - (\lambda + j\nu + j\mu) P_{i,j}^{(t)} \right) \\ &= \sum_{j=1}^{\infty} (j-1)(\lambda + (j-1)\nu) P_{i,j-1}^{(t)} + \sum_{j=1}^{\infty} (\lambda + (j-1)\nu) P_{i,j-1}^{(t)} \\ &\quad + \sum_{j=1}^{\infty} (j+1)(j+1)\mu P_{i,j+1}^{(t)} - \sum_{j=1}^{\infty} (j+1)\mu P_{i,j+1}^{(t)} \\ &\quad - \sum_{j=1}^{\infty} j(\lambda + j\nu) P_{i,j}^{(t)} - \sum_{j=1}^{\infty} j(j\mu) P_{i,j}^{(t)}. \end{aligned}$$

En numérotant de 1 à 6 sommations à droite de la dernière égalité, on peut les réduire de la façon suivante :

(1 + 5)

$$\begin{aligned} \sum_{j=1}^{\infty} (j-1)(\lambda + (j-1)\nu) P_{i,j-1}^{(t)} - \sum_{j=1}^{\infty} j(\lambda + j\nu) P_{i,j}^{(t)} &= - \sum_{j=1}^{\infty} \left[j(\lambda + j\nu) P_{i,j}^{(t)} - (j-1)(\lambda + (j-1)\nu) P_{i,j-1}^{(t)} \right] \\ &= -(1-1)(\lambda + (1-1)\nu) P_{i,1-1}^{(t)} = 0. \end{aligned}$$

(2)

$$\begin{aligned} \sum_{j=1}^{\infty} (\lambda + (j-1)\nu) P_{i,j-1}^{(t)} &= \lambda \sum_{j=0}^{\infty} P_{i,j}^{(t)} + \nu \sum_{j=0}^{\infty} j P_{i,j}^{(t)} \\ &= \lambda + \nu \mathbb{E}_i[X_t]. \end{aligned}$$

$$\begin{aligned}
(3+6) \quad \sum_{j=1}^{\infty} (j+1)(j+1)\mu P_{i,j+1}^{(t)} - \sum_{j=1}^{\infty} j(j\mu)P_{i,j}^{(t)} &= \sum_{j=1}^{\infty} \left[(j+1)(j+1)\mu P^{(t)} - i, j+1 - j(j\mu)P_{i,j}^{(t)} \right] \\
&= -\mu P_{i,1}^{(t)}.
\end{aligned}$$

$$\begin{aligned}
(4) \quad -\sum_{j=1}^{\infty} (j+1)\mu P_{i,j+1}^{(t)} &= -\sum_{j=2}^{\infty} j\mu P_{i,j}^{(t)} \\
&= -\mu \left(\sum_{j=1}^{\infty} jP_{i,j}^{(t)} - P_{i,1}^{(t)} \right) \\
&= -\mu \mathbb{E}_i [X_t] + \mu P_{i,1}^{(t)}.
\end{aligned}$$

Tout ceci ensemble donne finalement

$$\frac{d}{dt} \mathbb{E}_i [X_t] = \lambda + (\nu - \mu) \mathbb{E}_i [X_t];$$

si on écrit $M_i(t) = \mathbb{E}_i [X_t]$, on a en fait une équation différentielle inhomogène :

$$M_i'(t) = (\nu - \mu)M_i(t) + \lambda.$$

On a bien sûr que $M_i(0) = i$ comme condition initiale, puisque $\mathbb{E}_i [X_0] = i$.

On peut tenter une solution de la forme :

$$M_i(t) = F_i(t)e^{(\nu-\mu)t};$$

dans ce cas, on trouve que

$$\begin{aligned}
M_i'(t) &= F_i'(t)e^{(\nu-\mu)t} + (\nu - \mu)F_i(t)e^{(\nu-\mu)t} \\
&= (F_i'(t) + (\nu - \mu)F_i(t))e^{(\nu-\mu)t} \\
&= (\nu - \mu)F_i(t)e^{(\nu-\mu)t} + \lambda.
\end{aligned}$$

Pour que cette équation soit vérifiée, on doit avoir que

$$F_i'(t) = \lambda e^{-(\nu-\mu)t},$$

et bien sûr pour vérifier la condition initiale, on doit avoir $F_i(0) = i$. On a donc :

$$F_i(t) = i + \int_{u=0}^t \lambda e^{-(\nu-\mu)u} du = i - \frac{\lambda}{\nu - \mu} e^{-(\nu-\mu)t} + \frac{\lambda}{\nu - \mu}.$$

Finalement, on a donc :

$$M_i(t) = \left(i + \frac{\lambda}{\nu - \mu} \right) e^{(\nu-\mu)t} - \frac{\lambda}{\nu - \mu},$$

ou, en réarrangeant les termes :

$$M_i(t) = \frac{\lambda}{\mu - \nu} (1 - e^{(\nu-\mu)t}) + i e^{(\nu-\mu)t}.$$

On a donc finalement que

$$\lim_{t \rightarrow \infty} M_i(t) = \begin{cases} \infty & \text{si } \nu > \mu \\ \frac{\lambda}{\mu - \nu} & \text{si } \nu < \mu. \end{cases}$$

Autrement dit, à long terme, on s'attend à ce que la population se stabilise autour de $\lambda/(\mu - \nu)$.

4.3.3. Solutions et distributions stationnaires. Pour clore ce chapitre, nous allons discuter d'une solution stationnaire générale pour les processus de naissance et de morts.

PROPOSITION 4.9. *Soit $\mathbf{X} = (X_t)_{t \in \mathbb{R}^+}$ un processus de naissances et de morts avec taux de naissances et de morts $(\nu_i)_{i \in \mathbb{Z}^+}$ et $(\mu_i)_{i \in \mathbb{N}}$ respectivement.*

Alors, si on définit $\theta_0 = 1$ et

$$(4.3.2) \quad \theta_k = \frac{\prod_{i=0}^{k-1} \nu_i}{\prod_{j=1}^k \mu_j} = \frac{\nu_0 \cdot \nu_1 \cdot \nu_2 \cdots \nu_{k-1}}{\mu_1 \cdot \mu_2 \cdot \mu_3 \cdots \mu_k},$$

la famille $\theta = (\theta_k)_{k \in \mathbb{Z}^+}$ est une solution stationnaire de \mathbf{X} .

DÉMONSTRATION. Il suffit de vérifier les équations de stationnarité. Elles sont :

$$0 = -\nu_0 \theta_0 + \mu_1 \theta_1; \quad 0 = \nu_{k-1} \theta_{k-1} + \mu_{k+1} \theta_{k+1} - (\nu_k + \mu_k) \theta_k \quad \forall k \geq 1.$$

On trouve immédiatement que $\theta_1 = \frac{\nu_0}{\mu_0} \theta_0$.

Puis, en fixant $k \geq 1$, on peut additionner les équations de stationnarité pour tous $j \geq k$:

$$\begin{aligned} 0 &= \sum_{j \geq k} [\nu_{j-1} \theta_{j-1} - (\nu_j + \mu_j) \theta_j + \mu_{j+1} \theta_{j+1}] \\ &= \sum_{j \geq k} (\mu_{j+1} \theta_{j+1} - \mu_j \theta_j) - \sum_{j \geq k} (\nu_j \theta_j - \nu_{j-1} \theta_{j-1}) \\ &= -\mu_k \theta_k + \nu_{k-1} \theta_{k-1}. \end{aligned}$$

Ainsi, les θ_k satisfont :

$$\theta_k = \frac{\nu_{k-1}}{\mu_k} \theta_{k-1},$$

et en itérant, avec les conditions initiales trouvées, on a que

$$\theta_k = \theta_0 \cdot \frac{\nu_0 \cdot \nu_1 \cdot \nu_2 \cdots \nu_{k-1}}{\mu_1 \cdot \mu_2 \cdot \mu_3 \cdots \mu_k}.$$

On fixe $\theta_0 = 1$ et on a que notre famille $\theta = (\theta_k)_{k \in \mathbb{Z}^+}$ est une solution stationnaire de \mathbf{X} . □

COROLLAIRE 4.3. *La distribution stationnaire existe si et seulement si*

$$(4.3.3) \quad M = \sum_{k=0}^{\infty} \theta_k < +\infty.$$

Dans ce cas, elle est donnée par :

$$(4.3.4) \quad \pi_k = \frac{\theta_k}{M}.$$

EXEMPLE 4.7. On considère une file d'attente dans un magasin. Les client·e·s arrivent à des intervalles indépendants et sans mémoire d'espérance $1/\nu > 0$. Le préposé au comptoir sert chaque client·e, tour à tour, en prenant chaque fois un temps aléatoire sans mémoire d'espérance $1/\mu$. Lorsqu'un·e client·e arrive et qu'il y a déjà quelqu'un en file, le/ladit·e client·e se place en file et attend son tour.

On note X_t le nombre de personnes qui sont en file au temps t .

- (a) Donner les taux générateurs pour le processus X_t . De quel type de processus s'agit-il ?
- (b) Dans quels cas la distribution stationnaire existe-t-elle ? À quoi cela correspond-il ?
- (c) Si une distribution stationnaire existe, quelle est cette distribution ? Donner l'espérance du nombre de client·e·s en file.

SOLUTION. (a) Dans l'état 0, on transitionne vers l'état 1 avec un taux ν . Dans l'état k , pour $k \geq 1$, quelqu'un arrive dans la file (et on transitionne vers $k+1$) avec un taux ν . La personne au comptoir est servie (et s'en va) avec un taux μ (et on transitionne vers $k-1$).

Donc, on a :

$$G_{0,0} = -\nu, G_{0,1} = \nu; \quad G_{k,k-1} = \mu, G_{k,k} = -\mu - \nu, G_{k,k+1} = \nu.$$

- (b) On prend

$$\theta_k = \left(\frac{\nu}{\mu}\right)^k;$$

Pour que la distribution stationnaire existe, il faut que

$$M = \sum_{k=0}^{\infty} \theta_k < +\infty.$$

C'est le cas lorsque $\nu/\mu < 1$, soit lorsque $\nu < \mu$ – c'est à dire que si les client·e·s arrivent à un taux plus lent que le service. Dans ce cas, on a que

$$M = \frac{1}{1 - \nu/\mu} = \frac{\mu}{\mu - \nu},$$

et

$$\pi_k = \frac{\mu(\nu/\mu)^k}{\mu - \nu}$$

est une distribution stationnaire.

- (c) La distribution stationnaire est une distribution géométrique à partir de 0, avec paramètre $p = 1 - \frac{\nu}{\mu}$. l'espérance du nombre de client·e·s en file à long terme est

$$\sum_{k=0}^{\infty} k\pi_k = \frac{1-p}{p} = \frac{\nu/\mu}{1 - \nu/\mu} = \frac{\nu}{\mu - \nu}.$$

EXEMPLE 4.8. Expliquer pourquoi le nombre de personnes connectées en même temps sur un site web à un moment donné suit une loi de Poisson. Expliquer les hypothèses qu'il faut faire pour arriver à cette conclusion.

SOLUTION. On s'imagine que de nouveaux clients se connectent à notre serveur à intervalles indépendants et sans-mémoire à un taux de ν client·e·s par minute, par exemple. Ça semble une hypothèse raisonnable, vu que les gens ne se consultent pas avant de se connecter, et qu'on ne peut jamais prédire quand sera la prochaine connexion.

On peut aussi s'imaginer que chaque client, indépendamment des autres, passe un temps aléatoire d'espérance $1/\mu$. Ici, on va aussi supposer que ce temps est sans mémoire. Cette hypothèse-là est un peu plus stretchée, mais on va s'y tenir pour l'instant.

Dans ce contexte, le processus X_t qui donne le nombre de clients connectés au serveur au temps t est un processus de naissances et de morts. Les taux de naissance et de mort sont respectivement :

$$\nu_k = \nu, \quad \mu_k = k\mu.$$

La solution stationnaire donnée par la proposition 4.9 est :

$$\theta_k = \frac{(\nu/\mu)^k}{k!},$$

et donc on a toujours que

$$M = \sum_{k=0}^{\infty} \theta_k = e^{\nu/\mu}.$$

Finalement,

$$\pi_k = \frac{\theta_k}{M} = e^{-\nu/\mu} \frac{(\nu/\mu)^k}{k!},$$

et la distribution stationnaire est une distribution de Poisson de paramètre $\lambda = \nu/\mu$.

4.4. Exercices

EXERCICE 4.1 (LESSARD, ex. 2.11 – SOA M A06 #9). Un jeu dans un casino fait des paiements selon un processus de Poisson d'intensité 5 par heure, et le montant d'un paiement est de n dollars avec probabilité $(1/2)^n$. Les paiements sont indépendants les uns des autres.

- Quelle est la probabilité qu'il n'y ait aucun paiement de moins de 4 dollars dans une durée de 20 minutes ?
- Quelle est la probabilité qu'il n'y ait aucun paiement de plus de 3 dollars dans une période de 20 minutes ?
- Quelle est la probabilité qu'il n'y ait aucun paiement durant une période de 20 minutes ?

EXERCICE 4.2. Un **processus de morts** est un cas particulier d'un processus de naissances et de morts où $\nu_k = 0$ pour tout $k \in \mathbb{Z}^+$. Les éléments du générateur sont donc

$$G_{k,k} = -\mu_k, G_{k,k-1} = \mu_k.$$

- On dit qu'un processus de morts est linéaire lorsque $\mu_k = k\mu$ pour $\mu > 0$ et $k \in \mathbb{Z}^+$. Soit $\tau_0 = \inf \{t > 0 : X_t = 0\}$ le temps d'atteinte de l'état 0. Calculer $\mathbb{E}_n[\tau_0]$.
- Peut-il y avoir une explosion dans un processus de morts ? Pourquoi, ou pourquoi pas ?

EXERCICE 4.3. Une file d'attente $M/M/n$ est une file où

- les clients arrivent à intervalles indépendants sans mémoire de taux constant $\nu > 0$ (le premier M est pour *memoryless*) ;
 - les clients sont servis en des temps aléatoires sans mémoire de taux constant μ .
 - un maximum de n clients peuvent être servis en même temps.
- Donner les taux générateurs pour une file d'attente $M/M/n$.
 - Donner une solution stationnaire pour une file d'attente $M/M/2$. À quelle condition la distribution stationnaire existe-t-elle ?
 - Supposons que la distribution stationnaire existe ; on pose L le nombre moyen de clients en file, et W le temps moyen d'attente pour un client en régime stationnaire. Expliquer intuitivement la *formule de Little* :

$$L = \nu W.$$

EXERCICE 4.4 (LESSARD, ex. 2.21). On considère une station de taxis dans un aéroport où les taxis et les clients arrivent selon des processus de Poisson d'intensité 2 et 3 respectivement. On suppose qu'un taxi attend à la station, quel que soit le nombre de taxis présents lors de son arrivée. En revanche, si un client ne trouve pas de taxi à son arrivée, il décide d'utiliser un autre moyen de transport.

- Quelle est la proportion moyenne à long terme des clients qui prennent un taxi ?
- Quelle est le nombre moyen de taxis qui attendent à l'aéroport à long terme ?
- Lorsqu'un taxi arrive à un instant au hasard à la station, quel est le temps moyen que ce taxi doit attendre avant de servir un client ?

EXERCICE 4.5 (LESSARD, ex. 2.23). Déterminer, si elle existe, la distribution stationnaire d'un processus de naissances et de morts dont les taux de naissances et de morts sont donnés par

$$\nu_k = \nu, \quad \mu_k = \frac{k}{k+1}.$$

EXERCICE 4.6 (LESSARD, ex. 2.24). Une population reçoit des immigrants au taux de 1. De plus, chaque individu se reproduit à un taux constant égal à 2 et meurt à un taux donné par $1+k$ lorsqu'il y a k individus.

- (a) Donner les taux de naissances et de morts pour cette population.
- (b) Existe-t-il une distribution stationnaire ?

EXERCICE 4.7 (LESSARD, ex. 2.25). Des clients se présentent pour recevoir un service selon un processus de Poisson d'intensité λ . Ils sont servis un à la fois et le temps de service est exponentiel de paramètre μ . De plus, les clients qui attendent en file pour être servis deviennent impatients et ils quittent la file au taux δ indépendamment les uns des autres.

- (a) Donner la distribution stationnaire dans le cas où $\delta = \mu$.
- (b) Montrer que la distribution stationnaire existe pour tout $\delta > 0$.

EXERCICE 4.8 (Un principe de superposition général). Soient $\mathbf{X} = (X_t)_{t \in \mathbb{R}^+}$ et $\mathbf{Y} = (Y_t)_{t \in \mathbb{R}^+}$ deux processus de naissances et de morts dont les taux de naissance et de mort, pour \mathbf{X} puis pour \mathbf{Y} , sont donnés respectivement par $(\nu_k^{\mathbf{X}})_{k \in \mathbb{Z}^+}$, $(\mu_k^{\mathbf{X}})_{k \in \mathbb{Z}^+}$, $(\nu_k^{\mathbf{Y}})_{k \in \mathbb{Z}^+}$ et $(\mu_k^{\mathbf{Y}})_{k \in \mathbb{Z}^+}$. On définit $Z_t = X_t + Y_t$.

Montrer que $\mathbf{Z} = (Z_t)_{t \in \mathbb{R}^+}$ est aussi un processus de naissances et de morts. Quels sont les taux de naissance et de mort pour \mathbf{Z} ?

Deuxième partie

Structures de dépendance particulières

Chapitre 5

Processus de renouvellements

(À venir)

Les martingales

Dans les quatre premiers chapitres, nous nous sommes penchés sur les chaînes de Markov – c’est-à-dire un type particulier de processus stochastique, qui satisfait la propriété de Markov, telle que définie au chapitre 1. Dans le chapitre qui suit, nous nous intéresserons à une autre propriété intéressante : la propriété de *martingale*.

Un peu de contexte

Pour rappel, on s’intéresse à des processus stochastiques – c’est à dire des familles comme $\mathbf{X} = (X_t)_{t \geq 0}$ de variables aléatoires à valeur dans un espace d’états S , indicées par le temps t . Nous avons vu au chapitre 1 que l’on peut distinguer l’étude de ces processus selon :

- la topologie du *temps* : soit le temps est discret (c’est \mathbb{Z}^+), soit il est continu (c’est \mathbb{R}^+).
- la topologie de *l’espace* : soit l’espace S est discret, soit il est continu – nous n’avons encore pas touché à cette dernière possibilité, puisqu’il était beaucoup plus simple de travailler avec des fonctions de masse qu’avec des densités.

Dans la section 4.1.3, nous avons également introduit les notions d’incrément, et nous avons remarqué qu’il est souvent utile de préciser un peu plus les structures disponibles sur l’espace d’états S – en l’occurrence, pour définir les incréments et l’homogénéité spatiale, il fallait que S ait une structure « additive » – c’est-à-dire qu’on veut une définition adéquate de l’opérations d’addition dans S , avec un élément neutre, un inverse unique, etc.

Dans le contexte des martingales, le temps pourra être discret ou continu. Par contre, nous aurons besoin de quelques structures supplémentaires sur l’espace des états S :¹

- i. Nous aurons besoin que S soit muni d’opérations d’addition et de soustraction bien définies, comme précédemment.
- ii. Nous aurons besoin aussi d’une notion de « mesure » sur l’espace S , afin de pouvoir y définir une intégrale.²

Pour nos intérêts, il suffira amplement de choisir $S = \mathbb{R}^n$; dorénavant dans ce chapitre, nous travaillerons donc toujours dans l’espace des états \mathbb{R}^n . La variable X_t prendra donc des valeurs dans \mathbb{R}^n , et on aura donc accès aux notions et aux quantités suivantes :

- Les *incrément*s : $X_t - X_s$;
- La *norme* : $\|X_t\|$;
- L’*espérance* : $\mathbb{E}[X_t]$.

1. À noter qu’on peut également fournir une définition de martingale à temps continu dans le cas où S est une variété différentiable dont l’espace tangent possède en tout point ces propriétés. Pour en apprendre plus sur les variétés différentiables... il faudra faire un cours de géométrie différentielle!

2. Vous pouvez consulter [les notes d’André Giroux](#) au sujet de la théorie de la mesure.

Pour la grande majorité des exemples vus dans ce chapitre-ci, nous travaillerons en temps discret. Toutefois, *a priori*, la théorie est à peu près la même en temps continu. Lorsque les distinctions sont importantes, elles seront mentionnées explicitement.

6.1. Définition et exemples

DÉFINITION 6.1. Soient $\mathbf{X} = (X_t)_{t \geq 0}$ et $\mathbf{Y} = (Y_t)_{t \geq 0}$ deux familles de variables aléatoires indicées par le temps t (discret ou continu), à valeurs dans \mathbb{R}^n (les distributions peuvent être discrètes ou continues), avec $Y_0 = 0$.

- i. On dit que \mathbf{X} est **adaptée** à \mathbf{Y} si, pour tout $t \geq 0$, il existe une fonction $f : \mathbb{R}^{n \cdot t} \rightarrow \mathbb{R}^n$ telle que

$$X_t = f(\mathbf{Y}_{\leq t}).$$

- ii. On dit que \mathbf{X} est **intégrable** au temps t (ou simplement que X_t est intégrable) si

$$\mathbb{E}[\|X_t\|] < +\infty.$$

- iii. On dit que \mathbf{X} a la **propriété de martingale** par rapport à \mathbf{Y} si, pour tous temps s, t avec $0 \leq s \leq t$, on a

$$(6.1.1) \quad \mathbb{E}[X_t \mid \mathbf{Y}_{\leq s}] = X_s.$$

Le processus \mathbf{X} est une **martingale** par rapport à la famille \mathbf{Y} s'il est *adapté* à \mathbf{Y} , intégrable pour tout t et possède la propriété de martingale par rapport à \mathbf{Y} . On dit que \mathbf{X} est simplement une martingale (sans préciser) si elle est une martingale *par rapport à elle-même*.

6.1.1. Commentaires sur la définition.

Adaptabilité. Avant d'aller plus loin, il convient de fournir une intuition de ce qui se passe ici. Dans la définition 6.1, il y a deux familles de variables aléatoires à démêler.

- $\mathbf{X} = (X_t)_{t \geq 0}$ est le processus auquel on s'intéresse.
- $\mathbf{Y} = (Y_t)_{t \geq 0}$ représente une suite ou une famille de variables aléatoires qui encode, à chaque instant de temps, la « nouvelle information dont on dispose au temps t ».
- $\mathbf{Y}_{\leq t}$ représente donc l'ensemble des informations dont on dispose au temps t .

Intuitivement, \mathbf{X} est *adapté* à \mathbf{Y} seulement si on peut connaître entièrement la valeur de X_t à partir des informations contenues dans $\mathbf{Y}_{\leq t}$. Autrement dit, pour tout temps t , on peut calculer X_t en fonction des valeurs des Y_u pour $u \leq t$.

EXEMPLE 6.1. On joue à « Pile ou Face » de façon répétée. On définit $Y_t = 1$ si le t ème lancer donne Pile, et $Y_t = 0$ si le t ème lancer donne Face.

- (a) Soit X_t le nombre de Pile obtenus jusqu'à la t ème partie inclusivement. Est-ce que $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ est adapté à $\mathbf{Y} = (Y_t)_{t \in \mathbb{Z}^+}$?
- (b) Est-ce que \mathbf{Y} est adapté à \mathbf{X} ?
- (c) On définit Z_t le nombre de lancers, immédiatement après le t ème lancer, qu'il faudra réaliser pour obtenir Pile à nouveau. Est-ce que $\mathbf{Z} = (Z_t)_{t \geq 0}$ est adapté à \mathbf{Y} ?

SOLUTION. (a) Si X_t est le nombre de Pile obtenus jusqu'à la t ème partie inclusivement, alors on a que

$$X_t = \sum_{i=1}^t Y_i$$

(et $X_0 = 0$). On voit bien que X_t est entièrement déterminée par les valeurs de $\mathbf{Y}_{\leq t}$. Donc, \mathbf{X} est adapté à \mathbf{Y} .

- (b) De même, il est clair que $Y_t = X_t - X_{t-1}$, qui ne dépend que des valeurs de $\mathbf{X}_{\leq t}$. Donc, dans ce cas-ci, \mathbf{Y} est adapté à \mathbf{X} aussi. *Attention!* Ça n'est pas toujours le cas!
- (c) La variable Z_t donne le nombre de lancers, immédiatement après le t ème, qu'il faut encore réaliser pour obtenir Pile à nouveau. Comme nous avons vu à l'exemple 1.4, Z_t est bel et bien une chaîne de Markov; même si Z_t « dépend du futur » (on ne peut pas connaître Z_t sans savoir ce qui se passera dans les lancers subséquents), ce futur est indépendant du passé lorsqu'on connaît le présent.

Toutefois, \mathbf{Z} n'est *pas* adapté à \mathbf{Y} , puisque justement, dans le cas où $Z_{t-1} = 1$, Z_t est entièrement indépendant de $\mathbf{Y}_{\leq t}$; en effet, lorsqu'on vient juste d'obtenir Pile, connaître les parties qu'il y a eu *avant* le temps t ne nous aide en rien à prédire combien ça prendra de temps avant d'avoir Pile à nouveau.

On remarque en outre que toute famille $\mathbf{X} = (X_t)_{t \geq 0}$ est forcément adaptée à elle-même.

Intégrabilité. On requiert que X_t soit intégrable afin que l'espérance conditionnelle soit bien définie; si ce n'était pas le cas, on courrait le risque que la propriété de martingale n'ait pas vraiment de sens!

Équivalences pour la propriété de martingale par rapport à \mathbf{Y} . Il s'agit du gros morceau de notre définition. Nous allons maintenant voir un peu plus intuitivement ce que la propriété de martingale signifie, d'un point de vue plus intuitif. Supposons que $\mathbf{X} = (X_t)_{t \geq 0}$ est une martingale par rapport à $\mathbf{Y} = (Y_t)_{t \geq 0}$.

Le tout premier constat à faire, c'est que, par définition des espérances conditionnelles, on a que pour tout s ,

$$\mathbb{E}[X_s \mid \mathbf{Y}_{\leq s}] = X_s$$

Il suit donc que, lorsque l'on a la propriété de martingale par rapport à \mathbf{Y} , pour tous s, t avec $0 \leq s \leq t$, on a :

$$\mathbb{E}[X_t - X_s \mid \mathbf{Y}_{\leq s}] = 0.$$

En fait, on a même l'équivalence :

LEMME 6.1. *Soit $\mathbf{X} = (X_t)_{t \geq 0}$ un processus intégrable pour tout t et adapté à $\mathbf{Y} = (Y_t)_{t \geq 0}$.*

\mathbf{X} a la propriété de martingale par rapport à \mathbf{Y} si et seulement si, pour tous s, t tels que $0 \leq s \leq t$, on a :

$$(6.1.2) \quad \mathbb{E}[X_t - X_s \mid \mathbf{Y}_{\leq s}] = 0.$$

DÉMONSTRATION. On a :

$$\begin{aligned} \mathbb{E}[X_t \mid \mathbf{Y}_{\leq s}] &= \mathbb{E}[X_t - X_s + X_s \mid \mathbf{Y}_{\leq s}] \\ &= \mathbb{E}[X_t - X_s \mid \mathbf{Y}_{\leq s}] + X_s, \end{aligned}$$

où la seconde égalité est par linéarité et puisque \mathbf{X} est adapté à \mathbf{Y} .

Clairement, la propriété de martingale est donc logiquement équivalente à la condition que

$$\mathbb{E}[X_t - X_s \mid \mathbf{Y}_{\leq s}] = 0.$$

□

Mais on peut aller plus loin ! En effet, la propriété de martingale interagit de façon amusante avec la propriété de Markov :

LEMME 6.2. *Soit $\mathbf{X} = (X_t)_{t \geq 0}$ une chaîne de Markov sur \mathbb{R}^n intégrable pour tout t . Alors, \mathbf{X} est une **martingale** (par rapport à elle-même) si et seulement si elle satisfait*

$$(6.1.3) \quad \mathbb{E}[X_t \mid X_s] = X_s.$$

DÉMONSTRATION. \mathbf{X} est adaptée à elle-même et intégrable. Par la propriété de Markov,

$$\mathbb{E}[X_t \mid \mathbf{X}_{\leq s}] = \mathbb{E}[X_t \mid X_s].$$

Donc, on a la propriété de martingale (par rapport à elle-même) si et seulement si on a

$$\mathbb{E}[X_t \mid X_s] = X_s.$$

□

Finalement, la propriété d'homogénéité spatiale (introduite au paragraphe 4.1.3) permet d'aller encore plus loin dans les équivalences :

LEMME 6.3. *Soit $\mathbf{X} = (X_t)_{t \geq 0}$ une chaîne de Markov spatialement homogène sur \mathbb{R}^n et intégrable pour tout t .*

Alors, \mathbf{X} est une martingale si et seulement si, pour tous s, t tels que $0 \leq s \leq t$:

$$(6.1.4) \quad \mathbb{E}[X_t - X_s] = 0.$$

DÉMONSTRATION. Par la propriété de Markov, on a que

$$\mathbb{E}[X_t - X_s \mid \mathbf{X}_{\leq s}] = \mathbb{E}[X_t - X_s \mid X_s],$$

mais puisque \mathbf{X} est spatialement homogène, par le lemme 4.2³, on a que forcément, $X_t - X_s$ est indépendant de X_s , et

$$\mathbb{E}[X_t - X_s \mid X_s] = \mathbb{E}[X_t - X_s].$$

Par le lemme 6.1, on vérifie la propriété de martingale si et seulement si $\mathbb{E}[X_t - X_s \mid \mathbf{X}_{\leq s}] = \mathbb{E}[X_t - X_s] = 0$. □

3. Le lemme 4.2 est dans le cas d'un espace d'états discret S ; notre définition de l'homogénéité spatiale est aussi faite dans ce cadre ; pourtant, on peut éa définir de façon plus générale aussi pour des espace d'états continus ; l'intuition demeure la même.

Le cas du temps discret. Pour terminer cette discussion, nous allons introduire également une équivalence qui nous simplifiera beaucoup la vie dans le cas où le temps est discret :

LEMME 6.4. *Soit $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ un processus stochastique à temps discret, adapté à $\mathbf{Y} = (Y_t)_{t \in \mathbb{Z}^+}$ et intégrable en tout $t \in \mathbb{Z}^+$.*

Alors \mathbf{X} est une martingale par rapport à \mathbf{Y} si et seulement si, pour tout $t \in \mathbb{Z}^+$, on a

$$(6.1.5) \quad \mathbb{E}[X_{t+1} \mid \mathbf{Y}_{\leq t}] = X_t.$$

DÉMONSTRATION. Il est évident que si \mathbf{X} est une martingale par rapport à \mathbf{Y} , il suffit de choisir $t = s + 1$ dans la définition pour obtenir (6.1.5). Nous allons maintenant montrer que ça marche aussi dans l'autre sens.

Supposons qu'on a déjà vérifié que

$$\mathbb{E}[X_{t+k} \mid \mathbf{Y}_{\leq t}] = X_t$$

pour tout $k \leq n$ et pour tous t . On a alors⁴ :

$$\mathbb{E}[X_{t+n+1} \mid \mathbf{Y}_{\leq t}] = \mathbb{E}[\mathbb{E}[X_{t+1+n} \mid \mathbf{Y}_{\leq t+1}] \mid \mathbf{Y}_{\leq t}]$$

Or, bien sûr, par notre hypothèse d'induction, on a déjà que

$$\mathbb{E}[X_{t+1+n} \mid \mathbf{Y}_{\leq t+1}] = X_{t+1}.$$

On a donc :

$$\mathbb{E}[X_{t+n+1} \mid \mathbf{Y}_{\leq t}] = \mathbb{E}[X_{t+1} \mid \mathbf{Y}_{\leq t}].$$

On complète en utilisant (6.1.5).

On a donc montré que si on a que

$$\mathbb{E}[X_{t+k} \mid \mathbf{Y}_{\leq t}] = X_t$$

pour tout t et pour tout $k \leq n$, c'est aussi vrai pour tout t et pour $k = n + 1$. Par induction, puisque c'est vrai pour $k = 0$ et pour $k = 1$ (par (6.1.5)), c'est donc vrai pour tout $t \in \mathbb{Z}^+$ et pour tout $k \in \mathbb{Z}^+$; autrement dit, pour tout $s \in \mathbb{Z}^+$ et pour tout $t \geq s$, on a que

$$\mathbb{E}[X_t \mid \mathbf{Y}_{\leq s}] = X_s.$$

□

COROLLAIRE 6.1. *Les lemmes 6.1 à 6.3 sont encore vrais si on remplace t par $s + 1$ lorsque le temps est discret.*

La proposition suivante fait un bref résumé des équivalences pertinentes pour la condition de martingale :

PROPOSITION 6.1. *Soit $\mathbf{X} = (X_t)_{t \geq 0}$ un processus stochastique adapté à $\mathbf{Y} = (Y_t)_{t \geq 0}$ (avec $Y_0 = 0$) et intégrable pour tout $t \geq 0$.*

*\mathbf{X} est une **martingale** si et seulement si au moins l'une des conditions suivantes est remplie ; si l'une d'elle est remplie, toutes les suivantes le sont aussi.*

4. Souvenons-nous que, de façon générale, on a que

$$\mathbb{E}[\mathbb{E}[Z \mid X, Y] \mid X] = \mathbb{E}[Z \mid X].$$

Consulter les notes de MAT1720 pour de plus amples détails sur cette propriété des espérances conditionnelles.

- i. \mathbf{X} est une chaîne de Markov spatialement homogène, $\mathbf{Y} = \mathbf{X}$ et pour tous s, t avec $s \leq t$ (ou seulement $t = s + 1$ en temps discret) :

$$\mathbb{E}[X_t - X_s] = 0 \quad \Leftrightarrow \quad \mathbb{E}[X_t] = \mathbb{E}[X_s]$$

- ii. \mathbf{X} est une chaîne de Markov, $\mathbf{Y} = \mathbf{X}$ et pour tous s, t avec $s \leq t$ (ou seulement $t = s + 1$ en temps discret) :

$$\mathbb{E}[X_t \mid X_s] = X_s \quad \Leftrightarrow \quad \mathbb{E}[X_t - X_s \mid X_s] = 0.$$

- iii. Pour tous s, t avec $s \leq t$ (ou seulement $t = s + 1$ en temps discret) :

$$\mathbb{E}[X_t - X_s \mid \mathbf{Y}_{\leq s}] = 0 \quad \Leftrightarrow \quad \mathbb{E}[X_t \mid \mathbf{Y}_{\leq s}] = X_s.$$

DÉMONSTRATION. La preuve repose directement sur les lemmes 6.1 à 6.4 et le corollaire 6.1. \square

COROLLAIRE 6.2. Soit $\mathbf{X} = (X_t)_{t \geq 0}$ une martingale par rapport à $\mathbf{Y} = (Y_t)_{t \geq 0}$. Alors, $\mathbb{E}[X_t] = \mathbb{E}[X_0] = X_0$ pour tout $t \geq 0$.

DÉMONSTRATION. Il suffit de constater que pour tous s, t tels que $0 \leq s \leq t$:

$$\mathbb{E}[X_t] = \mathbb{E}[\mathbb{E}[X_t \mid \mathbf{Y}_{\leq s}]] = \mathbb{E}[X_s];$$

on a donc que l'espérance de X_t est constante en t , et on conclue en prenant $s = 0$. \square

REMARQUE. *Attention !* Il ne suit pas forcément que l'inverse est vrai en général. Un processus arbitraire n'est pas nécessairement une martingale pour autant lorsque ses incréments sont d'espérance nulle ; cela n'est vérifié que lorsque \mathbf{X} est une chaîne de Markov spatialement homogène.

6.1.2. Exemples. Nous allons maintenant nous pencher sur quelques exemples simples.

EXEMPLE 6.2. On définit $\mathbf{Y} = (Y_t)_{t \in \mathbb{N}}$ une suite de variables aléatoires indépendantes et identiquement distribuées sur \mathbb{R}^n . Puis, on définit :

$$X_t = \sum_{i=1}^t Y_i.$$

Le processus $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ est adapté à \mathbf{Y} ; il est également intégrable pourvu que $\mathbb{E}[\|Y_1\|] < +\infty$.

On appelle ce type de processus une *marche aléatoire spatialement homogène* sur \mathbb{R}^n ; c'est une chaîne de Markov homogène à temps discret, spatialement homogène. Donc, on a que \mathbf{X} est une martingale si et seulement si $\mathbb{E}[Y_1] = 0$.

Par exemple, La marche aléatoire symétrique sur \mathbb{Z}^d est une martingale.

EXEMPLE 6.3. On considère $\mathbf{Y} = (Y_t)_{t \in \mathbb{N}}$ une suite de variables aléatoires indépendantes et identiquement distribuées avec

$$\mathbb{P}\{Y_1 = 1\} = p \quad \mathbb{P}\{Y_1 = -1\} = q = 1 - p,$$

et on définit $X_t = \sum_{i=1}^t Y_i$.

Le processus $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ est une martingale lorsque $p = q$; autrement, non.

Or, si on pose $\rho = q/p$, Le processus $\mathbf{Z} = (Z_t = \rho^{X_t})_{t \in \mathbb{Z}^+}$ est une martingale par rapport à \mathbf{Y} . En effet,

$$\begin{aligned} \mathbb{E}[Z_{t+1} \mid \mathbf{Y}_{\leq t}] &= \mathbb{E}[Z_t \rho^{Y_{t+1}} \mid \mathbf{Y}_{\leq t}] \\ &= Z_t \mathbb{E}[\rho^{Y_{t+1}}] \\ &= Z_t(p\rho + q\rho^{-1}) \\ &= Z_t \end{aligned}$$

EXEMPLE 6.4. Soit $\mathbf{Z} = (Z_t)_{t \in \mathbb{Z}^+}$ un processus de Galton-Watson, avec ψ la fonction génératrice des probabilités pour le nombre d'enfants d'un individu, et $m = \psi'(1)$.

On définit $X_t = Z_t/m^t$. Alors, $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ est une martingale par rapport à \mathbf{Z} ; en effet, d'une part \mathbf{X} est adaptée à \mathbf{Z} , et on sait également que \mathbf{X} est intégrable en tout t puisque \mathbf{Z} l'est aussi. Finalement, on a que :

$$\begin{aligned} \mathbb{E}[X_{t+1}] &= \frac{1}{m^{t+1}} \mathbb{E}[Z_{t+1} \mid \mathbf{Z}_{\leq t}] \\ &= \frac{1}{m^{t+1}} \mathbb{E}\left[\sum_{i=1}^{Z_t} \xi_{t,i} \mid Z_t\right] \\ &= \frac{1}{m^{t+1}} Z_t \mathbb{E}[\xi_{0,1}] \\ &= \frac{Z_t}{m^t} = X_t. \end{aligned}$$

6.2. Temps d'arrêt

On introduit maintenant une notion primordiale pour la suite de notre étude des martingales – mais ça vaut la peine de s'y arrêter un instant.

DÉFINITION 6.2. Soit $\mathbf{Y} = (Y_t)_{t \geq 0}$ une famille de variables aléatoires.

On dit que la variable aléatoire τ est un **temps d'arrêt** par rapport à \mathbf{Y} lorsque, pour tout t , l'événement $\{\tau > t\}$ peut être entièrement décrit en fonction de $\mathbf{Y}_{\leq t}$.⁵

Avant d'aller plus loin, il convient de donner une intuition pour cette définition. Comme au début de la section 6.1.1, $\mathbf{Y} = (Y_t)_{t \in \mathbb{Z}^+}$ représente une famille de variables aléatoires qui encode à chaque temps la nouvelle information dont on dispose. Ainsi, $\mathbf{Y}_{\leq t}$ contient toute l'information dont on dispose au temps t .

Dans ce contexte, la condition que τ soit un *temps d'arrêt* signifie, intuitivement, qu'on peut savoir, au temps t , si le temps τ est déjà survenu ou pas. On dit que τ est un temps *d'arrêt* parce que c'est un temps auquel on peut « arrêter » un chronomètre imaginaire aussitôt qu'il survient, parce qu'on sait au temps t si le temps est déjà survenu ou pas.

EXEMPLE 6.5. (a) Disons qu'on note $Y_i = 1$ si il neige le jour i , et $Y_i = 0$ si il ne neige pas ce jour-là. On peut définir « la date de la première neige » :

$$\tau = \min \{i \in \mathbb{N} : Y_i = 1\}.$$

5. Dans le langage plus précis de la théorie de la mesure, on dira que l'événement $\{\tau \geq t\}$ est *mesurable* par rapport à la tribu engendrée par $\mathbf{Y}_{\leq t}$.

Alors, τ est un temps d'arrêt par rapport à $\mathbf{Y} = (Y_i)_{i \in \mathbb{N}}$; en effet, on a :

$$\{\tau > t\} = \{Y_1 = Y_2 = Y_3 = \dots = Y_t = 0\};$$

autrement dit, au jour t , on peut toujours déterminer si la date de la première neige est survenue ou pas ; on n'a pas besoin d'attendre le futur pour connaître la valeur exacte de τ – dès qu'il y a de la neige, on peut le marquer sur notre calendrier.

(b) Dans ce même scénario, si on définit par exemple

$$\tau_N = \min \{i \in \mathbb{N} : Y_i = Y_{i+1} = Y_{i+2} = \dots = Y_{i+N-1} = 1\},$$

qui correspond à la date du début de la toute première séquence de N journées de neige consécutives, pour $N > 1$, alors, τ_N n'est *pas* un temps d'arrêt par rapport à \mathbf{Y} . Pour bien comprendre pourquoi, imaginons qu'on fixe $N = 3$. S'il a neigé pour la première fois le 16 novembre, est-ce que $\tau_N = 16$ novembre ? On ne le sait pas ! Et si il neige aussi le 17 ? On ne le sait toujours pas ! Mais si il neige le 18, alors là, OUI ! On sait maintenant (rendus le 18 novembre) que $\tau_3 = 16$ novembre. τ_N n'est pas un temps d'arrêt parce qu'on n'aurait pas pu arrêter un chronomètre au temps τ_N .

(c) De même, la date de la dernière neige de l'hiver n'est pas un temps d'arrêt par rapport à \mathbf{Y} , parce qu'on ne peut pas savoir quelle date c'est avant la fin de l'hiver.

6.2.1. La propriété forte de Markov. On se souvient qu'on avait que la propriété *faible* de Markov était :

DÉFINITION (1.3). Soit $\mathbf{X} = (X_t)_{t \geq 0}$ un processus stochastique quelconque (à temps discret ou continu, sur un espace d'états S discret ou continu).

On dit que le processus \mathbf{X} possède la **propriété de Markov faible** si, pour tout temps $t \geq 0$, on a que

$$\mathbb{P}\{\mathbf{X}_{>t} \in A \mid X_t = i, \mathbf{X}_{<t} \in B\} = \mathbb{P}\{\mathbf{X}_{>t} \in A \mid X_t = i\},$$

pour tous $A \subseteq S^{\{u: u > t\}}$ et $B \subseteq S^{\{u: 0 \leq u < t\}}$.

La propriété *forte* de Markov, c'est la même chose, mais on peut remplacer le temps t par un temps d'arrêt !

DÉFINITION 6.3. Soit $\mathbf{X} = (X_t)_{t \geq 0}$ un processus stochastique quelconque (à temps discret ou continu, sur un espace d'états S discret ou continu).

On dit que le processus \mathbf{X} possède la **propriété de Markov forte** si, pour tout temps d'arrêt τ (par rapport à \mathbf{X}), on a que

$$\mathbb{P}\{\mathbf{X}_{>\tau} \in A \mid X_\tau = i, \mathbf{X}_{<\tau} \in B\} = \mathbb{P}\{\mathbf{X}_{>\tau} \in A \mid X_\tau = i\}.$$

Mais existe-t-il une réelle différence entre les propriétés de Markov faibles et fortes ? En temps discret, *non* ! Voici une proposition (laissée sans démonstration) :

PROPOSITION 6.2. Soit $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ un processus stochastique à temps discret sur l'espace des états S . Alors, \mathbf{X} a la propriété de Markov forte si et seulement si \mathbf{X} a la propriété de markov faible.

6.2.2. La formule de Wald. On se souvient que, en général, si on a une famille $\mathbf{Y} = (Y_i)_{i \in \mathbb{N}}$ de variables aléatoires indépendantes identiquement distribuées, et une variable aléatoire entière positive N indépendante de \mathbf{Y} , alors on a que

$$\begin{aligned} \mathbb{E} \left[\sum_{i=1}^N Y_i \right] &= \mathbb{E} \left[\mathbb{E} \left[\sum_{i=1}^N Y_i \mid N \right] \right] \\ &= \mathbb{E} \left[\sum_{i=1}^N \mathbb{E}[Y_i \mid N] \right] \\ &= \mathbb{E} \left[\sum_{i=1}^N \mathbb{E}[Y_i] \right] \\ &= \mathbb{E}[N \mathbb{E}[Y_1]] \\ &= \mathbb{E}[N] \mathbb{E}[Y_1]. \end{aligned}$$

Ce que nous allons regarder maintenant, c'est ce qu'on peut faire lorsque N n'est pas forcément indépendante de \mathbf{Y} – mais si c'est un temps d'arrêt.

LEMME 6.5. *Soit $\mathbf{Y} = (Y_t)_{t \in \mathbb{Z}^+}$ une suite de variables aléatoires indépendantes et identiquement distribuées. Soit N un temps d'arrêt par rapport à \mathbf{Y} .*

Alors, les variables aléatoires Y_t et $\mathbb{1}_{\{N \geq t\}}$ sont indépendantes. En particulier,

$$\mathbb{E}[Y_t \mathbb{1}_{\{N \geq t\}}] = \mathbb{E}[Y_t] \mathbb{P}\{N \geq t\}.$$

DÉMONSTRATION. L'événement $\{N \geq t\}$ est entièrement déterminé par $\mathbf{Y}_{\leq t-1}$, et puisque \mathbf{Y} est une famille indépendante, il suit que $\mathbb{1}_{\{N \geq t\}}$ est indépendant de Y_t . \square

Ce lemme a pour conséquence le résultat suivant :

PROPOSITION 6.3. *Soit $\mathbf{Y} = (Y_i)_{i \in \mathbb{Z}^+}$ une suite de variables aléatoires indépendantes et identiquement distribuées avec $\mathbb{E}[Y_1] = \mu$. Soit N un temps d'arrêt par rapport à \mathbf{Y} .*

Alors,

$$(6.2.1) \quad \mathbb{E} \left[\sum_{i=1}^N Y_i \right] = \mu \mathbb{E}[N].$$

DÉMONSTRATION. On a que

$$\begin{aligned} \mathbb{E} \left[\sum_{i=1}^N Y_i \right] &= \mathbb{E} \left[\sum_{i=1}^{\infty} Y_i \mathbb{1}_{\{N \geq i\}} \right] \\ &= \sum_{i=1}^{\infty} \mathbb{E}[Y_i \mathbb{1}_{\{N \geq i\}}] \\ &= \mu \sum_{i=1}^{\infty} \mathbb{P}\{N \geq i\} \\ &= \mu \mathbb{E}[N]. \end{aligned}$$

\square

L'équation (6.2.1) est appelée *formule de Wald*.

EXEMPLE 6.6. Une pièce de monnaie tombe sur Pile avec probabilité p et sur Face avec probabilité $1-p$. On joue jusqu'à obtenir k fois Pile. Soit N le nombre de lancers nécessaires. Calculer $\mathbb{E}[N]$.

SOLUTION. On pourrait faire deux arguments ; on pourrait voir N comme une somme de k variables géométriques indépendantes, chacune d'espérance $1/p$. Ou...

Si $Y_i = 1$ lorsque le i ème lancer retombe sur Pile, et 0 sinon, on voit que N est un temps d'arrêt par rapport à $\mathbf{Y} = (Y_i)_{i \in \mathbb{N}}$. On a :

$$\sum_{i=1}^N Y_i = k.$$

Par la formule de Wald, on a immédiatement :

$$p\mathbb{E}[N] = k,$$

d'où $\mathbb{E}[N] = k/p$.

6.3. Le théorème d'arrêt.

La notion de temps d'arrêt en elle-même n'a *a priori* rien à voir avec les martingales ; les définitions sont logiquement indépendantes l'une de l'autre, et on aurait très bien pu présenter la notion de temps d'arrêt beaucoup plus tôt – au premier chapitre, par exemple.

Si nous la présentons maintenant, c'est que cette notion a une application toute particulière et très puissante dans le contexte des martingales : le *théorème d'arrêt de Doob*, que l'on appelle parfois aussi le théorème d'arrêt optionnel.

THÉOREME 6.1 (d'arrêt de Doob). Soit $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ une martingale à temps discret par rapport à la suite $\mathbf{Y} = (Y_t)_{t \in \mathbb{Z}^+}$, et soit τ un temps d'arrêt par rapport à \mathbf{Y} .

On définit $M_t = X_{t \wedge \tau}$; le processus $\mathbf{M} = (M_t)_{t \in \mathbb{Z}^+}$ est appelé le **processus \mathbf{X} arrêté en τ** .

Alors, \mathbf{M} est aussi une martingale par rapport à \mathbf{Y} dès que l'une des deux conditions suivantes est remplie :

- i. $\mathbb{P}\{\tau < +\infty\} = 1$ et il existe $K > 0$ tel que $|M_t| \leq K$ pour tout $t \in \mathbb{Z}^+$; ou
- ii. $\mathbb{E}[\tau] < +\infty$ et il existe $K > 0$ tel que $|M_t - M_{t-1}| \leq K$ pour tout $t \in \mathbb{N}$.

DÉMONSTRATION. Il suffit de vérifier dans l'ordre les conditions énoncées dans la définition d'une martingale.

Étape 1 : \mathbf{M} est adaptée à \mathbf{Y} .

Pour montrer cela, il suffit de constater que l'on a :

$$M_t = X_{t \wedge \tau} = X_t \mathbb{1}_{\{\tau > t\}} + X_\tau \mathbb{1}_{\{\tau \leq t\}};$$

\mathbf{X} étant une martingale par rapport à \mathbf{Y} , elle est adaptée à celle-ci, et on a donc que la variable X_t est déterminée entièrement par $\mathbf{Y}_{\leq t}$, de même que les variables $\mathbb{1}_{\{\tau > t\}}$.

Il reste $X_\tau \mathbb{1}_{\{\tau \leq t\}}$. Ici c'est plus subtil. On a que

$$X_\tau \mathbb{1}_{\{\tau \leq t\}} = \begin{cases} X_u & \text{si } \tau = u \leq t ; \\ 0 & \text{si } \tau = u > t \end{cases}.$$

Puisque la connaissance de $\mathbf{Y}_{\leq t}$ permet de déterminer si $\tau \leq t$, et que, sachant cela, on peut aussi forcément déterminer X_u lorsque $\tau = u \leq t$, il suit forcément qu'on peut nécessairement déterminer $X_\tau \mathbb{1}_{\{\tau \leq t\}}$ en sachant $\mathbf{Y}_{\leq t}$.

Donc, \mathbf{M} est adaptée à \mathbf{Y} .

Étape 2 : Le processus \mathbf{M} est intégrable.

Pour cela, il faut montrer que $\mathbb{E}[|M_t|] < +\infty$ pour tout t . Bien sûr, lorsque $|M_t| \leq K$ pour tout t , c'est immédiat.

Lorsque $\mathbb{E}[\tau] < +\infty$ et que $|M_t - M_{t-1}| \leq K$ pour tout t , on a que

$$\begin{aligned} \mathbb{E}[|M_t|] &= \mathbb{E}[|X_t \mathbb{1}_{\{\tau > t\}} + X_\tau \mathbb{1}_{\{\tau \leq t\}}|] \\ &\leq \mathbb{E}[|X_t|] + \mathbb{E}[|X_\tau|] \\ &= \mathbb{E}[|X_t|] + \mathbb{E}\left[\left|X_0 + \sum_{u=1}^{\tau} (X_u - X_{u-1})\right|\right] \\ &\leq \mathbb{E}[|X_0|] + \mathbb{E}[|X_t|] + \mathbb{E}\left[\sum_{u=1}^{\tau} |X_u - X_{u-1}|\right]. \end{aligned}$$

Puisque \mathbf{X} est une martingale, les deux premiers termes après la dernière égalité sont finis ; il suffit donc de montrer que le dernier l'est aussi ; or, pour tout $u \leq \tau$, $M_u - M_{u-1} = X_u - X_{u-1}$, et on a donc que $|X_u - X_{u-1}| \leq K$; on trouve donc :

$$\begin{aligned} \mathbb{E}[|M_t|] &\leq \mathbb{E}[|X_0|] + \mathbb{E}[|X_t|] + \mathbb{E}\left[\sum_{u=1}^{\tau} K\right] \\ &= \mathbb{E}[|X_0|] + \mathbb{E}[|X_t|] + K\mathbb{E}[\tau]. \end{aligned}$$

Et puisque τ est intégrable, la variable M_t l'est aussi, pour tout t .

Étape 3 : Vérifier la condition de martingale.

Ça, c'est assez rapide :

$$\begin{aligned} \mathbb{E}[M_{t+1} \mid \mathbf{Y}_{\leq t}] &= \mathbb{E}[X_{t+1} \mathbb{1}_{\{\tau > t+1\}} + X_\tau \mathbb{1}_{\{\tau \leq t+1\}} \mid \mathbf{Y}_{\leq t}] \\ &= \mathbb{E}[X_{t+1} \mathbb{1}_{\{\tau > t+1\}} + X_\tau \mathbb{1}_{\{\tau = t+1\}} + X_\tau \mathbb{1}_{\{\tau \leq t\}} \mid \mathbf{Y}_{\leq t}] \\ &= \mathbb{E}[X_{t+1} \mathbb{1}_{\{\tau > t+1\}} + X_{t+1} \mathbb{1}_{\{\tau = t+1\}} \mid \mathbf{Y}_{\leq t}] + X_\tau \mathbb{1}_{\{\tau \leq t\}} \\ &= \mathbb{E}[X_{t+1} \mathbb{1}_{\{\tau > t\}} \mid \mathbf{Y}_{\leq t}] + X_\tau \mathbb{1}_{\{\tau \leq t\}} \\ &= \mathbb{E}[X_{t+1} \mid \mathbf{Y}_{\leq t}] \mathbb{1}_{\{\tau > t\}} + X_\tau \mathbb{1}_{\{\tau \leq t\}} \\ &= X_t \mathbb{1}_{\{\tau > t\}} + X_\tau \mathbb{1}_{\{\tau \leq t\}} = M_t. \end{aligned}$$

□

REMARQUE. Il existe une version de ce théorème pour les martingales à temps continu aussi, mais l'énoncé, les conditions et la preuve sont plus complexes et hors du cadre de ce cours.

REMARQUE. En fait, la première partie de la condition ?? est superflue ; en effet, on remarque que cette hypothèse n'a pas été utilisée pour montrer, ni la propriété de martingale, ni l'adaptabilité ou l'intégrabilité. À strictement parler, cette hypothèse n'est pas nécessaire pour conclure le résultat. Cependant elle est nécessaire pour conclure le corollaire suivant !

COROLLAIRE 6.3. Soit $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ une martingale par rapport à $\mathbf{Y} = (Y_t)_{t \in \mathbb{Z}^+}$, τ un temps d'arrêt par rapport à \mathbf{Y} et \mathbf{M} le processus \mathbf{X} arrêté en τ .

En supposant que \mathbf{M} et τ satisfont l'une des conditions du théorème d'arrêt de Doob, il suit que

$$\mathbb{E}[X_0] = \mathbb{E}[X_\tau].$$

DÉMONSTRATION. Puisque \mathbf{M} est aussi une martingale par rapport à \mathbf{Y} , son espérance est aussi constante. Or, bien sûr,

$$\lim_{t \rightarrow \infty} M_t = X_\tau,$$

et par conséquent, puisque $\mathbb{P}\{\tau < +\infty\} = 1$, on doit avoir que

$$\mathbb{E}[X_0] = \mathbb{E}[M_0] = \lim_{t \rightarrow \infty} \mathbb{E}[M_t] = \mathbb{E}[X_\tau].$$

□

Une interprétation de ce résultat. Avant de montrer quelques exemples, ça vaut la peine de faire une petite discussion des implications de ce résultat – spécifiquement dans le contexte qui a d'abord motivé l'étude des martingales : les jeux de hasard.

Dans ce contexte, une martingale représente les gains d'un·e participant·e à un jeu de hasard équilibré, où l'espérance du gain à chaque nouvelle partie est nulle ; ainsi, l'espérance des gains totaux après $t + 1$ parties est égale aux gains totaux qu'on apporte avec nous après la t ème partie.

Dans le contexte d'un jeu de hasard, un temps d'arrêt représente un temps où un·e joueur·se hypothétique pourrait décider de jouer (puisqu'il ou elle saurait que le temps est venu d'arrêter de jouer à temps pour ne pas s'engager dans une partie supplémentaire). Ainsi, le corollaire 6.3 garantit qu'aucune stratégie d'arrêt, quelle qu'elle soit, ne peut conduire à une espérance de gain positive.

Des exemples. Nous allons maintenant voir quelques exemples d'application du théorème d'arrêt de Doob – mais avant d'aller plus loin, nous allons surtout voir un exemple où le théorème ne s'applique pas !

EXEMPLE 6.7. Soit $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ une marche aléatoire simple symétrique sur \mathbb{Z} – c'est à dire que les probabilités de transition en un pas pour \mathbf{X} sont $P_{i,i \pm 1} = \frac{1}{2}$. On considère le temps d'atteinte $\tau_{-1} = \inf\{t \geq 0 : X_t = -1\}$, qui est un temps d'arrêt par rapport à \mathbf{X} .

La figure 6.1 montre une réalisation de la marche aléatoire, ainsi que le processus arrêté correspondant.

On remarque que

$$\mathbb{E}_0[X_{\tau_{-1}}] = -1 \neq \mathbb{E}_0[X_0] = 0.$$

Pourtant \mathbf{X} est une martingale et τ_{-1} est un temps d'arrêt par rapport à \mathbf{X} . Que se passe-t-il ?

Eh bien, comme on l'a déjà vu, $\mathbb{E}_0[\tau_{-1}] = +\infty$; de plus, $X_{t \wedge \tau}$ n'est pas bornée (puisqu'il existe une probabilité non-nulle que ce processus atteigne éventuellement n'importe quel nombre naturel).

Donc, on ne peut pas appliquer le théorème de Doob en vertu de l'une ou l'autre des conditions.

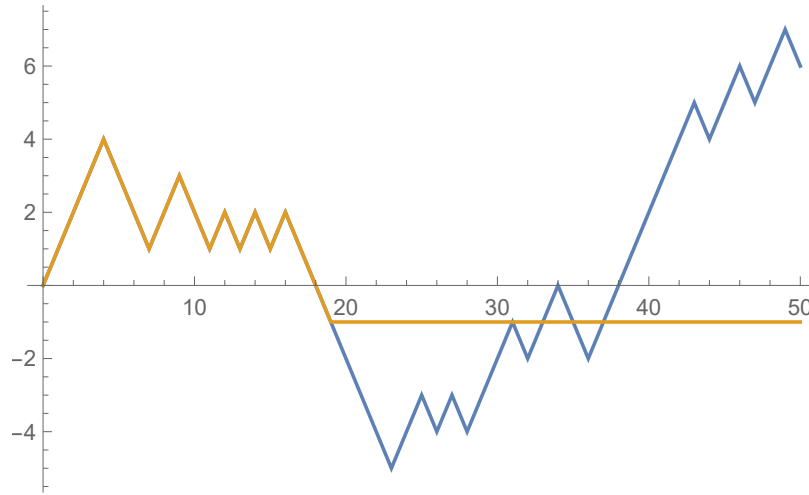
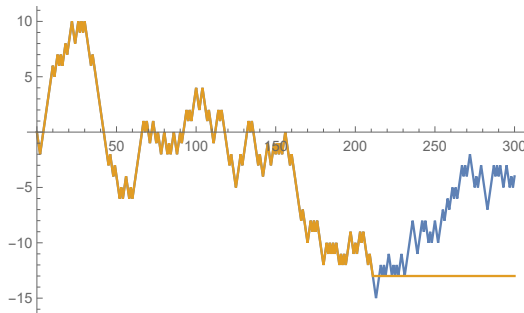


FIGURE 6.1 – Une réalisation d’une marche aléatoire symétrique (en bleu) et le processus arrêté au temps τ_{-1} (en jaune).

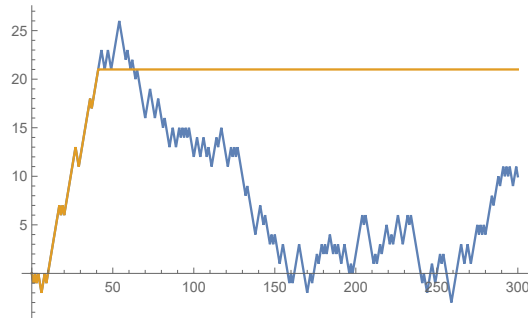
EXEMPLE 6.8 (la ruine du joueur, version martingale). Soit $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ une marche aléatoire simple symétrique sur \mathbb{Z} , comme à l’exemple précédent. Cette fois, on considère $\tau_{\{-b,a\}}$, le temps d’atteinte de l’un ou l’autre des états $-b$ ou a :

$$\tau_{\{-b,a\}} = \inf \{t \in \mathbb{Z}^+ : X_t \in \{-b, a\}\}.$$

La figure 6.2 montre deux réalisations du processus \mathbf{X} arrêté en $\tau_{\{-b,a\}}$.



(A) Une réalisation du processus qui s’est arrêté en $-b$.



(B) Une réalisation du processus qui s’est arrêté en a .

FIGURE 6.2 – Le processus \mathbf{X} arrêté en $\tau_{\{-b,a\}}$, comme à l’exemple 6.8. Ici, $a = 21$ et $b = 13$.

$\tau_{\{-b,a\}}$ est un temps d’arrêt. De plus, lorsque $X_0 = 0$, bien sûr, $X_{t \wedge \tau_{\{-b,a\}}}$ est contenu entre $-b$ et a ; donc c’est borné, et on est également garanti d’atteindre $-b$ ou a éventuellement, puisque \mathbf{X} est un processus transient. Donc $\mathbb{P}\{\tau_{\{-b,a\}} < +\infty\} = 1$. On peut donc appliquer le théorème d’arrêt de Doob – et son corollaire !

Par conséquent, on a que

$$\begin{aligned} 0 &= \mathbb{E}_0 \left[X_{\tau_{\{-b,a\}}} \right] \\ &= \mathbb{P}_0 \left\{ X_{\tau_{\{-b,a\}}} = -b \right\} \times (-b) + \mathbb{P}_0 \left\{ X_{\tau_{\{-b,a\}}} = a \right\} \times a. \end{aligned}$$

Or, bien sûr, $X_{\tau_{\{-b,a\}}} = b$ si et seulement si $\tau_{-b} < \tau_a$, et vice versa ; on trouve donc que :

$$0 = -b\mathbb{P}_0 \{ \tau_{-b} < \tau_a \} + a(1 - \mathbb{P}_0 \{ \tau_{-b} < \tau_a \}),$$

et on peut isoler $\mathbb{P}_0 \{ \tau_{-b} < \tau_a \}$ pour trouver :

$$\mathbb{P}_0 \{ \tau_{-b} < \tau_a \} = \frac{a}{a+b}.$$

Si on décale tout de b unités vers la droite (donc on additionne b à tous les états), et qu'on pose $b = k$ et $a + b = N$, alors on retrouve le fameux résultat de la ruine du parieur dans le cas symétrique (Comparer avec l'exemple 1.17) :

$$\mathbb{P}_k \{ \tau_0 < \tau_N \} = \frac{N-k}{N}.$$

Le raisonnement est rapide, et on n'a pas besoin de se casser la tête avec des récurrences !

EXEMPLE 6.9 (La ruine du parieur, biaisée.). Soit $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ une marche aléatoire simple *biaisée* sur \mathbb{Z} ; c'est-à-dire qu'on a

$$P_{i,i+1} = p; \quad P_{i,i-1} = q := 1 - p,$$

avec $p \neq q$. On prend $\tau_{\{-b,a\}}$ comme à l'exemple précédent.

Cette fois-ci, \mathbf{X} n'est pas une martingale. On prend $Z_t = \rho^{X_t}$, où $\rho = q/p$. Comme nous l'avons vu précédemment, $\mathbf{Z} = (Z_t)_{t \in \mathbb{Z}^+}$ est une martingale par rapport à \mathbf{X} . De plus, on a aussi que $M_t = Z_{t \wedge \tau_{\{-b,a\}}}$ est bornée, puisque d'une part $Z_t > 0$, et d'autre part, $Z_t \leq \max \{ \rho^{-b}, \rho^a \}$. Puisqu'on sait aussi qu'on va forcément atteindre l'état a éventuellement (puisque \mathbf{X} est transient et tend vers $+\infty$), on sait donc que $\mathbb{P}_0 \{ \tau_{\{-b,a\}} < +\infty \} = 1$, et on peut utiliser le théorème de Doob et son corollaire.

La conséquence, c'est bien sûr que

$$\begin{aligned} 1 &= \mathbb{E}_0 [Z_0] = \mathbb{E}_0 [\rho^{X_0}] \\ &= \mathbb{E}_0 [Z_{\tau_{\{-b,a\}}}] \\ &= \rho^a \mathbb{P}_0 \{ \tau_a < \tau_{-b} \} + \rho^{-b} \mathbb{P}_0 \{ \tau_{-b} < \tau_a \} \\ &= \rho^a + (\rho^{-b} - \rho^a) \mathbb{P}_0 \{ \tau_{-b} < \tau_a \}. \end{aligned}$$

Et on conclue que

$$\mathbb{P}_0 \{ \tau_{-b} < \tau_a \} = \frac{1 - \rho^a}{\rho^{-b} - \rho^a} = \frac{\rho^b - \rho^{a+b}}{1 - \rho^{a+b}}.$$

Encore une fois, avec $b = k$, $a + b = N$ et une petite translation de $k = b$ états vers la droite, on trouve donc le même résultat de ruine du parieur :

$$\mathbb{P}_k \{ \tau_0 < \tau_N \} = \frac{\rho^k - \rho^N}{1 - \rho^N}.$$

En prenant a qui tend vers $+\infty$, lorsque $p > q$ (donc $\rho < 1$), on trouve que

$$\mathbb{P}_0 \{ \tau_{-b} < +\infty \} = \rho^b.$$

Mais on peut même aller plus loin ! En utilisant la formule de Wald – attendu que $X_t = \sum_{i=1}^t \xi_i$, où les $(\xi_i)_{i \in \mathbb{N}}$ sont i.i.d., avec

$$\xi_i = X_i - X_{i-1} = \begin{cases} +1 & \text{avec probabilité } p \\ -1 & \text{avec probabilité } q, \end{cases}$$

alors, on trouve bien sûr que

$$\mathbb{E}_0 [X_{\tau_{\{-b,a\}}}] = \mathbb{E} [\tau_{\{-b,a\}}] \mathbb{E}_0 [\xi_1].$$

Or, $\mathbb{E}_0 [\xi_1] = p - q$. En mettant tout cela ensemble, on trouve donc :

$$\mathbb{E}_0 [\tau_{\{-b,a\}}] = \frac{1}{p - q} \left(a \frac{1 - \rho^b}{1 - \rho^{a+b}} - b \frac{\rho^b - \rho^{a+b}}{1 - \rho^{a+b}} \right).$$

Nous n'avions pas pu obtenir ce résultat par la méthode des récurrences ; c'est possible maintenant, entièrement grâce à notre théorème d'arrêt et à la formule de Wald !

6.4. Exercices

Les exercices de ce chapitre sont entièrement tirés de la section 4.3 de l'ouvrage de Sabin LESSARD.

EXERCICE 6.1. Une urne contient des boules rouges et des boules vertes. À chaque instant $n \geq 0$, on tire une boule au hasard de l'urne et on la remet dans l'urne en ajoutant une autre boule de la même couleur que celle tirée. On désigne par X_n la proportion de boules rouges dans l'urne à cet instant et on suppose qu'il y a une boule de chaque couleur dans l'urne initialement.

Montrer que $\mathbf{X} = (X_n)_{n \in \mathbb{Z}^+}$ est une martingale.

EXERCICE 6.2. Soit $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ une marche aléatoire simple sur \mathbb{Z} ; c'est-à-dire qu'on peut écrire

$$X_t = \sum_{i=1}^t Y_i,$$

avec $\mathbf{Y} = (Y_i)_{i \in \mathbb{N}}$ une famille de variables aléatoires indépendantes identiquement distribuées satisfaisant $\mathbb{P}\{Y_1 = \pm 1\} = \frac{1}{2}$.

- Soient τ_{-b} et τ_a respectivement les temps d'atteinte des états $-b$ et a par \mathbf{X} . On note $\tau_{-b,a} = \tau_{-b} \wedge \tau_a$. Est-ce un temps d'arrêt?
- Utiliser le théorème d'arrêt de Doob pour calculer $\mathbb{P}_0\{X_{\tau_{-b,a}}\} = a$.
- On définit $S_t = X_t^2 - t$. Montrer que S_t est une martingale.
- Utiliser le théorème d'arrêt avec S_t pour déterminer $\mathbb{E}[\tau_{-b,a}]$.

EXERCICE 6.3. On pose $X_0 = 1$ et $X_t = \sum_{i=1}^t Y_i$, où les Y_i sont indépendantes, identiquement distribuées, avec $\mathbb{P}\{Y_1 = -1\} = 2/3$ et $\mathbb{P}\{Y_1 = 1\} = 1/3$. On définit

$$\tau = \min\{t \in \mathbb{Z}^+ : X_t = 0 \text{ ou } X_t = 4\}.$$

- Montrer que $(2^{X_t})_{t \in \mathbb{Z}^+}$ est une martingale par rapport à $\mathbf{Y} = (Y_i)_{i \in \mathbb{Z}^+}$.
- Vérifier que τ est un temps d'arrêt par rapport à \mathbf{Y} .
- En utilisant le théorème d'arrêt, déterminer la fonction de masse de X_τ .
- Utiliser la formule de Wald, déterminer $\mathbb{E}[\tau]$.

EXERCICE 6.4. On considère une population de N individus dans laquelle, à chaque unité de temps, un individu choisi au hasard parmi les N individus existants produit une copie identique à lui-même et cette copie remplace l'un des N individus existants choisi au hasard. C'est le modèle de Moran en génétique des populations. Le nombre d'individus d'un type particulier A à l'instant $t \geq 0$, noté X_t , forme alors une chaîne de Markov homogène à temps discret, avec les probabilités de transition en un pas données par

$$P_{i,i-1} = P_{i,i+1} = \frac{i(N-i)}{N^2}, \quad P_{i,i} = 1 - 2\frac{i(N-i)}{N^2}$$

pour $i = 1, 2, 3, \dots, N-1$; par ailleurs,

$$P_{0,0} = P_{N,N} = 1.$$

- (a) Montrer que $\mathbf{X} = (X_t)_{t \in \mathbb{Z}^+}$ est une martingale.
- (b) Déterminer la probabilité que le type A disparaisse éventuellement de la population étant donné qu'il y a k individus de ce type initialement.

EXERCICE 6.5. Un joueur mise un dollar sur chacun des jeux d'une série de jeux identiques et indépendants. Son gain net par partie – le gain brut moins la mise d'un dollar – est d'espérance $\mu = -1/2$ et de variance $\sigma^2 = 1$, en plus d'être borné par une constante $K > 0$. Le gain net après $t \geq 1$ jeux est donc donné par

$$X_t = \sum_{i=1}^t Y_i,$$

où Y_i est le gain net sur l' i ème jeu (on assume $X_0 = 0$).

On introduit maintenant, pour tous $t \geq 0$:

$$M_t = X_t - \mu t; \quad W_t = M_t^2 - \sigma^2 t,$$

et on définit

$$\tau = \min \{t \geq 0 : X_t = -N\},$$

où N est un entier naturel quelconque.

- (a) Montrer que $\mathbf{M} = (M_t)_{t \in \mathbb{Z}^+}$ et $\mathbf{W} = (W_t)_{t \in \mathbb{Z}^+}$ sont des martingales par rapport à \mathbf{Y} .
- (b) Vérifier que τ est un temps d'arrêt par rapport à \mathbf{Y} .
- (c) Utiliser le théorème d'arrêt pour déterminer l'espérance et la variance de τ .

Troisième partie

Sujets avancés

Le mouvement Brownien

Pour clore la discussion des processus stochastiques, nous faisons une petite incursion dans le territoire des martingales à temps continu en explorant la reine parmi les reines des martingales à temps continu : le *processus de Wiener*.

Mais... et le mouvement Brownien ? L'appellation *mouvement Brownien*, que vous reconnaissez sûrement, fait référence au mouvement imprévisible de particules en suspens dans un fluide dont on ne peut voir les courants. Lucrèce, disciple d'Épicure, décrit par exemple le mouvement des grains de poussière dans l'air en ces termes¹ :

Lorsqu'à travers la nuit d'une chambre fermée
Le Soleil entre et darde une flèche enflammée
Regarde, et tu verras dans le champ du rayon
D'innombrables points d'or mêlés en tourbillons
Former leurs rangs, les rompre, encor, toujours, sans trêve
Et livrer un combat qui jamais ne s'achève.
[...]
Vois ces points sous des heurts que l'œil ne saisit pas,
Changer de route, aller, revenir sur leurs pas
Ici, là. En passant, quelqu'atome les dérange,
Et c'est ce qui reforme ou défait leur phalange :
Par lui-même en effet se meut tout corps premier.
Sur les groupes errants qui n'ont pu se lier
S'il tombe un poids égal, il les réduit en poudre.
L'imperceptible choc n'a-t-il pu les dissoudre ?
Ont-ils pu résister ? Ils tremblent seulement.
Ainsi des corps premiers part tout ce mouvement
Qui par degrés arrive à nos sens et rencontre
Enfin ces frêles grains que le rayon nous montre.
Nous voyons ondoyer leur poussière, et nos yeux
Ne peuvent point saisir la cause de leurs jeux.

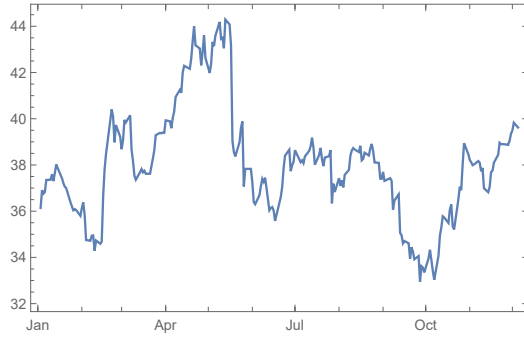
Près de 1900 ans plus tard, c'est le biologiste écossais Robert Brown remarque un mouvement similaire en observant des grains de pollen suspendus dans l'eau. C'est le nom de Brown qui restera accolé à ce type de mouvement. La description mathématique surviendra plus tard ; on la devra en partie à Albert Einstein, puis, plus tard, au mathématicien américain Norbert Wiener.

Ce qu'il faut retenir, donc, c'est que tandis que le nom *mouvement Brownien* fait référence, de façon plus ou moins vague, à un certain type de phénomène physique, le *processus*

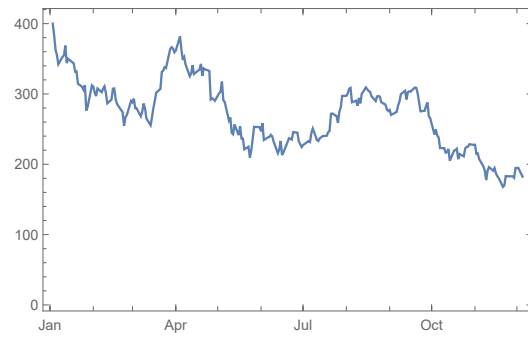
1. LUCRÈCE, *De la nature des choses*, trad. André LEFÈVRE. Société d'éditions littéraire, Paris, 1899. Extrait du Livre Deuxième, « Les atomes », vers 121 à 148.

de Wiener est le processus stochastique qui fournit un modèle mathématique qui permet de décrire ce mouvement – en une dimension.

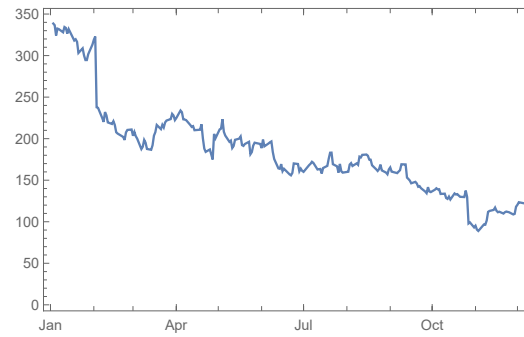
Et ça ressemble à quoi ? On se sert du processus de Wiener pour modéliser une panoplie de phénomènes qui se comporte aléatoirement en temps continu ; l'exemple le plus fréquemment employé est le cours d'une action en bourse. La figure 7.1 illustre l'évolution de la valeur de certains actifs financiers entre le 1er janvier et le 6 décembre 2022.²



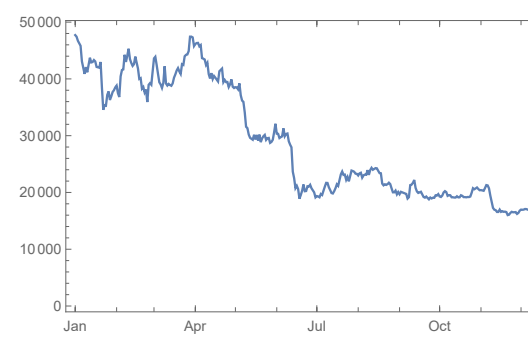
(A) Le cours de l'action du groupe *Kraft-Heinz Corporation*.



(B) Le cours de l'action de *Tesla Motors Inc.*



(C) Le cours de l'action de l'entreprise *Meta*.



(D) La valeur d'un Bitcoin en dollars américains.

FIGURE 7.1 – « *Schadenfreude* »

La figure 7.2 illustre plusieurs réalisations d'un processus de Wiener.

On remarque que, vu de loin, ça ressemble à des marches aléatoires... et nous allons voir que ce n'est pas pour rien !

7.1. Définition axiomatique du processus de Wiener

Nous allons commencer par introduire une définition du processus de Wiener qui se fonde sur une série de postulats. *Attention !* Avant de faire quoique ce soit, il faudra d'abord vérifier qu'il existe bel et bien un tel objet ! Mais pour l'instant, nous allons simplement nous contenter de donner une « liste d'épicerie » des caractéristiques que nous aimerions que le processus ait.

2. Ces données ont été récupérées grâce au module de données financières du logiciel *Wolfram Mathematica*, le 6 décembre 2022. Si vous avez investi au début de 2022, valaient mieux le Ketchup et le *Cheese-Wiz* qu'un gros achat de Bitcoin !

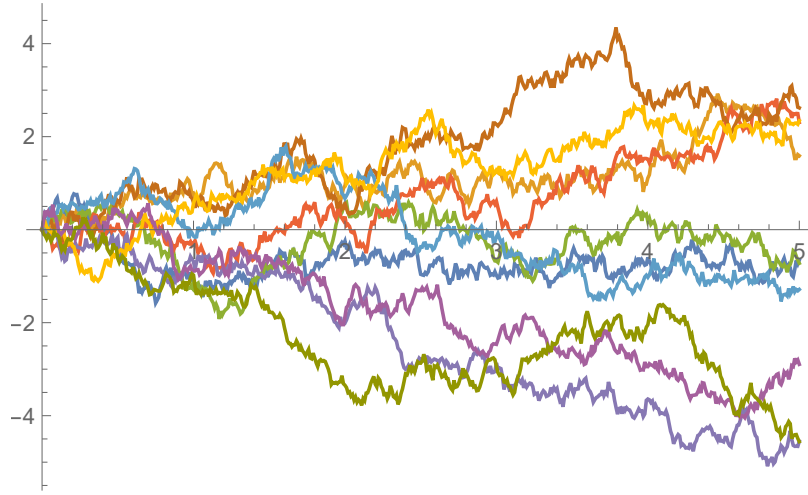


FIGURE 7.2 – Dix réalisations différentes d'un processus de Wiener sur l'intervalle $[0, 5]$

DÉFINITION 7.1 (Processus de Wiener). Soit $\mathbf{W} = (W_t)_{t \in \mathbb{R}^+}$ un processus stochastique sur l'espace des états \mathbb{R} . C'est un **processus de Wiener** si

- i.* $W_0 = 0$;
- ii.* L'application $t \mapsto W_t$ est continue en t presque sûrement.
- iii.* Les incréments sont indépendants (c'est-à-dire que le processus est spatialement homogène) : pour toute suite de temps croissante $(t_i)_{i \in \mathbb{Z}^+}$, la famille $(W_{t_i} - W_{t_{i-1}})_{i \in \mathbb{N}}$ est une famille de variables aléatoires indépendantes. Sans perdre de généralités, on demande que pour tous s, t tels que $0 \leq s \leq t$, les variables $W_t - W_s$ et W_s soient indépendantes.
- iv.* Les incréments suivent une distribution normale centrée de variance égale à la largeur de l'intervalle ; autrement dit, pour tous s, t avec $0 \leq s \leq t$:

$$W_t - W_s \sim \mathcal{N}(0, t - s)$$

7.1.1. Commentaires sur la définition. Avant d'aller plus loin, ça vaut tout de même la peine de faire quelques constats sur les conséquences des axiomes. Nous allons les présenter ici sous forme de petits lemmes.

Attention ! Nous n'avons pas encore montré que le processus de Wiener existe ! Donc, ces petits lemmes ne sont pas forcément utiles. Évidemment, on ne ferait pas tout ça si le processus de Wiener n'existait pas ! Mais pour l'instant, on n'en a pas de preuve. Ça viendra après.

Propriété de Markov et de martingale. Vous remarquerez qu'à priori, le fait que \mathbf{W} soit un processus de Wiener ne garantit pas qu'il soit une chaîne de Markov. C'est pourtant le cas ; en effet, on a le résultat suivant :

LEMME 7.1. Soit $\mathbf{X} = (X_t)_{t \geq 0}$ un processus stochastique spatialement homogène sur l'espace additif S .

Alors, \mathbf{X} a la propriété de Markov faible.

DÉMONSTRATION. *Nous n'entrerons pas dans les détails formels de la preuve parce qu'en temps et espace continus ça peut devenir difficile.*

Connaître $\mathbf{X}_{\leq s}$, ça revient à connaître n'importe quelle suite d'incrément consécutifs entre 0 et s , et de connaître X_s . Mais peu importe les incréments qu'on connaît, il est impossible qu'ils influent sur $X_t - X_s$. Dès lors, si on connaît X_s , $X_t = (X_t - X_s) + X_s$ est indépendante de quoique ce soit d'autre aussitôt qu'on connaît X_s ; c'est précisément ça que dit la propriété de Markov. \square

Ainsi donc, \mathbf{W} est bel et bien une chaîne de Markov; en fait, \mathbf{W} a aussi la propriété de Markov forte. \mathbf{W} est aussi homogène dans le temps, puisque la distribution des incréments ne dépend que de la largeur de l'intervalle.

\mathbf{W} est aussi une martingale!

LEMME 7.2. *Soit \mathbf{W} un processus de Wiener. Alors, \mathbf{W} est une martingale.*

DÉMONSTRATION. Pour prouver cela, il suffit de constater que \mathbf{W} est spatialement homogène, et que par conséquence, puisque $\mathbb{E}[W_t - W_s] = 0$ pour tous $s \leq t$ (puisque $W_t - W_s \sim \mathcal{N}(0, t - s)$), on peut conclure que c'est une martingale en vertu de la proposition 6.1; en effet, \mathbf{W} est forcément adapté à lui-même, et $W_t \sim \mathcal{N}(0, t)$ est intégrable pour tout t . \square

Motivation pour le choix de la distribution normale. Avant de terminer ce commentaire, arrêtons-nous un instant sur le choix de la distribution pour $W_t - W_s$; clairement, il faut une distribution qui a certaines propriétés d'additivité. Si on suppose $W_t - W_s$ suit une distribution $\mathcal{D}(t - s)$, alors il faut en effet que

$$\mathcal{D}(t + s) \sim \mathcal{D}(t) + \mathcal{D}(s).$$

Une telle distribution est appelée *infiniment divisible* – il existe plusieurs types de distributions qui sont infiniment divisibles; voici quelques exemples :

- La distribution de Poisson de paramètre $\lambda > 0$;
- La distribution normale centrée de variance $\sigma^2 > 0$;
- La distribution Gamma de paramètres α, λ est infiniment divisible en α .

A priori, donc, on aurait pu prendre une autre distribution que la distribution normale. En fait...

SCOOP! Si à la place de $\mathcal{D}(t) = \mathcal{N}(0, t)$, on prenait $\mathcal{D}(t)$ une Poisson de paramètre $\lambda t > 0$, le processus résultant serait un processus de Poisson!! En général, si on remplace l'hypothèse que $\mathcal{D}(t) = \mathcal{N}(0, t)$ par la simple exigence que $\mathcal{D}(t)$ soit infiniment divisible, les processus résultants sont une plus large classe de processus stochastiques s'appellent des *processus de Lévy*.

Mais pourquoi donc s'intéresser particulièrement à la distribution normale?

Eh bien en fait, c'est parce que la distribution normale centrée avec variance t satisfait l'équation de la chaleur. Si $f(t, x)$ est la densité de probabilité conditionnelle de W_t au point x sachant que $W_0 = 0$, alors on a que :

$$\frac{\partial}{\partial t} f(t, x) = \frac{1}{2} \frac{\partial^2}{\partial x^2} f(t, x).$$

L'équation de la chaleur est l'équation aux dérivées partielles qui gouverne tout processus physique de diffusion ; il est donc naturel de s'attendre à ce que, pour un processus symétrique comme le processus de Wiener, la densité de W_t satisfasse cette équation. Or, la solution à cette équation différentielle est :

$$f(t, x) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}}.$$

Cela correspond exactement à la densité d'une variable aléatoire normale centrée de variance t !

7.1.2. Construction du processus de Wiener. Comme mentionné précédemment, la discussion du processus de Wiener est, jusqu'ici, de nature complètement spéculative. On n'a pas encore vu ce processus. On a simplement donné une « liste d'épicerie » de propriétés plaisantes qu'on souhaiterait avoir. Mais est-ce bien le cas ?

Pour le prouver, il suffira de produire une *construction* ; c'est à dire une sorte de recette à suivre pour fabriquer de toutes pièces un processus de Wiener. Une fois que ce sera fait, on devra montrer que notre construction possède bien les propriétés désirées.

PROPOSITION 7.1. *Soit $(\xi_i)_{i \in \mathbb{N}}$ une suite de variables aléatoires indépendantes et identiquement distribuées avec $\xi_i = \pm 1$ avec probabilité $1/2$. On définit $S_0 = 0$ et $S_n = \sum_{i=1}^n \xi_i$. On pose maintenant pour tous $N \in \mathbb{N}$:*

$$W_t^{(N)} = \frac{1}{\sqrt{N}} S_{[Nt]}.$$

Les processus $\mathbf{W}^{(N)} = (W_t^{(N)})_{t \in \mathbb{R}^+}$ convergent en loi vers un processus de Wiener lorsque N tend vers l'infini.

DÉMONSTRATION. *Cette preuve ne sera pas complète car certains aspects sont difficiles à démontrer avec les outils disponibles à ce niveau. Nous allons principalement montrer que la distribution marginale limite pour $W_t^{(N)}$ est normale.*

On a que

$$W_t^{(N)} = \frac{1}{\sqrt{N}} S_{[Nt]} = \sqrt{\frac{[Nt]}{N}} \cdot \frac{S_{[Nt]}}{\sqrt{[Nt]}}.$$

Par le théorème central de la limite, on a que

$$\frac{S_{[Nt]}}{\sqrt{[Nt]}} \rightarrow \mathcal{N}(0, 1)$$

en distribution lorsque N tend vers l'infini. Or, on a aussi que

$$\lim_{N \rightarrow \infty} \frac{[Nt]}{N} = t;$$

donc, on a que dans la limite où N tend vers l'infini,

$$W_t^{(N)} \rightarrow \sqrt{t} \mathcal{N}(0, 1) \sim \mathcal{N}(0, t)$$

en distribution lorsque N tend vers l'infini.

En principe il faudrait aussi montrer que la limite en distribution jointe de $\mathbf{W}^{(N)}$ est un processus continu avec des incréments indépendants ; c'est hors de notre portée à ce niveau-ci. \square

REMARQUE. Cette construction est simple mais elle n'est pas entièrement satisfaisante puisque l'on n'a qu'une limite en distribution. On peut tout de même s'en servir pour faire des déductions intéressantes, comme on le verra plus loin. Il existe cependant des constructions (légèrement plus élaborées) qui permettent de construire le processus de Wiener comme une limite presque sûre ; dans ces cas-là, on peut littéralement faire des approximations de W_t , directement.

Finalement, pour clore cette section, il convient d'ajouter quelques mots à propos de l'*unicité* de la distribution jointe pour le processus de Wiener. Après tout, nous avons maintenant montré qu'il existe des processus stochastiques qui satisfont les axiomes de la définition 7.1, mais ce n'est *a priori* pas évident que tous ces processus ont la même distribution jointe.

Le résultat suivant (laissé sans preuve) nous le garantit :

PROPOSITION 7.2. Soient \mathbf{W} et \mathbf{W}' deux processus de Wiener. Alors $\mathbf{W} \sim \mathbf{W}'$; autrement dit, les distributions jointes de \mathbf{W} et \mathbf{W}' coïncident. Autrement dit, les axiomes de la définition 7.1 caractérisent complètement la distribution jointe d'un processus de Wiener.

7.2. Le processus de Wiener avec dérive et variance.

On introduit maintenant le processus de Wiener avec dérive et variance :

DÉFINITION 7.2. Soit $\mathbf{W} = (W_t)_{t \in \mathbb{R}^+}$ un processus de Wiener. On définit maintenant

$$X_t = \mu t + \sigma^2 W_t,$$

où $\mu \in \mathbb{R}$ et $\sigma > 0$.

Alors, $\mathbf{X} = (X_t)_{t \in \mathbb{R}^+}$ est un **processus de Wiener avec dérive et variance**. Les paramètres μ et σ sont respectivement appelés la **dérive** et l'**écart-type** de \mathbf{X} .

Le processus \mathbf{X} a les propriétés suivantes :

$$\mathbb{E}[X_t] = \mu t; \quad \text{Var}[X_t] = \sigma^2 t.$$

Il s'agit donc d'une forme légèrement plus générale qui permet de décrire un mouvement brownien aléatoire qui serait biaisé dans une direction, avec une certaine variance qui peut être plus ou moins importante. La figure 7.3 montre 10 réalisations de processus de Wiener avec différentes dérives et variances.

7.2.1. Construction directe. Comme pour le processus de Wiener standard, on peut faire une construction qui va nous permettre d'obtenir directement le processus de Wiener avec dérive et variance. Pour ce faire, on pose $(\xi_i)_{i \in \mathbb{N}}$ une suite de variables aléatoires indépendantes et identiquement distribuées avec la distribution suivante :

$$\xi_i = \begin{cases} +1 & \text{avec probabilité } p \\ -1 & \text{avec probabilité } q = 1 - p. \end{cases}$$

On va ensuite poser $S_n = \sum_{i=1}^n \xi_i$ et $S_0 = 0$, et on va également prendre :

$$X_t^{(N)} = \frac{\sigma}{\sqrt{N}} S_{[Nt]}.$$

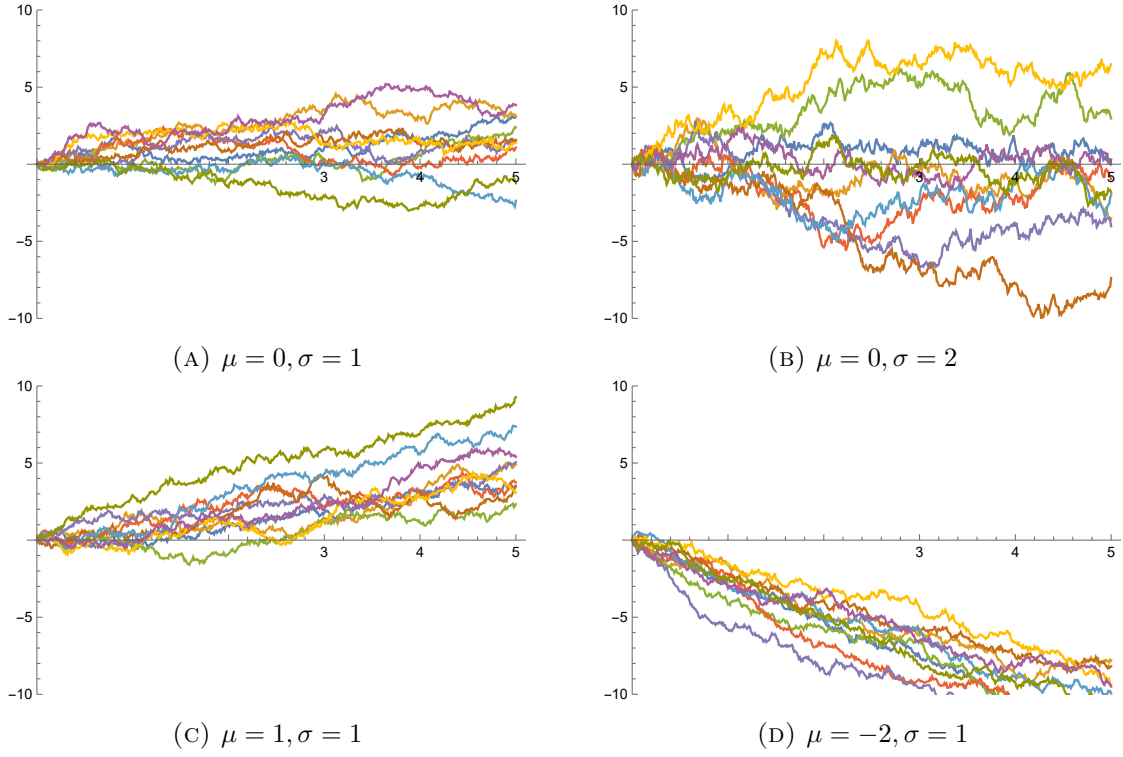


FIGURE 7.3 – Processus de Wiener avec dérives et variances différentes.

Pour que tout se passe bien, il faudra bien sûr que $\mathbb{E}[X_t^{(N)}] \sim \mu t$ lorsque N tend vers l'infini. Pas de souci ; en effet, on a bien sûr que

$$\mathbb{E}[X_t^{(N)}] = \frac{\sigma}{\sqrt{N}} [Nt] (p - q),$$

et si $p - q = \frac{\mu}{\sigma\sqrt{N}}$, on a gagné puisque

$$\begin{aligned} \lim_{N \rightarrow \infty} \mathbb{E}[X_t^{(N)}] &= \frac{\sigma}{\sqrt{N}} [Nt] \frac{\mu}{\sigma\sqrt{N}} \\ &= \mu \lim_{N \rightarrow \infty} \frac{[Nt]}{N} \\ &= \mu t. \end{aligned}$$

On prend donc $p = \frac{1}{2} \left(1 + \frac{\mu}{\sigma\sqrt{N}}\right)$. Par la suite, on pourra montrer que le processus ainsi construit atteint, en distribution, la limite souhaitée lorsque N tend vers l'infini.

Quelques résultats. On peut utiliser cette construction pour profiter de nos connaissances sur la marche aléatoire simple (biaisée ou pas) afin d'obtenir des résultats sur le processus de Wiener, par le truchement des limites !

PROPOSITION 7.3. *Soit \mathbf{X} un processus de Wiener avec dérive μ et écart-type $\sigma > 0$. Soit $a \in \mathbb{R}$, et soit τ_a le temps d'atteinte de l'état a pour \mathbf{X} .*

Alors si $a\mu < 0$, on a que :

$$(7.2.1) \quad \mathbb{P}\{\tau_a < +\infty\} = e^{\frac{2\mu a}{\sigma^2}}.$$

De plus, si $a\mu > 0$, on a que :

$$(7.2.2) \quad \mathbb{E}[\tau_a] = \frac{a}{\mu}.$$

DÉMONSTRATION. Si on définit $\tau_a^{(N)}$ le temps d'atteinte de a pour $\mathbf{X}^{(N)}$ (avec $a > 0$) :

$$\tau_a^{(N)} = \inf \left\{ t \in \mathbb{R}^+ : X_t^{(N)} \geq a \right\},$$

avec $T_{a\sqrt{N}/\sigma}$ le temps d'atteinte de la partie entière de $a\sqrt{N}/\sigma$ pour $(S_n)_{n \in \mathbb{Z}^+}$ (qui dépend de N) :

$$\begin{aligned} \mathbb{P}\{\tau_a < +\infty\} &= \lim_{N \rightarrow \infty} \mathbb{P}\{\tau_a^{(N)} < +\infty\} \\ &= \lim_{N \rightarrow \infty} \mathbb{P}\{T_{a\sqrt{N}/\sigma} < +\infty\}. \end{aligned}$$

Or, $\mathbb{P}\{T_{a\sqrt{N}/\sigma} < +\infty\}$ est simplement donné (par la ruine du parieur) par :

$$\mathbb{P}\{T_{a\sqrt{N}/\sigma} < +\infty\} = \left(\frac{(1 + \mu/(\sigma\sqrt{N}))}{(1 - \mu/(\sigma\sqrt{N}))} \right)^{a\sqrt{N}/\sigma}.$$

Or, on a :

$$\lim_{N \rightarrow \infty} \left(1 \pm \frac{\mu/\sigma}{\sqrt{N}} \right)^{a\sqrt{N}/\sigma} = e^{\pm \mu a / \sigma^2}.$$

L'équation 7.2.1 découle de ces limites.

D'autre part, on sait que $T_{a\sqrt{N}/\sigma}/N$ converge vers τ_a en distribution aussi; on a pu calculer $\mathbb{E}[T_{a\sqrt{N}/\sigma}]$ par un résultat du chapitre précédent; il suffit maintenant donc de remplacer et de prendre la limite. Le résultat qu'on trouve à l'équation 7.2.2 n'est pas très surprenant. Si μ est une « vitesse moyenne », le temps moyen qu'il faut pour passer de 0 à a est assez clairement a/μ ... \square

7.3. Le processus de Wiener géométrique

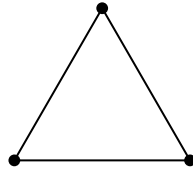
(À venir)

Quatrième partie

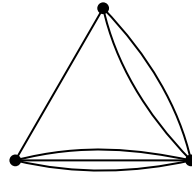
Annexes

Lexique de la théorie des graphes

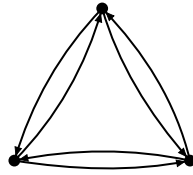
A.1. Description de la structure des graphes



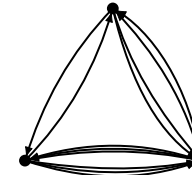
(A) Un graphe simple : les sommets peuvent être reliés par au plus une arête.



(B) Un multigraphe : il peut y avoir plus qu'une arête distincte entre deux sommets.



(C) Un graphe orienté : chaque arête porte une orientation.



(D) Un multigraphe orienté.

FIGURE A.1 – Quatre types de graphes : les graphes simples, les multigraphes, les graphes orientés et les multigraphes orientés.

DÉFINITION A.1 (Vocabulaire général des graphes). *i.* Un **graphe** $G = (V, E, f)$ est un objet mathématique constitué d'un ensemble de **sommets** V , d'un ensemble d'**arêtes** E et d'une fonction f dont le domaine est E .

Chaque arête relie deux sommets du graphe ; la fonction f définie sur E nous permet de déterminer quels sont les sommets reliés par toute arête $e \in E$.

ii. Si le graphe est **orienté** (figures A.1c, A.1d), la fonction f a des images dans $V \times V$; dans ce type de graphe, on dit que l'arête e *part de* u et *se termine en* v (ou *arrive en* v) si et seulement si

$$f(e) = (u, v).$$

La fonction f peut être décomposée dans ses coordonnées composantes :

$$f_1(e) = u \quad f_2(e) = v.$$

On note $u \rightarrow v$ s'il existe une arête qui va de u à v , et on dit que v est **adjacent à** u . Dans un graphe orienté, $u \rightarrow v$ n'implique pas $v \rightarrow u$.

On dit que l'arête e **sort de** u (noté $u \rightarrow e$ si $f_1(e) = u$; on dit qu'elle **entre en** u (noté $e \rightarrow u$ si $f_2(e) = u$). On dit qu'une arête e est **incidente** au sommet u (noté $e \sim u$ ou $u \sim e$) indistinctement si elle sort de u ou si elle entre en u . (figure A.2b)

On dit que les arêtes e et \bar{e} sont **conjuguées** si $f_1(e) = f_2(\bar{e})$ et $f_2(e) = f_1(\bar{e})$. On dit qu'un graphe orienté est **symétrique** si les arêtes (excluant les boucles) peuvent être partitionnées en paires d'arêtes conjuguées.

- iii. Si le graphe est **non-orienté** (figures A.1a, A.1b), la fonction f a des images dans $\binom{V}{1} \cup \binom{V}{2}$, c'est à dire l'ensemble des singletons et des paires de V .¹ Puisque le graphe est non-orienté, l'ordre importe peu. On dit qu'une arête *relie* les sommets u et v si et seulement si

$$f(e) = \{u, v\}.$$

On note $u \sim v$ s'il existe une arête qui relie u et v , et on dit que v est **adjacent à** u . Dans un graphe non-orienté, $u \sim v$ est strictement équivalent à $v \sim u$.

On dit que l'arête e **est incidente** à u si $u \in f(e)$. (figure A.2a)

Qu'un graphe soit orienté ou non, $u \in V$ est une **extrémité** (ou une **terminaison**) de $e \in E$ si et seulement si e est incidente à v .

- iv. Qu'un graphe soit orienté ou non, une arête qui relie un sommet à lui-même est appelée une **boucle**. On dit qu'un graphe *admet les boucles* si de telles arêtes sont permises.
- v. Qu'un graphe soit orienté ou non, deux arêtes e et e' sont **adjacentes** si elles ont au moins une extrémité en commun. On note $e \sim e'$.

Dans un graphe orienté, les arêtes e et e' se **s'enchaînent** ou **se suivent** si elles $f_2(e) = f_1(e')$. On note $e \rightarrow e'$.

- vi. Qu'un graphe soit orienté ou non, le graphe est **simple** si la fonction f est injective – dans ce cas, il existe au plus une arête reliant toute paire (possiblement ordonnée) de sommets. Si la fonction f n'est pas injective, c'est un **multigraphe**.
- vii. Qu'un graphe soit orienté ou non, le graphe est **complet** si la fonction f est surjective – dans ce cas, toutes les paires de sommets possibles (ordonnées ou non) sont reliées par au moins une arête.

En pratique, on travaillera presque toujours avec des graphes simples; dans ce cas, on pourra simplement étiqueter les arêtes directement par la paire (resp. la paire ordonnée) de sommets que l'arête relie. On aurait alors que $E \subseteq \binom{V}{1} \cup \binom{V}{2}$ (resp. $V \times V$), et la fonction f serait simplement l'identité. On noterait alors

$$e = \{u, v\} \quad (\text{resp. } (u, v))$$

si l'arête e relie les sommets u et v (resp. va de u à v). Dans le cas d'un graphe orienté, spécifiquement, si on a $e = (u, v)$, alors

$$u = e_1, \quad v = e_2.$$

1. De façon générale, pour un ensemble X quelconque, on utilise la notation

$$\binom{X}{k} = \{Y \subseteq X : |Y| = k\},$$

soit que $\binom{X}{k}$ est l'ensemble des parties de X à exactement k éléments distincts.

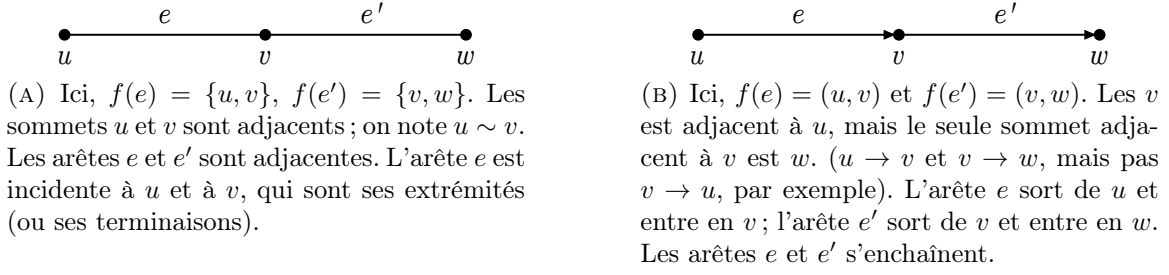


FIGURE A.2 – Le vocabulaire de l'adjacence dans les graphes orientés et non-orientés.

DÉFINITION A.2 (Degré). Dans un graphe quelconque (V, E) , le **degré** du sommet $v \in V$, noté $\deg(v)$, correspond au nombre d'arêtes incidentes à v – plus précisément, le degré correspond au nombre de fois où v est une terminaison. En particulier, une boucle de v à v contribue deux fois au degré de v .

Dans un graphe orienté, on distingue le **degré sortant** $\deg^+(v)$, correspondant au nombre d'arêtes qui sortent de v , ainsi que le **degré entrant** $\deg^-(v)$, correspondant au nombre d'arêtes qui entrent en v . On a dès lors le résultat évident :

$$(A.1.1) \quad \deg(v) = \deg^+(v) + \deg^-(v).$$

Dans tous les graphes, on a également le résultat suivant :

$$(A.1.2) \quad |E| = \frac{1}{2} \sum_{v \in V} \deg(v).$$

DÉFINITION A.3 (Sous-graphe). Un graphe $G' = (V', E')$ est un **sous-graphe** de $G = (V, E)$ si $V' \subseteq V$ et $E' \subseteq E$.

Soit $V' \subseteq V$ une partie quelconque de V . Le **sous-graphe induit par V'** est le graphe $G' = (V', E')$ où E' est l'ensemble des arêtes de E qui ont toutes les extrémités dans V' :

$$E' = \left\{ e \in E : f(e) \in \binom{V'}{1} \cup \binom{V'}{2} \right\} \quad (\text{resp. } V \times V).$$

Dans le cas de graphes simples, on peut simplement choisir

$$E' = E \cap \left(\binom{V'}{1} \cup \binom{V'}{2} \right) \quad (\text{resp. } V \times V).$$

Soit plutôt $E' \subseteq E$ une partie quelconque de E . Le **sous-graphe induit par E'** est le graphe $G' = (V', E')$ où V' est l'ensemble de toutes les extrémités des arêtes de E' . Pour les multigraphes non-orientés, c'est :

$$V' = \bigcup_{e \in E'} f(e);$$

pour les multigraphes orientés, c'est :

$$V' = \{f_1(e) : e \in E'; f_2(e) : e \in E'\}.$$

A.2. Connexité, topologie et métrique des graphes

DÉFINITION A.4 (Vocabulaire de la connexité). Soit $G = (V, E)$ un graphe et $u, v \in V$ deux sommets.

- i. Un **chemin** C de longueur n entre u et v est une suite finie $C = (e_1, e_2, e_3, \dots, e_n)$ d'arêtes adjacentes, la première étant incidente à u , et la dernière incidente à v .

Un **chemin demi-infini** C partant de u est une suite $C = (e_1, e_2, e_3, \dots)$ d'arêtes adjacentes, la première étant incidente à u .

Un **chemin infini** C passant par u est une suite doublement infinie $C = (\dots, e_{-2}, e_{-1}, e_0, e_1, e_2, \dots)$ dont au moins une arête est incidente à u .

- ii. Dans un graphe orienté, un **chemin orienté** de longueur n de u à v est une suite finie $(e_1, e_2, e_3, \dots, e_n)$ d'arêtes qui s'enchaînent, la première sortant de u et la dernière entrant dans v .

Les définitions de chemins orientés demi-infinis et infinis sont faites de façon analogue.

- iii. La **trace** ou la **trajectoire** d'un chemin (ou d'un chemin orienté) est la suite finie $(u = w_0, w_1, w_2, \dots, w_{n-1}, w_n = v)$ des sommets visités par le chemin, dans l'ordre.

Un chemin **visite** le sommet v si v est un élément de la trace de ce chemin.

- iv. Un **sous-chemin** d'un chemin $(e_1, e_2, e_3, \dots, e_n)$ entre u et v (ou d'un chemin orienté de u à v) est un le chemin $(e_k, e_{k+1}, e_{k+2}, \dots, e_l)$, où $1 \leq k \leq l \leq n$.

- v. Soient $C_1 = (e_1, e_2, e_3, \dots, e_m)$ un chemin entre u et v , et $C_2 = (e'_1, e'_2, e'_3, \dots, e'_n)$ un chemin entre v et w (ou des chemins orientés de u à v et de v à w respectivement).

Le chemin $(e_1, e_2, e_3, \dots, e_m, e'_1, e'_2, \dots, e'_n)$ est la **concaténation** des chemins C_1 et C_2 , qu'on note $C_1 \leftrightarrow C_2$ (ou $C_1 \rightarrow C_2$ pour des chemins orientés).

- vi. Un chemin est un **sentier** (on dit parfois **simple**) si aucune arête n'est répétée deux fois. Un sentier est un **circuit** (parfois **sommet-simple**) s'il ne passe pas deux fois par le même sommet.

- vii. Dans un graphe non-orienté, on dit que u et v sont **connectés** si $u = v$ ou s'il existe un chemin entre u et v . On note $u \longleftrightarrow v$.

Dans un graphe orienté, on dit que v est **accessible depuis** u si $u = v$ ou s'il existe un chemin orienté de u à v . On note $u \longrightarrow v$. On dit que u et v sont **connectés** si u est accessible depuis v et v est accessible depuis u ; on emploie alors la même notation : $u \longleftrightarrow v$.

- viii. La relation de connexion \longleftrightarrow est une relation d'équivalence.

La partition de $V / \longleftrightarrow = \{V_1, V_2, V_3, \dots\}$ en classes d'équivalences induit les sous-graphes correspondants $G_1 = (V_1, E_1), G_2 = (V_2, E_2), G_3 = (V_3, E_3), \dots$

Dans un graphe non-orienté, ces sous-graphes sont appelés les **composantes connexes** de G . On a alors que

$$\bigcup_{i \geq 1} E_i = E,$$

ce qui n'est pas forcément le cas pour les sous-graphes induits dans un graphe orienté.

DÉFINITION A.5 (Typologie des chemins). Soit C un chemin dans le graphe $G = (V, E)$.

- i. Le chemin C est **hamiltonien** s'il visite chaque sommet de V exactement une fois.

- ii. Le chemin C est **eulérien** s'il visite chaque arête de E exactement une fois.

DÉFINITION A.6 (Cycles). Soit $v \in V$ un sommet du graphe $G = (V, E)$.

- i. Un chemin entre v et v est un **cycle**. Un chemin orienté de v à v est un **cycle orienté**.
 ii. Un cycle est **hamiltonien** s'il visite chaque sommet exactement une fois (à l'exception du sommet de départ).

Un cycle est **eulérien** s'il visite chaque arête exactement une fois.

DÉFINITION A.7 (Genre, graphe planaire). Soit $G = (V, E)$ un graphe.

- i. Un **plongement** de G dans une surface Σ est « un dessin » du graphe G , où les sommets sont représentés par des points, et les arêtes sont représentées par des arcs de courbes reliant ces points de telle façon qu'aucune paire d'arête ne s'intersecte.
 ii. Le **genre** du graphe G correspond au genre minimal d'une surface Σ qui admet un plongement de G .²
 iii. Les graphes **planaires** sont les graphes de genre 0 – ils peuvent être plongés dans une sphère ou dans un plan. Intuitivement, les graphes planaires sont ceux qu'il est possible de tracer sur du papier sans que deux arêtes ne se croisent.

DÉFINITION A.8 (Distance). Soit $G = (V, E)$ un graphe.

- i. La **distance** $d(u, v)$ entre les sommets u et v est la longueur minimale d'un chemin reliant u et v . Dans les graphes non-orientés et les graphes symétriques, c'est une *métrique* sur V .³
 ii. Le **diamètre** d'un graphe est la plus grande distance possible entre deux sommets de ce graphe.

2. En topologie, le **genre** d'une surface compacte correspond au « nombre de trous » formés par la surface. Pour les surfaces compactes, le genre identifie les classes d'équivalence par homéomorphisme. Par exemple, une sphère possède un genre 0 puisqu'elle ne forme aucun trou. Un tore (un « beigne » possède un genre 1, etc.

3. Une **métrique** sur un espace V est une fonction $m : V \times V \rightarrow \mathbb{R}^+$ qui respecte les axiomes suivants :

- i. $m(u, v) = 0$ si et seulement si $u = v$;
 ii. $m(u, v) = m(v, u)$ (symétrie) ;
 iii. $m(u, w) \leq m(u, v) + m(v, w)$ (inégalité du triangle).

Résultats d'analyse

Le présent annexe regroupe des résultats généraux d'analyse utilisés dans les présentes notes. Veuillez prendre note que toutes les preuves ne sont pas présentées.

B.1. Convergence dominée et limites de séries

À plusieurs reprises, on utilise le *théorème de la convergence dominée* ; il s'agit d'un résultat important de la *théorie de l'intégration de Lebesgue* :

THÉORÈME B.1. Soit $(f_n : \Omega \rightarrow \mathbb{R})_{n \in \mathbb{N}}$ une suite de fonctions

- intégrables (au sens de Lebesgue) sur Ω avec la mesure μ ;
- dominées (bornées en valeur absolue) par une fonction positive $g : \Omega \rightarrow \mathbb{R}$ également intégrable ;
- convergeant ponctuellement vers une limite $f : \Omega \rightarrow \mathbb{R}$ intégrable.

Alors, on a que

$$(B.1.1) \quad \lim_{n \rightarrow \infty} \int_{\Omega} f_n d\mu = \int_{\Omega} f d\mu.$$

REMARQUE. Ce théorème est un résultat important de la théorie de la mesure et de l'intégration ; il indique que, lorsqu'une suite de fonctions converge ponctuellement vers une limite, et que toutes les fonctions sont dominées par une fonction intégrable, alors la limite des intégrales est l'intégrale de la limite (pour autant que ces intégrales existent toutes). C'est un théorème qui nous permet donc d'intervertir une intégrale (au sens de Lebesgue) avec une limite.

Pour en apprendre plus et/ou voir la preuve, vous pouvez consulter les *notes de cours d'André Giroux* (il s'agit du théorème 39).

Dans le cas particulier où notre espace mesuré est l'ensemble $\Omega = \mathbb{N}$, et la mesure est simplement la cardinalité $\mu(A) = |A|$ pour les parties de \mathbb{N} , on trouve que si on voit une suite $(a_n)_{n \in \mathbb{N}}$ de nombre réels comme une fonction $f : \mathbb{N} \rightarrow \mathbb{R}$ définie par $f(n) = a_n$, alors on a :

$$\int_{\mathbb{N}} f d\mu = \sum_{n=1}^{\infty} a_n;$$

autrement dit, la sommation sur \mathbb{N} est carrément une intégrale.

En particulier, on a donc le corollaire suivant :

COROLLAIRE B.1. Soit $(a_{n,k})_{n,k \in \mathbb{N}}$ une famille de nombres réels telle que

- pour tout $k \in \mathbb{N}$, il existe un a_k tel que $\lim_{n \rightarrow \infty} a_{n,k} = a_k$;
- il existe une suite positive $(b_k)_{k \in \mathbb{N}}$ telle que pour tout $n, k \in \mathbb{N}$, $|a_{n,k}| \leq b_k$;

- les séries $\sum_{k \geq 1} a_{n,k}$ pour tout $n \in \mathbb{N}$, $\sum_{k \geq 1} a_k$ et $\sum_{k \geq 1} b_k$ sont absolument convergentes.

Alors, il suit que

$$(B.1.2) \quad \lim_{n \rightarrow \infty} \sum_{k=1}^{\infty} a_{n,k} = \sum_{k=1}^{\infty} a_k.$$

DÉMONSTRATION. Il suffit d'appliquer le théorème de convergence dominée. On note $f_n(k) = a_{n,k}$ pour tout $n, k \in \mathbb{N}$. Si $\Omega = \mathbb{N}$ et la mesure μ est le cardinal (on l'appelle parfois la mesure de comptage), alors nos hypothèses reviennent à dire que

- les f_n convergent ponctuellement vers une limite $f(k) = a_k$;
- elles sont bornées en valeur absolue par $g(k) = b_k$;
- les f_n sont intégrables, ainsi que f et g (puisque les séries correspondantes sont absolument convergentes).

On a donc :

$$\begin{aligned} \lim_{n \rightarrow \infty} \sum_{k=1}^{\infty} a_{n,k} &= \lim_{n \rightarrow \infty} \int_{\mathbb{N}} f_n d\mu \\ &= \int_{\mathbb{N}} f d\mu \\ &= \sum_{k=1}^{\infty} a_k \end{aligned}$$

□

B.2. Convergence de suites

B.2.1. Limites supérieures et inférieures.

DÉFINITION B.1. Soit $(a_n)_{n \in \mathbb{N}}$ une suite de nombres réels.

La **limite supérieure** de la suite (a_n) , notée $\limsup_{n \rightarrow \infty} a_n$, est la limite de la suite $(\sup \{a_k : k \geq n\})_{n \in \mathbb{N}}$, ou $+\infty$ si cette limite n'existe pas.

La **limite inférieure** de (a_n) , notée $\liminf_{n \rightarrow \infty} a_n$, est la limite de la suite $(\inf \{a_k : k \geq n\})_{n \in \mathbb{N}}$, ou $-\infty$ si cette limite n'existe pas.

La figure B.1 illustre ces notions.

PROPOSITION B.1. On a les propriétés suivantes :

- $\limsup_n (a_n + b_n) \leq \limsup_n a_n + \limsup_n b_n$;
- $\liminf_n (a_n + b_n) \geq \liminf_n a_n + \liminf_n b_n$;
- $\liminf_n (a_n + b_n) \leq \liminf_n a_n + \limsup_n b_n$;
- Il existe une sous-suite de a_n qui converge vers $\limsup_n a_n$ si celle-ci est finie,
- et idem pour la \liminf .

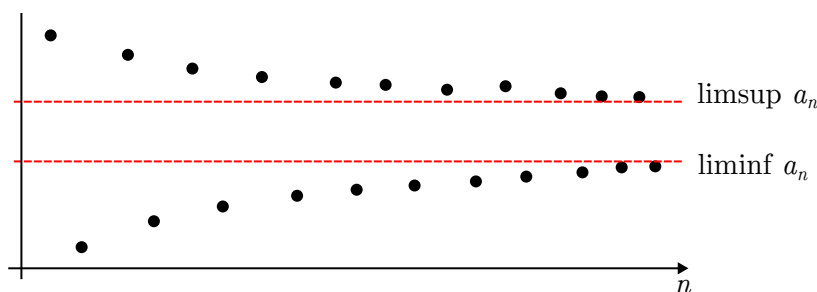


FIGURE B.1 – Les notions de limites supérieures et inférieures sont très utiles.

B.2.2. Lemme de Cesàro.

PROPOSITION B.2 (Lemme de Cesàro). *Soit (a_n) une suite avec $\lim_n a_n = a$. Alors, on a que*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n a_k = a.$$

DÉMONSTRATION. Si a_n converge vers a , pour tout $\epsilon > 0$ il existe N tel que lorsque $n > N$, $|a_n - a| < \epsilon/2$

Également, puisque N/n tend vers 0 quand n tend vers l'infini, pour tout $\epsilon > 0$ et $M > 0$, il existe un $N_2 \in \mathbb{N}$ tel que pour tout $n > N_2$, on a que $|\frac{N}{n}| < \epsilon/(4M)$.

Et puisque la suite a_n converge, elle est bornée en valeur absolue par une certaine constante $M > 0$. Clairement $|a| \leq M$ également.

On a donc que pour $\epsilon > 0$, si $n > \max N, N_2$,

$$\begin{aligned} \left| \frac{1}{n} \sum_{k=1}^n a_k - a \right| &= \left| \frac{1}{n} \sum_{k=1}^n (a_k - a) \right| \\ &\leq \frac{1}{n} \sum_{k=1}^n |a_k - a| \\ &= \frac{1}{n} \sum_{k=1}^N |a_k - a| + \frac{1}{n} \sum_{k=N+1}^n |a_k - a| \\ &\leq \frac{N}{n} 2M + \frac{\epsilon}{2} \\ &\leq \epsilon. \end{aligned}$$

et on conclue que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n a_k = a.$$

□

DÉFINITION B.2. On dit qu'une suite $(a_n)_n$ converge vers a **au sens de Cesàro** si la limite

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n a_k = a.$$

Toutes les suites convergentes convergent au sens de Cesàro, mais l'inverse n'est pas vrai en général.

B.3. Convexité

DÉFINITION B.3. Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ une fonction continue.

Elle est dite **convexe** (resp. **concave**) si pour n'importe quels deux points sur son graphe, la corde qui les relie est *au-dessus* (resp. *en-dessous*) du graphe. Autrement dit, f est convexe (resp. concave) si, pour tout $x < y$ et pour tout $t \in [0, 1]$, on a que :

$$(B.3.1) \quad f(tx + (1-t)y) \leq tf(x) + (1-t)f(y). \quad (\text{resp. } \geq)$$

On dit qu'elle est **strictement convexe** (resp. concave) si l'inégalité est stricte pour tout $t \in (0, 1)$.

PROPOSITION B.3. Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ une fonction continue dérivable deux fois continûment.

Alors, f est convexe (resp. strictement convexe) si et seulement si $f''(x) \geq 0$ (resp. > 0).

B.4. Mesures, fonctions généralisées et distributions.

On rappelle d'abord, pour les curieux/ses, la définition d'une mesure positive sur un espace muni d'une tribu de parties mesurables :

DÉFINITION B.4. Soit Ω un ensemble muni d'une tribu \mathcal{F} de parties mesurables. Une **mesure** μ sur (Ω, \mathcal{F}) est une fonction d'ensembles $\mu : \mathcal{F} \rightarrow \mathbb{R}^+$ qui attribue, pour chaque partie mesurable $E \in \mathcal{F}$, une valeur réelle non-négative. Les mesures sont caractérisées par les axiomes suivants :

- i. Pour toute partie $E \in \mathcal{F}$, $\mu(E) \geq 0$.
- ii. Pour toute famille $(E_i)_{i \in \mathbb{N}}$ dénombrable de parties mesurables disjointe de Ω , on doit avoir :

$$\mu \left(\bigsqcup_{i \in \mathbb{N}} E_i \right) = \sum_{i \in \mathbb{N}} \mu(E_i).$$

Voici quelques exemples fameux de mesures sur \mathbb{R}^n , avec comme ensembles mesurables la *tribu borélienne* \mathcal{B} .¹

- La *mesure de Lebesgue* est la mesure uniforme sur \mathbb{R}^n : pour toute partie A mesurable dans \mathbb{R}^n , on a que

$$\mu(A) = \int_A d^n \mathbf{x} = \text{Vol}_n(A),$$

1. La tribu borélienne est la plus petite tribu qui contient tous les ouverts. Une *tribu* est un ensemble de parties d'un ensemble qui satisfait les axiomes énoncés après la définition 2.2 dans les notes de cours de MAT1720. (Paragraphe 2.1.2 – Les événements). Pour plus de détails, vous pouvez consulter le premier chapitre de l'excellent *Knowing the Odds*, par John B. WALSH (Graduate texts in Mathematics, AMS).

où $\text{Vol}_n(A)$ est le « n -volume » de l'ensemble A – la longueur de l'intervalle si $n = 1$, l'aire de A si $n = 2$, le volume géométrique si $n = 3$, etc.

- Pour une partie $C \subseteq \mathbb{R}^n$ discrète, la *mesure de comptage* de l'ensemble C donne, pour toute partie mesurable A de \mathbb{R}^n , le nombre d'éléments de C qui sont dans A :

$$\mu_C(A) = |A \cap C|.$$

- Si $g : \mathbb{R}^n \rightarrow \mathbb{R}^+$ est une fonction intégrable à valeurs positives, alors

$$\mu(A) = \int_A g(\mathbf{x}) d^n \mathbf{x}$$

est une mesure sur \mathbb{R}^n aussi.

Ce dernier exemple est particulièrement intéressant ; en effet, il introduit l'idée qu'on peut représenter n'importe quelle mesure sur \mathbb{R}^n par une intégrale d'une fonction quelconque²

En se servant de fonctions ordinaires, ce n'est pas toujours possible à proprement parler ; en effet, par exemple, il est impossible de représenter la mesure de comptage μ_C mentionnée précédemment comme l'intégrale d'une fonction.

Pour pallier ce problème, on introduit les *fonctions généralisées* ; ce sont des objets qui « ressemblent » à des fonctions, mais qu'on ne peut pas évaluer en un point. Les fonctions généralisées ne peuvent être bien définies que sous le signe d'intégrale, pour désigner une mesure.

B.4.1. La fonction δ de Kronecker.

DÉFINITION B.5. On introduit la *fonction δ de Kronecker* $\delta_{\mathbf{x}}(\mathbf{y})$; il s'agit d'une fonction généralisée sur \mathbb{R}^n , entièrement caractérisée par les propriétés suivantes :

- i. Pour tout $A \subseteq \mathbb{R}^n$ tel que $\mathbf{x} \notin A$, et pour toute fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$, on doit avoir

$$\int_A f(\mathbf{y}) \delta_{\mathbf{x}}(\mathbf{y}) d^n \mathbf{y} = 0.$$

- ii. Pour tout $A \subseteq \mathbb{R}^n$ tel que $\mathbf{x} \in A$, et pour toute fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$, on doit avoir

$$\int_A f(\mathbf{y}) \delta_{\mathbf{x}}(\mathbf{y}) d^n \mathbf{y} = f(\mathbf{x}).$$

On remarque que la fonction $\delta_{\mathbf{x}}(\mathbf{y})$ de Kronecker représente en fait la mesure de comptage pour le singleton $\{\mathbf{x}\}$: en effet, si on prend $f(\mathbf{y}) = 1$ pour tout $\mathbf{y} \in \mathbb{R}^n$, alors on doit avoir que

$$\int_A \delta_{\mathbf{x}}(\mathbf{y}) d^n \mathbf{y} = \begin{cases} 0 & \text{si } \mathbf{x} \notin A \\ 1 & \text{si } \mathbf{x} \in A. \end{cases}$$

On remarque que notre fonction généralisée $\delta_{\mathbf{x}}$ agit en quelque sorte comme la fonction g dans notre exemple plus haut – elle nous permet de représenter notre mesure par une *densité*, même si la densité n'est pas réellement une fonction à proprement parler.

2. Pour les plus curieux/ses d'entre vous, renseignez-vous sur l'intégrale de Lebesgue et le théorème de Radon-Nikodym.

EXEMPLE B.1. Si X est une variable aléatoire réelle discrète pouvant prendre les valeurs $V = \{v_1, v_2, v_3, \dots\}$, avec la fonction de masse $p(v_i) = p_i$ pour tout i (et $\sum_{i=1}^{\infty} p_i = 1$), alors la fonction généralisée

$$f(x) = \sum_{i=1}^{\infty} p_i \delta_{v_i}(x)$$

est la « densité » de la mesure de probabilité induite par X sur \mathbb{R} .

En effet, par exemple, on voit que

$$\mathbb{P}\{X = v_i\} = \int_{\{v_i\}} \left(\sum_{j=1}^{\infty} p_j \delta_{v_j}(x) \right) dx = \sum_{j=1}^{\infty} p_j \int_{v_i} \delta_{v_j}(x) dx = p_i,$$

puisque v_j est dans $\{v_i\}$ si et seulement si $i = j$.

Notions avancées en probabilités

C.1. Tribus et sous-tribus

DÉFINITION C.1 (Tribu). Une **tribu** sur un ensemble Ω est une partie $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ de l'ensemble des parties de Ω qui satisfait les axiomes suivants :

- i. $\emptyset \in \mathcal{F}$;
- ii. pour tout $E \in \mathcal{F}$, $E^c \in \mathcal{F}$;
- iii. pour toute famille dénombrable $(E_i)_{i \in \mathbb{N}}$ d'éléments de \mathcal{F} , on a que $\bigcup_{i \in \mathbb{N}} E_i$ et $\bigcap_{i \in \mathbb{N}} E_i$ sont aussi éléments de \mathcal{F} ;
- iv. $\Omega \in \mathcal{F}$.

Une **sous-tribu** $\mathcal{G} \subseteq \mathcal{F}$ est une tribu sur Ω dont tous les éléments sont aussi des éléments de la tribu \mathcal{F} .

Dans un espace de probabilités, la tribu des événements (notée \mathcal{E} dans les notes de cours de MAT1720) est une tribu.

D'un point de vue intuitif, une tribu d'événements représente l'information qui peut être déterminée par un aléa – c'est l'ensemble des événements dont on peut décider, sachant l'aléa, si ils sont réalisés ou non. Sous cet angle, les axiomes sont très logiques : si on sait que E est réalisé, on peut inférer logiquement que E^c ne l'est pas, et vice versa. Donc, si E est dans notre tribu, E^c doit l'être aussi. De même pour les réunions et les intersections dénombrables d'événements.

DÉFINITION C.2 (Tribus engendrées). La **tribu engendrée par une variable aléatoire** X , notée $\sigma(X)$, est la plus petite tribu qui contient tous les événements de la forme $\{X \leq u\}$.

La tribu engendrée par une variable aléatoire X correspond à l'information que nous donne la variable X ; ça représente l'ensemble des événements dont on peut déterminer si ils se sont réalisés seulement en connaissant la valeur de X – et pas forcément le résultat complet de l'expérience aléatoire.

Bien entendu, $\sigma(X) \subseteq \mathcal{E}$; c'est une sous-tribu de la tribu des événements.

DÉFINITION C.3 (Mesurabilité par rapport à une tribu). On dit que la variable aléatoire \mathbf{X} est **mesurable** par rapport à la tribu \mathbf{F} si $\sigma(X) \subseteq \mathcal{F}$.

On a le lemme suivant :

LEMME C.1. Y est mesurable par rapport à $\sigma(X)$ si et seulement si il existe une fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ telle que $f(X) = Y$ presque sûrement.

COROLLAIRE C.1. $\sigma(X) = \sigma(Y)$ si et seulement si alors il existe une fonction f inversible telle que $f(X) = Y$ et $f^{-1}(Y) = X$ presque sûrement.

On a également la notion d'indépendance pour les tribus d'événements :

DÉFINITION C.4 (Indépendance de tribus). Deux tribus d'événements \mathcal{F} et \mathcal{G} sont indépendantes si et seulement si pour tout $A \in \mathcal{F}$ et $B \in \mathcal{G}$, les événements A et B sont indépendants.

On a le lemme suivant :

LEMME C.2. X et Y sont des variables aléatoires indépendantes si et seulement si $\sigma(X)$ et $\sigma(Y)$ sont des tribus indépendantes.

C.2. Filtrations, processus adaptés

DÉFINITION C.5 (Filtration). Une **filtration** sur les événements \mathcal{E} est une suite monotone croissante $(\mathcal{F}_t)_{t \geq 0}$ de sous-tribus de \mathcal{E} . C'est une filtration *discrète* si $t \in \mathbb{Z}^+$; c'est une filtration *continue* si $t \in \mathbb{R}^+$.

L'exemple le plus courant d'une filtration est définie le long d'un processus stochastique : si $\mathbf{X} = (X_t)_{t \geq 0}$ est un processus stochastique, on définit $\mathcal{F}_t = \sigma(X_u : u \leq t)$ la tribu engendrée par le processus $\mathbf{X}_{\leq t}$. Alors, les \mathcal{F}_t forment naturellement une filtration.

DÉFINITION C.6 (Processus adapté à une filtration). On dit que le processus stochastique $\mathbf{X} = (X_t)_{t \geq 0}$ est **adapté** à la filtration $(\mathcal{F}_t)_{t \geq 0}$ si X_t est mesurable par rapport à \mathcal{F}_t pour tout t .

EXEMPLE C.1. Soit Y_1, Y_2, Y_3, \dots les résultats successifs de parties de pile ou face. La filtration $\mathcal{F}_n = \sigma(Y_1, Y_2, Y_3, \dots, Y_n)$ est une filtration pratique.

Le processus X_t qui compte le nombre de pile obtenus au temps t est adapté à \mathcal{F}_t

Le processus Z_t qui compte combien de lancers depuis la dernière fois qu'on a eu pile est adapté à \mathcal{F}_t .

Le processus W_t qui compte combien de lancers il reste jusqu'à la prochaine fois qu'on aura pile *n'est pas* adapté à \mathcal{F}_t . C'est tout de même une chaîne de Markov ! Mais elle n'est pas adaptée à la même filtration.

DÉFINITION C.7 (Tribu-limite). Soit $(\mathcal{F}_t)_{t \geq 0}$ une famille de tribus.

La **Tribu-limite** de cette famille est donnée par

$$\mathcal{F}_\infty = \bigcap_{t \geq 0} \sigma \left(\bigcup_{u \geq t} \mathcal{F}_u \right)$$

Dans le contexte d'une famille de tribus engendrées respectivement par un processus stochastique $\mathbf{X} = (X_t)_{t \geq 0}$, la tribu-limite \mathcal{F}_∞ correspond à l'ensemble des événements qui sont déterminés par le processus $\mathbf{X}_{\geq T}$ peu importe T . Ce sont les événements qu'on peut déterminer en ignorant un intervalle de temps arbitrairement grand (mais fini).

EXEMPLE C.2. Des événements comme « X_t converge lorsque t tend à l'infini » sont dans la tribu-limite des tribus engendrées par les X_t .

C.3. Résultats sur les limites.

THÉORÈME C.1 (Loi zéro-un de Kolmogorov). *Soient $(\mathcal{F}_n)_{n \in \mathbb{N}}$ une suite de tribus indépendantes, et soit \mathcal{F}_∞ sa tribu limite. Alors, pour tout $E \in \mathcal{F}_\infty$, soit $\mathbb{P}\{E\} = 0$ ou $\mathbb{P}\{E\} = 1$.*

La loi zéro-un de Kolmogorov permet de dire que lorsqu'un événement se trouve dans la tribu-limite d'une suite de tribus indépendantes, on n'a pas le choix d'admettre que cet événement se produit soit presque sûrement, ou presque sûrement pas.

On a également le résultats uivant :

THÉORÈME C.2 (Lemme de Borel-Cantelli). *Soit $(E_i)_{i \in \mathbb{N}}$ une suite d'événements. On note*

$$\limsup_{n \rightarrow \infty} E_n = \bigcap_{k \geq 1} \left(\bigcup_{i \geq k} E_i \right).$$

Il s'agit de l'événement $\{ \text{Une infinité de } E_i \text{ sont réalisés.} \}$.

Alors, si $\sum_{n=1}^{\infty} \mathbb{P}\{E_n\} < +\infty$, on a que

$$\mathbb{P} \left\{ \limsup_{n \rightarrow \infty} E_n \right\} = 0.$$

COROLLAIRE C.2. *Lorsque les événements sont aussi indépendants, on peut utiliser la loi zéro-un de Kolmogorov pour conclure que, si $\sum_{n=1}^{\infty} \mathbb{P}\{E_n\}$ diverge, alos*

$$\mathbb{P} \left\{ \limsup_{n \rightarrow \infty} E_n \right\} = 1.$$