Université Paris-Est-Créteil
ESIPE

CHAÎNES DE MARKOV

Spécialité : INGENIEUR, $1^{\text{ère}}$ année

Béatrice de Tilière

La partie "Rappels de probabilités" est basée sur des notes écrites en collaboration avec Frédérique Petit pour la préparation au CAPES. Je tiens à remercie Bernard Marchal de m'avoir donné ses notes de cours de "Processus stochastiques", je m'en suis largement inspirée et en ai tiré tous les dessins de graphes pour les chaînes de Markov.

TABLE DES MATIÈRES

I Rappels de probabilités **3**

1 Modélisation des phénomènes aléatoires **5**
1.1 Introduction 5
1.2 L'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ 6
1.2.1 Espace des états 6
1.2.2 Événements 7
1.2.3 Tribu 8
1.2.4 Probabilité 9

2 Construction d'espaces probabilisés **13**
2.1 Caractérisation d'une probabilité : cas fini ou dénombrable 13
2.2 Cas où l'univers est fini 13
2.2.1 Dénombrement, modèle d'urne 14
2.3 Cas où l'univers est infini dénombrable 18
2.4 Cas où l'univers est infini non-dénombrable 19

3 Conditionnement et indépendance **27**
3.1 Probabilité conditionnelle 27
3.1.1 Définition 27
3.1.2 Formule des probabilités totales et formule de Bayes 29
3.2 Indépendance des événements 31
3.2.1 Nombre infini de jets de dés 32

4 Variables aléatoires **43**
4.1 Définition et loi d'une variable aléatoire 43
4.2 Fonction de répartition 45
4.3 Variables aléatoires discrètes 45
4.3.1 Définitions et exemples classiques 45
4.3.2 Fonction de répartition 50
4.3.3 Espérance 51
4.3.4 Variance, moments d'ordres supérieurs 54
4.3.5 Inégalité de Markov et de Bienaymé Tchebychev 58
4.4 Vecteurs aléatoires discrets 59
4.4.1 Définition et lois des vecteurs aléatoires 59
4.4.2 Espérance, covariance, matrice de covariance 61
4.4.3 Variables aléatoires indépendantes 63
4.5 Suites de variables aléatoires réelles 70
4.5.1 Loi faible des grands nombres 71

4.5.2 Théorème central limite 72

II Chaînes de Markov **75**

5 Introduction et définitions **77**
5.1 Processus stochastiques 77
5.2 A.A. Markov 78
5.3 Définitions 78
5.3.1 Propriété de Markov 78
5.3.2 Probabilités et matrices de transition 79
5.3.3 Matrices stochastiques 80
5.3.4 Graphe associé à une chaîne de Markov homogène 82
5.4 Exercices : Introduction aux chaînes de Markov 83

6 Dynamique d'une chaîne de Markov **91**
6.1 Caractérisation d'une chaîne de Markov 91
6.2 Transitions d'ordre n et loi à l'instant n 93
6.3 Exercices : dynamique d'une chaîne de Markov 95

7 Classification des états **99**
7.1 Classes de communication 99
7.2 Période 100
7.3 Exercices : classes de communication, période 101
7.4 Récurrence et transience 105
7.4.1 Définitions 105
7.4.2 Critères de récurrence/transience 105
7.4.3 Classes récurrentes/transientes 108
7.5 Exercices : récurrence/transience 111

8 Mesures stationnaires **117**
8.1 Mesures stationnaires et réversibles 117
8.2 Existence 119
8.3 Unicité 120
8.4 Caractérisation des chaînes de Markov récurrentes positives 122
8.5 Exercices : mesures stationnaires et invariantes 124
8.6 Convergence vers l'équilibre 127
8.7 Exercices : convergence vers l'équilibre 129

Bibliography **130**

Première partie

Rappels de probabilités

Chapitre 1

Modélisation des phénomènes aléatoires

1.1 Introduction

Une *expérience (ou phénomène) aléatoire* consiste en une expérience pour laquelle toutes les issues possibles sont connues, mais où interviennent de nombreux facteurs, dont nous ne connaissons ou maîtrisons qu'une petite partie. Dans ce cas, l'issue n'est pas prévisible avec certitude. La *théorie des probabilités* consiste en l'étude de ces expériences aléatoires.

Citons quelques exemples : le résultat d'un jeu de hasard (pile ou face, jet de dé, roulette etc.) ; durée de vie d'un atome radioactif, d'un individu, d'une ampoule ; les instants de passage d'un bus à un arrêt donné ; la promenade d'un ivrogne dans la rue ; la trajectoire d'une poussière à la surface de l'eau etc.

Les applications de la théorie des probabilités sont nombreuses : base de la statistique, outil puissant en finance, dans les assurances, théorie des jeux. Elle permet également de modéliser de nombreux phénomènes complexes en biologie, médecine, sciences humaines, climatologie. Elle s'est aussi révélée utile dans de nombreux domaines des mathématiques pures. Mais surtout, elle a acquis une place importante au sein des mathématiques en tant que discipline à part entière, de part son intérêt intrinsèque.

Historiquement, les jeux des hasards sont présents en Égypte, en Grèce et à Rome dès l'Antiquité. Il est cependant intéressant de constater qu'un traitement systématique n'est apparu qu'au XVI$^{\text{e}}$ siècle dans le livre *Liber de Ludo Alea* de Gerolamo Cardano (1501-1576). La véritable étincelle se trouve dans la correspondance entre Blaise Pascal (1623-1662) et Pierre de Fermat ($\sim$1605-1665), au sujet de problèmes posés par le chevalier de Méré. Encouragé par Pascal, Christian Huygens (1629-1695) publie *De ratiocinis in ludo aleae* (raisonnements sur les jeux de dés) en 1657. Ce livre est le premier ouvrage important sur les probabilités. Il y définit la notion d'espérance et y développe plusieurs problèmes de partages de gains lors de jeux ou de tirages dans des urnes. Deux ouvrages fondateurs sont également à noter : *Ars Conjectandi* de Jacques Bernoulli (1654-1705) qui définit la notion de variable aléatoire et donne la première version de la loi des grands nombres, et *The Doctrine of Chance* d'Abraham de Moivre (1668-1754) qui généralise l'usage de la combinatoire. On mentionnera également Pierre-Simon de Laplace (1749-1827), Leonhard Euler (1707-1783) et Johann Carl Friedrich Gauss (1777-1855).

La théorie des probabilités classique ne prend réellement son essor qu'avec les notions de mesure et d'ensembles mesurables qu'Émile Borel (1871-1956) introduit en 1897. Cette notion de mesure est complétée par Henri Léon Lebesgue (1875-1941) et sa théorie de l'intégration. La première version moderne du théorème central limite est donnée par Alexandre Liapounov en 1901 et la première preuve du théorème moderne est due à Paul Lévy en 1910. Il faudra attendre 1933 pour que la théorie des probabilités sorte d'un ensemble de méthodes et d'exemples divers et devienne une véritable théorie, axiomatisée par Andreï Nikolaïevitch Kolmogorov (1903-1987).

1.2 L'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$

Le but de la théorie des probabilités est de fournir un modèle mathématique pour décrire les expériences aléatoires. Sous sa forme moderne, la formulation de cette théorie contient trois ingrédients : l'*espace des états*, les *événements*, et la *loi de probabilité* ou simplement la *probabilité*. Dans toute la suite, nous considérons une expérience aléatoire que nous cherchons à modéliser.

1.2.1 Espace des états

Définition. L'*espace des états* appelé aussi *univers*, noté Ω, est l'ensemble des résultats possibles de l'expérience.

Exemple 1.1. Voici quelques exemples de choix d'univers.

1. Lancer d'une pièce de monnaie. $\Omega = \{P, F\}$.
2. Deux lancers successifs d'une même pièce de monnaie. $\Omega = \{PP, PF, FP, FF\}$.
3. Lancer d'un dé. $\Omega = \{1, 2, 3, 4, 5, 6\}$.
4. Deux lancers successifs d'un même dé, et on s'intéresse à la somme des nombres obtenus. Dans ce cas, il y a trois choix raisonnables :

$$\begin{aligned}
\Omega_1 &= \{(i,j) \,:\, i \in \{1, \cdots, 6\},\, j \in \{1, \cdots, 6\}\} = \{1, \cdots, 6\}^2, \\
\Omega_2 &= \{2, 3, 4, \cdots, 12\}, \\
\Omega_3 &= \{\{i,j\} \,:\, i \in \{1, \cdots, 6\},\, j \in \{1, \cdots, 6\},\, i \le j\}.
\end{aligned}$$

5. Lancer d'un même dé indéfiniment.

$$\Omega = \{(u_n)_{n \ge 1} \,:\, \forall n \in \mathbb{N}^*,\, u_n \in \{1, \cdots, 6\}\} = \{1, \cdots, 6\}^{\mathbb{N}^*}.$$

6. Durée de vie d'un individu. $\Omega = \{x \in \mathbb{R}^+ \,:\, 0 \le x \le 120\}$.
7. Promenade d'un ivrogne dans une rue (un pas en avant, un pas en arrière).

$$\Omega = \{(u_n)_{n \ge 1} \,:\, \forall n \in \mathbb{N}^*,\, u_n \in \{-1, 1\}\} = \{-1, 1\}^{\mathbb{N}^*}.$$

8. Trajectoire d'une poussière à la surface de l'eau pendant un intervalle de temps $[0, T]$. $\Omega = C([0, T]; \mathbb{R}^2)$.

1.2.2 Événements

Définition 1.1. Un *événement* est une propriété dont on peut dire si elle est réalisée ou non, une fois l'issue de l'expérience connue. Un événement correspond un sous-ensemble A de l'univers Ω.

Un singleton, c'est-à-dire un événement réduit à un seul élément de Ω, est appelé un *événement élémentaire*, sinon on parle d'*événement composite.*

On note un événement par une lettre majuscule A, B, C... et l'ensemble de tous les événements de Ω par $\mathcal{A}$.

Remarque. Nous verrons au paragraphe suivant la définition (mathématique) d'un événement. Pour l'instant, essayons de voir à quelles propriétés doivent satisfaire les événements.

Exemple 1.2. On reprend la numérotation de l'exemple 1.1. Voici quelques exemples d'événements écrits d'abord en mots, puis en tant que sous-ensembles de l'espace des états Ω.

2. "Le premier jet donne pile" est le sous-ensemble $\{PP, PF\}$ de Ω.
4. "La somme des résultats obtenus est égale à 4" est le sous-ensemble $\{(1,3),(2,2),(3,1)\}$ de Ω_1, au sous-ensemble $\{4\}$ de Ω_2, et au sous-ensemble $\{\{1,3\},\{2,2\}\}$ de Ω_3.
5. "Le premier 1 est obtenu au N-ième lancer" est le sous-ensemble
$$\{(u_n)_{n\geq 1} \in \Omega \,:\, u_1 \geq 2, \cdots, u_{N-1} \geq 2,\ u_N = 1\}.$$
6. "L'individu atteint au moins 50 ans" est le sous-ensemble :
$$\{x \in \mathbb{R}^+ \,:\, 50 \leq x \leq 120\}.$$
7. "L'ivrogne avance au N-ième pas" est le sous-ensemble :
$$\{(u_n)_{n\geq 1} \in \Omega \,:\, u_N = 1\}.$$

Remarque. Les événements, qui sont par définition des sous-ensembles de l'univers, sont en général décrits à l'aide de phrases dans un premier temps. En effet, on commence par se poser une question liée à une expérience aléatoire, puis on introduit un modèle probabiliste pour y répondre. Par exemple, on cherche la probabilité que la somme de deux dés lancés au hasard soit égale à 4 ; l'événement considéré est alors "la somme des dés est égale à 4".

Une fois fixé le choix de l'univers, un événement correspond à un *unique* sous-ensemble de ce dernier. Comme il n'y a pas forcément unicité du modèle et qu'alors les événements peuvent s'écrire en termes de sous-ensembles sous des formes différentes, la phrase qui décrit un événement permet de se comprendre, quel que soit le modèle choisi, voir par exemple les exemples 1.1 et 1.2 numéro 4. Remarquons aussi que, étant donné un sous-ensemble d'un univers, il est souvent possible de le décrire par différentes phrases, qui représentent toutes le même événement. Par exemple l'événement $\{PF, FF\}$ de l'exemple 1.2 numéro 2 peut se traduire par "le premier jet donne pile" ou "le premier jet ne donne pas face".

Puisque les événements sont des sous-ensembles, on peut effectuer les opérations habituelles, avec la correspondance suivante entre les terminologies ensembliste et probabiliste.

Notation	Terminologie ensembliste	Terminologie probabiliste
Ω	ensemble entier	espace des états, événement certain
ω	élément de Ω	événement élémentaire
A	sous-ensemble de Ω	événement
$\omega \in A$	ω appartient à A	A est réalisé si ω est le résultat de l'expérience
$A \subset B$	A est inclu dans B	si A est réalisé alors B aussi
$A \cup B$	réunion de A et B	l'événement "A ou B" (ou non exclusif !)
$A \cap B$	intersection de A et B	l'événement "A et B"
A^c	complémentaire de A	l'événement contraire de A
$\emptyset$	ensemble vide	événement impossible
$A \cap B = \emptyset$	A et B sont disjoints	A et B sont incompatibles

Exemple. Deux lancers successifs d'une même pièce de monnaie. Soient $A = \{PP\}$, $B = \{PF\}$, $C = \{FP, FF\}$. Alors,

— $A \cup B = \{PP, PF\} = C^c$, est l'événement "le premier jet donne pile" ;
— $A \cap B = \emptyset$, est l'événement impossible, A et B sont incompatibles.

Propriété 1.2. *Les opérations sur les événements satisfont aux règles suivantes. Pour tout événements A, B, C, on a*

— *commutativité :* $A \cup B = B \cup A$ *;*
— *associativité :* $(A \cup B) \cup C = A \cup (B \cup C)$ *;*
— *distributivité :* $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$ *;*
— *lois de De Morgan :* $(A \cup B)^c = A^c \cap B^c$*, et* $(A \cap B)^c = (A^c \cup B^c)$.

1.2.3 TRIBU

L'ensemble des événements $\mathcal{A}$ associés à une expérience aléatoire est donc un sous-ensemble des parties de Ω, $\mathcal{A} \subset \mathcal{P}(\Omega)$. Il semblerait naturel de prendre $\mathcal{A} = \mathcal{P}(\Omega)$, mais il y a alors des exemples où il est impossible d'associer à chaque événement une probabilité de façon cohérente. Dans ces cas-là, il est donc nécessaire de se restreindre à un sous-ensemble strict de $\mathcal{P}(\Omega)$ contenant les événements "intéressants".

L'ensemble des événements que l'on considère en probabilité doivent satisfaire à quelques propriétés naturelles, ils doivent former une tribu, dont voici la définition.

Définition. Un ensemble $\mathcal{A}$ de parties de Ω est une *tribu*, ou *σ-algèbre*, s'il satisfait aux conditions suivantes :

1. $\Omega \in \mathcal{A}$;
2. $\forall A \in \mathcal{P}(\Omega)$, $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$;
3. $\mathcal{A}$ est stable par réunion finie ou dénombrable.

Exemple 1.3.

- $\{\emptyset, \Omega\}$ est une tribu et c'est la plus petite (au sens de l'inclusion).
- $\mathcal{P}(\Omega)$ est une tribu et c'est la plus grande.
- Soit $\mathcal{C}$ un ensemble arbitraire de parties de Ω, alors la plus petite tribu contenant $\mathcal{C}$, notée $\sigma(\mathcal{C})$ est appelée la *tribu engendrée par* $\mathcal{C}$. On admet l'existence de cette tribu.
- Soit $A \in \mathcal{P}(\Omega)$, $A \neq \emptyset$, $A \neq \Omega$, alors la tribu engendrée par A est $\sigma(A) = \{\emptyset, A, A^c, \Omega\}$.
- Sur $\mathbb{R}$, on utilise la tribu engendrée par les ouverts de $\mathbb{R}$, appelée *tribu borélienne* de $\mathbb{R}$. On admet le fait qu'elle soit différente de $\mathcal{P}(\mathbb{R})$.
- Dans le cas où l'espace des états est fini ou dénombrable, on prend toujours $\mathcal{A} = \mathcal{P}(\Omega)$.

Exercice 1.1. Soit $\Omega = \{1, 2, 3\}$.
- Quelle est la tribu engendrée par $A = \{1, 2\}$?
- Quelle est la tribu engendrée par $\mathcal{C} = \{\{1\}, \{2\}\}$?

Solution de l'exercice 1.1.
- D'après l'Exemple 1.3, la tribu engendrée par $A = \{1, 2\}$ est $\{\emptyset, \{1, 2\}, \{3\}, \{1, 2, 3\}\}$.
- La tribu engendrée par $\mathcal{C}$ est stable par réunion et complémentation, elle doit donc contenir $\{1, 2\}$ et $\{3\}$; en particulier elle contient $\{1\}, \{2\}, \{3\}$. Il est alors facile de voir que $\sigma(\mathcal{C}) = \mathcal{P}(\Omega)$.

Définition. L'*ensemble des événements* associé à une expérience est la tribu $\mathcal{A}$ choisie sur Ω.

Remarque. Dans le cas où l'espace des états est fini ou dénombrable, puisque $\mathcal{A} = \mathcal{P}(\Omega)$, un événement est donc simplement n'importe quel sous-ensemble de Ω.

1.2.4 Probabilité

Nous souhaitons maintenant associer à chacun des événements une *probabilité*, qui mesure la vraisemblance que l'on accorde a priori à l'événement avant la réalisation de l'expérience. C'est une des données du modèle, que l'on peut comprendre intuitivement de différentes manières, en voici deux.

Approche utilisant les symétries. On considère un dé non-pipé. Il est alors naturel de supposer que chacune des issues possibles ait la même probabilité égale à $1/6$. Il faut cependant être prudent avec cette approche. En effet, supposons que nous souhaitions déterminer la probabilité du sexe d'un nouveau né. Il n'y a aucune raison de penser qu'il y a plus de chances d'avoir un garçon ou une fille, de sorte qu'il est naturel d'associer une probabilité $1/2$ à chacun des événements élémentaires. Cependant, les statistiques montrent que la proportion de garçons nouvellement né est de $51,2\%$ (INED, France métropolitaine).

Approche fréquentiste. On suppose qu'une expérience d'univers Ω est exécutée plusieurs fois sous les mêmes conditions. Pour chaque événement A de Ω, on définit $n_N(A)$ comme le nombre de fois où l'événement A survient lors des N premières répétitions de l'expérience. Alors la *probabilité de l'événement* A, notée $\mathbb{P}(A)$, est définie comme la limite, dans un sens à préciser, du quotient $n_N(A)/N$.

Cela veut dire que $\mathbb{P}(A)$ est définie comme la limite du pourcentage du nombre de fois où A survient par rapport au nombre total des répétitions. C'est donc la fréquence limite de A. Bien que cette définition soit intuitivement commode, elle présente un sérieux inconvénient. En effet, il faut justifier de l'existence de la limite, ce qui est difficile a priori.

Il est plus raisonnable d'admettre que les probabilités satisfont à un ensemble d'axiomes simples et intuitivement acceptables, pour ensuite démontrer qu'une telle fréquence limite existe dans un certain sens (il s'agit de la *loi des grands nombres*).

Définition. Étant donnés un espace d'états Ω et $\mathcal{A}$ l'ensemble des événements, une *probabilité* $\mathbb{P}$ *sur* $(\Omega, \mathcal{A})$, est une application de $\mathcal{A}$ dans $[0,1]$, possédant les propriétés suivantes.

1. L'événement certain est de probabilité 1 : $\mathbb{P}(\Omega) = 1$.
2. *Axiome de σ-additivité* : pour toute famille dénombrable $(A_n)_{n\geq 0}$ d'événements de $\mathcal{A}$, deux-à-deux disjoints, on a
$$\mathbb{P}\Big(\bigcup_{n\geq 0} A_n\Big) = \sum_{n=0}^{+\infty} \mathbb{P}(A_n).$$

Le triplet $(\Omega, \mathcal{A}, \mathbb{P})$ est alors appelé un *espace probabilisé.*

On a les conséquences immédiates suivantes.

Proposition 1.3. *Soit* $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé et soient deux événements* $A \in \mathcal{A}$*,* $B \in \mathcal{A}$*.*

1. Additivité. *Si A et B sont disjoints, alors $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$. En particulier, $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$, et $\mathbb{P}(\emptyset) = 0$.*
2. *Si $A \subset B$, alors : $\mathbb{P}(A) \leq \mathbb{P}(B)$.*
3. *$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.*
4. *Plus généralement, on a la* formule de Poincaré *: Soit $(A_n)_{n=1}^N$ une famille d'événements de $\mathcal{A}$, alors :*
$$\mathbb{P}\Big(\bigcup_{n=1}^N A_n\Big) = \sum_{n=1}^N (-1)^{n-1} \sum_{\substack{J \subset \{1, \cdots, N\} \\ \mathrm{card}(J) = n}} \mathbb{P}\Big(\bigcap_{k\in J} A_k\Big).$$

Voici une conséquence plus abstraite, qui est fréquemment utilisée.

Proposition 1.4. *Soit* $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé.*

— *Soit $(A_n)_{n\geq 1}$ une suite croissante d'événements de $\mathcal{A}$, c'est-à-dire, pour tout $n \geq 1$, $A_n \subset A_{n+1}$. Soit $A = \bigcup_{n=1}^{+\infty} A_n$, alors :*
$$\mathbb{P}(A) = \lim_{n\to+\infty} \mathbb{P}(A_n).$$

— *Soit $(B_n)_{n\geq 1}$ une suite décroissante d'événements de $\mathcal{A}$, c'est-à-dire, pour tout $n \geq 1$, $B_n \supset B_{n+1}$. Soit $B = \bigcap_{n=1}^{+\infty} B_n$, alors :*
$$\mathbb{P}(B) = \lim_{n\to+\infty} \mathbb{P}(B_n).$$

Remarque 1.5. Les suites réelles $(\mathbb{P}(A_n))_{n\geq 1}$ et $(\mathbb{P}(B_n))_{n\geq 1}$ sont respectivement croissante et décroissante, bornées, donc convergentes, justifiant ainsi l'existence de la limite. Par ailleurs, la suite $\left(\bigcup_{n=1}^{N} A_n\right)_{N\geq 1}$, égale à $(A_N)_{N\geq 1}$, est une suite croissante au sens de l'inclusion, et l'ensemble $\bigcup_{n=1}^{+\infty} A_n$ est habituellement noté $\lim_{n\to+\infty} A_n$, de sorte que le résultat de la proposition précédente peut encore s'écrire :

$$\mathbb{P}\Big(\lim_{n\to+\infty} A_n\Big) = \lim_{n\to+\infty} \mathbb{P}(A_n).$$

De même, la suite $\left(\bigcap_{n=1}^{N} B_n\right)_{N\geq 1}$, égale à $(B_N)_{N\geq 1}$, est une suite décroissante au sens de l'inclusion, et l'ensemble $\bigcap_{n=1}^{+\infty} B_n$ est habituellement noté $\lim_{n\to+\infty} B_n$, de sorte que le résultat de la proposition précédente s'écrit :

$$\mathbb{P}\Big(\lim_{n\to+\infty} B_n\Big) = \lim_{n\to+\infty} \mathbb{P}(B_n).$$

Chapitre 2

Construction d'espaces probabilisés

2.1 Caractérisation d'une probabilité : cas fini ou dénombrable

On suppose que l'univers Ω est fini ou dénombrable. L'ensemble des événements est donc $\mathcal{P}(\Omega)$.

Avec la définition de probabilité du chapitre précédent, il est n'est pas si aisé de définir concrètement une probabilité. Dans le cas où Ω est fini ou dénombrable, on dispose de la caractérisation suivante.

Proposition 2.1. *Une probabilité sur $(\Omega, \mathcal{P}(\Omega))$ est caractérisée par sa valeur sur les singletons $\{\omega\}$, pour tout $\omega \in \Omega$. Réciproquement, à toute famille $(p_\omega)_{\omega\in\Omega}$ telle que :*

1. *pour tout $\omega \in \Omega$, $0 \le p_\omega \le 1$,*
2. $\sum_{\omega\in\Omega} p_\omega = 1$,

on peut associer une unique probabilité $\mathbb{P}$ sur $(\Omega, \mathcal{P}(\Omega))$ définie par : $\mathbb{P}(\{\omega\}) = p_\omega$. On étend ensuite $\mathbb{P}$ à $\mathcal{P}(\Omega)$ par σ-additivité : pour tout $A \in \mathcal{P}(\Omega)$,

$$\mathbb{P}(A) = \sum_{\omega\in A} p_\omega.$$

2.2 Cas où l'univers est fini

Un exemple particulier et très classique de choix de probabilité nous amène à toutes les questions de dénombrements.

Définition. Une probabilité $\mathbb{P}$ sur $(\Omega, \mathcal{A})$ est dite *uniforme*, si $\mathbb{P}(\{\omega\})$ ne dépend pas de $\omega \in \Omega$. On dit alors que l'on est en situation d'*équiprobabilité*.

Corollaire 2.2. *Dans ce cas, pour tout $\omega \in \Omega$, $\mathbb{P}(\{\omega\}) = \frac{1}{\mathrm{card}(\Omega)}$, et, pour tout événement $A \in \mathcal{P}(\Omega)$, on a :*

$$\mathbb{P}(A) = \frac{\mathrm{card}(A)}{\mathrm{card}(\Omega)}.$$

Exemple. Reprenons la question 4 des exemples 1.1 et 1.2 du Chapitre 1 et calculons la probabilité de l'événement A "la somme des dés est égale à 4".

Supposons que l'on ait choisi l'espace Ω_1. Alors, on est en situation d'équiprobabilité et la probabilité $\mathbb{P}_1$ sur Ω_1 est uniforme, de sorte que pour tout $(i,j) \in \{1, \cdots, 6\}^2$:

$$\mathbb{P}_1[\{(i,j)\}] = \frac{1}{36}.$$

Ainsi, $\mathbb{P}_1(A) = \mathbb{P}_1[\{(1,3),(2,2),(3,1)\}] = \frac{\text{card}(A)}{\text{card}(\Omega_1)} = \frac{3}{36}$.

Supposons maintenant que l'on ait choisi l'espace Ω_2. Alors, on n'est plus en situation d'équiprobabilité. Au vu des conditions de l'expérience, on définit $\mathbb{P}_2$ ainsi :

$$\mathbb{P}_2(\{2\}) = \mathbb{P}_2(\{12\}) = \frac{1}{36}, \quad \mathbb{P}_2(\{3\}) = \mathbb{P}_2(\{11\}) = \frac{1}{18}, \quad \mathbb{P}_2(\{4\}) = \mathbb{P}_2(\{10\}) = \frac{1}{12},$$
$$\mathbb{P}_2(\{5\}) = \mathbb{P}_2(\{9\}) = \frac{1}{9}, \quad \mathbb{P}_2(\{6\}) = \mathbb{P}_2(\{8\}) = \frac{5}{36}, \quad \mathbb{P}_2(\{7\}) = \frac{1}{6}.$$

Ainsi, $\mathbb{P}_2(A) = \mathbb{P}_2(\{4\}) = \frac{1}{12}$ mais $\frac{\text{card}(A)}{\text{card}(\Omega_2)} = \frac{1}{11}$, d'où $\mathbb{P}_2(A) \neq \frac{\text{card}(A)}{\text{card}(\Omega_2)}$.

Remarquer que s'il on avait choisi l'univers Ω_3, on ne serait pas non plus en situation d'équiprobabilité.

Cet exemple montre qu'il est très important de spécifier le choix d'univers et de probabilité. Bien que les résultats finaux ne changent pas, les raisonnements pour y arriver sont différents et doivent être explicités.

Lorsque l'espace des états est fini, les calculs de probabilités se ramènent essentiellement à des problèmes de dénombrement, sujet de la section suivante.

2.2.1 Dénombrement, modèle d'urne

Définition. Une *population* de taille N est un ensemble $\mathcal{S} = \{s_1, \cdots, s_N\}$ de N éléments. Les éléments de $\mathcal{S}$ sont les *individus* de la population $\mathcal{S}$. La *taille* N est le nombre d'éléments de $\mathcal{S}$.

Tirages ordonnés

Un *échantillon* de taille r est un r-uplet $(s_{i_1}, \cdots, s_{i_r})$ d'éléments de $\mathcal{S}$. Deux procédures sont possibles.

• *Le tirage avec remise.* Dans ce cas, chaque élément de l'ensemble peut être choisi à plusieurs reprises. On parle alors d'*échantillon de taille r avec répétitions.* Soit Ω_1 l'ensemble de ces échantillons, alors $\Omega_1 = \{s_1, \cdots, s_N\}^r$, et :

$$\text{card}(\Omega_1) = N^r.$$

• *Le tirage sans remise.* Dans ce cas, chaque élément de l'ensemble peut être choisi au plus une fois. On parle alors d'*échantillon de taille r sans répétition*, ou d'*arrangement des éléments de $\mathcal{S}$ pris r à r*. Naturellement, on impose les conditions supplémentaires $r \leq N$, et $\forall\, j \neq k,\ i_j \neq i_k$. Soit Ω_2 l'ensemble de ces échantillons, on a alors :

$$\text{card}(\Omega_2) = N(N-1)\cdots(N-r+1) = \frac{N!}{(N-r)!}$$

Ce nombre a deux notations usuelles : A_N^r ou $(N)_r$ (*symbole de Pochhammer*).

Exemple. On considère une population de taille N et un échantillon aléatoire de taille r avec répétition. On choisit alors comme univers Ω_1 que l'on munit de la probabilité uniforme, notée $\mathbb{P}$. On s'intéresse à l'événement A "aucun individu n'a été choisi plus d'une fois" qui est la même chose que "tous les individus sont distincts". Alors on a, $A = \Omega_2$ et :

$$\mathbb{P}(A) = \frac{\mathrm{card}(\Omega_2)}{\mathrm{card}(\Omega_1)} = \frac{A_N^r}{N^r}.$$

Donnons quelques applications de ce résultat.

1. On jette un dé six fois de suite. Alors la probabilité d'obtenir six nombres distincts est $\frac{6!}{6^6} \sim 0,015$.
2. Supposons que dans une ville, il y ait sept accidents par semaine. Alors la probabilité d'avoir exactement un accident chaque jour de la semaine est $\frac{7!}{7^7} \sim 0,00612$.

TIRAGES NON ORDONNÉS

Une *sous-population* de taille r est un sous-ensemble $\{s_{i_1}, \cdots, s_{i_r}\}$ d'éléments de $\mathcal{S}$. De manière similaire aux tirages ordonnés, deux procédures sont possibles.

• *Le tirage sans remise.* On parle alors de *sous-population de taille r sans répétition*, ou de *combinaison de r éléments.* On impose à nouveau les conditions supplémentaires, $r \leq N$, et $\forall\, j \neq k$, $i_j \neq i_k$. Soit Ω_3 l'ensemble de ces populations, on a alors :

$$\mathrm{card}(\Omega_3) = \frac{N!}{(N-r)!r!}.$$

Ce nombre, appelé *coefficient binomial*, a deux notations usuelles : C_N^r (notation qui n'est plus tellement en vigueur) ou $\binom{N}{r}$.

Démonstration. Chacun des sous-ensembles à r éléments fournit $r!$ échantillons de taille r sans répétition, de sorte que $\mathrm{card}(\Omega_2) = r!\,\mathrm{card}(\Omega_3)$. □

Exemple 2.1.

1. On appelle main de poker l'ensemble des 5 cartes que chacun des quatre joueurs reçoit lors de la distribution d'un jeu qui en contient 32. Alors il existe $\binom{32}{5}$ mains différentes. Soit A l'événement "les hauteurs des 5 cartes sont différentes", calculons $\mathrm{card}(A)$. On peut choisir ces hauteurs de $\binom{8}{5}$ manières différentes. Il faut ensuite choisir la couleur (trèfle, carreau, cœur, pique) de chacune des hauteurs. Ainsi :

$$\mathrm{card}(A) = \binom{8}{5} 4^5.$$

Étant donné que toutes les mains sont supposées équiprobables, la probabilité d'obtenir une main dont les 5 cartes ont une hauteur différente est :

$$\mathbb{P}(A) = \frac{\binom{8}{5}4^5}{\binom{32}{5}}.$$

2. Une urne contient N_b boules blanches et N_n boules noires. Posons $N = N_b + N_n$. On tire r boules avec remise dans l'urne, il y a alors N^r tirages possibles. Soit A_k l'événement "on a tiré exactement k boules blanches", calculons $\mathrm{card}(A_k)$. L'événement A_k est réalisé lorsque l'issue est constituée de k boules blanches et $r-k$ boules noires. Il y a $\binom{r}{k}$ façons de choisir la position des boules blanches, la position des boules noires est ensuite fixée. Pour chacune des positions de boule blanche, il y a ensuite N_b choix de boules blanches possibles, et pour chacune des positions de boule noire, il y a N_n choix possibles, ainsi :
$$\mathrm{card}(A_k) = \binom{r}{k} N_b^k N_n^{r-k}.$$
Étant donné que tous les tirages sont supposés équiprobables, la probabilité d'obtenir exactement k boules blanches lors d'un tirage de r boules avec remise est :
$$\mathbb{P}(A_k) = \frac{\binom{r}{k} N_b^k N_n^{r-k}}{N^r} = \binom{r}{k} \left(\frac{N_b}{N}\right)^k \left(\frac{N_n}{N}\right)^{r-k}.$$
Ceci est un exemple de la *loi binomiale*, que nous reverrons plus tard.

3. Soit $\mathcal{S}$ une population de taille N (*ex.* des étudiants), que l'on range en deux catégories a et b incompatibles (*ex.* filles et garçons), de tailles respectives N_a et $N_b = N - N_a$. On choisit "au hasard" une sous-population de taille r sans répétition, il y a alors $\binom{N}{r}$ choix possibles. Soit A_k l'événement "on a choisi exactement k individus de la catégorie a", calculons $\mathrm{card}(A_k)$. L'événement est réalisé lorsque l'issue est constituée de k individus de la catégorie a et $r-k$ de la catégorie b. Il y a $\binom{N_a}{k}$ façons de choisir les k individus de la catégorie a et pour chacune il y a $\binom{N-N_a}{r-k}$ façons de choisir les individus restants dans la catégorie b, ainsi :
$$\mathrm{card}(A_k) = \binom{N_a}{k}\binom{N-N_a}{r-k}.$$
Remarquer que pour que ceci ait un sens, il faut que $0 \leq k \leq \min\{r, N_a\}$. Étant donné que tous les tirages sont supposés équiprobables, la probabilité d'obtenir k individus de la catégorie a lors de ce tirage est :
$$\mathbb{P}(A_k) = \frac{\binom{N_a}{k}\binom{N_b}{r-k}}{\binom{N}{r}}.$$
Ceci est un exemple de la *loi hypergéométrique.*

Remarque. Supposons que $N_a = N_a(N)$ soit une fonction de N et que le nombre total de boules tende vers l'infini, de sorte que la proportion $\frac{N_a}{N}$ tende vers p (et donc que $\frac{N_b}{N}$ tende vers $1-p$), avec $0 < p < 1$. Ainsi, N_a et N_b tendent vers $+\infty$ avec N. Fixons $r \geq 0$ et k compris entre 0 et r. Alors, pour N assez grand, on a $N_a \geq k$, $N_b \geq r-k$ et $\mathbb{P}(A_k)$

peut s'écrire :

$$\begin{aligned}
\mathbb{P}(A_k) &= \frac{N_a(N_a-1)\dots(N_a-k+1)}{k!}\frac{N_b(N_b-1)\dots(N_b-r+k+1)}{(r-k)!}\frac{r!}{N(N-1)\dots(N-r+1)}\\
&= \frac{r!}{k!(r-k)!}\frac{N_a(N_a-1)\dots(N_a-k+1)\,N_b(N_b-1)\dots(N_b-r+k+1)}{N(N-1)\dots(N-r+1)}\\
&= \frac{r!}{k!(r-k)!}\frac{N^k\,\frac{N_a}{N}(\frac{N_a}{N}-\frac{1}{N})\dots(\frac{N_a}{N}-\frac{k-1}{N})\,N^{r-k}\,\frac{N_b}{N}(\frac{N_b}{N}-\frac{1}{N})\dots(\frac{N_b}{N}-\frac{r-k-1}{N})}{N^r\,1(1-\frac{1}{N})\dots(1-\frac{r-1}{N})}\\
&= \binom{r}{k}\frac{\frac{N_a}{N}(\frac{N_a}{N}-\frac{1}{N})\dots(\frac{N_a}{N}-\frac{k-1}{N})\,\frac{N_b}{N}(\frac{N_b}{N}-\frac{1}{N})\dots(\frac{N_b}{N}-\frac{r-k-1}{N})}{1(1-\frac{1}{N})\dots(1-\frac{r-1}{N})}\\
&\xrightarrow[N\to+\infty]{} \frac{\binom{r}{k}\,p^k\,(1-p)^{r-k}}{1}.
\end{aligned}$$

Ainsi, $\mathbb{P}(A_k)$ tend vers $\binom{r}{k}p^k(1-p)^{r-k}$. On a donc obtenu la loi binomiale comme limite de lois hypergéométriques. Ce résultat est intuitif, car lorsque le nombre de boules est très grand, que le tirage s'effectue avec ou sans remise ne change pas grand chose : on a peu de chance de tirer deux fois la même boule.

• *Partitionnement*

Soient $r_1, \cdots, r_k$ des entiers positifs (éventuellement nuls) tels que, $r_1 + \cdots + r_k = N$. Le nombre de façons de répartir N objets dans k familles de sorte que la i-ième famille contienne r_i éléments est égal à :

$$\frac{N!}{r_1!\cdots r_k!}.$$

Ce nombre se note $\binom{N}{r_1\cdots r_k}$ et s'appelle *coefficient multinomial.*

Démonstration. Pour remplir la première famille, il faut choisir r_1 objets parmi N, ce qui peut se faire de $\binom{N}{r_1}$ façons. Pour remplir la seconde famille, il faut choisir r_2 objets parmi $N-r_1$, soit $\binom{N-r_1}{r_2}$. En continuant ainsi, on obtient que le nombre de telles répartitions est de :

$$\binom{N}{r_1}\binom{N-r_1}{r_2}\cdots\binom{N-r_1-\cdots-r_{k-1}}{r_k} = \frac{N!}{r_1!\cdots r_k!}.$$

□

Exemple. Le nombre d'anagrammes du mot CHERCHER est $\frac{8!}{2!2!2!2!}$.

• *Le tirage avec remise.* On parle alors de *sous-population de taille r avec répétitions.* Soit Ω_4 l'ensemble de ces populations, on a alors :

$$\mathrm{card}(\Omega_4) = \binom{N+r-1}{N-1} = \binom{N+r-1}{r}.$$

Démonstration. Ce problème revient à placer r boules indistinguables dans N urnes. En effet, le nombre de boules dans la i-ième urne compte le nombre de répétitions de l'individu i lors du tirage. Représentons les r boules par r étoiles alignées, avec une cloison à chacune des extrémités. Par exemple, lorsque $r = 7$,

$$|*******|$$

Répartir les r boules dans N urnes revient à rajouter $N-1$ cloisons formant les N urnes. Par exemple, lorsque $r=7$, $N=3$,

$$|**||*****|,$$

représente le tirage : $s_1, s_1, s_3, s_3, s_3, s_3, s_3$. Ainsi, ce problème revient à placer $N-1$ cloisons sur $N+r-1$ positions, les positions restantes étant occupées par des $*$.

$\square$

Exemple. Soient $r \in \mathbb{N}^*$ et $n \in \mathbb{N}^*$. On cherche à compter le nombre de suites d'entiers naturels $r_1, \cdots, r_n$, telles que :

$$r_1 + \cdots + r_n = r.$$

Ce problème revient à placer r boules indistinguables dans n urnes, où le nombre de boules dans la i-ième urne représente r_i. Ainsi, le nombre de ces suites est $\binom{n+r-1}{n-1}$. Par exemple, si $r=10$, $n=3$,

$$|**|*****|***|$$

représente la partition $(2,5,3)$ de 10. Remarquer que ces suites sont naturellement ordonnées de sorte que l'on distingue $(2,5,3)$ de $(5,3,2)$.

2.3 Cas où l'univers est infini dénombrable

Lorsque Ω est infini dénombrable, on procède de la même manière que dans le cas fini : on prend $\mathcal{A} = \mathcal{P}(\Omega)$, et on associe à chaque événement élémentaire $\omega \in \Omega$, sa probabilité, $\mathbb{P}(\{\omega\}) = p_\omega \in [0,1]$, avec :

$$\sum_{\omega \in \Omega} p_\omega = 1.$$

Exemple. On jette une pièce de monnaie jusqu'à l'obtention du premier pile. On peut choisir $\Omega = \mathbb{N}^* \cup \{\infty\}$ où, pour tout $k \in \mathbb{N}^*$, $\{k\}$ représente l'événement "le premier pile est obtenu au k-ième jet", et $\{\infty\}$ représente l'événement "pile ne sort jamais". Si la pièce est équilibrée, on aura, pour tout $k \in \mathbb{N}^*$:

$$\mathbb{P}(\{k\}) = \frac{1}{2^k}.$$

Comme $\mathbb{N}^*$ et $\{\infty\}$ sont disjoints, $1 = \mathbb{P}[\Omega] = \mathbb{P}[\{\infty\}] + \mathbb{P}[\underset{k\in\mathbb{N}^*}{\cup}\{k\}]$. Ainsi, la probabilité que pile ne sorte jamais est donnée par :

$$\begin{aligned}
\mathbb{P}(\{\infty\}) &= 1 - \mathbb{P}[\underset{k\in\mathbb{N}^*}{\cup}\{k\}] \\
&= 1 - \sum_{k=1}^{+\infty} \mathbb{P}[\{k\}], \text{ car les événements sont disjoints} \\
&= 1 - \sum_{k=1}^{+\infty} \frac{1}{2^k} = 0.
\end{aligned}$$

La probabilité que le premier pile sorte après un nombre pair de lancers est :
$\mathbb{P}(\{2,4,6,\cdots\}) = \sum\limits_{k=1}^{+\infty} \mathbb{P}(\{2k\}) = \sum\limits_{k=1}^{+\infty} \frac{1}{2^{2k}} = \frac{1}{3}.$

2.4 Cas où l'univers est infini non-dénombrable

Cette situation est beaucoup plus subtile que dans les cas précédents. On ne peut plus définir une probabilité en définissant celle des singletons (événements élémentaires) car celle-ci est nulle. La procédure est alors la suivante :

— on détermine une algèbre [1] d'événements intéressants sur laquelle on définit une probabilité ;

— on utilise un théorème fondamental de la théorie de la mesure, le "théorème d'extension de Carathéodory", qui affirme qu'une probabilité sur une algèbre s'étend de façon unique en une probabilité sur la tribu engendrée par l'algèbre.

Exemple. Souvenez-vous que lorsque l'espace des états est $\mathbb{R}$, on prend comme tribu celle des boréliens (la plus petite tribu engendrée par les ouverts de $\mathbb{R}$). On admettra que cette tribu est engendrée par les intervalles de la forme $]-\infty, x]$, $x \in \mathbb{R}$, et qu'une mesure de probabilité $\mathbb{P}$ sur $\mathbb{R}$ est entièrement caractérisée par la valeur qu'elle attribue aux intervalles de cette forme.

1. Un ensemble $\mathcal{A}$ de parties de Ω est une *algèbre*, si $\Omega \in \mathcal{A}$, $\mathcal{A}$ est stable par complémentation et par réunions finies. À la différence d'une tribu on ne demande pas la stabilité par réunions dénombrables.

Exercices

Exercice 2.1. Combien de menus différents peut-on composer si l'on a le choix entre 5 entrées, 2 plats et 4 desserts ?

Solution de l'exercice 2.1. On peut composer $5 \times 2 \times 4 = 40$ menus différents.

Exercice 2.2. Une femme a dans sa garde-robe 6 pantalons, 5 hauts et 3 vestes. Elle choisit au hasard un pantalon, un haut et une veste. De combien de façons différentes peut-elle s'habiller ?

Solution de l'exercice 2.2. Elle peut s'habiller de $6 \times 5 \times 3 = 90$ façons différentes.

Exercice 2.3. Un questionnaire à choix multiples, autorisant une seule réponse par question, comprend 20 questions. Pour chaque question, on propose 4 réponses possibles. De combien de façons peut-on répondre à ce questionnaire ?

Solution de l'exercice 2.3. Il y a 4^{20} façons de répondre à ce questionnaire.

Exercice 2.4. Un clavier de 9 touches (A, B, C, 1 ,2, 3, 4, 5 et 6) permet de composer le code d'entrée d'un immeuble, à l'aide d'une lettre suivie d'un nombre de 3 chiffres distincts ou non.

1. Combien de codes différents peut-on former ?
2. Combien y a-t-il de codes sans le chiffre 4 ?
3. Combien y a-t-il de codes comportant au moins une fois le chiffre 4 ?
4. Combien y a-t-il de codes comportant des chiffres distincts ?
5. Combien y a-t-il de codes comportant au moins deux chiffres identiques ?

Solution de l'exercice 2.4.

1. Il y a 3×6^3 codes différents.
2. Il y a 3×5^3 codes différents sans le chiffre 4.
3. Il y a $3 \times 6^3 - 3 \times 5^3$ codes contenant au moins une fois le chiffre 4.
4. Il y a $3 \times 6 \times 5 \times 4$ codes comportant des chiffres distincts.
5. Il y a $3 \times 6^3 - 3 \times 6 \times 5 \times 4$ codes comportant au moins deux chiffres identiques.

Exercice 2.5. Au service du personnel, on compte 12 célibataires parmi les 30 employés. On désire faire un sondage : pour cela on choisit un échantillon de quatre personnes dans ce service.

1. Quel est le nombre d'échantillons différents possibles ?
2. Quel est le nombre d'échantillons ne contenant aucun célibataire ?
3. Quel est le nombre d'échantillons contenant au moins un célibataire ?

Solution de l'exercice 2.5.

1. Le nombre d'échantillons différents possibles est $\binom{30}{4}$.
2. Le nombre d'échantillons ne contenant aucun célibataire est $\binom{18}{4}$.

3. Le nombre d'échantillons contenant au moins un célibataire est $\binom{30}{4} - \binom{18}{4}$.

Exercice 2.6. Soient A et B deux évènements de probabilités, $\mathbb{P}(A) = 3/4$ et $\mathbb{P}(B) = 1/3$. Montrer que

$$\frac{1}{12} \leq \mathbb{P}(A \cap B) \leq \frac{1}{3}.$$

Solution de l'exercice 2.6. Montrons l'inégalité de gauche. Comme conséquence de la définition d'une probabilité, on a :

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

Or $\mathbb{P}(A \cup B) \leq 1$, ce qui implique $\mathbb{P}(A \cap B) \geq \mathbb{P}(A) + \mathbb{P}(B) - 1$. Dans notre cas cela donne

$$\mathbb{P}(A \cap B) \geq \frac{3}{4} + \frac{1}{3} - 1 = \frac{1}{12},$$

et montre la première inégalité. Pour la deuxième, on utilise que $A \cap B \subset B$, ce qui implique

$$\mathbb{P}(A \cap B) \leq \mathbb{P}(B) = \frac{1}{3}.$$

Exercice 2.7. Supposons que 23 personnes sont dans une même salle. Quelle est la probabilité qu'au moins deux d'entre elles aient l'anniversaire le même jour ? (On ne considérera pas les années bissextiles.)

Solution de l'exercice 2.7. L'univers Ω est formé de tous les 23-uplets de jours d'anniversaire. On a donc $\Omega = \{1, \cdots, 365\}^{23}$ et $\mathrm{card}(\Omega) = 365^{23}$. On suppose que les dates d'anniversaire sont distribuées "au hasard" sur l'année, de sorte que l'on munit Ω de la probabilité uniforme, notée $\mathbb{P}$. Ainsi, si $A \in \mathcal{P}(\Omega)$ est un événement, $\mathbb{P}(A) = \frac{\mathrm{card}(A)}{\mathrm{card}(\Omega)}$.
On considère l'événement A "les personnes ont toutes leur anniversaire un jour différent". L'événement A est formé de tous les échantillons de taille 23 sans répétition, donc $\mathrm{card}(A) = A_{365}^{23}$, et

$$\mathbb{P}(A) = \frac{A_{365}^{23}}{(365)^{23}}.$$

La probabilité qu'au moins deux personnes aient leur anniversaire le même jour est la probabilité de l'événement complémentaire, et est donc égale à $1 - \frac{A_{365}^{23}}{(365)^{23}} = 0,507...$ Cette probabilité est d'environ 0,97 s'il y a 50 personnes, et d'environ 0,999 s'il y en a 100.

Exercice 2.8. Dans une course, n chevaux sont au départ. On suppose qu'ils ont tous la même chance de gagner. Calculer la probabilité de gagner le tiercé avec un ticket :

1. dans l'ordre,
2. dans l'ordre ou dans un ordre différent,
3. dans un ordre différent ?

Solution de l'exercice 2.8. L'univers est l'ensemble des tiercés possibles, soit

$$\Omega = \{(i, j, k) \, : \, i, \, j, \, k \in \{1, \cdots, n\}, i \neq j, \, j \neq k, \, k \neq i\}.$$

Alors, $\mathrm{card}(\Omega) = A_n^3 = n(n-1)(n-2)$. On suppose que tous les tiercés ont la même chance de se réaliser, de sorte que l'on munit Ω de la probabilité uniforme, notée $\mathbb{P}$. Ainsi, pour tout $A \in \mathcal{P}(\Omega)$, $\mathbb{P}(A) = \frac{\mathrm{card}(A)}{\mathrm{card}(\Omega)}$.

1. Soit A l'événement "obtenir le tiercé gagnant dans l'ordre". Comme il existe un unique tiercé gagnant, $\mathrm{card}(A) = 1$, d'où la probabilité cherchée est :
$$\mathbb{P}(A) = \frac{1}{n(n-1)(n-2)}.$$
2. Soit B l'événement "obtenir le tiercé gagnant dans l'ordre ou dans un ordre différent". Comme il y a 3! manières d'ordonner le tiercé gagnant, $\mathrm{card}(B) = 6$, d'où la probabilité cherchée est :
$$\mathbb{P}(B) = \frac{6}{n(n-1)(n-2)}.$$
3. Soit C l'événement "obtenir le tiercé gagnant dans un ordre différent". Comme il y a $3! - 1 = 5$ manières d'ordonner le tiercé gagnant dans un ordre différent, $\mathrm{card}(C) = 5$, d'où la probabilité cherchée est :
$$\mathbb{P}(C) = \frac{5}{n(n-1)(n-2)}.$$

Exercice 2.9. Un joueur de poker reçoit une main de 5 cartes d'un jeu de 32 cartes. Quelle est la probabilité qu'il reçoive :

1. une seule paire (deux cartes de même hauteur) ;
2. deux paires ;
3. un brelan (trois cartes de même hauteur et pas de paire ni de carré) ;
4. un carré (quatre cartes de même hauteur) ;
5. un full (une paire et un brelan) ?

Solution de l'exercice 2.9. L'univers Ω est l'ensemble des mains de 5 cartes possibles, c'est-à-dire l'ensemble des combinaisons de 5 éléments pris parmi 32, de sorte que, $\mathrm{card}(\Omega) = \binom{32}{5}$. On suppose que toutes les mains sont équiprobables et on munit Ω de la probabilité uniforme, notée $\mathbb{P}$. Ainsi, si $A \in \mathcal{P}(\Omega)$ est un événement, $\mathbb{P}(A) = \frac{\mathrm{card}(A)}{\mathrm{card}(\Omega)}$.

1. Soit A l'événement "le joueur possède une seule paire", calculons $\mathrm{card}(A)$.
- Choix de la hauteur de la paire : $\binom{8}{1}$.
- Choix de la couleur (trèfle, carreau, cœur, pique) de chacune des cartes de la paire : $\binom{4}{2}$.
- Choix des hauteurs des trois autres cartes : $\binom{7}{3}$.
- Choix de la couleur de chacune des trois autres cartes : 4^3.
Ainsi, $\mathrm{card}(A) = \binom{8}{1}\binom{4}{2}\binom{7}{3}4^3$, et
$$\mathbb{P}(A) = \frac{\binom{8}{1}\binom{4}{2}\binom{7}{3}4^3}{\binom{32}{5}}.$$

2. Soit B l'événement “le joueur possède deux paires”, calculons card(B).
   - Choix des hauteurs des deux paires : $\binom{8}{2}$.
   - Choix de la couleur de chacune des cartes de chacune des paires : $\binom{4}{2}^2$.
   - Choix de la hauteur de la dernière carte : $\binom{6}{1}$.
   - Choix de la couleur de la dernière carte : 4.

$$\text{Ainsi, } \mathbb{P}(B) = \frac{\binom{8}{2}\binom{4}{2}^2\binom{6}{1}4}{\binom{32}{5}}.$$

3. Soit C l'événement “le joueur possède un brelan”.
   - Choix de la hauteur du brelan : $\binom{8}{1}$.
   - Choix de la couleur de chacune des cartes du brelan : $\binom{4}{3}$.
   - Choix de la hauteur des deux cartes restantes : $\binom{7}{2}$.
   - Choix de la couleur de chacune des deux cartes restantes : 4^2.

$$\text{Ainsi, } \mathbb{P}(C) = \frac{\binom{8}{1}\binom{4}{3}\binom{7}{2}4^2}{\binom{32}{5}}.$$

4. Soit D l'événement “le joueur possède un carré”.
   - Choix de la hauteur du carré : $\binom{8}{1}$.
   - Choix de la couleur de chacune des cartes du carré : $\binom{4}{4} = 1$.
   - Choix de la hauteur de la carte restante : 7.
   - Choix de la couleur de la carte restante : 4.

$$\text{Ainsi, } \mathbb{P}(D) = \frac{8.7.4}{\binom{32}{5}}.$$

5. Soit E l'événement “le joueur possède un full”.
   - Choix de la hauteur de la paire : 8.
   - Choix de la couleur de la paire : $\binom{4}{2}$.
   - Choix de la hauteur du brelan : 7.
   - Choix de la couleur du brelan : $\binom{4}{3}$.

$$\text{Ainsi, } \mathbb{P}(E) = \frac{8.6.7.4}{\binom{32}{5}}.$$

Exercice 2.10. (D'après C. Bouzitat et G. Pagès, En passant par hasard... Chapitre XI. Ed. Vuibert (1999)). Au loto, le joueur doit cocher 6 numéros dans une grille en comportant 49. Un tirage consiste à extraire, sans remise, 6 boules numérotées d'une urne, dont les numéros sont dits gagnants, et une 7-ième boule fournissant le numéro dit complémentaire. Est gagnant du premier rang, toute grille sur laquelle sont cochés les 6 numéros gagnants. Est gagnante du 2-ième rang, toute grille sur laquelle sont cochés 5 des 6 numéros gagnants et dont le 6-ième numéro est le numéro complémentaire. Est gagnant du 3-ième rang, toute grille sur laquelle sont exactement cochés 5 des 6 numéros gagnants.

Considérons une grille validée et notons

$$p_k = \mathbb{P}(\text{la grille est gagnante au } k\text{-ième rang}).$$

Calculer p_k pour $k \in \{1, 2, 3\}$.

Solution de l'exercice 2.10. L'univers Ω est l'ensemble des combinaisons de 6 numéros choisis parmi 49, de sorte que, $\text{card}(\Omega) = \binom{49}{6}$. On suppose que toutes ces combinaisons sont équiprobables et on munit Ω de la probabilité uniforme, notée $\mathbb{P}$. Ainsi, si $A \in \mathcal{P}(\Omega)$ est un événement, $\mathbb{P}(A) = \frac{\text{card}(A)}{\text{card}(\Omega)}$.

Pour $k \in \{1, 2, 3\}$, soit A_k l'événement "la grille du joueur est une grille gagnante du k-ième rang".

1. A_1 est formé des grilles qui correspondent exactement au tirage, donc $\text{card}(A_1) = 1$ et : $p_1 = \frac{1}{\binom{49}{6}} \sim 7,15.10^{-8}$.
2. A_2 est formé des grilles qui contiennent 5 des 6 numéros du tirage et exactement le numéro complémentaire, donc : $p_2 = \frac{\binom{6}{5}}{\binom{49}{6}} \sim 4,29.10^{-7}$.
3. A_3 est formé des grilles qui contiennent 5 des 6 numéros du tirage et un numéro autre (que les 7 tirés), donc :
$$p_3 = \frac{\binom{6}{5}\binom{49-7}{1}}{\binom{49}{6}} = \frac{\binom{6}{5}42}{\binom{49}{6}} \sim 1,8.10^{-5}.$$

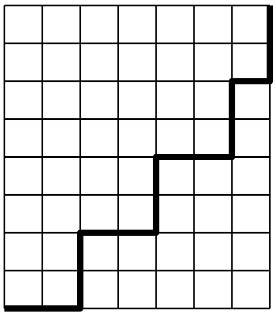
Exercice 2.11. On considère la distribution aléatoire de r boules dans n urnes. Quelle est la probabilité qu'une urne donnée contienne exactement k boules ? ($k \leq r$)

Solution de l'exercice 2.11. L'espace des états Ω est l'ensemble de toutes les manières de distribuer r boules dans n urnes. Pour la première boule, il y a n choix d'urnes, pour la deuxième également, etc., donc :
$$\text{card}(\Omega) = n^r.$$
Comme la distribution des boules est aléatoire, on munit Ω de la probabilité uniforme, de sorte que si A est un événement de Ω, on a $\mathbb{P}(A) = \frac{\text{card}(A)}{\text{card}(\Omega)}$. Pour tout $k \in \{0, \cdots, r\}$, on introduit A_k l'événement "une urne donnée contient k boules". Alors il y a $\binom{r}{k}$ choix possibles pour ces k boules, et comme aucune autre boule ne peut être dans l'urne donnée, chacune a le choix entre $(n-1)$ urnes. Ainsi,
$$\text{card}(A_k) = \binom{r}{k}(n-1)^{r-k},$$
et
$$\mathbb{P}(A_k) = \binom{r}{k}\frac{(n-1)^{r-k}}{n^r} = \binom{r}{k}\left(1 - \frac{1}{n}\right)^{r-k}\left(\frac{1}{n}\right)^k.$$
On retrouve un exemple de *loi binomiale*.

Exercice 2.12. Combien l'équation $x_1 + x_2 + x_3 = 15$ a-t-elle de solutions entières et non négatives ?

Solution de l'exercice 2.12. On cherche l'ensemble des suites de 3 entiers naturels de somme 15. Il y a donc $\binom{17}{2}$ solutions.

Exercice 2.13. Un homme travaille à Manhattan, dans un quartier où les avenues sont orientées nord-sud et les rues est-ouest. Il travaille à 7 pâtés de maison à l'est et 8 pâtés de maisons au nord de son domicile. Pour aller à son travail chaque jour il parcourt donc la longueur de 15 pâtés de maison (il ne se dirige ni vers le sud, ni vers l'ouest). On suppose qu'il existe une voie le long de chaque pâté de maisons et qu'il peut prendre n'importe lesquelles de ce schéma rectangulaire. La figure ci-dessous illustre la situation ; un exemple de trajet est représenté en ligne grasse.



1. Proposer un codage permettant de décrire le trajet représenté.
2. Combien de trajets différents l'homme peut-il emprunter ?
3. L'homme prétend que le nombre de trajets est aussi le nombre de suites de 8 entiers naturels dont la somme est 8. A-t-il raison ?

Solution de l'exercice 2.13.

1. À tout chemin possible on peut associer la succession de choix que l'on fait à chaque intersection rencontrée. Ces choix peuvent être codés par "E" (est) et "N" (nord), ces deux directions étant les seules répondant aux conditions proposées. Le chemin dessiné devient par exemple,
$$EENNEENNEENNENN.$$
Il est évident que, vice-versa, la donnée de cette suite permet de connaître sans ambiguïté le chemin correspondant puisqu'à chaque intersection, on saura où se diriger.
2. Les chemins ayant le point de départ et d'arrivée donnés ont tous en commun qu'il faut aller 8 fois vers le nord et 7 fois vers l'est. Dans ce codage, on trouvera nécessairement 7 fois le "E" et 8 fois le "N". Donc tout codage est une suite de 15 symboles composée de 7 "E" et 8 "N" et réciproquement. Il apparaît donc une bijection entre l'ensemble des chemins et l'ensemble des suites de 15 caractères décrites ci-dessus. Le nombre de solutions est alors :
$$\binom{15}{7} = \binom{15}{8}.$$
3. À chaque chemin on peut associer la longueur des parcours que l'on fait vers le nord dans chacune des avenues que l'on rencontre. Il y a 8 avenues possibles, ce qui donnera, quel que soit le chemin, une suite de 8 entiers naturels dont la somme est 8. Par exemple, pour le chemin dessiné, on a :
$$0, 0, 2, 0, 2, 0, 2, 2.$$

À partir de ce nouveau codage, on retrouve l'ancien en remplaçant k par :

$$\underbrace{N \cdots N}_{k \text{ fois}} E,$$

Le dernier nombre est uniquement remplacé par la suite de k fois "N". Dans l'exemple ci-dessus, on remplace 0 par "E" et 2 par "NNE" (sauf, pour le dernier 2 qui est remplacé par "NN") et on retrouve le chemin codé en "N" et "E". Nous avons donc explicité une bijection entre l'ensemble des suites de 8 entiers naturels de somme 8 et celui des chemins.

Remarque. Les questions 1 et 3 sont deux interprétations de l'expérience qui consiste à placer 8 boules indistinguables dans 8 urnes. Dans le premier codage les parois des urnes (sauf la première et la dernière) sont remplacées par des "E" et les boules par des "N". Dans le deuxième codage, pour $i \in \{1, \cdots, 8\}$, r_i représente le nombre de boules dans la i-ième urne. La distribution de boules qui correspond à la séquence "$EENNEENNEENNENN$" ou "$0, 0, 2, 0, 2, 0, 2, 2$" est :

$$||| * *|| * *|| * *| * *|$$

Exercice 2.14. Un étudiant s'habille très vite le matin et prend, au hasard dans la pile d'habits, un pantalon, un tee-shirt, une paire de chaussettes ; il y a ce jour-là dans l'armoire 5 pantalons dont 2 noirs, 6 tee-shirt dont 4 noirs, 8 paires de chaussettes, dont 5 paires noires. Combien y-a-t-il de façons de s'habiller ? Quelles sont les probabilités des événements suivants :

1. il est tout en noir ;
2. une seule pièce est noire sur les trois.

Solution de l'exercice 2.14. Il y a $5 \times 6 \times 8$ façons de s'habiller.

1. Soit A l'événement "il est tout en noir".
$$\mathbb{P}(A) = \frac{2 \times 4 \times 5}{5 \times 6 \times 8}$$
2. Soit B l'événement "une seule pièce est noire sur les trois".
$$\mathbb{P}(B) = \frac{2 \times (6-4) \times (8-5) + (5-2) \times 4 \times (8-5) + (5-2) \times (6-4) \times 5}{5 \times 6 \times 8}$$

Exercice 2.15. On lance un dé équilibré à six faces. Calculer les probabilités d'obtenir un nombre pair, un multiple de 3, un nombre pair et supérieur à 3.

Solution de l'exercice 2.15. La probabilité d'obtenir un nombre pair est $\frac{3}{6}$. La probabilité d'obtenir un multiple de 3 est $\frac{2}{6}$. La probabilité d'obtenir un nombre pair et supérieur à 3 est $\frac{1}{6}$.

Exercice 2.16. Lors d'une loterie de Noël, 300 billets sont vendus aux enfants de l'école ; 4 billets sont gagnants. J'achète 10 billets, quelle est la probabilité pour que je gagne au moins un lot ?

Solution de l'exercice 2.16. La réponse est 1 moins la probabilité de ne gagner aucun lot, soit :

$$1 - \frac{\binom{296}{10}}{\binom{300}{10}}.$$

Chapitre 3

Conditionnement et indépendance

3.1 Probabilité conditionnelle

3.1.1 Définition

Motivons la définition de probabilité conditionnelle sur un exemple.

Soit Ω une population partitionnée en $\Omega = S \cup S^c$, où S représente l'ensemble des individus fumeurs. Soit F l'ensemble des femmes, de sorte que $F \cup F^c$ représente une autre partition de Ω. On suppose que l'on choisit un individu “au hasard", de sorte que l'on munit Ω de la probabilité uniforme, notée $\mathbb{P}$. Ainsi, la probabilité que l'individu choisi soit fumeur est :

$$\mathbb{P}(S) = \frac{\operatorname{card}(S)}{\operatorname{card}(\Omega)}.$$

Si maintenant on choisit un individu avec l'information supplémentaire qu'il s'agit d'une femme, tout se passe comme si l'univers considéré est F, et que l'on a partitionné F (et non plus Ω) en $S \cap F$ et $S^c \cap F$. Ainsi la probabilité que l'individu choisi soit fumeur, étant donné l'information que c'est une femme, est égale à :

$$\frac{\operatorname{card}(S \cap F)}{\operatorname{card}(F)} = \frac{\operatorname{card}(S \cap F)/\operatorname{card}(\Omega)}{\operatorname{card}(F)/\operatorname{card}(\Omega)},$$

quantité encore égale, avec les notations précédentes, à

$$\frac{\mathbb{P}(S \cap F)}{\mathbb{P}(F)}.$$

Définition. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, et B un événement tel que $\mathbb{P}(B) > 0$. Pour tout $A \in \mathcal{A}$, on définit *la probabilité conditionnelle de A sachant B*, notée $\mathbb{P}_B(A)$, par :

$$\mathbb{P}_B(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}. \tag{3.1}$$

Lemme 3.1. *Sous les hypothèses de la définition ci-dessus, l'application $\mathbb{P}_B$ est une probabilité sur $(\Omega, \mathcal{A})$, ainsi $(\Omega, \mathcal{A}, \mathbb{P}_B)$ est un espace probabilisé.*

Démonstration. Il faut montrer que $\mathbb{P}_B$ satisfait aux trois axiomes caractérisant une probabilité.

— Par hypothèse, $\mathbb{P}(B) > 0$. De plus, pour tout $A \in \mathcal{A}$, on a $A \cap B \subset B$, d'où $0 \leq \mathbb{P}(A \cap B) \leq \mathbb{P}(B)$, et donc :
$$0 \leq \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \leq 1.$$
— Comme $\Omega \cap B = B$, on a $\mathbb{P}_B(\Omega) = 1$.
— Soit $(A_n)_{n\geq 1}$ une famille dénombrable d'événements deux à deux disjoints. Alors on a l'égalité $(\underset{n\geq 1}{\cup} A_n) \cap B = \underset{n\geq 1}{\cup}(A_n \cap B)$. Comme les événements $(A_n)_{n\geq 1}$ sont deux à deux disjoints, il en est de même pour les événements $(A_n \cap B)_{n\geq 1}$. Ainsi la σ-additivité de $\mathbb{P}$ implique $\mathbb{P}\left(\underset{n\geq 1}{\cup}(A_n \cap B)\right) = \sum_{n\geq 1} \mathbb{P}(A_n \cap B)$, et on conclut :
$$\mathbb{P}_B\left(\underset{n\geq 1}{\cup} A_n\right) = \frac{\mathbb{P}\left((\underset{n\geq 1}{\cup} A_n) \cap B\right)}{\mathbb{P}(B)} = \sum_{n\geq 1} \frac{\mathbb{P}(A_n \cap B)}{\mathbb{P}(B)} = \sum_{n\geq 1} \mathbb{P}_B(A_n).$$
□

Remarque.
— On utilise aussi la notation $\mathbb{P}(A|B)$. La notation $\mathbb{P}_B(A)$ met en valeur le fait que $\mathbb{P}_B$ soit une probabilité sur $(\Omega, \mathcal{A})$.
— Comme $(\Omega, \mathcal{A}, \mathbb{P}_B)$ est un espace probabilisé, $\mathbb{P}_B$ satisfait à toutes les propriétés de la proposition 1.3.
— Si $\mathbb{P}(A) > 0$ et $\mathbb{P}(B) > 0$, on peut réécrire l'équation (3.1) sous la forme :
$$\mathbb{P}(A \cap B) = \mathbb{P}_A(B)\,\mathbb{P}(A) = \mathbb{P}_B(A)\,\mathbb{P}(B).$$
— De manière plus générale, si $(A_k)_{1\leq k\leq n+1}$ sont des événements qui satisfont à $\mathbb{P}\left(\bigcap_{k=1}^{n} A_k\right) > 0$, alors :
$$\mathbb{P}\Big(\bigcap_{k=1}^{n+1} A_k\Big) = \mathbb{P}(A_1)\,\mathbb{P}_{A_1}(A_2)\,\mathbb{P}_{A_1\cap A_2}(A_3)\cdots\mathbb{P}_{A_1\cap\cdots\cap A_n}(A_{n+1}).$$

Exemple 3.1. On considère deux lancers successifs d'un même dé. Sachant que le premier jet donne 3, on souhaite calculer la probabilité que la somme soit strictement supérieure à 6.

Supposons que l'on ait choisi l'univers $\Omega = \{(i,j)\,:\, i \in \{1,\cdots,6\},\, j \in \{1,\cdots,6\}\}$, que l'on munit de la tribu $\mathcal{A} = \mathcal{P}(\Omega)$. On est alors en situation d'équiprobabilité, et on munit Ω de la probabilité uniforme, notée $\mathbb{P}$. Ainsi, pour tout événement A de $\mathcal{A}$, $\mathbb{P}(A) = \frac{\mathrm{card}(A)}{\mathrm{card}(\Omega)} = \frac{\mathrm{card}(A)}{36}$.

Soit B l'événement "le premier jet donne 3" ; B est le sous-ensemble $\{(3,j)\,:\, j \in \{1,\cdots,6\}\}$ de $\mathcal{A}$, son cardinal est égal à 6, d'où $\mathbb{P}(B) = \frac{6}{36} = \frac{1}{6} > 0$. La probabilité $\mathbb{P}(B)$ étant strictement positive, on peut conditionner par l'événement B.

Soit A l'événement "la somme des dés est strictement supérieure à 6" ; A est le sous-ensemble $\{(i,j) \in \Omega\,:\, i+j > 6\}$ de $\mathcal{A}$, et $A \cap B = \{(3,4),(3,5),(3,6)\}$, d'où : $\mathbb{P}(A \cap B) = \frac{3}{36} = \frac{1}{12}$. On conclut que :
$$\mathbb{P}_B(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{1}{2}.$$

3.1.2 Formule des probabilités totales et formule de Bayes

Définition. Soit I une partie finie ou dénombrable de $\mathbb{N}$. Une famille $(B_i)_{i\in I}$ d'événements de Ω forme un *système complet d'événements* de Ω, si

$$\forall\, i \neq j,\ B_i \cap B_j = \emptyset, \text{ et } \bigcup_{i\in I} B_i = \Omega.$$

Autrement dit, la famille $(B_i)_{i\in I}$ est une *partition* de Ω.

Théorème 3.2. *Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé.*

— ***Formule des probabilités totales**. Soit $(B_i)_{i\in I}$ un système complet d'événements de Ω, telle que pour tout $i \in I$, $\mathbb{P}(B_i) > 0$, et soit $A \in \mathcal{A}$. Alors :*

$$\mathbb{P}(A) = \sum_{i\in I} \mathbb{P}(A \cap B_i) = \sum_{i\in I} \mathbb{P}_{B_i}(A)\,\mathbb{P}(B_i).$$

Dans le cas particulier où $I = \{1, \cdots, n\}$, on a :

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A \cap B_i) = \sum_{i=1}^{n} \mathbb{P}_{B_i}(A)\,\mathbb{P}(B_i).$$

— ***Formule de Bayes**. Soient $A \in \mathcal{A}$ et $B \in \mathcal{A}$, tels que $0 < \mathbb{P}(A) < 1$ et $\mathbb{P}(B) > 0$, alors :*

$$\mathbb{P}_B(A) = \frac{\mathbb{P}_A(B)\mathbb{P}(A)}{\mathbb{P}_A(B)\mathbb{P}(A) + \mathbb{P}_{A^c}(B)\mathbb{P}(A^c)}.$$

Démonstration. Comme $(B_i)_{i\in I}$ est une partition de Ω, on a :

$$A = A \cap \Omega = A \cap (\underset{i\in I}{\cup} B_i) = \underset{i\in I}{\cup} (A \cap B_i),$$

où les événements $(A\cap B_i)_{i\in I}$ sont deux-à-deux disjoints. Ainsi, en utilisant la σ-additivité de $\mathbb{P}$:

$$\mathbb{P}(A) = \sum_{i\in I} \mathbb{P}(A \cap B_i) = \sum_{i\in I} \mathbb{P}_{B_i}(A)\,\mathbb{P}(B_i).$$

Par définition de la probabilité conditionnelle, on a :

$$\mathbb{P}_B(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}_A(B)\,\mathbb{P}(A)}{\mathbb{P}(B)}.$$

La formule de Bayes est obtenue en appliquant la formule des probabilités totales au dénominateur avec la partition $\{A, A^c\}$ puisque $\mathbb{P}(A) \neq 0$ et $\mathbb{P}(A^c) \neq 0$.

□

Exemple 3.2. Un sondage est effectué dans une société comprenant 40% de cadres et 60% d'ouvriers. On sait que 20% des cadres et 10% des ouvriers de cette société savent parler anglais. On interroge une personne “au hasard". Quelle est la probabilité que ce soit :

— un cadre sachant parler anglais ?
— un ouvrier sachant parler anglais ?
— une personne sachant parler anglais ?

L'individu interrogé sait parler anglais. Quelle est la probabilité que ce soit un ouvrier ?

On prend pour univers Ω l'ensemble des employés de la société, que l'on munit de la tribu $\mathcal{A} = \mathcal{P}(\Omega)$. Étant donné que l'on interroge une personne "au hasard", on munit Ω de la probabilité uniforme, notée $\mathbb{P}$.

Soit C l'événement "l'employé est un cadre", O "l'employé est un ouvrier", A "l'employé parle anglais". Alors, la donnée du problème nous dit que :

$$\mathbb{P}(C) = \frac{4}{10},\ \mathbb{P}(O) = \frac{6}{10}.$$

Comme ces deux probabilités sont strictement positives, on peut conditionner par rapport aux événements C et O. D'après la donnée nous savons que :

$$\mathbb{P}_C(A) = \frac{2}{10},\ \mathbb{P}_O(A) = \frac{1}{10}.$$

À la première question, on cherche $\mathbb{P}(C \cap A)$. D'après la définition des probabilités conditionnelles, on a :

$$\mathbb{P}(C \cap A) = \mathbb{P}_C(A)\mathbb{P}(C) = \frac{2}{10}\frac{4}{10} = 0,08.$$

De manière similaire pour la deuxième question :

$$\mathbb{P}(O \cap A) = \mathbb{P}_O(A)\mathbb{P}(O) = \frac{1}{10}\frac{6}{10} = 0,06.$$

Étant donné que $\{C, O\}$ forme une partition de Ω, d'après la formule des probabilités totales, on a :

$$\mathbb{P}(A) = \mathbb{P}(A \cap C) + \mathbb{P}(A \cap O) = 0,08 + 0,06 = 0,14.$$

Pour répondre à la dernière question, on utilise la formule de Bayes :

$$\mathbb{P}_A(O) = \frac{\mathbb{P}_O(A)\mathbb{P}(O)}{\mathbb{P}(A)} = \frac{1.6}{10.10}\frac{100}{14} = \frac{3}{7}.$$

Exemple 3.3. (Paradoxe de Simpson). Cet exemple réel[1] montre un paradoxe surprenant qui s'explique grâce aux probabilités conditionnelles et à la formule des probabilités totales. Il vous convaincra de l'importance de bien comprendre ce concept pour interpréter correctement les résultats d'études statistiques. Il provient d'une étude médicale sur le succès de deux traitements contre les calculs rénaux. Le traitement A a été effectué dans les années 1972-1980, et le traitement B dans les années 1980-1985.

La première table montre le succès global et le nombre de traitements pour chaque méthode.

Succès (taux de succès)

Traitement A	Traitement B
273/350 (78%)	289/350 (83%)

Cela semble révéler que traitement B, qui est nouveau, est plus efficace. Maintenant, en ajoutant des données concernant la taille des calculs, la comparaison prend une autre tournure :

1. Charig CR, ; Webb DR, ; Payne SR, ; Wickham OE . Comparison of treatment of renal calculi by operative surgery, percutaneous nephrolithotomy, and extracorporeal shock wave lithotripsy. BMJ 1986 ;292 : 879–82. 3

Résultats en fonction de la taille des calculs

petits calculs		grands calculs	
Traitement A	Traitement B	Traitement A	Traitement B
(81/87) 93%	(234/270) 87%	(192/263) 73%	(55/80) 69%

L'information au sujet de la taille des calculs a inversé les conclusions concernant l'efficacité de chaque traitement. Le traitement A est maintenant considéré comme plus efficace dans les deux cas. Le rebroussement de cette inégalité, qui conduit au paradoxe, se produit à cause de deux effets concurrents :

1. la variable supplémentaire (ici la taille) a un impact significatif sur les rapports ;
2. les tailles des groupes qui sont combinés quand la variable supplémentaire est ignorée sont très différentes. (Les groupes utilisés pour le traitement A et B ont la même taille, mais n'ont pas la même répartition de petits et grands calculs).

Vérifions les calculs pour le traitement A. On choisit comme univers Ω les 350 patients de l'échantillon, que l'on munit de la tribu $\mathcal{P}(\Omega)$. Appelons S l'événement "le traitement est un succès", P l'événement "les calculs sont petits". Alors d'après le premier tableau, $\mathbb{P}(S) = \frac{273}{350}$, et d'après le deuxième, $\mathbb{P}(P) = \frac{87}{350}$, $\mathbb{P}(P^c) = \frac{263}{350}$. Ces deux probabilités étant strictement positives, on peut conditionner par les événements correspondants. D'après le deuxième tableau toujours, on a $\mathbb{P}_P(S) = \frac{81}{87}$, $\mathbb{P}_{P^c}(S) = \frac{192}{263}$. Utilisons la formule des probabilités totales pour calculer $\mathbb{P}(S)$.

$$\begin{aligned}\mathbb{P}(S) &= \mathbb{P}_P(S)\mathbb{P}(P) + \mathbb{P}_{P^c}(S)\mathbb{P}(P^c)\\ &= \frac{81}{87}\frac{87}{350} + \frac{192}{263}\frac{263}{350} = \frac{273}{350}.\end{aligned}$$

On retrouve bien le résultat du premier tableau. Des calculs similaires permettent de vérifier les résultats pour le traitement B. Ainsi, ces résultats apparemment contradictoires s'expliquent aisément grâce aux probabilités conditionnelles.

3.2 Indépendance des événements

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé.

Définition. Deux événements A et B de $\mathcal{A}$ sont dits *indépendants*, si

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

Exemple 3.4.

1. Les événements $\emptyset$ et Ω sont indépendants. En effet,

$$\mathbb{P}(\emptyset \cap \Omega) = \mathbb{P}(\emptyset) = 0, \text{ et } \mathbb{P}(\emptyset)\mathbb{P}(\Omega) = 0.1 = 0,$$

   d'où, $\mathbb{P}(\emptyset \cap \Omega) = \mathbb{P}(\emptyset)\,\mathbb{P}(\Omega)$.

2. On jette un dé parfaitement équilibré. Soit A l'événement "obtenir 1,2 ou 3", et B l'événement "obtenir 1,2,4 ou 5". On choisit comme espace des états, $\Omega = \{1, \cdots, 6\}$, que l'on munit de la probabilité uniforme. On a alors,

$$\mathbb{P}(A) = \frac{1}{2},\ \mathbb{P}(B) = \frac{2}{3},\ \mathbb{P}(A \cap B) = \mathbb{P}(\{1,2\}) = \frac{1}{3}.$$

Ainsi, comme $\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)$, on déduit que les événements A et B sont indépendants.

Remarque.
- L'indépendance n'a rien à voir avec le fait que les événements soient disjoints ou non. Dans l' Exemple 2 ci-dessus, les événements sont indépendants, mais non disjoints ($A \cap B \neq \emptyset$).
- Si A et B sont deux événements de probabilité non nulle, alors :

$$A \text{ et } B \text{ sont indépendants} \Leftrightarrow \mathbb{P}(A) = \mathbb{P}_B(A) \Leftrightarrow \mathbb{P}(B) = \mathbb{P}_A(B).$$

Le fait d'avoir une information supplémentaire, à savoir que B est réalisé, ne modifie pas la probabilité de A (de même pour B lorsqu'on sait que A est réalisé) ce qui justifie la terminologie d'*indépendance*. Ces critères ne sont cependant pas utilisés comme définition car ils nécessitent l'hypothèse supplémentaire, $\mathbb{P}(A) > 0$ et $\mathbb{P}(B) > 0$.

Proposition 3.3. *Si les événements A et B sont indépendants, il en est de même des événements A^c et B, A et B^c, A^c et B^c.*

Démonstration. Démontrons que A^c et B sont indépendants si A et B le sont.

$$\begin{aligned}
\mathbb{P}(A^c \cap B) &= \mathbb{P}(B) - \mathbb{P}(A \cap B), \text{ d'après la formule des probabilités totales} \\
&= \mathbb{P}(B)\,(1 - \mathbb{P}(A)), \text{ car } A \text{ et } B \text{ sont indépendants} \\
&= \mathbb{P}(B)\,\mathbb{P}(A^c), \text{ donc } A^c \text{ et } B \text{ sont indépendants.}
\end{aligned}$$

En remplaçant A et B par B et A^c, l'implication que l'on vient de prouver entraîne que si A^c et B sont indépendants (et donc B et A^c indépendants), alors B^c et A^c sont indépendants. Et ainsi de suite pour les autres cas. □

Définition.
- Des événements $A_1, \cdots, A_n$ de $\mathcal{A}$ sont dits *mutuellement indépendants*, si pour tout entier $1 \leq k \leq n$ et tout k-uplet d'entiers $(i_1, \cdots, i_k)$, tels que, $1 \leq i_1 < \cdots < i_k \leq n$,

$$\mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \cdots \mathbb{P}(A_{i_k}).$$

- Les événements $A_1, \cdots, A_n$ sont dits *deux-à-deux indépendants*, si pour tout $i \in \{1, \cdots, n\}$, $j \in \{1, \cdots, n\}$, tels que $i \neq j$, on a :

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\,\mathbb{P}(A_j).$$

3.2.1 Nombre infini de jets de dés

Dans cette section, nous abordons de manière approfondie un exemple classique qui utilise la notion d'indépendance : la description mathématique d'une expérience aléatoire répétée, dans les mêmes conditions, et de manière indépendante, un nombre fini ou infini de fois. Afin de fixer les idées, nous prendrons comme exemple des jets indépendants d'un dé non pipé, mais vous pourriez aussi imaginer prendre les jets d'une pièce de monnaie, etc.

Un jet de dé

On choisit comme espace des états, $\Omega_1 = \{1, \cdots, 6\}$ et comme tribu $\mathcal{A}_1 = \mathcal{P}(\Omega_1)$. Étant donné que le dé est non-pipé, on munit Ω_1 de la probabilité uniforme, notée $\mathbb{P}_1$, de sorte que :

$$\forall\, i \in \{1, \cdots, 6\},\ \mathbb{P}_1(\{i\}) = \frac{1}{6}.$$

Deux jets de dés indépendants

On choisit comme espace des états,

$$\Omega_2 = \{(i,j) \,|\, i \in \{1, \cdots, 6\},\, j \in \{1, \cdots, 6\}\} = \{1 \cdots, 6\}^2.$$

Pour les mêmes raisons que dans le cas d'un dé, on choisit comme tribu $\mathcal{A}_2 = \mathcal{P}(\Omega_2)$ et l'on munit Ω_2 de la probabilité uniforme, notée $\mathbb{P}_2$, de sorte que :

$$\forall\, (i,j) \in \Omega_2,\ \mathbb{P}_2(\{(i,j)\}) = \frac{1}{36}.$$

Il y a une autre manière de voir ceci qui est la suivante : nous souhaitons modéliser l'indépendance des expériences successives, par conséquent deux événements portant l'un sur la première expérience, l'autre sur la deuxième, doivent être indépendants.

Pour $k \in \{1, \cdots, 6\}$, $\ell \in \{1, 2\}$, on considère l'"événement E_k^ℓ "le ℓ-ième jet donne k". Lorsque $\ell = 1$, cet événement est le sous-ensemble $A_k^1(\Omega_2) = \{k\} \times \Omega_1$ de Ω_2, et lorsque $\ell = 2$, c'est le sous-ensemble $A_k^2(\Omega_2) = \Omega_1 \times \{k\}$ de Ω_2. Remarquons de plus que l'événement élémentaire $\{(i,j)\}$ de Ω_2 s'écrit :

$$\{(i,j)\} = A_i^1(\Omega_2) \cap A_j^2(\Omega_2).$$

Ainsi, on souhaite que :

$$\mathbb{P}_2(\{(i,j)\}) = \mathbb{P}_2[A_i^1(\Omega_2)] \cap A_j^2[(\Omega_2)] = \mathbb{P}_2[A_i^1(\Omega_2)]\mathbb{P}_2[A_j^2(\Omega_2)].$$

D'autre part, comme l'événement E_i^1 ne dépend que du premier jet et E_j^2 que du deuxième, on peut les écrire comme des sous-ensemble de Ω_1 (représentant la première et la deuxième expérience respectivement) : E_i^1 est le sous-ensemble $A_i^1(\Omega_1) = \{i\}$ de Ω_1 et E_j^2 est le sous-ensemble $A_j^2(\Omega_1) = \{j\}$ de Ω_1. Donc, il est naturel de demander que :

$$\begin{aligned}
\mathbb{P}_2(\{(i,j)\}) &= \mathbb{P}_2[A_i^1(\Omega_2) \cap A_j^2(\Omega_2)] = \mathbb{P}_2[A_i^1(\Omega_2)]\mathbb{P}_2[A_j^2(\Omega_2)] \\
&= \mathbb{P}_1[A_i^1(\Omega_1)]\,\mathbb{P}_1[A_j^2(\Omega_1)] = \mathbb{P}_1(\{i\})\,\mathbb{P}_1(\{j\}) \\
&= \frac{1}{6^2} = \frac{1}{36}.
\end{aligned}$$

n jets de dés indépendants

Ceci se généralise naturellement au cas de n jets.

$$\Omega_n = \{1, \cdots, 6\}^n,\ \mathcal{A}_n = \mathcal{P}(\Omega_n),$$

Pour tout $(i_1, \cdots, i_n) \in \Omega_n$, $\mathbb{P}_n[\{(i_1, \cdots, i_n)\}] = \frac{1}{6^n}$.

Répétition infinie de jets de dés indépendants

Ceci sort du programme à proprement parlé, cependant c'est un exemple qui apparaît souvent sous une forme déguisée. On choisit comme espace des états $\Omega_\infty = \{1, \cdots, 6\}^{\mathbb{N}^*}$. Attention Ω_∞ n'est pas dénombrable, donc on ne choisit pas comme tribu $\mathcal{P}(\Omega_\infty)$.

La tribu $\mathcal{A}_\infty$ sur Ω_∞ est la tribu engendrée par les événements ne dépendant que d'un nombre fini de jets. Soit $n \geq 1$, et pour tout $k \in \{1, \cdots, n\}$, soit $i_k \in \{1, \cdots, 6\}$. On définit l'"événement" $E_{i_1,\cdots,i_n}$ "l'issue du premier tirage est $i_1,\cdots$, l'issue du n-ième tirage est i_n". Alors $E_{i_1,\cdots,i_n}$ est le sous-ensemble $A_{i_1,\cdots,i_n}(\Omega_\infty) = \{i_1\} \times \cdots \times \{i_n\} \times \Omega_1 \times \Omega_1 \cdots$ de Ω_∞, et le sous-ensemble $A_{i_1,\cdots,i_n}(\Omega_n) = \{(i_1, \cdots, i_n)\}$ de Ω_n.

En utilisant le théorème d'extension de Carthéodory, on montre qu'il existe une unique probabilité $\mathbb{P}_\infty$ sur $(\Omega_\infty, \mathcal{A})$, telle que :

$$\begin{aligned}\mathbb{P}_\infty[A_{i_1,\cdots,i_n}(\Omega_\infty)] &= \mathbb{P}_n[A_{i_1,\cdots,i_n}(\Omega_n)] = \mathbb{P}_n(\{(i_1, \cdots, i_n)\}) \\ &= \mathbb{P}_1(\{i_1\}) \cdots \mathbb{P}_1(\{i_n\}) = \frac{1}{6^n}.\end{aligned}$$

Cette probabilité s'appelle la *probabilité produit* sur Ω_∞. Vous ne devez pas connaître les détails de la construction, cependant l'idée à retenir est la suivante : si l'univers est Ω_∞, la probabilité choisie sur Ω_∞ permet de calculer la probabilité de tous les événements qui ne dépendent que d'un nombre fini de jets et cette probabilité est donnée par le produit des probabilités pour chacun des jets.

Exercice 3.1. On lance un dé à 6 faces non-truqué, indéfiniment.

1. Décrire l'univers Ω_∞ associé à cette expérience aléatoire.
2. Pour $n \in \mathbb{N}^*$, soit A_n l'événement "on obtient 1 pour la première fois au n-ième jet". Calculer la probabilité de l'événement A_n.
3. Pour $n \in \mathbb{N}^*$, soit B_n l'événement "on obtient 1 aux n premiers jets, et soit B l'événement "on obtient toujours 1". Calculer la probabilité de B_n et B.

Solution de l'exercice 3.1.

1. L'univers Ω_∞ choisi est
$$\Omega_\infty = \{1, \cdots, 6\}^{\mathbb{N}^*} = \{(i_n)_{n \geq 1} \,:\, \forall\, n \geq 1,\, i_n \in \{1, \cdots, 6\}\}.$$
2. Pour tout $n \geq 1$, l'événement A_n est le sous-ensemble
$$\{i_1 \geq 2\} \times \cdots \times \{i_{n-1} \geq 2\} \times \{i_n = 1\},$$
de Ω_n, ainsi
$$\mathbb{P}(A_n) = \mathbb{P}(\{i_1 \geq 2\}) \cdots \mathbb{P}(\{i_{n-1} \geq 2\})\mathbb{P}(\{i_n = 1\}) = \Big(\frac{5}{6}\Big)^{n-1} \cdot \frac{1}{6}.$$
Ceci est un exemple de *loi géométrique* que nous reverrons plus tard. Étant donné que la mesure produit n'est pas au programme, on ne l'écrit pas $\mathbb{P}_\infty$ et on passe sous silence sa construction (contrairement à ce qui se fait dans le cas où Ω est fini ou dénombrable !)
3. Pour tout $n \geq 1$, l'événement B_n est l'élément
$$\{1\} \times \cdots \times \{1\} = (1, \cdots, 1)$$

de Ω_n, ainsi,

$$\mathbb{P}(B_n) = \mathbb{P}(\{1\}) \cdots \mathbb{P}(\{1\}) = \frac{1}{6^n}.$$

L'événement B est l'élément $\{(i_n)_{n\geq 1} \,:\, \forall n \geq 1,\, i_n = 1\}$ de Ω_∞. Remarquons que B s'écrit aussi $B = \bigcap\limits_{n\geq 1} B_n$. Or les $(B_n)_{n\geq 1}$ forment une suite décroissante d'événements, donc d'après la Proposition 1.4 :

$$\mathbb{P}(B) = \mathbb{P}(\bigcap_{n\geq 1} B_n) = \lim_{n\to+\infty} \mathbb{P}(B_n) = \lim_{n\to+\infty} \frac{1}{6^n} = 0.$$

Exercices

Exercice 3.2. On choisit une famille “au hasard" parmi toutes les familles ayant deux enfants.

1. Sachant que la famille choisie a au moins un garçon, quelle est la probabilité qu'elle ait deux garçons ?
2. Sachant que l'aîné de la famille choisie est un garçon, quelle est la probabilité que le plus jeune soit aussi un garçon ?

Solution de l'exercice 3.2. On choisit comme univers, $\Omega = \{(G,G),(G,F),(F,G),(F,F)\}$, où la première coordonnée (resp. la deuxième) donne le sexe de l'aîné (du cadet) des enfants. On munit Ω de la tribu $\mathcal{A} = \mathcal{P}(\Omega)$. Comme la famille est choisie au “hasard”, on munit Ω de la probabilité uniforme, notée $\mathbb{P}$, de sorte que pour tout événement A de $\mathcal{A}$, $\mathbb{P}(A) = \frac{\text{card}(A)}{\text{card}(\Omega)} = \frac{\text{card}(A)}{4}$.

Soit A l'événement “la famille a au moins un garçon”, B “l'aîné de la famille est un garçon, C “les deux enfants sont des garçons”. Alors, $A \cap C = C$, $B \cap C = C$, et

$$\begin{aligned}
\mathbb{P}(A) &= \mathbb{P}[\{(G,G),(G,F),(F,G)\}] = \frac{3}{4} > 0,\\
\mathbb{P}(B) &= \mathbb{P}[\{(G,G),(G,F)\}] = \frac{1}{2} > 0,\\
\mathbb{P}(C) &= \mathbb{P}[\{(G,G)\}] = \frac{1}{4}.
\end{aligned}$$

En particulier les probabilités conditionnées par rapport à A ou B sont bien définies. La probabilité recherchée au Point 1. est :

$$\mathbb{P}_A(C) = \frac{\mathbb{P}(A \cap C)}{\mathbb{P}(A)} = \frac{\mathbb{P}(C)}{\mathbb{P}(A)} = \frac{4}{12} = \frac{1}{3}.$$

La probabilité recherchée au Point 2. est :

$$\mathbb{P}_B(C) = \frac{\mathbb{P}(B \cap C)}{\mathbb{P}(B)} = \frac{\mathbb{P}(C)}{\mathbb{P}(B)} = \frac{2}{4} = \frac{1}{2}.$$

Exercice 3.3. *Un exemple d'urne de Polya.* Une urne contient au départ 5 boules blanches et 7 noires. Chaque fois que l'on tire une boule, on la réintroduit en rajoutant deux nouvelles boules de la même couleur que celle tirée. Quelle est la probabilité que les deux premières boules tirées soient noires ? Que la deuxième boule tirée soit noire ?

Remarque : les urnes de Polya peuvent servir pour modéliser la propagation de maladies infectieuses. En effet, chaque réalisation d'un événement augmente la probabilité des réalisations suivantes.

Solution de l'exercice 3.3. On choisit comme univers $\Omega = \{(B,B),(B,N),(N,B),(N,N)\}$, où la première (resp. deuxième) coordonnée représente l'issue du premier (resp. deuxième) tirage. On munit Ω de la tribu $\mathcal{A} = \mathcal{P}(\Omega)$. On note $\mathbb{P}$ la probabilité sur $(\Omega, \mathcal{A})$ correspondant à cette expérience. Soit $k \in \{1,2\}$, on note B_k (resp. N_k) l'événement “la boule tirée au k-ième tirage est blanche (resp. noire)”. Alors, la donnée de l'exercice nous dit que :

$$\mathbb{P}(B_1) = \frac{5}{12}, \quad \mathbb{P}(N_1) = \frac{7}{12}.$$

Comme ces deux probabilités sont strictement positives, on peut conditionner par rapport aux événements B_1 et N_1. D'après l'énoncé, nous savons que :

$$\mathbb{P}_{B_1}(B_2) = \frac{7}{14}, \quad \mathbb{P}_{B_1}(N_2) = \frac{7}{14},$$
$$\mathbb{P}_{N_1}(B_2) = \frac{5}{14}, \quad \mathbb{P}_{N_1}(N_2) = \frac{9}{14}.$$

La réponse à la première question est $\mathbb{P}(N_1 \cap N_2)$ et à la deuxième, $\mathbb{P}(N_2)$. Calculons ces probabilités. D'après la définition des probabilités conditionnelles, nous savons que :

$$\mathbb{P}(N_1 \cap N_2) = \mathbb{P}_{N_1}(N_2)\mathbb{P}(N_1) = \frac{9.7}{14.12} = \frac{3}{8}.$$

De manière analogue, nous pouvons calculer :

$$\mathbb{P}(B_1 \cap N_2) = \mathbb{P}_{B_1}(N_2)\mathbb{P}(B_1) = \frac{7.5}{14.12} = \frac{5}{24}.$$

Comme $\{B_1, N_1\}$ forme une partition de Ω, nous déduisons de la formule des probabilités totales :

$$\mathbb{P}(N_2) = \mathbb{P}(N_1 \cap N_2) + \mathbb{P}(B_1 \cap N_2) = \frac{9}{24} + \frac{5}{24} = \frac{7}{12}.$$

Exercice 3.4. On considère trois cartes à jouer de même forme. Les deux faces de la première carte ont été colorées en noir, les deux faces de la deuxième en rouge tandis que la troisième porte une face noire et une face rouge. On mélange les trois cartes au fond d'un chapeau puis une carte est tirée au hasard et placée sur la table. Si la face apparente est rouge, quelle est la probabilité que l'autre soit noire ?

Solution de l'exercice 3.4. Chaque carte possède deux faces, que l'on va distinguer. Choisissons l'univers $\Omega = \{R_1, R_2, N_1, N_2, R_3, N_3\}$, où par exemple l'événement élémentaire R_1 est réalisé si c'est la première face de la carte unicolore rouge qui est apparente. On munit Ω de la tribu $\mathcal{A} = \mathcal{P}(\Omega)$. Étant donné que la carte est tirée "au hasard", on munit Ω de la probabilité uniforme, notée $\mathbb{P}$, de sorte que pour tout événement A de $\mathcal{A}$, $\mathbb{P}(A) = \frac{\mathrm{card}(A)}{\mathrm{card}(\Omega)} = \frac{\mathrm{card}(A)}{6}$.
On souhaite calculer la probabilité conditionnelle de l'événement R_3 sachant que l'événement R_1 ou R_2 ou R_3 est réalisé. Soit $A = R_1 \cup R_2 \cup R_3$, alors $\mathbb{P}(A) = \frac{1}{2}$, et la probabilité conditionnelle $\mathbb{P}_A(R_3)$ est bien définie. D'après la définition de la probabilité conditionnelle, nous avons :

$$\mathbb{P}_A(R_3) = \frac{\mathbb{P}(A \cap R_3)}{\mathbb{P}(A)} = \frac{\mathbb{P}(R_3)}{\mathbb{P}(A)} = \frac{1/6}{1/2} = \frac{1}{3}.$$

Sans faire attention, on pourrait penser que cette probabilité est de $1/2$, pensant qu'à partir du moment où le côté rouge apparaît, il reste deux situations équiprobables.

Exercice 3.5. Le test de dépistage d'un certain virus n'est pas infaillible :

— 1 fois sur 100, il est positif, alors que l'individu n'est pas contaminé,

— 2 fois sur 100, il est négatif alors que l'individu est contaminé.

D'autre part, on sait que sur la population totale, la fraction de porteurs est approximativement de $1/1000$.

1. Étant donné que le test est positif, quelle est la probabilité que l'individu ne soit pas porteur du virus ?
2. Étant donné que son test est négatif, quelle est la probabilité qu'un individu soit porteur du virus ?

Solution de l'exercice 3.5. On choisit comme espace des états l'ensemble des individus de la population. On appelle V l'événement "l'individu est contaminé", et T l'événement "le test est positif". D'après la donnée de l'exercice nous connaissons les probabilités suivantes :

$$\mathbb{P}(V) = \frac{1}{1000}, \text{ d'où } \mathbb{P}(V^c) = 1 - \frac{1}{1000} = \frac{999}{1000}.$$

Ces deux probabilités étant strictement positives, on peut conditionner par les événements V et V^c. D'après l'énoncé, nous savons que :

$$\mathbb{P}_{V^c}(T) = \frac{1}{100}, \quad \mathbb{P}_V(T^c) = \frac{2}{100}.$$

On souhaite calculer $\mathbb{P}_T(V^c)$ et $\mathbb{P}_{T^c}(V)$. Comme $0 < \mathbb{P}(V) < 1$, nous pouvons utiliser la formule de Bayes, de sorte que :

$$\mathbb{P}_T(V^c) = \frac{\mathbb{P}_{V^c}(T)\mathbb{P}(V^c)}{\mathbb{P}_{V^c}(T)\mathbb{P}(V^c) + \mathbb{P}_V(T)\mathbb{P}(V)} \sim 0,91.$$

Un calcul similaire montre que :

$$\mathbb{P}_{T^c}(V) = \frac{\mathbb{P}_V(T^c)\mathbb{P}(V)}{\mathbb{P}_{V^c}(T^c)\mathbb{P}(V^c) + \mathbb{P}_V(T^c)\mathbb{P}(V)} \sim 0,00002.$$

Exercice 3.6. Un dé à six faces n'est pas bien équilibré, un échantillonnage a permis d'obtenir le tableau suivant.

Score	1	2	3	4	5	6	Total
Fréquence	0,1	0,2	0,1	0,4	0,1	0,1	1

On cherche à savoir si, avec ce dé, il est plus probable de faire un score d'au moins 5 lorsque le score est pair ou lorsque le score est impair.

1. Déterminer un choix d'espace probabilisé adapté à cette expérience.
2. Calculer la probabilité conditionnelle que le score soit d'au moins 5 sachant que le score est pair. Calculer la probabilité conditionnelle que le score soit d'au moins 5 sachant que le score est impair. Conclure.
3. Calculer la probabilité conditionnelle que le score soit impair sachant qu'il est d'au moins 5. Calculer la probabilité conditionnelle que le score soit pair sachant qu'il est d'au moins 5. Interpréter.

Solution de l'exercice 3.6.

1. On choisit comme univers $\Omega = \{1,2,3,4,5,6\}$, que l'on munit de la tribu $\mathcal{A} = \mathcal{P}(\Omega)$. On associe la probabilité $\mathbb{P}$ naturellement définie par le tableau, de sorte que :

$$\mathbb{P}[\{1\}] = 0,1;\ \mathbb{P}[\{2\}] = 0,2;\ \mathbb{P}[\{3\}] = 0,1;\ \mathbb{P}[\{4\}] = 0,4;\ \mathbb{P}[\{5\}] = 0,1; = \mathbb{P}[\{6\}] = 0,1.$$

   Soit A l'événement "le score est pair", B l'événement "le score est impair", C "le score est d'au moins 5". Alors,

$$\begin{aligned}
&\mathbb{P}[A] = \mathbb{P}[\{2,4,6\}] = \mathbb{P}[\{2\}] + \mathbb{P}[\{4\}] + \mathbb{P}[\{6\}] = 0,2 + 0,4 + 0,1 = 0,7\\
&\mathbb{P}[B] = \mathbb{P}[A^c] = 1 - \mathbb{P}[A] = 0,3\\
&\mathbb{P}[C] = \mathbb{P}[\{5,6\}] = 0,1 + 0,1 = 0,2.
\end{aligned}$$

   Ces trois probabilités étant strictement positives, il est possible de conditionner par les événements A, B ou C.
2. On cherche $\mathbb{P}_A[C]$. D'après la formule des probabilités conditionnelles :

$$\mathbb{P}_A[C] = \frac{\mathbb{P}[A \cap C]}{\mathbb{P}[A]} = \frac{\mathbb{P}[\{6\}]}{\mathbb{P}[A]} = \frac{0,1}{0,7} = \frac{1}{7}.$$

   On cherche ensuite $\mathbb{P}_B[C]$. D'après la formule des probabilités conditionnelles :

$$\mathbb{P}_B[C] = \frac{\mathbb{P}[B \cap C]}{\mathbb{P}[B]} = \frac{\mathbb{P}[\{5\}]}{\mathbb{P}[B]} = \frac{0,1}{0,3} = \frac{1}{3}.$$

   Ainsi il est plus probable de faire un score d'au moins 5 si le score est impair.
3. On cherche $\mathbb{P}_C[A]$. D'après la formule de Bayes, on a :

$$\mathbb{P}_C[A] = \frac{\mathbb{P}_A[C]\mathbb{P}[A]}{\mathbb{P}[C]} = \frac{\frac{1}{7}\frac{7}{10}}{\frac{2}{10}} = \frac{1}{2}.$$

   On cherche ensuite $\mathbb{P}_C[B] = \mathbb{P}_C[A^c]$. Comme une probabilité conditionnelle est une probabilité, on a en particulier :

$$\mathbb{P}_C[B] = \mathbb{P}_C[A^c] = 1 - \mathbb{P}_C[A] = \frac{1}{2}.$$

   S'il on sait que le score est d'au moins 5, alors il est aussi probable d'obtenir un score pair qu'un score impair.

Exercice 3.7. Soit $\Omega = \{1,2,3,4\}$ muni de la tribu $\mathcal{A} = \mathcal{P}(\Omega)$ et de la probabilité uniforme, notée $\mathbb{P}$. On considère les événements $A = \{1,2\}$, $B = \{2,3\}$, $C = \{1,3\}$.
— Montrer que les événements A, B, C, sont deux-à-deux indépendants,
— Montrer que les événements A, B, C, ne sont pas mutuellement indépendants.
Ainsi l'indépendance deux-à-deux est plus faible que l'indépendance mutuelle.

Solution de l'exercice 3.7. Comme $(\Omega, \mathcal{A})$ est munit de la probabilité uniforme, on a pour tout événement A de $\mathcal{A}$, $\mathbb{P}(A) = \frac{\mathrm{card}(A)}{\mathrm{card}(\Omega)} = \frac{\mathrm{card}(A)}{4}$. Ainsi,

$$\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = \frac{2}{4} = \frac{1}{2}$$

D'autre part, $A \cap B = \{2\}$, $B \cap C = \{3\}$, $A \cap C = \{1\}$, $A \cap B \cap C = \{\emptyset\}$. D'où :

$$\mathbb{P}(A \cap B) = \mathbb{P}(B \cap C) = \mathbb{P}(A \cap C) = \frac{1}{4}, \quad \text{et } \mathbb{P}(A \cap B \cap C) = 0.$$

— D'après les calculs ci-dessus, on a :

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) = \frac{1}{4},\ \mathbb{P}(B \cap C) = \mathbb{P}(B)\mathbb{P}(C) = \frac{1}{4},\ \mathbb{P}(A \cap C) = \mathbb{P}(A)\mathbb{P}(C) = \frac{1}{4},$$

d'où les événements A, B, C sont deux-à-deux indépendants.

— D'après les calculs ci-dessus, $\mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) = \frac{1}{8}$ et $\mathbb{P}(A \cap B \cap C) = 0$, d'où les événements A, B, C ne sont pas mutuellement indépendants.

Exercice 3.8. Une urne contient 9 boules indiscernables, numérotée de 1 à 9. On tire une boule "au hasard". Les événements suivants sont-ils indépendants ?

1. A : "la boule tirée porte un numéro pair",
2. B : "le numéro tiré est multiple de 3".

Répondre à la même question lorsque l'urne contient 12 boules.

Solution de l'exercice 3.8. Soit k le nombre de boules contenues dans l'urne, avec $k = 9$ ou $k = 12$. On choisit comme espace des états $\Omega_k = \{1, 2, \cdots, k\}$, et la tribu $\mathcal{A} = \mathcal{P}(\Omega_k)$. Comme la boule est tirée "au hasard", on munit Ω_k de la probabilité uniforme, notée $\mathbb{P}_k$, de sorte que pour tout événement A de $\mathcal{A}$,

$$\mathbb{P}_k(A) = \frac{\operatorname{card}(A)}{\operatorname{card}(\Omega_k)} = \frac{\operatorname{card}(A)}{k}.$$

Dans le cas où $k = 9$, les événements $A, B, A \cap B$ s'écrivent :

$$A = \{2, 4, 6, 8\},\ B = \{3, 6, 9\},\ A \cap B = \{6\}.$$

Ainsi, $\mathbb{P}_9(A) = \frac{4}{9}$, $\mathbb{P}_9(B) = \frac{1}{3}$, $\mathbb{P}_9(A \cap B) = \frac{1}{9}$. Comme $\mathbb{P}_9(A \cap B) \neq \mathbb{P}_9(A)\,\mathbb{P}_9(B)$, on en déduit que les événements A et B ne sont pas indépendants.

Dans le cas où $k = 12$, les événements $A, B, A \cap B$ s'écrivent :

$$A = \{2, 4, 6, 8, 10, 12\},\ B = \{3, 6, 9, 12\},\ A \cap B = \{6, 12\},$$

Ainsi, $\mathbb{P}_{12}(A) = \frac{1}{2}$, $\mathbb{P}_{12}(B) = \frac{1}{3}$, $\mathbb{P}_{12}(A \cap B) = \frac{1}{6}$. Comme $\mathbb{P}_{12}(A \cap B) = \mathbb{P}_{12}(A)\,\mathbb{P}_{12}(B)$, on en déduit que les événements A et B sont indépendants.

Exercice 3.9. Soit $\Omega = \{1, 2, 3, 4\}$ muni de la tribu $\mathcal{A} = \mathcal{P}(\Omega)$ et de la probabilité uniforme, notée $\mathbb{P}$. On considère les événements $A = \{1, 2\}$, $B = \{2, 3\}$, $C = \{1, 3\}$.

— Montrer que les événements A, B, C, sont deux-à-deux indépendants,

— Montrer que les événements A, B, C, ne sont pas mutuellement indépendants.

Ainsi l'indépendance deux-à-deux est plus faible que l'indépendance mutuelle.

Solution de l'exercice 3.9. Comme $(\Omega, \mathcal{A})$ est munit de la probabilité uniforme, on a pour tout événement A de $\mathcal{A}$, $\mathbb{P}(A) = \frac{\operatorname{card}(A)}{\operatorname{card}(\Omega)} = \frac{\operatorname{card}(A)}{4}$. Ainsi,

$$\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = \frac{2}{4} = \frac{1}{2}$$

D'autre part, $A \cap B = \{2\}$, $B \cap C = \{3\}$, $A \cap C = \{1\}$, $A \cap B \cap C = \{\emptyset\}$. D'où :

$$\mathbb{P}(A \cap B) = \mathbb{P}(B \cap C) = \mathbb{P}(A \cap C) = \frac{1}{4}, \quad \text{et } \mathbb{P}(A \cap B \cap C) = 0.$$

- D'après les calculs ci-dessus, on a :

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) = \frac{1}{4},\ \mathbb{P}(B \cap C) = \mathbb{P}(B)\mathbb{P}(C) = \frac{1}{4},\ \mathbb{P}(A \cap C) = \mathbb{P}(A)\mathbb{P}(C) = \frac{1}{4},$$

  d'où les événements A, B, C sont deux-à-deux indépendants.
- D'après les calculs ci-dessus, $\mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) = \frac{1}{8}$ et $\mathbb{P}(A \cap B \cap C) = 0$, d'où les événements A, B, C ne sont pas mutuellement indépendants.

Chapitre 4

Variables aléatoires

4.1 Définition et loi d'une variable aléatoire

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Plutôt que de travailler avec des événements de $\mathcal{A}$, il est souvent plus commode d'associer une valeur numérique aux résultats d'une expérience aléatoire. Par exemple, lors de n jets de pile ou face, il sera intéressant d'étudier le nombre de piles obtenus. Cela motive l'introduction de la notion de *variable aléatoire*, qui est une application X de Ω dans un ensemble E qui sera typiquement, $\mathbb{N}^d$, $\mathbb{Z}^d$ ou $\mathbb{R}^d$ $(d \geq 1)$.

Lorsque X ne prend qu'un nombre dénombrable de valeurs $X(\Omega) = \{x_j \,:\, j \in J\}$, où J est une partie non-vide finie ou dénombrable de $\mathbb{N}$, alors X est appelée une variable aléatoire *discrète*.

Remarque.

- La terminologie de variable aléatoire peut être trompeuse, car il s'agit en fait d'une fonction de Ω dans E.
- Afin de pouvoir définir une probabilité de manière cohérente, il faut supposer de plus que pour une classe importante de parties B de E, l'ensemble $\{\omega \in \Omega \,:\, X(\omega) \in B\}$ appartient à $\mathcal{A}$. Formellement, on munit E d'une tribu $\mathcal{E}$. Une *variable aléatoire* X est alors une application de $(\Omega, \mathcal{A})$ dans $(E, \mathcal{E})$ telle que :

  $$\forall B \in \mathcal{E},\ \{\omega \in \Omega \,:\, X(\omega) \in B\} \in \mathcal{A}.$$

  Une variable aléatoire X est en fait une application *mesurable* de $(\Omega, \mathcal{A})$ dans $(E, \mathcal{E})$.
- Soit B un sous-ensemble de E. Rappelez-vous les notations :

  $$X^{-1}(B) = \{\omega \in \Omega \,:\, X(\omega) \in B\} = \{X \in B\}.$$

  $\{X \in B\}$ est l'événement "X prend une valeur appartenant à B". Attention, $X^{-1}(B)$ désigne l'image réciproque de la partie B par l'application X. Cela ne sous-entend nullement que X est bijective !

Définition. Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et X une variable aléatoire définie sur Ω, à valeurs dans E. On définit la *loi de probabilité de* X, notée $\mathbb{P}^X$ de la manière suivante. Pour tout sous-ensemble B de E tel que $\{X \in B\} \in \mathcal{A}$:

$$\mathbb{P}^X(B) = \mathbb{P}(\{X \in B\}).$$

Remarque. Reprenons le cadre plus formel, où l'on munit E d'une tribu $\mathcal{E}$. La loi de probabilité de X est alors une probabilité sur $(E, \mathcal{E})$. C'est la transposition de la structure abstraite $(\Omega, \mathcal{A}, \mathbb{P})$ sur $(E, \mathcal{E})$. L'espace E muni de la tribu $\mathcal{E}$ et de la probabilité $\mathbb{P}^X$ devient un espace probabilisé $(E, \mathcal{E}, \mathbb{P}^X)$.

Proposition 4.1.

— *Dans le cas où X est une variable aléatoire discrète, $X(\Omega) = \{x_j : j \in J\}$, où J est une partie non-vide, finie ou dénombrable de $\mathbb{N}$, alors la loi de probabilité de X est caractérisée par la donnée :*
$$\forall j \in J,\ \mathbb{P}^X(\{x_j\}) = \mathbb{P}(\{X = x_j\}).$$

— *Dans le cas où $X(\Omega) = \mathbb{R}$, la loi de probabilité de X est caractérisée par la donnée de :*
$$\forall (a,b) \in \mathbb{R}^2,\ a < b,\ \mathbb{P}^X([a,b[) = \mathbb{P}(\{X \in [a,b[\}).$$

Exemple 4.1.

1. On considère deux lancers successifs d'un même dé, et on note S la variable aléatoire correspondant à la somme des valeurs obtenues, ainsi S prend ses valeurs dans $\{2, 3, \cdots, 12\}$. Calculons la loi de S. Choisissons comme univers $\Omega = \{1, \cdots, 6\}^2$, que l'on munit de la tribu $\mathcal{A} = \mathcal{P}(\Omega)$, et de la probabilité uniforme, notée $\mathbb{P}$. Ainsi pour tout sous-ensemble A de $\mathcal{P}(\Omega)$, $\mathbb{P}(A) = \frac{\mathrm{card}(A)}{\mathrm{card}(\Omega)}$. Étant donné que S prend ses valeurs dans $\{2, \cdots, 12\}$, pour déterminer la loi de probabilité $\mathbb{P}^S$ de S, il suffit de calculer, pour tout $i \in \{2, \cdots, 12\}$, $\mathbb{P}^S(\{i\})$:
$$\begin{aligned}\mathbb{P}^S(\{2\}) &= \mathbb{P}(\{S = 2\})\\ &= \mathbb{P}(\{\omega \in \Omega : S(\omega) = 2\}) = \mathbb{P}(\{(1,1)\}) = \frac{1}{36},\\ \mathbb{P}^S(\{3\}) &= \mathbb{P}(\{\omega \in \Omega : S(\omega) = 3\}) = \mathbb{P}(\{(1,2),(2,1)\}) = \frac{1}{18},\\ &\text{etc.}\end{aligned}$$

2. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, et $A \in \mathcal{A}$ un événement. Alors l'*indicatrice de A*, notée $\mathbb{I}_A$, est la variable aléatoire définie sur Ω par :
$$\forall \omega \in \Omega,\ \mathbb{I}_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A,\\ 0 & \text{sinon.}\end{cases}$$
Ainsi l'indicatrice de A prend ses valeurs dans $\{0, 1\}$. En probabilité, il est important de bien savoir manipuler cette notation. En particulier l'indicatrice satisfait aux conditions suivantes. Si A et B sont deux événements de $\mathcal{A}$, alors :
$$\mathbb{I}_{A^c} = 1 - \mathbb{I}_A, \quad \mathbb{I}_{A \cap B} = \mathbb{I}_A \mathbb{I}_B, \quad \mathbb{I}_{A \cup B} = \mathbb{I}_A + \mathbb{I}_B - \mathbb{I}_{A \cap B}.$$
Calculons la loi de l'indicatrice de A. Étant donné qu'elle prend ses valeurs dans $\{0, 1\}$, il suffit de déterminer, pour $i \in \{0, 1\}$, $\mathbb{P}^{\mathbb{I}_A}(\{i\})$:
$$\begin{aligned}\mathbb{P}^{\mathbb{I}_A}(\{1\}) &= \mathbb{P}(\mathbb{I}_A \in \{1\}) = \mathbb{P}(\{\omega \in \Omega : \mathbb{I}_A(\omega) = 1\})\\ &= \mathbb{P}(\{\omega \in \Omega : \omega \in A\}) = \mathbb{P}(A),\\ \mathbb{P}^{\mathbb{I}_A}(\{0\}) &= \mathbb{P}(\mathbb{I}_A \in \{0\}) = \mathbb{P}(\{\omega \in \Omega : \mathbb{I}_A(\omega) = 0\})\\ &= \mathbb{P}(\{\omega \in \Omega : \omega \notin A\}) = \mathbb{P}(A^c) = 1 - \mathbb{P}(A).\end{aligned}$$

4.2 Fonction de répartition

Définition. Soit X une variable aléatoire de loi $\mathbb{P}^X$ à valeurs dans $\mathbb{R}$ (ou une partie de $\mathbb{R}$). On appelle *fonction de répartition de la loi* $\mathbb{P}^X$ ou encore, par abus, *fonction de répartition de* X, l'application F_X définie sur $\mathbb{R}$ par :

$$\forall t \in \mathbb{R},\ F_X(t) = \mathbb{P}^X(]-\infty, t]) = \mathbb{P}(X \leq t).$$

Propriété 4.2. *La fonction de répartition de la loi* $\mathbb{P}^X$ *satisfait aux propriétés suivantes :*

1. F_X *prend ses valeurs dans* $[0, 1]$*,*
2. F_X *est une application croissante,*
3. F_X *est continue à droite et admet une limite à gauche,*
4. $\lim\limits_{t\to-\infty} F_X(t) = 0,\ \lim\limits_{t\to+\infty} F_X(t) = 1.$

La proposition suivante nous dit qu'une fonction de répartition caractérise la loi de la variable aléatoire.

Proposition 4.3. *Toute application définie de* $\mathbb{R}$ *dans* $[0, 1]$ *qui possède les propriétés* $2, 3, 4$, *est la fonction de répartition d'une unique loi de probabilité sur* $\mathbb{R}$.

La fonction de répartition permet de calculer les probabilités suivantes :

Propriété 4.4. *Soit* $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé, et* X *une variable aléatoire sur* Ω *de fonction de répartition* F_X*, alors :*
- $\mathbb{P}(X > x) = 1 - F_X(x)$,
- $\mathbb{P}(x < X \leq y) = F_X(y) - F_X(x)$,
- $\mathbb{P}(X < x) = F_X(x^-)$.

Pour ce cours, qui a pour but d'introduire les chaînes de Markov, nous n'avons besoin que de variables et vecteurs aléatoires discrets, nous nous restreignons donc à ces cas.

4.3 Variables aléatoires discrètes

4.3.1 Définitions et exemples classiques

Voici une liste de lois discrètes classiques.

<u>Loi uniforme</u>. Soit $n \in \mathbb{N}^*$ et une variable aléatoire, $X : \Omega \to \{x_1, \cdots, x_n\}$, où pour tout $i \neq j$, $x_i \neq x_j$. Supposons que la loi de probabilité de X est donnée par :

$$\forall j \in \{1, \cdots, n\},\ \mathbb{P}^X(\{x_j\}) = \frac{1}{n}.$$

Alors la loi de X est appelée la *loi uniforme discrète* sur $\{x_1, \cdots, x_n\}$.

Exemple 4.2. On lance une pièce équilibrée. Soit $\Omega = \{P, F\}$ que l'on munit de la probabilité uniforme $\mathbb{P}$. On définit la variable aléatoire $X : \Omega \to \{-1, 1\}$, par $X(\{P\}) = 1$, $X(\{F\}) = -1$. Ainsi, la loi de probabilité de X est :

$$\mathbb{P}^X(\{1\}) = \mathbb{P}(\{X = 1\}) = \mathbb{P}(\{P\}) = \frac{1}{2},\ \mathbb{P}^X(\{-1\}) = \mathbb{P}(\{X = -1\}) = \mathbb{P}(\{F\}) = \frac{1}{2},$$

et la variable aléatoire X suite une loi uniforme sur $\{-1, 1\}$.

<u>Loi de Bernoulli</u>. Soit $0 < p < 1$ et une variable aléatoire $X : \Omega \to \{0, 1\}$ de loi de probabilité :

$$\mathbb{P}^X(\{1\}) = p,\ \mathbb{P}^X(\{0\}) = 1 - p.$$

Alors la loi de X est appelée *loi de Bernoulli de paramètre* p.

Exemple 4.3. Une *épreuve de Bernoulli de paramètre* p est une expérience aléatoire admettant deux issues succès/échec, telle que p est la probabilité d'obtenir un succès. Soit Ω représentant les issues de l'expérience, $\mathbb{P}$ une probabilité sur Ω et A l'événement représentant le succès. D'après la description de l'expérience, on a $\mathbb{P}(A) = p$. Dans ce cas, l'indicatrice de A, $\mathbb{I}_A$, suit une loi de Bernoulli de paramètre p. En effet, $\mathbb{I}_A$ est à valeurs dans $\{0, 1\}$ et nous avons déjà vu que :

$$\mathbb{P}^{\mathbb{I}_A}(\{1\}) = \mathbb{P}(A), \quad \mathbb{P}^{\mathbb{I}_A}(\{0\}) = 1 - \mathbb{P}(A).$$

On conclut en utilisant le fait que $\mathbb{P}(A) = p$.

Par exemple, on jette un dé équilibré une fois. On appelle "succès" l'événement A "obtenir un nombre plus grand ou égal à 2". On choisit comme univers $\Omega = \{1, \cdots, 6\}$, alors $\mathbb{P}(A) = \mathbb{P}(\{2, 3, 4, 5, 6\}) = \frac{5}{6}$ et $\mathbb{I}_A$ suit une loi de Bernoulli de paramètre $5/6$.

<u>Loi binomiale</u>. Soit $n \in \mathbb{N}^*$, $0 < p < 1$ et une variable aléatoire $X : \Omega \to \{0, \cdots, n\}$ de loi de probabilité :

$$\forall k \in \{0, \cdots, n\},\ \mathbb{P}^X(\{k\}) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

Alors la loi de X est appelée *loi binomiale de paramètres* n *et* p.

Exemple 4.4. On appelle *schéma de Bernoulli de paramètres* n *et* p l'expérience qui consiste à répéter n fois de manière indépendante une épreuve de Bernoulli de paramètre p. On choisit comme univers Ω^n, que l'on munit d'une probabilité $\mathbb{P}_n$. Pour $i \in \{1, \cdots, n\}$, on note A_i l'événement "obtenir un succès lors de la i-ième expérience". Soit $k \in \{0, \cdots, n\}$, comme les expériences sont indépendantes, la probabilité d'obtenir un succès lors des k premières expériences et un échec lors des $n - k$ dernières est :

$$\begin{aligned} \mathbb{P}_n(A_1 \cap \cdots \cap A_k \cap A_{k+1}^c \cdots \cap A_n^c) &= \mathbb{P}(A_1) \cdots \mathbb{P}(A_k) \mathbb{P}(A_{k+1}^c) \cdots \mathbb{P}(A_n^c) \\ &= p^k (1 - p)^{n-k}. \end{aligned}$$

Soit X la variable aléatoire qui compte le nombre de succès dans un schéma de Bernoulli, alors X suit une loi binomiale de paramètres n et p. En effet,

$$\mathbb{P}^X(\{k\}) = \mathbb{P}_n(X = k) = \mathbb{P}_n(\{\text{obtenir } k \text{ succès sur les } n \text{ expériences}\}).$$

Il y a $\binom{n}{k}$ façons d'obtenir ces succès, et pour chacune des façons la probabilité est $p^k(1-p)^{n-k}$. Ainsi,

$$\mathbb{P}^X(\{k\}) = \binom{n}{k} p^k(1-p)^{n-k}.$$

Par exemple, si on jette n fois une pièce de monnaie équilibrée, et on appelle X la variable aléatoire qui compte le nombre de fois où l'on obtient un nombre plus grand ou égal à 2, alors X suit une loi binomiale de paramètres n et $5/6$.

Loi géométrique. Soit $0 < p < 1$ et une variable aléatoire $X : \Omega \to \mathbb{N}^*$ de loi de probabilité :

$$\forall k \in \mathbb{N}^*,\ \mathbb{P}^X(\{k\}) = (1-p)^{k-1}p.$$

Alors la loi de X est appelée *loi géométrique de paramètre* p. En translatant de 1 les valeurs de la variable aléatoire, on obtient la loi géométrique sur $\mathbb{N}$.

Exemple 4.5. Considérons un schéma de Bernoulli où l'expérience est répétée indéfiniment. Si X est la variable aléatoire égale au temps passé jusqu'au premier succès, alors X suit une loi géométrique de paramètre p. En effet, choisissons comme univers $\Omega^{\mathbb{N}^*}$ et calculons la loi de probabilité de X. Pour tout $k \in \mathbb{N}^*$,

$$\begin{aligned}
\mathbb{P}^X(\{k\}) &= \mathbb{P}_k(A_1^c \cap \cdots \cap A_{k-1}^c \cap A_k)\\
&= \mathbb{P}(A_1^c)\cdots\mathbb{P}(A_{k-1}^c)\mathbb{P}(A_k), \quad \text{par indépendance}\\
&= (1-p)^{k-1}p.
\end{aligned}$$

Loi de Poisson. Soit $\theta > 0$ et une variable aléatoire $X : \Omega \to \mathbb{N}$ de loi :

$$\forall k \in \mathbb{N},\ \mathbb{P}^X(\{k\}) = e^{-\theta}\,\frac{\theta^k}{k!}.$$

Alors la loi de X est appelée *loi de Poisson de paramètre* θ.

Exemple 4.6. La loi de Poisson modélise le nombre d'autobus passés à un arrêt avant un instant T donné, et θ représente le nombre moyen d'arrivées dans cet intervalle.

Exercices

Exercice 4.1. Pour chacune des lois ci-dessus, montrer que la somme des probabilités des événements élémentaires vaut 1.

Exercice 4.2. Soit $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5\}$ un espace fini à 5 éléments, de probabilités respectives 1/4, 1/4, 1/6, 1/6, 1/6. On note X la variable aléatoire définie par :

$$X(\omega_1) = X(\omega_2) = 0, \quad X(\omega_3) = X(\omega_4) = 1, \quad X(\omega_5) = 2.$$

Déterminer la loi de la variable aléatoire X.

Solution de l'exercice 4.2. La variable aléatoire X est à valeurs dans $\{0, 1, 2\}$. Sa loi de probabilité $\mathbb{P}^X$ est caractérisée par :

$$\begin{aligned}
\mathbb{P}^X[\{0\}] &= \mathbb{P}[\{X = 0\}] = \mathbb{P}[\{\omega_1, \omega_2\}] = \mathbb{P}[\{\omega_1\}] + \mathbb{P}[\{\omega_2\}] = \frac{1}{4} + \frac{1}{4} = \frac{1}{2} \\
\mathbb{P}^X[\{1\}] &= \mathbb{P}[\{X = 1\}] = \mathbb{P}[\{\omega_3, \omega_4\}] = \frac{1}{6} + \frac{1}{6} = \frac{1}{3} \\
\mathbb{P}^X[\{2\}] &= \mathbb{P}[\{X = 2\}] = \mathbb{P}[\{\omega_5\}] = \frac{1}{6}.
\end{aligned}$$

Exercice 4.3. Une urne contient n boules numérotées de 1 à n, $n \geq 3$. On tire 3 boules d'un seul coup. Soit X la variable aléatoire égale à 1 si on a tiré la boule no 1, égale à 0 dans le cas contraire. Donner la loi de la variable aléatoire X.

Solution de l'exercice 4.3. On choisit comme univers Ω, l'ensemble des tirages possibles. Il s'agit d'un tirage non ordonné, sans remise de 3 boules parmi n, ainsi $\mathrm{card}(\Omega) = \binom{n}{3}$. Étant donné que Ω est de cardinal fini, on le munit de la tribu $\mathcal{A}$ des parties de Ω. Comme le tirage s'effectue "au hasard", on munit $(\Omega, \mathcal{A})$ de la probabilité uniforme, notée $\mathbb{P}$.

Soit A l'événement "on a tiré la boule numéro 1". Il y a 1 manière de choisir la boule numéro 1 dans un tirage non ordonné, puis $\binom{n-1}{2}$ manières de choisir les 2 autres boules. Ainsi, $\mathrm{card}(A) = \binom{n-1}{2}$, et

$$\mathbb{P}[A] = \frac{\mathrm{card}(A)}{\mathrm{card}(\Omega)} = \frac{\binom{n-1}{2}}{\binom{n}{3}} = \frac{(n-1)(n-2)3.2}{n(n-1)(n-2)2} = \frac{3}{n}.$$

Remarquons que l'événement $\{X = 1\} = A$. Ainsi, la loi de la variable aléatoire X est donnée par :

$$\mathbb{P}^X[\{1\}] = \mathbb{P}[\{X = 1\}] = \mathbb{P}[A] = \frac{3}{n}, \quad \mathbb{P}^X[\{0\}] = 1 - \mathbb{P}^X[\{1\}] = \frac{n-3}{n}.$$

Remarque. Si l'énoncé disait que les boules étaient tirées successivement, il aurait été naturel de choisir le tirage comme étant ordonné. Dans ce cas, $\mathrm{card}(\Omega) = n(n-1)(n-2)$. On aurait muni Ω de la probabilité uniforme et on aurait considéré le même événement A. Alors, $\mathrm{card}(A) = 3(n-1)(n-2)$, de sorte que $\mathbb{P}[A] = \frac{3}{n}$.

Exercice 4.4. On lance deux dés équilibrés. On note X le plus grand des numéros obtenus. Déterminer la loi de la variable aléatoire X.

Solution de l'exercice 4.4. On choisit comme espace des états $\Omega = \{1, \cdots, 6\}^2$. Comme Ω est discret, on le munit de la tribu $\mathcal{A}$ des parties de Ω. Étant donné que les dés sont équilibrés, on munit Ω de la probabilité uniforme, notée $\mathbb{P}$, de sorte que pour tout $A \in \mathcal{A}$, $\mathbb{P}[A] = \frac{\mathrm{card}(A)}{\mathrm{card}(\Omega)} = \frac{\mathrm{card}(A)}{36}$.

La variable aléatoire X prend ses valeurs dans $\{1, \cdots, 6\}$. La loi de X est donnée par :

$$\begin{aligned}
&\mathbb{P}^X[\{1\}] = \mathbb{P}[\{X = 1\}] = \mathbb{P}[\{(1,1)\}] = \frac{1}{36},\\
&\mathbb{P}^X[\{2\}] = \mathbb{P}[\{X = 2\}] = \mathbb{P}[\{(1,2); (2,1); (2,2)\}] = \frac{1}{12},\\
&\mathbb{P}^X[\{3\}] = \mathbb{P}[\{X = 3\}] = \mathbb{P}[\{(1,3); (3,1); (2,3); (3,2); (3,3)\}] = \frac{5}{36},\\
&\mathbb{P}^X[\{4\}] = \mathbb{P}[\{X = 4\}] = \mathbb{P}[\{(1,4); (4,1); (2,4); (4,2); (3,4); (4,3); (4,4)\}] = \frac{7}{36},\\
&\mathbb{P}^X[\{5\}] = \frac{9}{36},\\
&\mathbb{P}^X[\{6\}] = \frac{11}{36}.
\end{aligned}$$

Exercice 4.5. Une urne contient N_b boules blanches et N_n boules noires. Posons $N = N_b + N_n$. On tire r boules avec remise dans l'urne. Soit X la variable aléatoire égale au nombre de boules blanches tirées. Déterminer la loi de la variable aléatoire X. Reconnaître la loi de X.

Solution de l'exercice 4.5. On choisit comme univers Ω l'ensemble des tirages possibles. Il s'agit d'un tirage ordonné avec remise de r boules parmi N, ainsi $\text{card}(\Omega) = N^r$. Étant donné que Ω est de cardinal fini, on le munit de la tribu $\mathcal{A} = \mathcal{P}(\Omega)$. Comme le tirage s'effectue au hasard, on munit $(\Omega, \mathcal{A})$ de la probabilité uniforme, notée $\mathbb{P}$, de sorte que pour tout $A \in \mathcal{A}$, $\mathbb{P}(A) = \frac{\text{card}(A)}{\text{card}(\Omega)}$.
Pour $k \in \{0, \dots, r\}$, on définit l'évènement A_k "on a tiré exactement k boules blanches. D'après l'exemple 4.4, nous savons que $\text{card}(A_k) = \binom{r}{k} N_b^k N_n^{r-k}$, et que :

$$\mathbb{P}[A_k] = \frac{\binom{r}{k} N_b^k N_n^{r-k}}{N^r} = \binom{r}{k} \left(\frac{N_b}{N}\right)^k \left(\frac{N_n}{N}\right)^{r-k}.$$

Remarquons que la variable aléatoire X est à valeurs dans $\{0, \dots, r\}$, et que pour tout $k \in \{0, \dots, r\}$, l'évènement $\{X = k\} = A_k$. Ainsi, la loi de la variable aléatoire X est donnée, pour tout $k \in \{0, \cdots, r\}$, par :

$$\mathbb{P}^X[\{k\}] = \mathbb{P}[\{X = k\}] = \mathbb{P}[A_k] = \binom{r}{k} \left(\frac{N_b}{N}\right)^k \left(\frac{N_n}{N}\right)^{r-k}.$$

La variable aléatoire X suite une loi binomiale de paramètres r et $\frac{N_b}{N}$.

Exercice 4.6. On lance un dé à 6 faces non truqué, indéfiniment. Soit X la variable aléatoire égale au temps passé jusqu'à ce que le premier 1 soit obtenu. Déterminer la loi de la variable aléatoire X. Reconnaître la loi de X.

Solution de l'exercice 4.6. On choisit comme univers $\Omega = \{1, \dots, 6\}^{\mathbb{N}^*}$. Souvenez-vous que l'on ne munit pas Ω de la tribu $\mathcal{P}(\Omega)$ et que, exceptionnellement, on passe sous silence le choix de la tribu. Soit $n \geq 1$ et soit $(i_1, \dots, i_n) \in \{1, \dots, 6\}^n$, notons $B_{i_1, \cdots, i_n}$ l'évènement "l'issue du premier tirage est $i_1, \dots,$ l'issue du n-ième tirage est i_n". Le dé étant non truqué et les jets étant indépendants, on munit Ω de la probabilité $\mathbb{P}$, telle que :

$$\mathbb{P}[B_{i_1, \cdots, i_n}] = \mathbb{P}[B_{i_1}] \dots \mathbb{P}[B_{i_n}] = \frac{1}{6^n}.$$

Pour $n \geq 1$, définissons l'évènement A_n "on obtient pile pour la première fois au n-ième jet". D'après l'Exercice 3.1, nous savons que :

$$\mathbb{P}[A_n] = \frac{1}{6}\Big(\frac{5}{6}\Big)^{n-1}.$$

Remarquons que la variable aléatoire X est à valeurs dans $\mathbb{N}^*$, et que pour tout $n \geq 1$, l'évènement $\{X = n\} = A_n$. Ainsi, pour tout $n \geq 1$:

$$\mathbb{P}^X[\{n\}] = \mathbb{P}[\{X = n\}] = \mathbb{P}[A_n] = \frac{1}{6}\Big(\frac{5}{6}\Big)^{n-1}.$$

On reconnaît que la variable aléatoire X suit une loi géométrique de paramètre $\frac{1}{6}$.

4.3.2 Fonction de répartition

Dans le cas des variables aléatoires discrètes, la fonction de répartition vérifie les propriétés suivantes.

Propriété 4.5. *Soit* $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé et soit* X *une variable aléatoire discrète réelle définie sur* Ω. *Alors la fonction de répartition* F_X *de la loi* $\mathbb{P}^X$, *vérifie :*

1. $F_X(x) = \sum_{\{y \in X(\Omega)\,:\,y \leq x\}} \mathbb{P}^X(\{y\})$;
2. *si* x *et* y *sont deux points consécutifs de* $X(\Omega)$, *alors* F_X *est constante sur* $[x, y[$;
3. *la hauteur du saut en* $x \in X(\Omega)$ *est donnée par* $\mathbb{P}^X(\{x\})$.

Exercice 4.7. Calculer la fonction de répartition de :

1. la loi uniforme sur $\{1, \cdots, n\}$,
2. la loi de Bernoulli de paramètre p,
3. la loi binomiale de paramètres 4 et $\frac{1}{2}$,
4. la loi géométrique de paramètre p.

Solution de l'exercice 4.7.

1. Si X suit une loi uniforme sur $\{1, \cdots, n\}$, alors pour tout $k \in \{1, \cdots, n\}$, $\mathbb{P}^X(\{k\}) = \frac{1}{n}$. Ainsi,

$$F_X(y) = \begin{cases} 0 & \text{si } y < 1 \\ \frac{k}{n} & \text{si } y \in [k, k+1[,\ k \in \{1, \cdots, n-1\} \\ 1 & \text{si } y \geq n. \end{cases}$$

2. Si X suit une loi de Bernoulli de paramètre p, alors $\mathbb{P}^X(\{1\}) = p$, et $\mathbb{P}^X(\{0\}) = 1 - p$. Ainsi,

$$F_X(y) = \begin{cases} 0 & \text{si } y < 0 \\ 1 - p & \text{si } y \in [0, 1[\\ 1 & \text{si } y \geq 1. \end{cases}$$

3. Si X suit une loi binomiale de paramètres 4 et $\frac{1}{2}$, alors pour tout $k \in \{0, \cdots, 4\}$,

$$\mathbb{P}^X(\{k\}) = \binom{4}{k}\left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{4-k} = \binom{4}{k}\frac{1}{16}.$$

D'où,

$$\mathbb{P}^X(\{0\}) = \frac{1}{16},\ \mathbb{P}^X(\{1\}) = \frac{4}{16},\ \mathbb{P}^X(\{2\}) = \frac{6}{16},\ \mathbb{P}^X(\{3\}) = \frac{4}{16},\ \mathbb{P}^X(\{4\}) = \frac{1}{16}.$$

Ainsi,

$$F_X(y) = \begin{cases} 0 & \text{si } y < 0 \\ \frac{1}{16} & \text{si } y \in [0, 1[\\ \frac{5}{16} & \text{si } y \in [1, 2[\\ \frac{11}{16} & \text{si } y \in [2, 3[\\ \frac{15}{16} & \text{si } y \in [3, 4[\\ 1 & \text{si } y \geq 4. \end{cases}$$

4. Si X suit une loi géométrique de paramètre p, alors pour tout $k \in \mathbb{N}^*$, $\mathbb{P}^X(\{k\}) = (1-p)^{k-1}p$. Ainsi,

$$F_X(y) = \begin{cases} 0 & \text{si } y < 1 \\ \sum\limits_{i=1}^{k}(1-p)^{i-1}p = 1-(1-p)^k & \text{si } y \in [k, k+1[,\ k \in \mathbb{N}^* \end{cases}$$

4.3.3 Espérance

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et X une variable aléatoire discrète définie sur Ω.

L'espérance de la variable aléatoire X représente sa moyenne pondérée par la probabilité de chacune des issues. Lors d'un jeu de hasard par exemple, elle permet de déterminer si le jeu est équitable ou non.

Définition.

— La variable aléatoire discrète X est dite *intégrable*, si la série de terme général $|x_j|\, p_j$ converge.

— Si X est une variable aléatoire discrète intégrable, on définit son *espérance*, notée $\mathbb{E}(X)$, par :

$$\mathbb{E}(X) = \sum_{j \in J} x_j\, \mathbb{P}^X(\{x_j\}).$$

Remarque.

— Lorsque l'univers Ω est fini, toute variable aléatoire définie sur Ω est intégrable.

— Si X n'est pas intégrable, elle n'admet pas d'espérance.

L'espérance satisfait aux propriétés suivantes.

Propriété 4.6.

1. *Une variable aléatoire discrète X est intégrable si et seulement si $|X|$ est intégrable.*

2. *Si la variable aléatoire X est bornée, alors elle est intégrable.*
3. (Linéarité). *Si X et Y sont deux variables aléatoires discrètes intégrables, et si a, b sont deux constantes réelles, alors $aX + bY$ est intégrable et on a :*

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y).$$

4. (Monotonie). *Si X et Y sont deux variables aléatoires discrètes intégrables telles que $X \leq Y$, alors $\mathbb{E}(X) \leq \mathbb{E}(Y)$. En particulier si X est intégrable, alors $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$.*
5. *Si X est une variable aléatoire constante, $X \equiv a$, alors X est intégrable et $\mathbb{E}(X) = a$.*
6. (Théorème de transfert). *Si X est une variable aléatoire discrète et si f est une application définie sur $X(\Omega)$ telle que la série de terme général $|f(x_j)|\mathbb{P}^X(\{x_j\})$ converge, alors $f(X)$ est une variable aléatoire discrète intégrable et*

$$\mathbb{E}(f(X)) = \sum_{j \in J} f(x_j)\mathbb{P}^X(\{x_j\}).$$

Définition. Soit X une variable aléatoire discrète intégrable. Si X est d'espérance nulle, on dit que X est *centrée.*

Exercices

Exercice 4.8. Soit X une variable aléatoire discrète intégrable. Montrer que la variable aléatoire $X - \mathbb{E}(X)$ est centrée.

Solution de l'exercice 4.8. Si X est intégrable, alors par la propriété de linéarité de l'espérance, $X - \mathbb{E}(X)$ est intégrable et,

$$\mathbb{E}(X - \mathbb{E}(X)) = \mathbb{E}(X) - \mathbb{E}(X) = 0,$$

d'où on déduit que la variable aléatoire $X - \mathbb{E}(X)$ est centrée.

Exercice 4.9. Calculer, si elle existe, l'espérance des variables aléatoires ayant pour loi :

1. loi uniforme sur $\{1, \cdots, n\}$,
2. loi de Bernoulli de paramètre p. Soit $A \in \mathcal{A}$ un événement de Ω, déduire que

$$\mathbb{E}(\mathbb{I}_A) = \mathbb{P}(A).$$

3. loi binomiale de paramètres n et p,
4. loi géométrique de paramètre p, $0 < p < 1$,
5. loi de Poisson de paramètre $\theta > 0$.

Solution de l'exercice 4.9.

1. Soit X une variable aléatoire suivant une loi uniforme sur $\{1, \cdots, n\}$. Comme X ne prend qu'un nombre fini de valeurs, alors X est intégrable et admet donc une espérance. Par définition :
$$\mathbb{E}(X) = \sum_{k=1}^{n} k\frac{1}{n} = \frac{n(n+1)}{2n} = \frac{n+1}{2}.$$

2. Soit X une variable aléatoire suivant une loi de Bernoulli de paramètre p. Comme X ne prend qu'un nombre fini de valeurs, alors X est intégrable et admet donc une espérance. Par définition :
$$\mathbb{E}(X) = 1.p + 0.(1-p) = p.$$
Soit A une événement de Ω, alors l'indicatrice de A, $\mathbb{I}_A$, suit une loi de Bernoulli de paramètre $\mathbb{P}(A)$, ainsi $\mathbb{I}_A$ admet une espérance et
$$\mathbb{E}(\mathbb{I}_A) = \mathbb{P}(A).$$

3. Soit X une variable aléatoire suivant un loi binomiale de paramètres n et p. Comme X ne prend qu'un nombre fini de valeurs, alors X est intégrable et admet donc une espérance. Nous allons calculer l'espérance de X de deux manières. Par définition de l'espérance,
$$\begin{aligned}\mathbb{E}(X) &= \sum_{k=0}^{n} k\binom{n}{k}p^k(1-p)^{n-k} = \sum_{k=1}^{n} \frac{k.n!}{k!(n-k)!}p^k(1-p)^{n-k}\\ &= \sum_{k=1}^{n} \frac{(n-1)!n}{(k-1)!(n-1-(k-1))!}p^k(1-p)^{n-k}\\ &= \sum_{j=0}^{n-1} n\binom{n-1}{j}p^{j+1}(1-p)^{n-j-1}, \quad \text{(avec le changement d'indice } j=k-1)\\ &= np\sum_{j=0}^{n-1}\binom{n-1}{j}p^j(1-p)^{n-1-j}\\ &= np(p+1-p)^{n-1} = np, \quad \text{(en utilisant la formule du binôme de Newton).}\end{aligned}$$
D'autre part, la variable aléatoire X a même loi que $\sum\limits_{k=1}^{n} X_k$, où les $(X_k)_{1\leq k\leq n}$ sont des variables aléatoires indépendantes de Bernoulli de paramètre p. Ainsi, en utilisant la linéarité de l'espérance et le Point 2, nous obtenons :
$$\mathbb{E}(X) = \mathbb{E}\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n}\mathbb{E}(X_k) = \sum_{k=1}^{n} p = np.$$

4. Soit X une variable aléatoire suivant une loi géométrique de paramètre p. Pour montrer que X est intégrable, nous devons prouver que la série de terme général $|k|\,\mathbb{P}^X(\{k\}) = kp(1-p)^{k-1}$ est convergente. Soit $x \in \mathbb{R}$, considérons la série de terme général x^k. On reconnaît la série géométrique, qui est absolument convergente pour tout x tel que $|x| < 1$, et vaut alors, $\phi(x) = \frac{1}{1-x}$. Par un théorème du cours, la série est infiniment dérivable sur l'intérieur

de son intervalle de convergence et les dérivées successives sont données par les dérivées successives terme à terme. En particulier, pour tout x tel que $|x| < 1$,

$$\phi'(x) = \frac{1}{(1-x)^2} = \sum_{k=1}^{+\infty} k\, x^{k-1}.$$

Comme $0 < p < 1$, on en déduit que la série de terme général $p[k(1-p)^{k-1}]$ est absolument convergente et qu'elle vaut $p\frac{1}{p^2}$. Ainsi, si X suit une loi géométrique de paramètre p, X admet une espérance et

$$\mathbb{E}(X) = \frac{1}{p}.$$

5. Soit X une variable aléatoire suivant une loi de Poisson de paramètre θ. Afin de montrer que X est intégrable, nous devons prouver que la série de terme général $|k|\,\mathbb{P}^X(\{k\}) = k\,e^{-\theta}\frac{\theta^k}{k!}$ est convergente. Or pour tout $k \geq 1$, $k\,e^{-\theta}\frac{\theta^k}{k!} = e^{-\theta}\theta\frac{\theta^{k-1}}{(k-1)!}$. On reconnaît, à la constante $e^{-\theta}\theta$ près, le développement en série de e^{θ}. Ainsi, si X suit une loi de Poisson de paramètre θ, X admet une espérance, et

$$\mathbb{E}(X) = \theta e^{-\theta} e^{\theta} = \theta.$$

4.3.4 Variance, moments d'ordres supérieurs

Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et X une variable aléatoire discrète définie sur Ω.

Définition. La variable aléatoire X et dite de *carré intégrable*, si X^2 admet une espérance. La quantité $\mathbb{E}(X^2)$ est alors bien définie et est appelée *moment d'ordre 2* de X.

Remarque.

1. Grâce au théorème de transfert, la variable aléatoire X est de carré intégrable si et seulement si la série de terme général $x_j^2\,\mathbb{P}^X(\{x_j\})$ converge.
2. Si la variable aléatoire X est de carré intégrable, alors elle est intégrable.
   En effet, supposons que la variable aléatoire X prenne les valeurs $\{x_i : i \in \mathbb{N}\}$. Fixons $n \in \mathbb{N}^*$, et notons I_n l'ensemble des entiers compris entre 0 et n. Alors,
$$\begin{aligned}\sum_{i\in I_n} |x_i|\,\mathbb{P}^X(\{x_i\}) &= \sum_{\{i\in I_n : |x_i|\leq 1\}} |x_i|\,\mathbb{P}^X(\{x_i\}) + \sum_{\{i\in I_n : |x_i|>1\}} |x_i|\,\mathbb{P}^X(\{x_i\}) \\ &\leq \sum_{\{i\in I_n : |x_i|\leq 1\}} 1\cdot\mathbb{P}^X(\{x_i\}) + \sum_{\{i\in I_n : |x_i|>1\}} x_i^2\,\mathbb{P}^X(\{x_i\}) \\ &\leq 1 + \sum_{i\in I_n} x_i^2\,\mathbb{P}^X(\{x_i\}).\end{aligned}$$
   On conclut en utilisant le critère de comparaison des séries à termes positifs.
3. Soit $\lambda \in \mathbb{R}$. Si la variable aléatoire X est de carré intégrable, il en est de même pour les variables aléatoires $(X+\lambda)$ et $(|X|+\lambda)$.

Propriété 4.7. *Si X est une variable aléatoire discrète de carré intégrable, alors :*

$$\mathbb{E}(|X|)^2 \leq \mathbb{E}(X^2).$$

Démonstration. Soit $\lambda \in \mathbb{R}$. Si la variable aléatoire X est de carré intégrable, alors il en est de même pour $|X| + \lambda$ et nous posons : $f(\lambda) = \mathbb{E}[(|X| + \lambda)^2]$ En utilisant la linéarité de l'espérance,

$$f(\lambda) = \mathbb{E}(X^2) + 2\lambda\mathbb{E}(|X|) + \lambda^2.$$

Ainsi, f est un polynôme de degré 2, positif ou nul. Son discriminant est donc négatif ou nul, ce qui implique le résultat. □

Définition. Si X est une variable aléatoire discrète de carré intégrable, on définit sa *variance*, notée $\mathrm{Var}(X)$, par :

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2].$$

L'*écart-type*, noté $\sigma(X)$, est défini par $\sigma(X) = \sqrt{\mathrm{Var}(X)}$.

Remarque.

— La variance de X mesure la façon dont X s'écarte de sa moyenne $\mathbb{E}(X)$.

— L'écart-type s'utilise surtout en statistique. Remarquer que si la variable aléatoire X a une unité, alors l'écart-type a la même unité, tandis que la variance a l'unité au carré, d'où l'intérêt dans la pratique de travailler avec l'écart-type.

Propriété 4.8. *Soit X une variable aléatoire discrète de carré intégrable, alors :*

1. $\mathrm{Var}(X) \geq 0$,
2. $\forall\, a \in \mathbb{R}$, $\mathrm{Var}(X + a) = \mathrm{Var}(X)$,
3. $\forall\, a \in \mathbb{R}$, $\mathrm{Var}(aX) = a^2 \,\mathrm{Var}(X)$,
4. $\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$,
5. $\mathrm{Var}(X) = 0 \Leftrightarrow \forall\, j \in J$ *tel que* $\mathbb{P}^X(\{x_j\}) > 0,\ x_j = \mathbb{E}[X]$.

Démonstration.

1. Découle de la propriété de monotonie de l'espérance et du fait que $(X - \mathbb{E}(X))^2 \geq 0$.

2. 3. et 4. En utilisant la définition de la variance et la linéarité de l'espérance, on a :

$$\begin{aligned}
\mathrm{Var}(X + a) &= \mathbb{E}[(X + a - \mathbb{E}(X + a))^2] = \mathbb{E}[(X - \mathbb{E}(X))^2] = \mathrm{Var}(X)\\
\mathrm{Var}(aX) &= \mathbb{E}[(aX - \mathbb{E}(aX))^2] = \mathbb{E}[a^2(X - \mathbb{E}(X))^2] = a^2\,\mathrm{Var}(X)\\
\mathrm{Var}(X) &= \mathbb{E}[(X - \mathbb{E}(X))^2] = \mathbb{E}[X^2 - 2\mathbb{E}(X)X + \mathbb{E}(X)^2]\\
&= \mathbb{E}(X^2) - 2\mathbb{E}(X)^2 + \mathbb{E}(X)^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2.
\end{aligned}$$

5. D'après le théorème de transfert,

$$\mathrm{Var}(X) = \sum_{j \in J} (x_j - \mathbb{E}(X))^2\, \mathbb{P}^X(\{x_j\}).$$

C'est une somme de termes positifs. Ainsi, cette somme est nulle si et seulement si chacun de ses termes l'est, c'est-à-dire si et seulement si,

$$\forall\, j \in J,\ \text{tel que } \mathbb{P}^X(\{x_j\}) > 0,\ x_j = \mathbb{E}(X).$$

□

Définition. Soit X une variable aléatoire discrète de carré intégrable. Si X est de variance égale à 1, on dit que X est *réduite*.

Définition. Soit X une variable aléatoire discrète, et soit $n \in \mathbb{N}^*$. Si X^n est intégrable, la quantité $\mathbb{E}(X^n)$ est bien définie et on l'appelle *moment d'ordre* n de la variable aléatoire X.

Remarque. D'après le théorème de transfert, la variable aléatoire X^n est intégrable si et seulement si la série de terme général $|x_j|^n \, \mathbb{P}^X(\{x_j\})$ converge.

EXERCICES

Exercice 4.10. Soit X une variable aléatoire discrète de carré intégrable et de variance non nulle. Montrer que la variable aléatoire $\frac{X}{\sigma(X)}$ est réduite, et que la variable aléatoire $\frac{X-\mathbb{E}(X)}{\sigma(X)}$ est centrée réduite.

Solution de l'exercice 4.10. Utilisant les propriétés de la variance et de l'espérance, nous déduisons que :

$$\begin{aligned}
\mathrm{Var}\left(\frac{X}{\sigma(X)}\right) &= \frac{1}{\sigma(X)^2}\,\mathrm{Var}(X) = 1\\
\mathrm{Var}\left(\frac{X-\mathbb{E}(X)}{\sigma(X)}\right) &= \mathrm{Var}\left(\frac{X}{\sigma(X)}\right) = 1\\
\mathbb{E}\left(\frac{X-\mathbb{E}(X)}{\sigma(X)}\right) &= \frac{1}{\sigma(X)}\mathbb{E}[X-\mathbb{E}(X)] = 0.
\end{aligned}$$

Exercice 4.11. Calculer, si elle existe, la variance de chacune des lois discrètes classiques.

Solution de l'exercice 4.11. Pour les Points 1, 2, 3, comme X ne prend qu'un nombre fini de valeurs, elle est de carré intégrable et admet donc une variance.

1. Calculons d'abord $\mathbb{E}(X^2)$ en utilisant le théorème de transfert.

$$\mathbb{E}(X^2) = \sum_{k=1}^{n} k^2 \frac{1}{n} = \frac{1}{n}\frac{n(n+1)(2n+1)}{6} = \frac{(n+1)(2n+1)}{6},$$

d'où

$$\begin{aligned}
\mathrm{Var}(X) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2\\
&= \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4}\\
&= \frac{(n+1)}{2}\frac{(n-1)}{6} = \frac{n^2-1}{12}.
\end{aligned}$$

2. Dans ce cas, $\mathbb{E}(X^2) = p$, d'où $\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = p(1-p)$.

3. Calculons d'abord $\mathbb{E}[X(X-1)]$ en utilisant le théorème de transfert.

$$\begin{aligned}\mathbb{E}[X(X-1)] &= \sum_{k=2}^{n} k(k-1)\binom{n}{k} p^k(1-p)^{n-k} \\ &= \sum_{k=2}^{n} \frac{n(n-1)(n-2)!}{(k-2)!(n-2-(k-2))!} p^k(1-p)^{n-k} \\ &= n(n-1)\sum_{j=0}^{n-2}\binom{n-2}{j} p^{j+2}(1-p)^{n-j-2}, \text{ (en posant, } j=k-2) \\ &= n(n-1)p^2\sum_{j=0}^{n-2}\binom{n-2}{j} p^{j}(1-p)^{n-2-j} \\ &= n(n-1)p^2(p+1-p)^{n-2} = n(n-1)p^2, \quad \text{(binôme de Newton).}\end{aligned}$$

Ainsi,

$$\begin{aligned}\mathrm{Var}(X) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 \\ &= \mathbb{E}[X(X-1)] + \mathbb{E}(X) - \mathbb{E}(X)^2 \\ &= n(n-1)p^2 + np - n^2p^2 = np(1-p).\end{aligned}$$

4. Soit X suivant une loi géométrique de paramètre p. Afin de montrer que X est de carré intégrable, il suffit de montrer que $X(X-1)$ est intégrable. Ainsi, d'après le théorème de transfert, nous devons prouver que la série de terme général

$$|k(k-1)|\mathbb{P}(\{X=k\}) = k(k-1)p(1-p)^{k-1} = p(1-p)[k(k-1)(1-p)^{k-2}],$$

est convergente. En utilisant ce que nous avons fait pour le calcul de l'espérance, nous savons que pour tout x, $|x|<1$:

$$\phi''(x) = \frac{2}{(1-x)^3} = \sum_{k=1}^{+\infty} k(k-1)x^{k-2}.$$

On en déduit que la série de terme général $p(1-p)[k(k-1)(1-p)^{k-2}]$ est absolument convergente, et vaut $\frac{2p(1-p)}{p^3}$. Ainsi, si X suit une loi géométrique de paramètre p, $X(X-1)$ admet une espérance et

$$\mathbb{E}[X(X-1)] = \frac{2(1-p)}{p^2},$$

d'où :

$$\mathrm{Var}(X) = \mathbb{E}[X(X-1)] + \mathbb{E}(X) - \mathbb{E}(X)^2 = \frac{2(1-p)}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2}.$$

5. Soit X une variable aléatoire suivant une loi de Poisson de paramètre θ. Afin de montrer que X est de carré intégrable, il suffit de montrer que $X(X-1)$ est intégrable. Ainsi, d'après le théorème de transfert, nous devons montrer que la série de terme général,

$$k(k-1)\mathbb{P}(\{X=k\}) = k(k-1)e^{-\theta}\frac{\theta^k}{k!},$$

est convergente. Or, pour tout $k \geq 2$, $k(k-1)e^{-\theta}\frac{\theta^k}{k!} = \theta^2 e^{-\theta}\frac{\theta^{k-2}}{(k-2)!}$. On reconnaît, à la constante $\theta^2 e^{-\theta}$ près, le développement en série de e^{θ}. Ainsi, si X suit une loi de Poisson de paramètre θ, $X(X-1)$ admet une espérance et

$$\mathbb{E}[X(X-1)] = \theta^2 e^{-\theta} e^{\theta} = \theta^2,$$

d'où

$$\mathrm{Var}(X) = \mathbb{E}[X(X-1)] + \mathbb{E}(X) - \mathbb{E}(X)^2 = \theta^2 + \theta - \theta^2 = \theta.$$

4.3.5 Inégalité de Markov et de Bienaymé Tchebychev

Voici deux inégalités classiques.

Proposition 4.9 (Inégalité de Markov). *Soit X une variable aléatoire admettant un moment d'ordre $n \geq 1$. Alors,*

$$\forall\, a > 0, \quad \mathbb{P}\left(|X| \geq a\right) \leq \frac{\mathbb{E}[|X|^n]}{a^n}.$$

Démonstration. Une manière courte d'écrire cette preuve est d'utiliser une variable aléatoire indicatrice. Remarquer que l'on peut écrire la fonction constante égale à 1 sur Ω de la manière suivante : pour tout $\omega \in \Omega$, pour tout réel $a > 0$,

$$1(\omega) = \mathbb{I}_{\{|X| \geq a\}}(\omega) + \mathbb{I}_{\{|X| < a\}}(\omega).$$

Ainsi, pour tout $\omega \in \Omega$, pour tout $a > 0$, on a :

$$\begin{aligned} |X(\omega)|^n &= |X(\omega)|^n \mathbb{I}_{\{|X| \geq a\}}(\omega) + |X(\omega)|^n \mathbb{I}_{\{|X| < a\}}(\omega) \\ &\geq |X(\omega)|^n \mathbb{I}_{\{|X| \geq a\}}(\omega), \quad \text{car } |X(\omega)|^n \mathbb{I}_{\{|X| < a\}}(\omega) \geq 0 \\ &\geq a^n \mathbb{I}_{\{|X| \geq a\}}(\omega), \quad \text{car } a \text{ est positif.} \end{aligned}$$

Autrement dit, on a l'inégalité $|X|^n \geq a^n \mathbb{I}_{\{|X| \geq a\}}$. On conclut en utilisant la monotonie de l'espérance :

$$\mathbb{E}[|X|^n] \geq \mathbb{E}[a^n \mathbb{I}_{\{|X| \geq a\}}] = a^n \mathbb{P}[|X| \geq a].$$

□

Proposition 4.10 (Inégalité de Bienaymé-Tchebychev). *Soit X une variable aléatoire discrète de carré intégrable. Alors,*

$$\forall a > 0, \quad \mathbb{P}[|X - \mathbb{E}(X)| \geq a] \leq \frac{\mathrm{Var}(X)}{a^2}.$$

Démonstration. L'inégalité de Bienaymé-Tchébychev est une conséquence de l'inégalité de Markov. La variable aléatoire X étant de carré intégrable, il en est de même pour la variable aléatoire $X - \mathbb{E}(X)$. On applique l'inégalité de Markov avec la variable aléatoire $Y = X - \mathbb{E}(X)$, et $n = 2$. Pour tout $a > 0$,

$$\begin{aligned} \mathbb{P}\left(|Y| \geq a\right) &\leq \frac{\mathbb{E}[|Y|^2]}{a^2} \\ \Leftrightarrow \mathbb{P}\left(|X - \mathbb{E}(X)| \geq a\right) &\leq \frac{\mathbb{E}[(X - \mathbb{E}(X))^2]}{a^2} = \frac{\mathrm{Var}(X)}{a^2}. \end{aligned}$$

□

4.4 Vecteurs aléatoires discrets

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, et soient X, Y deux variables aléatoires définies sur Ω. Alors la loi de probabilité de X et Y encode toute l'information pour chacune des variables aléatoires, par contre, elle n'encode aucune information sur les propriétés relativement l'une à l'autre. Une façon de résoudre ce problème est de considérer X et Y non pas comme deux variables aléatoires, mais comme les composantes d'un *vecteur aléatoire* (X, Y) prenant ses valeurs dans $\mathbb{R}^2$.

4.4.1 Définition et lois des vecteurs aléatoires

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé.

Définition. Soient X, Y deux variables aléatoires définies sur Ω. Le couple aléatoire $Z = (X, Y)$ est dit *discret* si chacune des variables aléatoires X et Y est discrète. Plus généralement, soient $X_1, \cdots, X_n$ des variables aléatoires définies sur Ω. Le vecteur aléatoire $\mathbf{X} = (X_1, \cdots, X_n)$ est dit *discret*, si chacune des variables aléatoires $X_1, \cdots, X_n$ l'est.

Notations. Pour un couple de variable aléatoire (X, Y), on note :

$$E = X(\Omega),\ F = Y(\Omega).$$

Pour un vecteur aléatoire $\mathbf{X} = (X_1, \cdots, X_n)$, on note :

$$\forall i \in \{1, \cdots, n\},\ E_i = X_i(\Omega).$$

Le vecteur $\mathbf{X} = (X_1, \cdots, X_n)$ est donc un vecteur aléatoire défini sur Ω, à valeurs dans l'ensemble discret $E_1 \times \cdots \times E_n$.

Proposition 4.11 (Définitions)**.**

1. *Loi des vecteurs aléatoires*
   — *Si $Z = (X, Y)$ est un couple aléatoire discret défini sur Ω, alors la loi de probabilité de Z est caractérisée par la donnée des nombres $(\mathbb{P}^Z(\{(x,y)\}))_{(x,y)\in E\times F}$, définis par $\forall (x,y) \in E \times F$:*

   $$\begin{aligned}\mathbb{P}^Z(\{(x,y)\}) &= \mathbb{P}(\{Z = (x,y)\}) = \mathbb{P}(\{X = x\} \cap \{Y = y\}) \\ &= \mathbb{P}(\{\omega \in \Omega \,:\, X(\omega) = x \text{ et } Y(\omega) = y\}).\end{aligned}$$

   — *Si $\mathbf{X} = (X_1, \cdots, X_n)$ est un vecteur aléatoire discret défini sur Ω. Alors la loi de probabilité de $\mathbf{X}$ est caractérisée par la donnée des nombres $(\mathbb{P}^{\mathbf{X}}(\{(x_1, \cdots, x_n)\}))_{(x_1,\cdots,x_n)\in E_1\times\cdots\times E_n}$, définis par :*

   $$\begin{aligned}&\forall (x_1, \cdots, x_n) \in E_1 \times \cdots \times E_n, \\ &\qquad \mathbb{P}^{\mathbf{X}}(\{(x_1, \cdots, x_n)\}) = \mathbb{P}(\{X_1 = x_1\} \cap \cdots \cap \{X_n = x_n\}).\end{aligned}$$

2. *Lois marginales d'un couple aléatoire. Connaissant la loi du couple aléatoire $Z = (X, Y)$, on retrouve la loi des variables aléatoires X et Y, dites* lois marginales *de Z, grâce aux formules suivantes :*

   $$\forall x \in E,\ \mathbb{P}^X(\{x\}) = \mathbb{P}(\{X = x\}) = \sum_{y\in F} \mathbb{P}^Z(\{(x,y)\}) = \sum_{y\in F} \mathbb{P}(\{X = x\} \cap \{Y = y\}),$$

   $$\forall y \in F,\ \mathbb{P}^Y(\{y\}) = \mathbb{P}(\{Y = y\}) = \sum_{x\in E} \mathbb{P}^Z(\{(x,y)\}) = \sum_{x\in E} \mathbb{P}(\{X = x\} \cap \{Y = y\}).$$

3. *Loi conditionnelle. Soit $x \in E$ tel que $\mathbb{P}(\{X = x\}) > 0$. La* loi conditionnelle de Y sachant que X prend la valeur x, *est caractérisée par la donnée des nombres :*

$$\forall y \in F,\ \mathbb{P}_{\{X=x\}}(\{Y = y\}) = \frac{\mathbb{P}(\{Y = y\} \cap \{X = x\})}{\mathbb{P}(\{X = x\})} = \frac{\mathbb{P}^Z(\{(x, y)\})}{\mathbb{P}^X(\{x\})}.$$

Remarque.

— La loi du vecteur aléatoire Z permet de connaître la loi des variables aléatoires X et Y, mais la réciproque est fausse ! La connaissance de chacune des lois X et Y n'entraîne pas la connaissance de la loi du couple. Voir l'exemple ci-dessous.

— Comme conséquence du Point 3. on sait que si $x \in E$ et $y \in F$ sont tels que $\mathbb{P}(\{X = x\}) > 0$ et $\mathbb{P}(\{Y = y\}) > 0$, alors :

$$\mathbb{P}^Z(\{(x, y)\}) = \mathbb{P}_{\{X=x\}}(\{Y = y\})\mathbb{P}(\{X = x\}) = \mathbb{P}_{\{Y=y\}}(\{X = x\})\mathbb{P}(\{Y = y\}).$$

Ainsi, dans ce cas, la connaissance de la loi de Z est équivalente à la connaissance de celle de X et de la loi conditionnelle de Y sachant X, ou encore est équivalente à la connaissance de la loi Y et de la loi conditionnelle de X sachant Y.

Exemple 4.7. Soit $Z = (X, Y)$ un couple aléatoire à valeurs dans

$$\{(-1, -1), (-1, 1), (1, -1), (1, 1)\},$$

respectivement avec les probabilités $\frac{1}{2} - p$, p, p, $\frac{1}{2} - p$, où $0 \leq p \leq \frac{1}{2}$. Calculons les lois marginales du couple aléatoire Z et la loi conditionnelle de Y sachant $\{X = 1\}$. Remarquons que X et Y sont à valeurs dans $\{-1, 1\}$. De la Proposition 4.11, nous déduisons :

$$\begin{aligned}
\mathbb{P}(X = 1) &= \mathbb{P}(Z = (1, 1)) + \mathbb{P}(Z = (1, -1)) = \frac{1}{2} - p + p = \frac{1}{2}\\
\mathbb{P}(X = -1) &= \mathbb{P}(Z = (-1, 1)) + \mathbb{P}(Z = (-1, -1)) = p + \frac{1}{2} - p = \frac{1}{2}\\
\mathbb{P}(Y = 1) &= \mathbb{P}(Z = (1, 1)) + \mathbb{P}(Z = (-1, 1)) = \frac{1}{2} - p + p = \frac{1}{2}\\
\mathbb{P}(Y = -1) &= \mathbb{P}(Z = (1, -1)) + \mathbb{P}(Z = (-1, -1)) = p + \frac{1}{2} - p = \frac{1}{2},
\end{aligned}$$

donc X et Y suivent une loi uniforme sur $\{-1, 1\}$. D'après la Proposition 4.11, comme $\mathbb{P}(X = 1) = \frac{1}{2} > 0$, la loi conditionnelle est bien définie et est donnée par :

$$\begin{aligned}
\mathbb{P}_{\{X=1\}}(Y = 1) &= \frac{\mathbb{P}(\{Y = 1\} \cap \{X = 1\})}{\mathbb{P}(X = 1)}\\
&= \frac{\mathbb{P}(Z = (1, 1))}{\mathbb{P}(X = 1)} = \frac{\frac{1}{2} - p}{\frac{1}{2}} = 1 - 2p.\\
\mathbb{P}_{\{X=1\}}(Y = -1) &= \frac{\mathbb{P}(\{Y = -1\} \cap \{X = 1\})}{\mathbb{P}(X = 1)}\\
&= \frac{\mathbb{P}(Z = (1, -1))}{\mathbb{P}(X = 1)} = \frac{p}{\frac{1}{2}} = 2p.
\end{aligned}$$

4.4.2 Espérance, covariance, matrice de covariance

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Toutes les variables aléatoires et vecteurs aléatoires que l'on considère sont définis sur Ω.

Proposition 4.12.

- (Théorème de transfert pour les couples). *Si $Z = (X, Y)$ est un couple aléatoire discret et que f est une application réelle définie sur $E \times F$, telle que la série $\sum_{(x,y)\in E\times F} |f(x,y)|\,\mathbb{P}^Z(\{(x,y)\})$ converge, alors $f(X,Y)$ est intégrable et :*
$$\mathbb{E}(f(X,Y)) = \sum_{(x,y)\in E\times F} f(x,y)\,\mathbb{P}^Z(\{(x,y)\}).$$
- (Théorème de transfert pour les vecteurs). *Si $\mathbf{X} = (X_1, \cdots, X_n)$ est un vecteur aléatoire discret et que f est une application réelle définie sur $E_1 \times \cdots \times E_n$, telle que la série*
$$\sum_{(x_1,\cdots,x_n)\in E_1\times\cdots\times E_n} |f(x_1,\cdots,x_n)|\,\mathbb{P}^{\mathbf{X}}(\{(x_1,\cdots,x_n)\})$$
converge, alors $f(x_1, \cdots, x_n)$ est intégrable et :
$$\mathbb{E}(f(X_1,\cdots,X_n)) = \sum_{(x_1,\cdots,x_n)\in E_1\times\cdots\times E_n} f(x_1,\cdots,x_n)\,\mathbb{P}^{\mathbf{X}}(\{(x_1,\cdots,x_n)\}).$$

Remarque. Avec le théorème de transfert pour les couples, nous avons les outils nécessaires pour démontrer la propriété de linéarité de l'espérance.

Exemple 4.8. Vérifions sur l'exemple que $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$. D'après le théorème de transfert :

$$\begin{aligned}\mathbb{E}(X+Y) = (1+1)\mathbb{P}(Z = (1,1)) + (1-1)\mathbb{P}(Z = (1,-1)) +& \\ + (-1+1)\mathbb{P}(Z = (-1,1)) + (-1-1)\mathbb{P}(Z = (-1,-1))& \\ = 2\left(\frac{1}{2} - p\right) - 2\left(\frac{1}{2} - p\right)& \\ = 0.&\end{aligned}$$

D'autre part, nous avons montré que X et Y suivent une loi uniforme sur $\{-1, 1\}$, donc :

$$\mathbb{E}(X) = 1\frac{1}{2} - 1\frac{1}{2} = 0 = \mathbb{E}(Y),$$

d'où $\mathbb{E}(X) + \mathbb{E}(Y) = 0$.

Proposition 4.13 (Inégalité de Cauchy-Schwartz)**.** *Soient X, Y deux variables aléatoires discrètes. Si X et Y sont de carré intégrable, alors XY est intégrable, et :*

$$|\mathbb{E}(XY)| \leq \mathbb{E}(X^2)^{\frac{1}{2}}\mathbb{E}(Y^2)^{\frac{1}{2}}.$$

Démonstration. L'idée de la preuve est la même que pour montrer qu'une variable aléatoire de carré intégrable est intégrable. $\square$

Définition. Soient X, Y deux variables aléatoires discrètes de carré intégrable, alors on définit la *covariance de X et Y*, notée $\mathrm{Cov}(X,Y)$, par :

$$\mathrm{Cov}(X,Y) = \mathbb{E}[(X-\mathbb{E}(X))(Y-\mathbb{E}(Y))] = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

Exemple 4.9. Calculons la covariance des variables aléatoires X et Y de l'exemple. Nous avons déjà calculé $\mathbb{E}(X) = \mathbb{E}(Y) = 0$. D'après le théorème de transfert, nous savons que :

$$\begin{aligned}\mathbb{E}(XY) &= 1.1\,\mathbb{P}(Z=(1,1)) + 1(-1)\,\mathbb{P}(Z=(1,-1)) + (-1)(1)\,\mathbb{P}(Z=(-1,1)) + (-1)(-1)\,\mathbb{P}(Z=(-1,-1))\\ &= 2\left(\frac{1}{2}-p\right) - 2p = 1-4p.\end{aligned}$$

Ainsi,

$$\mathrm{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 1-4p.$$

Propriété 4.14. *Soient X, Y deux variables aléatoires discrètes, de carré intégrable.*

1. $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$ *et* $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$.
2. *Si a, b, c, d, sont des constantes réelles, alors :*

$$\mathrm{Cov}(aX+b, cY+d) = ac\,\mathrm{Cov}(X,Y).$$

3. *La variance et la covariance sont reliées par l'égalité :*

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y).$$

4. *La covariance vérifie l'inégalité :*

$$|\,\mathrm{Cov}(X,Y)| \leq \sigma(X)\sigma(Y).$$

Démonstration.

2. Par définition de la covariance, et en utilisant la linéarité de l'espérance,

$$\begin{aligned}&\mathrm{Cov}(aX+b, cY+d) = \mathbb{E}[(aX+b)(cY+d)] - \mathbb{E}[aX+b]\mathbb{E}[cY+d]\\ &= ac\mathbb{E}(XY) + bc\mathbb{E}(Y) + ad\mathbb{E}(X) + bd - ac\mathbb{E}(X)\mathbb{E}(Y) - ad\mathbb{E}(X) - bc\mathbb{E}(Y) - bd\\ &= ac[\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)] = ac\,\mathrm{Cov}(X,Y).\end{aligned}$$

La covariance est une forme bilinéaire symétrique semi-définie positive sur les variables aléatoires de carré intégrable.

3. Par définition de la variance, et en utilisant la linéarité de l'espérance,

$$\begin{aligned}\mathrm{Var}(X+Y) &= \mathbb{E}[(X+Y)^2] - [\mathbb{E}(X+Y)]^2\\ &= \mathbb{E}(X^2) + 2\mathbb{E}(XY) + \mathbb{E}(Y^2) - [\mathbb{E}(X)^2 + 2\mathbb{E}(X)\mathbb{E}(Y) + \mathbb{E}(Y)^2]\\ &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 + \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 + 2[\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)]\\ &= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y).\end{aligned}$$

4. Il suffit d'appliquer l'inégalité de Cauchy-Schwartz avec $X-\mathbb{E}(X)$ et $Y - \mathbb{E}(Y)$.

□

Définition. Soient X, Y deux variables aléatoires discrètes de carré intégrable, de variances non nulles. Alors, le *coefficient de corrélation de* X *et* Y, noté $\rho(X,Y)$, est défini par :

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma(X)\sigma(Y)}.$$

Remarque. Le coefficient de corrélation est sans unité et est très utilisé en statistique. Comme conséquence de la Propriété 4. de la covariance, nous savons que :

$$-1 \leq \rho(X,Y) \leq 1.$$

Définition. Si $\mathbf{X} = (X_1, \cdots, X_n)$ est un vecteur aléatoire discret, telle que chacune des composantes est une variable aléatoire de carré intégrable, alors on définit la *matrice de covariance du vecteur* $\mathbf{X}$, notée $V(\mathbf{X})$, par ses coefficients :

$$\forall (i,j) \in \{1, \cdots, n\}^2, \quad (V(\mathbf{X}))_{i,j} = \mathrm{Cov}(X_i, X_j).$$

Propriété 4.15.

1. *La matrice de covariance $V(\mathbf{X})$ est une matrice réelle symétrique, dont la diagonale est formée des variances des variables aléatoires $X_1, \cdots, X_n$.*
2. $\mathrm{Var}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \mathrm{Cov}(X_i, X_j)$.

Démonstration. Le Point 1. est une conséquence directe de la Propriété 4.14. Le Point 2. se démontre par récurrence sur le nombre de variables aléatoires considérées. □

4.4.3 Variables aléatoires indépendantes

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Toutes les variables aléatoires et tous les vecteurs aléatoires que l'on considère sont définis sur Ω.

Définition.

— Soit (X,Y) un couple de variables aléatoires discrètes. On dit que les variables aléatoires X et Y sont *indépendantes*, si pour tout sous-ensemble A de E et tout sous-ensemble B de F, tels que $\{X \in A\} \in \mathcal{A}$ et $\{Y \in B\} \in \mathcal{A}$:

$$\mathbb{P}(\{X \in A\} \cap \{Y \in B\}) = \mathbb{P}(\{X \in A\})\,\mathbb{P}(\{Y \in B\}).$$

— Soit $\mathbf{X} = (X_1, \cdots, X_n)$ un vecteur aléatoire discret. On dit que $(X_1, \cdots, X_n)$ est un *n-uplet de variables aléatoires indépendantes*, ou un *vecteur aléatoire indépendant* si, pour tout sous-ensemble $(A_1, \cdots, A_n)$ de $E_1 \times \cdots \times E_n$ tel que pour tout $i \in \{1, \cdots, n\}$, $\{X_i \in A_i\} \in \mathcal{A}$:

$$\mathbb{P}(\{X_1 \in A_1\} \cap \cdots \cap \{X_n \in A_n\}) = \mathbb{P}(\{X_1 \in A_1\}) \cdots \mathbb{P}(\{X_n \in A_n\}).$$

Proposition 4.16. *Soit (X,Y) un couple de variables aléatoires discrètes. Les assertions suivantes sont équivalentes.*

1. *Les variables aléatoires X et Y sont indépendantes.*
2. *$\forall (x,y) \in E \times F$, $\mathbb{P}(\{X = x\} \cap \{Y = y\}) = \mathbb{P}(\{X = x\})\mathbb{P}(\{Y = y\})$.*

3. $\forall (x, y) \in E \times F$, $\mathbb{P}(\{X = x\}) > 0 \Rightarrow \mathbb{P}_{\{X=x\}}(\{Y = y\}) = \mathbb{P}(\{Y = y\})$.
4. $\forall (x, y) \in E \times F$, $\mathbb{P}(\{Y = y\}) > 0 \Rightarrow \mathbb{P}_{\{Y=y\}}(\{X = x\}) = \mathbb{P}(\{X = x\})$.
5. *Pour toutes fonctions f et g, respectivement définies sur E et F, telles que $f(X)$ et $g(Y)$ soient intégrables et telles que le produit $f(X)g(Y)$ est intégrable, et on a :*

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)].$$

Soit $(X_1, \cdots, X_n)$ un vecteur aléatoire discret. Les assertions suivantes sont équivalentes.

1. $(X_1, \cdots, X_n)$ *est un n-uplet de variables aléatoires indépendantes.*
2. $\forall (x_1, \cdots, x_n) \in E_1 \times \cdots \times E_n$,

$$\mathbb{P}(\{X_1 = x_1\} \cap \cdots \cap \{X_n = x_n\}) = \mathbb{P}(\{X_1 = x_1\}) \cdots \mathbb{P}(\{X_n = x_n\}).$$

3. *Pour toutes fonctions $f_1, \cdots, f_n$ respectivement définies sur $E_1, \cdots, E_n$, telles que $f_1(X_1), \cdots, f_n(X_n)$ soient intégrables et telles que le produit $f_1(X_1) \cdots f_n(X_n)$ est intégrable, et on a :*

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

Remarque 4.17. On déduit de la Proposition 4.16 le fait suivant : si les variables aléatoires X et Y sont indépendantes, alors pour toutes fonctions f et g suffisament régulières, les variables aléatoires $f(X)$ et $g(Y)$ sont aussi indépendantes.

Exemple 4.10. Étudions l'indépendance des variables aléatoires X et Y de l'exemple. Les variables aléatoires X et Y sont indépendantes, si et seulement si :

$$\begin{aligned} \mathbb{P}(\{X = 1\} \cap \{Y = 1\}) &= \mathbb{P}(X = 1)\mathbb{P}(Y = 1), \text{ et} \\ \mathbb{P}(\{X = 1\} \cap \{Y = -1\}) &= \mathbb{P}(X = 1)\mathbb{P}(Y = -1), \text{ et} \\ \mathbb{P}(\{X = -1\} \cap \{Y = 1\}) &= \mathbb{P}(X = -1)\mathbb{P}(Y = 1), \text{ et} \\ \mathbb{P}(\{X = -1\} \cap \{Y = -1\}) &= \mathbb{P}(X = -1)\mathbb{P}(Y = -1), \end{aligned}$$

$$\Leftrightarrow \frac{1}{2} - p = \frac{1}{4} \text{ et } p = \frac{1}{4} \Leftrightarrow p = \frac{1}{4}.$$

Proposition 4.18.

— *Soient X et Y deux variables aléatoires discrètes de carré intégrable. Alors, si X et Y sont indépendantes,*
  1. $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$,
  2. $\mathrm{Cov}(X, Y) = 0$,
  3. $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.

— *Soit $(X_1, \cdots, X_n)$ un n-uplet de variables aléatoires discrètes de carré intégrable et indépendantes, alors :*
  1. $\mathbb{E}(\prod_{i=1}^{n} X_i) = \prod_{i=1}^{n} \mathbb{E}(X_i)$,
  2. $\forall (i, j) \in \{1, \cdots, n\}^2$, $i \neq j$, $\mathrm{Cov}(X_i, X_j) = 0$. *Autrement dit, la matrice de covariance est diagonale.*
  3. $\mathrm{Var}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mathrm{Var}(X_i)$.

Démonstration. Montrons la proposition dans le cas des couples de variables aléatoires. D'après le théorème de transfert,

$$\begin{aligned}
\mathbb{E}(XY) &= \sum_{(x,y)\in E\times F} xy\,\mathbb{P}(\{X=x\}\cap\{Y=y\}) \\
&= \sum_{(x,y)\in E\times F} xy\,\mathbb{P}(X=x)\,\mathbb{P}(Y=y), \text{ (par indépendance)} \\
&= \Big(\sum_{x\in E} x\,\mathbb{P}(X=x)\Big)\Big(\sum_{y\in F} y\,\mathbb{P}(Y=y)\Big) \\
&= \mathbb{E}(X)\,\mathbb{E}(Y).
\end{aligned}$$

Ainsi, si X et Y sont indépendantes,

$$\mathrm{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 0,$$

et

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y) = \mathrm{Var}(X) + \mathrm{Var}(Y).$$

□

Exemple 4.11. Recalculons la variance d'une variable aléatoire qui suit une loi binomiale de paramètres n et p. Alors X a même loi que $\sum_{k=1}^{n} X_k$, où les $(X_k)_{k=1}^n$ sont des variables aléatoires indépendantes de Bernoulli de paramètre p. Ainsi, par la Proposition 4.18 :

$$\mathrm{Var}(X) = \sum_{k=1}^{n} \mathrm{Var}(X_k) = np(1-p).$$

Remarque. Attention, si $\mathrm{Cov}(X,Y) = 0$, cela n'implique pas que les variables aléatoires X et Y sont indépendantes. Par exemple, considérons le couple $Z = (X,Y)$ de loi uniforme sur $\{(1,0),(-1,0),(0,1),(0,-1)\}$. D'après la Proposition 4.11, nous avons :

$$\begin{aligned}
&\mathbb{P}(X=1) = \frac{1}{4},\ \mathbb{P}(X=-1) = \frac{1}{4},\ \mathbb{P}(X=0) = \frac{1}{2}, \\
&\mathbb{P}(Y=1) = \frac{1}{4},\ \mathbb{P}(Y=0) = \frac{1}{2},\ \mathbb{P}(Y=-1) = \frac{1}{4}.
\end{aligned}$$

Ainsi :

$$\begin{aligned}
&\mathbb{E}(X) = 0,\ \mathbb{E}(Y) = 0,\ \mathbb{E}(XY) = 0, \\
&\text{d'où } \mathrm{Cov}(X,Y) = 0.
\end{aligned}$$

Mais, X et Y ne sont pas indépendantes. En effet,

$$\mathbb{P}(X=1, Y=0) = \frac{1}{4} \text{ et } \mathbb{P}(X=1)\mathbb{P}(Y=0) = \frac{1}{8},$$
$$\text{d'où } \mathbb{P}(X=1, Y=0) \neq \mathbb{P}(X=1)\mathbb{P}(Y=0).$$

Définition. Soit $(X_k)_{k\geq 1}$ une suite de variables aléatoires discrètes. On dit que la suite $(X_k)_{k\geq 1}$ est une *suite de variables aléatoires discrètes indépendantes*, si pour tout entier n, $(X_1,\cdots,X_n)$ est un n-uplet de variables aléatoires indépendantes.

Exercices

Exercice 4.12. Soient X et Y deux variables aléatoires indépendantes, à valeurs dans $\mathbb{N}$.

1. Calculer la loi de $Z = X + Y$.
2. Calculer la loi de $T = \min(X, Y)$.

Solution de l'exercice 4.12.

1. La variable aléatoire Z est à valeurs dans $\mathbb{N}$ et la loi de Z est entièrement déterminée par la donnée des nombres :

$$\mathbb{P}^Z(\{n\}) = \mathbb{P}(Z = n) = \mathbb{P}(X + Y = n),\ n \in \mathbb{N}.$$

Soit n un entier naturel. D'après la formule des probabilités totales :

$$\begin{aligned}\mathbb{P}(Z = n) &= \sum_{k=0}^{n} \mathbb{P}(X + Y = n, X = k) \\ &= \sum_{k=0}^{n} \mathbb{P}(Y = n - k, X = k) \\ &= \sum_{k=0}^{n} \mathbb{P}(Y = n - k)\, \mathbb{P}(X = k) \quad \text{ car } X \text{ et } Y \text{ sont indépendantes.}\end{aligned}$$

2. La variable aléatoire T est à valeurs dans $\mathbb{N}$ et la loi de T est entièrement déterminée par la donnée des nombres :

$$\mathbb{P}^T(\{n\}) = \mathbb{P}(T = n) = \mathbb{P}(\min(X, Y) = n),\ n \in \mathbb{N}.$$

Soit $n \in \mathbb{N}$, calculons $\mathbb{P}(T > n - 1)$:

$$\begin{aligned}\mathbb{P}(T > n - 1) &= \mathbb{P}(\min(X, Y) > n - 1) \\ &= \mathbb{P}(\{X > n - 1\} \cap \{Y > n - 1\}) \\ &= \mathbb{P}(X > n - 1)\, \mathbb{P}(Y > n - 1), \text{ car } X \text{ et } Y \text{ sont indépendantes}\end{aligned}$$

Ainsi, pour tout $n \in \mathbb{N}$:

$$\begin{aligned}\mathbb{P}(T = n) &= \mathbb{P}(T > n - 1) - \mathbb{P}(T > n) \\ &= \mathbb{P}(X > n - 1)\mathbb{P}(Y > n - 1) - \mathbb{P}(X > n)\mathbb{P}(Y > n).\end{aligned}$$

Exercice 4.13.

1. Montrer que la loi de la somme de n variables aléatoires indépendantes de Bernoulli de paramètre p est une loi binomiale de paramètres n et p.
2. Montrer que la loi de la somme de n variables aléatoires indépendantes de Poisson de paramètres $\lambda_1, \cdots, \lambda_n$, respectivement est une loi de Poisson de paramètre $\sum\limits_{i=1}^{n} \lambda_i$.

3. Montrer que la variable aléatoire $T = \min(X, Y)$, où X et Y sont indépendantes de même loi géométrique de paramètres p ($p \in]0, 1[$), suit une loi géométrique de paramètre $1 - (1 - p)^2$.

Solution de l'exercice 4.13.

1. Pour tout $n \in \mathbb{N}^*$, on définit la propriété $\mathcal{P}(n)$: "la somme $Y_n = \sum\limits_{k=1}^{n} X_k$ de n variables indépendantes de Bernoulli de paramètre p suit une loi binomiale de paramètres n et p".
Si X_1 suit une loi de Bernoulli de paramètre p, alors X_1 suit une loi binomiale de paramètres 1 et p et la propriété $\mathcal{P}(1)$ est vérifiée.
Supposons que pour un entier $n \in \mathbb{N}^*$, la propriété $\mathcal{P}(n)$ soit vraie. La variable aléatoire $Y_{n+1} = \sum\limits_{k=1}^{n+1} X_k$ prend ses valeurs dans $\{0, \cdots, n+1\}$, et d'après l'exercice précédent nous savons que, pour tout $k \in \{0, \cdots, n+1\}$:
$$\begin{aligned}
\mathbb{P}(Y_{n+1} = k) &= \sum_{\ell=0}^{k} \mathbb{P}(Y_n = k - \ell)\, \mathbb{P}(X_{n+1} = \ell) \\
&= \mathbb{P}(Y_n = k)(1-p) + \mathbb{P}(Y_n = k-1)p, \quad \text{car } X_{n+1} \in \{0,1\} \\
&= \binom{n}{k} p^k (1-p)^{n+1-k} + \binom{n}{k-1} p^k (1-p)^{n-k+1}, \;\text{ par hypothèse} \\
&= \left(\binom{n}{k} + \binom{n}{k-1}\right) p^k \,(1-p)^{n+1-k} \\
&= \binom{n+1}{k} p^k \,(1-p)^{n+1-k},
\end{aligned}$$
et la propriété $\mathcal{P}(n+1)$ est vraie. On a donc montré que la propriété est vraie au rang initial et héréditaire. Du principe de raisonnement par récurrence on en déduit que la propriété $\mathcal{P}(n)$ est vraie pour tout $n \in \mathbb{N}^*$.
2. Nous ne rédigerons pas la récurrence en détail cette fois-ci. La propriété est clairement vraie au rang initial. Supposons qu'elle soit vraie pour un certain n, et soit $Y_{n+1} = \sum\limits_{k=1}^{n+1} X_k$ comme dans l'énoncé. Alors, Y_{n+1} est à valeurs dans $\mathbb{N}$ et d'après l'exercice précédent, pour tout $k \in \mathbb{N}$:
$$\begin{aligned}
\mathbb{P}(Y_{n+1} = k) &= \sum_{\ell=0}^{k} \mathbb{P}(Y_n = k - \ell)\mathbb{P}(X_{n+1} = \ell) \\
&= \sum_{\ell=0}^{k} e^{\left(-\sum\limits_{i=1}^{n} \lambda_i\right)} \frac{\left(\sum\limits_{i=1}^{n} \lambda_i\right)^{k-\ell}}{(k-\ell)!} e^{-\lambda_{n+1}} \frac{(\lambda_{n+1})^\ell}{\ell!}, \;\text{ par hypothèse} \\
&= e^{\left(-\sum_{i=1}^{n+1} \lambda_i\right)} \frac{1}{k!} \sum_{\ell=0}^{k} \binom{k}{\ell} \Big(\sum_{i=1}^{n} \lambda_i\Big)^{k-\ell} (\lambda_{n+1})^\ell \\
&= e^{\left(-\sum\limits_{i=1}^{n+1} \lambda_i\right)} \frac{1}{k!} \Big(\sum_{i=1}^{n+1} \lambda_i\Big)^k, \;\text{ d'après la formule du binôme de Newton,}
\end{aligned}$$
et la propriété est vraie au rang $n+1$.

3. À faire à la maison.

Exercice 4.14. Soient X et Y deux variables aléatoires indépendants de loi uniforme sur $\{1, \cdots, n\}$. Calculer $\mathbb{P}[X = Y]$.

Solution de l'exercice 4.14. Nous avons :

$$
\begin{aligned}
\mathbb{P}[X = Y] &= \mathbb{P}[\bigcup_{k=1}^{n} \{\{X = k\} \cap \{Y = k\}\}] \\
&= \sum_{k=1}^{n} \mathbb{P}[\{X = k\} \cap \{Y = k\}] \\
&\qquad \text{car les événements } (\{X = k\} \cap \{Y = k\}; 1 \leq k \leq n) \text{ sont incompatibles} \\
&= \sum_{k=1}^{n} \mathbb{P}[\{X = k\}]\, \mathbb{P}[\{Y = k\}], \text{ car les variables aléatoires } X \text{ et } Y \text{ sont indépendantes} \\
&= \sum_{k=1}^{n} \frac{1}{n^2}, \text{ car } X \text{ et } Y \text{ suivent la loi uniforme sur } \{1, \cdots, n\} \\
&= \frac{1}{n}.
\end{aligned}
$$

Exercice 4.15. Soient X et Y deux variables aléatoires indépendantes, de même loi de Bernoulli de paramètre p, $0 < p < 1$. On pose $S = X + Y$ et $D = XY$.

1. Déterminer la loi du couple (S, D).
2. En déduire les lois marginales de S et D.
3. Calculer de trois manières différentes $\mathbb{E}(S)$ et $\mathbb{E}(D)$.
4. Calculer $\mathrm{Cov}(S, D)$. Les variables aléatoires S et D sont-elles indépendantes ?

Solution de l'exercice 4.15.

1. La variable aléatoire S est à valeurs dans $E = \{0, 1, 2\}$ et la variable aléatoire D est à valeurs dans $F = \{0, 1\}$, ainsi le couple $Z = (S, D)$ est à valeurs dans $E \times F$. Sa loi est caractérisée par la donnée des nombres :

$$\forall\, (x_1, x_2) \in E \times F, \quad \mathbb{P}^Z[\{(x_1, x_2)\}] = \mathbb{P}[\{S = x_1\} \cap \{D = x_2\}].$$

Calculons. Pour tout $x_2 \in \{0,1\}$,

$$\begin{aligned}
\mathbb{P}^Z[\{(0,x_2)\}] &= \mathbb{P}[\{X+Y=0\} \cap \{XY=x_2\}],\\
&= \mathbb{P}[\{X=0\} \cap \{Y=0\} \cap \{XY=x_2\}]\\
&= \begin{cases} \mathbb{P}[\{X=0\} \cap \{Y=0\}] = (1-p)^2, & \text{si } x_2 = 0\\ \mathbb{P}[\emptyset] = 0, & \text{si } x_2 = 1 \end{cases}\\
\mathbb{P}^Z[\{(1,x_2)\}] &= \mathbb{P}[\{X+Y=1\} \cap \{XY=x_2\}],\\
&= \mathbb{P}[\{X=0\} \cap \{Y=1\} \cap \{XY=x_2\}] + \mathbb{P}[\{X=1\} \cap \{Y=0\} \cap \{XY=x_2\}]\\
&= \begin{cases} \mathbb{P}[\{X=0\} \cap \{Y=1\}] + \mathbb{P}[\{X=1\} \cap \{Y=0\}] = 2p(1-p), & \text{si } x_2 = 0\\ \mathbb{P}[\emptyset] = 0, & \text{si } x_2 = 1 \end{cases}\\
\mathbb{P}^Z[\{(2,x_2)\}] &= \mathbb{P}[\{X+Y=2\} \cap \{XY=x_2\}],\\
&= \mathbb{P}[\{X=1\} \cap \{Y=1\} \cap \{XY=x_2\}]\\
&= \begin{cases} \mathbb{P}[\emptyset] = 0, & \text{si } x_2 = 0\\ \mathbb{P}[\{X=1\} \cap \{Y=1\}] = p^2, & \text{si } x_2 = 1. \end{cases}
\end{aligned}$$

2. D'après la formule des probabilités totales, la loi marginale de la variable aléatoire S est donnée par :

$$\forall\, x_1 \in E, \quad \mathbb{P}^S[x_1] = \mathbb{P}[\{S = x_1\}] = \sum_{x_2 \in F} \mathbb{P}^Z[\{(x_1,x_2)\}].$$

Ainsi, la loi marginale de S est :

$$\begin{aligned}
\mathbb{P}^S[\{0\}] &= \mathbb{P}^Z[\{(0,0)\}] + \mathbb{P}^Z[\{(0,1)\}] = (1-p)^2\\
\mathbb{P}^S[\{1\}] &= \mathbb{P}^Z[\{(1,0)\}] + \mathbb{P}^Z[\{(1,1)\}] = 2p(1-p)\\
\mathbb{P}^S[\{2\}] &= \mathbb{P}^Z[\{(2,0)\}] + \mathbb{P}^Z[\{(2,1)\}] = p^2.
\end{aligned}$$

Remarquons que S étant somme de deux variables aléatoires de Bernoulli indépendantes, on sait d'après l'exercice précédent, qu'elle suit une loi binomiale de paramètres 2 et p.
De manière analogue, la loi marginale de la variable aléatoire D est donnée par :

$$\begin{aligned}
\mathbb{P}^D[\{1\}] &= \sum_{x_1=0}^{2} \mathbb{P}^Z[\{(x_1,1)\}] = p^2\\
\mathbb{P}^D[\{0\}] &= 1 - \mathbb{P}^D[\{1\}] = 1 - p^2.
\end{aligned}$$

3. Les variables aléatoires S et D ne prenant qu'un nombre fini de valeurs, elles admettent une espérance.
Première méthode. A la question précédente, nous avons déterminé les lois des variables aléatoires S et D. Ainsi, d'après la définition de l'espérance, nous avons :

$$\begin{aligned}
\mathbb{E}[S] &= 0 \cdot \mathbb{P}^S[\{0\}] + 1 \cdot \mathbb{P}^S[\{1\}] + 2 \cdot \mathbb{P}^S[\{2\}]\\
&= 2p(1-p) + 2p^2 = 2p.\\
\mathbb{E}[D] &= 0 \cdot \mathbb{P}^D[\{0\}] + 1 \cdot \mathbb{P}^D[\{1\}]\\
&= p^2.
\end{aligned}$$

Deuxième méthode. D'après la linéarité de l'espérance,

$$\mathbb{E}[S] = \mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y] = 2p.$$

Les variables aléatoires étant indépendantes :

$$\mathbb{E}[D] = \mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y] = p^2.$$

Troisième méthode. D'après le théorème de transfert appliqué au couple (X, Y), on a :

$$\begin{aligned}
\mathbb{E}[X+Y] &= \sum_{(x,y)\in\{0,1\}^2} (x+y)\mathbb{P}[\{X=x\}\cap\{Y=y\}] \\
&= \sum_{(x,y)\in\{0,1\}^2} (x+y)\mathbb{P}[\{X=x\}]\mathbb{P}[\{Y=y\}], \text{ car } X \text{ et } Y \text{ sont indépendantes} \\
&= 0(1-p)^2 + 1[p(1-p) + (1-p)p] + 2.p^2 = 2p \\
\mathbb{E}[XY] &= \sum_{(x,y)\in\{0,1\}^2} (xy)\mathbb{P}[\{X=x\}]\mathbb{P}[\{Y=y\}] \\
&= 0(1-p)^2 + 0.2p(1-p) + 1.p^2 = p^2.
\end{aligned}$$

4. Par définition,

$$\begin{aligned}
\mathrm{Cov}(S, D) &= \mathbb{E}[SD] - \mathbb{E}[S]\mathbb{E}[D] \\
&= \mathbb{E}[(X+Y)XY] - 2p^3, \text{ d'après la question 2} \\
&= \mathbb{E}[X^2Y] + \mathbb{E}[Y^2X] - 2p^3 \\
&= \mathbb{E}[X^2]\mathbb{E}[Y] + \mathbb{E}[Y^2]\mathbb{E}[X] - 2p^3, \text{ car } X \text{ et } Y \text{ sont indépendantes} \\
&= 2p^2 - 2p^3 = 2p^2(1-p).
\end{aligned}$$

Pour que les variables aléatoires soient indépendantes, il faut (mais il ne suffit pas) que $\mathrm{Cov}(S, D) = 0$; il faut donc que $p = 0$ ou $p = 1$. Or par hypothèse $0 < p < 1$, les variables aléatoires S et D ne sont donc pas indépendantes.

4.5 Suites de variables aléatoires réelles

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Toutes les variables aléatoires considérées dans ce paragraphe sont définies sur Ω, à valeurs réelles, discrètes ou à densité.

Définition. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires. On dit que la suite $(X_n)_{n\geq 1}$ *converge en probabilité* vers la variable aléatoire X si :

$$\forall \varepsilon > 0, \lim_{n\to+\infty} \mathbb{P}[|X - X_n| > \varepsilon] = 0.$$

Dans ce cas, on note : $X_n \xrightarrow[n\to+\infty]{\mathbb{P}} X$.

Proposition 4.19. *Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires convergeant en probabilité vers X, et soit f une fonction continue définie sur $\mathbb{R}$.*

1. *La suite $(f(X_n))_{n\geq 1}$ converge en probabilité vers $f(X)$.*
2. *Si de plus f est bornée, on a : $\lim_{n\to+\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)]$.*

4.5.1 Loi faible des grands nombres

Théorème 4.20 (Loi faible des grands nombres.)**.** *Soit* $(X_n)_{n\geq 1}$ *une suite de variables aléatoires indépendantes et de même loi. On suppose que ces variables sont de carré intégrable, d'espérance commune* m*. Alors la suite* $\left(\frac{1}{n}\sum\limits_{k=1}^{n} X_k\right)_{n\geq 1}$ *converge en probabilité vers* m*.*

Démonstration. Notons m l'espérance et σ^2 la variance commune des variables aléatoires X_k, $k \geq 1$.

Rappelons l'inégalité de Bienaymé-Tchebychev. Si Z est une variable aléatoire de carré intégrable, alors pour tout $a > 0$:

$$\mathbb{P}[|Z - \mathbb{E}[Z]| > a] \leq \frac{\mathrm{Var}(Z)}{a^2}.$$

Soit $\varepsilon > 0$ et définissons $Z = \frac{1}{n}\sum\limits_{k=1}^{n} X_k$. Alors la variable aléatoire Z est de carré intégrable car somme finie de variables aléatoires de carré intégrable. De plus :

$$\begin{aligned}
\mathbb{E}[Z] &= \mathbb{E}\left[\frac{1}{n}\sum_{k=1}^{n} X_k\right] = \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}[X_k] = m, \text{ par linéarité de l'espérance}\\
\mathrm{Var}[Z] &= \mathrm{Var}\left[\frac{1}{n}\sum_{k=1}^{n} X_k\right] = \frac{1}{n^2}\,\mathrm{Var}\Big[\sum_{k=1}^{n} X_k\Big]\\
&= \frac{1}{n^2}\sum_{k=1}^{n}\mathrm{Var}[X_k] = \frac{\sigma^2}{n}, \text{ car les variables } X_k \text{ sont indépendantes.}
\end{aligned}$$

En appliquant l'inégalité de Bienaymé-Tchebychev à Z, on obtient :

$$0 \leq \mathbb{P}\left[\left|\Big(\frac{1}{n}\sum_{k=1}^{n} X_k\Big) - m\right| > \varepsilon\right] \leq \frac{\sigma^2}{n\,\varepsilon^2}.$$

D'après le théorème d'encadrement, pour tout $\varepsilon > 0$:

$$\lim_{n\to+\infty}\mathbb{P}\left[\left|\Big(\frac{1}{n}\sum_{k=1}^{n} X_k\Big) - m\right| > \varepsilon\right] = 0,$$

et on en déduit que la suite $\left(\frac{1}{n}\sum\limits_{k=1}^{n} X_k\right)_{n\geq 1}$ converge en probabilité vers m. □

Remarque.

1. On peut démontrer que ce résultat reste vrai si on suppose seulement les variables aléatoires intégrables (au lieu de carré intégrable).
2. Il existe également une version "forte" de la loi des grands nombres.

Exemple. Considérons ce que l'on pourrait appeler un schéma de Bernoulli de paramètres ∞ et p, *i.e.* une expérience qui consiste à répéter indéfiniment une même expérience admettant deux issues : succès/échec, telle que p est la probabilité d'obtenir un succès. Si on note A l'événement "obtenir un succès", alors $\mathbb{P}(A) = p$. Par exemple, on lance indéfiniment un dé non pipé et on considère l'événement A "obtenir un 6", alors $p = 1/6$.

Pour $k \geq 1$, on définit A_k l'événement "A est réalisé lors de la k-ième expérience" et $X_k = \mathbb{I}_{A_k}$. Alors la suite $(X_k)_{k\geq 1}$ est une suite de variables aléatoires indépendantes de même loi de Bernoulli de paramètre p, de carré intégrable étant donné qu'elles ne prennent que deux valeurs. De plus, pour tout $k \geq 1$, $\mathbb{E}(X_k) = \mathbb{P}(A_k) = \mathbb{P}(A) = p$. D'après la loi des grands nombres :

$$\frac{1}{N}\sum_{k=1}^{N} X_k \xrightarrow[N\to+\infty]{\mathbb{P}} \mathbb{E}(X_1) = p.$$

Interprétons ce résultat. Pour tout $N \in \mathbb{N}$, la variable aléatoire $\sum\limits_{k=1}^{N} X_k = \sum\limits_{k=1}^{N} \mathbb{I}_{A_k} = n_N(A)$ représente le nombre de fois où A est réalisé lors des N premières épreuves, et $\frac{1}{N}\sum\limits_{k=1}^{N} X_k$ représente donc la fréquence de A lors des N premières épreuves. Ainsi la loi des grands nombres peut se réécrire :

$$\frac{n_N(A)}{N} \xrightarrow[N\to+\infty]{\mathbb{P}} \mathbb{P}(A).$$

Cela justifie ainsi l'approche intuitive dont nous avons parlé au début de ce cours (cf. chapitre 1). Dans l'exemple du lancer de dé, cela signifie que la fréquence du nombre de 6 obtenus tend vers 1/6 en probabilité.

4.5.2 Théorème central limite

Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes de même loi, de moyenne et variance commune m et σ^2 respectivement. On définit $S_n = \sum\limits_{k=1}^{n} X_k$. Alors d'après la loi des grands nombres,

$$\frac{S_n - nm}{n} \xrightarrow[n\to+\infty]{\mathbb{P}} 0.$$

Supposons que l'on divise par quelque chose de plus petit que n, peut-on alors obtenir une limite ? La réponse est oui, si on normalise par une quantité proportionnelle à $\sqrt{n}$. Cela précise en particulier la vitesse de convergence dans la loi faible des grands nombres.

Théorème 4.21 (Théorème central limite)**.** *Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires de carré intégrables, indépendantes, de même loi, de moyenne et variance commune m et σ^2 respectivement. On pose $S_n = \sum\limits_{k=1}^{n} X_k$. Alors pour tout réel t :*

$$\lim_{n\to+\infty} \mathbb{P}\left[\frac{S_n - nm}{\sigma\sqrt{n}} \leq t\right] = \lim_{n\to+\infty} \mathbb{P}\left[\frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathrm{Var}[S_n]}} \leq t\right] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-\frac{x^2}{2}}\,dx.$$

En posant $\tilde{S}_n = \frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathrm{Var}[S_n]}}$, cela peut s'écrire :

$$\forall\, t \in \mathbb{R}, \quad \lim_{n\to+\infty} F_{\tilde{S}_n}(t) = \Pi(t),$$

où Π est la fonction de répartition de la loi normale centrée réduite.

Remarque.

1. On dit que la variable aléatoire $\tilde{S}_n$ *converge en loi* vers une variable aléatoire gaussienne centrée réduite.
2. De manière équivalente, on a que pour tous réels a et b, $a \leq b$,
$$\begin{aligned} &\lim_{n\to+\infty} \mathbb{P}[a \leq \tilde{S}_n \leq b] = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}}\,dx \\ \Leftrightarrow &\lim_{n\to+\infty} F_{\tilde{S}_n}(b) - F_{\tilde{S}_n}(a) = \Pi(b) - \Pi(a). \end{aligned}$$
3. Ce théorème a d'abord été démontré lorsque la loi commune des variables est une loi de Bernoulli de paramètre p (le théorème est dans ce cas connu sous le nom de théorème de Moivre-Laplace). Dans ce cas, la variable S_n suit une loi binomiale de paramètres n et p. L'approximation de cette loi binomiale par la loi de Gauss est déjà très bonne lorsque $n \geq 10$, et que np et $n(1-p)$ dépassent quelques unités.
   Plus précisément dans ce cas, notons F_n la fonction de répartition de la variable aléatoire centrée réduite $\tilde{S}_n = \frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathrm{Var}[S_n]}}$. Alors, on a :
$$\sup_{x\in\mathbb{R}} |F_n(x) - \Pi(x)| \leq \frac{p^2+q^2}{\sqrt{npq}},$$
   où $q = 1-p$; (cf. Shiryayev, Probability, Springer).
   On a donc vu deux façons d'approcher la loi binomiale : par la loi de Gauss, et par la loi de Poisson.
4. Corrections de continuité (programme BTS).
   Si a et b sont des nombres entiers compris entre 0 et n (avec $a \leq b$), la probabilité que la variable S_n de loi binomiale de paramètres n et p soit comprise entre a et b est égale à $\mathbb{P}[a-\varepsilon \leq S_n \leq b+\varepsilon]$ pour tout $\varepsilon \in]0,1[$ puisqu'elle ne prend que des valeurs entières. Il se trouve que l'approximation que l'on obtient (pour les grandes valeurs de n) à l'aide du théorème central limite est la meilleure pour $\varepsilon = \frac{1}{2}$. On approchera donc $\mathbb{P}[a \leq S_n \leq b]$ par l'approximation de
$$\mathbb{P}[a - \frac{1}{2} \leq S_n \leq b + \frac{1}{2}] = \mathbb{P}\left[\frac{a-\frac{1}{2}-np}{\sqrt{npq}} \leq \frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathrm{Var}[S_n]}} \leq \frac{b+\frac{1}{2}-np}{\sqrt{npq}}\right],$$
   c'est-à-dire par $\Pi\left(\frac{b+\frac{1}{2}-np}{\sqrt{npq}}\right) - \Pi\left(\frac{a-\frac{1}{2}-np}{\sqrt{npq}}\right)$ (cf. Feller W., On the normal approximation to the binomial distribution, *Ann. Math. Statist.*, vol. 16, p. 319–329, 1945).
5. Dans les programmes du secondaire, on trouve parfois ce théorème sous le nom de *théorème de la limite centrée.* Mais aucun probabiliste sérieux n'utilise cette dénomination que l'on n'utilisera donc pas.

Exercice 4.16. [*D'après le polycopié de cours de P. Priouret.*] Un joueur pense qu'un dé (à six faces) est pipé. Il le lance 720 fois, et obtient le six 150 fois. Quelle conclusion le joueur peut-il en tirer ?

Solution de l'exercice 4.16. Supposons que le dé ne soit pas pipé. Notons A_k l'événement "le joueur obtient un six au k-ième lancer", et posons $X_k = \mathbb{I}_{A_k}$. La variable $S = \sum\limits_{k=1}^{720} X_k$ suit alors la loi binomiale de paramètres $n = 720$ et $p = \frac{1}{6}$ puisque le dé n'est pas pipé.

On a $\mathbb{E}[S] = np = \frac{720}{6} = 120$ et $\mathrm{Var}[S] = np(1-p) = \frac{720.5}{6.6} = 100$. Évaluons la probabilité de l'événement (qui s'est produit) : "on a obtenu au moins 150 fois le six", *i.e.* $\mathbb{P}[S \geq 150]$ en effectuant une correction de continuité. Puisqu'une variable de loi binomiale ne prend que des valeurs entières :

$$\begin{aligned}\mathbb{P}[S \geq 150] = \mathbb{P}[S \geq 149,5] &= P\left[\frac{S - \mathbb{E}[S]}{\sqrt{\mathrm{Var}[S]}} \geq \frac{149,5 - 120}{10}\right] \\ &= P\left[\frac{S - \mathbb{E}[S]}{\sqrt{\mathrm{Var}[S]}} \geq 2,95\right] \simeq 1 - \Pi(2,95) \simeq 0,0016,\end{aligned}$$

d'après le théorème central limite.

Ainsi, si le dé n'est pas pipé, il s'est produit un événement de probabilité très faible. Il y a donc de très grandes chances pour que le dé soit pipé (mais on ne peut en être certain !)

Deuxième partie

Chaînes de Markov

Chapitre 5

Introduction et définitions

5.1 Processus stochastiques

Très souvent, lorsque que nous étudions un phénomène qui dépend du hasard, il y a lieu de prendre en compte l'évolution de ce phénomène au cours du temps. Nous avons vu que chaque observation d'un phénomène réel est modélisée par une variable aléatoire réelle ; l'étude de phénomènes évoluant dans le temps va donc être modélisée par une famille de variables aléatoires, appelée processus stochastique.

Définition 5.1. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et soit $(E, \mathcal{E})$ un espace muni d'une tribu, appelé *espace des états*. Un *processus stochastique* est une famille de variables aléatoires réelles, $(X_t)_{t \in T}$, définies sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans $(E, \mathcal{E})$.

L'ensemble T représente le temps. Par suite la variable aléatoire X_t correspond à l'état du phénomène à l'instant t. Si l'ensemble T est :

- dénombrable, le processus stochastique est dit *discret* ; par exemple $T = \mathbb{N}$ ou tout ensemble possédant un nombre fini d'éléments. Nous employons le terme de *chaîne* et indexons la variable aléatoire par la lettre n.
- continu, le processus stochastique est dit *continu* ; par exemple $T = \mathbb{R}^+, [0, t_0]$ ou tout sous-ensemble de réels positifs. Nous indexons la variable aléatoire par la lettre t.

Définition 5.2. Soit $(X_t)_{t \in T}$ un processus stochastique défini sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans $(E, \mathcal{E})$ et soit $\omega \in \Omega$ un événement élémentaire. La *trajectoire* du processus associée à ω est l'application,

$$\begin{aligned} T &\to E \\ t &\mapsto X_t(\omega). \end{aligned}$$

Dans le cas particulier des chaînes où le temps est l'ensemble des entiers positifs $\mathbb{N}$, la *trajectoire* associée à ω est l'application,

$$\begin{aligned} \mathbb{N} &\to E \\ n &\mapsto X_n(\omega). \end{aligned}$$

5.2 A.A. Markov

Les premiers processus étudiés sont, bien-sûr, les suites de variables aléatoires indépendantes, ce qui a conduit à *la loi des grands nombres* et au *théorème central limite*. Le mathématicien russe Andreï Andreïevitch Markov [1] poussé vers la théorie des probabilités par son maître, Pafnoutï Lvovitch Tchebychev, qui démontra, sous des conditions assez générales, le théorème central limite, chercha à généraliser ce théorème à des suites de variables aléatoires dépendantes, pour répondre à une polémique entre lui et un mathématicien russe clérical et monarchiste. Il est amené ainsi à considérer des variables faiblement dépendantes, c'est à dire que l'évolution future ne dépend que de l'état présent, fondement de la théorie des processus auxquels fût donné son nom.

A.A. Markov est né à Riazan en 1856 et mort à Saint Petersbourg en 1922, où il devint professeur en 1886 et où il fut membre de l'Académie des Sciences à partir de 1896. Il succéda à son maître, dont il fut le meilleur étudiant, comme professeur de probabilités. Ses premiers travaux de recherche se situent dans le domaine de la théorie des nombres et en analyse, ses recherches concernent tout particulièrement les fractions continues, les limites d'intégrales, la convergence des séries et la théorie de l'approximation.

En 2006 était célébré le centenaire [2] de l'article lançant l'idée de base des chaînes dites de Markov. À noter que, passionné de littérature, il développa en 1912 une application des chaînes à l'étude statistique des textes. Il montre que l'occurrence d'une voyelle immédiatement après une consonne est un processus de Markov en étudiant les 20 000 premières lettres du roman de Pouchkine : *Eugène Onéguine.*

5.3 Définitions

Dans ce cours, nous allons étudier les processus de Markov, qui constituent les processus stochastiques les plus simples une fois l'hypothèse d'indépendance abandonnée. Nous supposons que l'ensemble T est l'ensemble des entiers naturels $\mathbb{N}$; nous étudions donc les *chaînes de Markov.* De plus, le maniement des probabilités discrètes étant plus simple que celui des probabilités continues et les exemples étant déjà très riches, nous étudions exclusivement des chaînes de Markov à valeurs dans un espace des états E *dénombrable* (fini ou infini), c'est-à-dire discret. L'espace E est muni de la tribu $\mathcal{E} = \mathcal{P}(E)$ des parties de E, ce que nous ne précisons plus dans la suite.

5.3.1 Propriété de Markov

Définition 5.3. Une suite de variables aléatoires $(X_n)_{n\geq 0}$ définie sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, à valeurs dans un espace E discret est appelée *chaîne de Markov* si, pour tout $n \geq 0$ et pour toute suite $(x_0, x_1, \ldots, x_n, x_{n+1})$ d'éléments de E, telle que la probabilité $\mathbb{P}[X_0 = x_0, \ldots, X_n = x_n] > 0$, nous avons :

$$\mathbb{P}[X_{n+1} = x_{n+1} | X_n = x_n, \ldots, X_0 = x_0] = \mathbb{P}[X_{n+1} = x_{n+1} | X_n = x_n]. \tag{5.1}$$

1. voir le site « http ://www-groups.dcs.st-and.ac.uk/ history »
2. voir l'article de Laurent Mazliak « Markov et ses chaînes »paru dans Pour la Science de février 2006

L'indice de la variable aléatoire étant assimilé au temps, X_n représente l'observation du processus à l'instant n. L'indice 0 représente l'instant de départ, il est appelé *l'instant initial* et l'état X_0 du processus en cet instant correspond à *l'état initial*.

L'égalité (5.1) est interprétée de la façon suivante : l'état x_{n+1} du processus à l'instant $n+1$ ne dépend pas du déroulement passé, $x_0, \ldots, x_{n-1}$, mais seulement de l'état présent x_n. Ceci se traduit encore par : le déroulement futur est le même quel que soit le déroulement passé, s'il se retrouve dans le même état présent. Cette propriété est connue sous le nom de *propriété de Markov* .

5.3.2 Probabilités et matrices de transition

Nous voyons qu'il est fondamental de connaître la probabilité d'être dans un état y à l'instant suivant sachant que l'on est dans un état x à l'instant présent. Ceci justifie la définition suivante.

Définition 5.4. Soit $(X_n)_{n\geq 0}$ une chaîne de Markov à valeurs dans un espace E. On appelle *probabilités de transition* la donnée, pour tout $n \geq 0$, pour tout $(x, y) \in E^2$, des

$$\mathbb{P}[X_{n+1} = y | X_n = x] := p_n(x, y).$$

Une chaîne de Markov $(X_n)_{n\geq 0}$ est dite *homogène* si $p_n(x, y)$ ne dépend pas de l'instant n :

$$\forall n \geq 0,\ \forall\, (x, y) \in E^2, \quad p_n(x, y) \equiv p(x, y). \tag{5.2}$$

En mots cela signifie que la probabilité de transiter d'un état x à un état y ne dépend pas de l'instant auquel la transition se fait.

Dans la suite nous supposons toujours que la chaîne de Markov est homogène et nous ne le mentionnons plus explicitement. Ceci n'est pas une restriction car, si une chaîne n'est pas homogène, nous pouvons toujours considérer le temps comme une dimension supplémentaire de l'espace des états, en prenant comme nouvel espace des états : $E' = E \times \mathbb{N}$, et se ramener ainsi à l'étude d'une chaîne homogène.

Définition 5.5. Étant donné une chaîne de Markov $(X_n)_{n\geq 0}$ à valeurs dans un espace dénombrable $E = \{x_1, x_2, \ldots\}$, nous lui associons une matrice appelée *matrice de transition*, notée $P = (p(x_i, x_j))$, dont les coefficients,

$$p(x_i, x_j) = \mathbb{P}[X_1 = x_j | X_0 = x_i],$$

sont les probabilités de transiter de l'état x_i vers l'état x_j :

$$P = \begin{pmatrix} p(x_1, x_1) & p(x_1, x_2) & p(x_1, x_3) & \ldots \\ p(x_2, x_1) & p(x_2, x_2) & p(x_2, x_3) & \ldots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Remarque. La matrice de transition est une matrice carrée de taille $\mathrm{card}(E) \times \mathrm{card}(E)$. Si l'espace E est fini, la matrice est de taille finie ; si l'espace des états est dénombrable infini, elle est alors infinie (voir par exemple le cas de la *marche aléatoire* où l'espace des états est l'ensemble $\mathbb{Z}$ des entiers relatifs). Cela ne pose pas de problème supplémentaire, la raison étant que cette matrice est à coefficients positifs ou nuls.

Une matrice de transition, de par sa construction, possède une propriété remarquable, qui est que chaque ligne correspond à une loi de probabilité. Nous détaillons ceci dans la section suivante.

5.3.3 Matrices stochastiques

Soit P une matrice de transition associée à une chaîne de Markov. Étant donnée une ligne correspondant à un état x, les valeurs des colonnes décrivent tous les états possibles à l'étape suivante ; par suite chaque ligne est une probabilité discrète. De telles matrices ont une très grande importance, au delà du calcul des probabilités, d'où l'introduction des définitions qui suivent.

Définition 5.6. Une matrice carrée $P = (p(x,y))$, dont les lignes et colonnes sont indexées par un ensemble d'états E dénombrable (fini ou infini), telle que toutes les lignes sont des probabilités, s'appelle une *matrice stochastique.* Une matrice stochastique est donc caractérisée par les propriétés suivantes,

- tous les coefficients sont positifs ou nuls : $\forall\,(x,y) \in E^2, \quad p(x,y) \geq 0,$
- les coefficients de chacune des lignes somment à 1 : $\forall\, x \in E, \quad \displaystyle\sum_{y\in E} p(x,y) = 1.$

Exercice 5.1. Montrer que la matrice de transition $P = (p(x,y))$ associée à une chaîne de Markov $(X_n)_{n\geq 0}$ à valeurs dans E est une matrice stochastique.

Solution de l'exercice 5.1. Par définition de la matrice de transition P, pour tout $(x,y) \in E^2$,

$$p(x,y) = \mathbb{P}[X_{n+1} = y | X_n = x].$$

D'après le lemme 3.1, une probabilité conditionnelle est une probabilité. Ainsi, $\mathbb{P}_{\{X_n=x\}} = \mathbb{P}[\,\cdot\,|\{X_n = x\}]$ est une probabilité. Ceci implique immédiatement la positivité de $p(x,y)$. Aussi, nous avons que pour tout $x \in E$,

$$\sum_{y\in E} p(x,y) = \sum_{y\in E} \mathbb{P}[X_{n+1} = y|X_n = x] = \sum_{y\in E} \mathbb{P}_{\{X_n=x\}}[X_{n+1} = y] = 1.$$

Il est peut-être utile de redémontrer cette partie du résultat :

$$\begin{aligned}\sum_{y\in E} \mathbb{P}_{\{X_n=x\}}[X_{n+1} = y] &= \sum_{y\in E} \frac{\mathbb{P}[\{X_{n+1} = y\} \cap \{X_n = x\}]}{\mathbb{P}[X_n = x]}, \text{ par définition des probabilité conditionnelles} \\ &= \frac{\sum_{y\in E} \mathbb{P}[\{X_{n+1} = y\} \cap \{X_n = x\}]}{\mathbb{P}[X_n = x]}, \text{ le dénominateur étant indépendant de } y \\ &= \frac{\mathbb{P}[X_n = x]}{\mathbb{P}[X_n = x]}, \text{ d'après la formule des probabilités totales au numérateur} \\ &= 1.\end{aligned}$$

Ceci démontre que les coefficients de chacune des lignes somment à 1 et conclut la preuve.

Exercice 5.2. Montrer que toute matrice stochastique $P = (p(x,y))$ admet une valeur propre égale à 1 et comme vecteur propre associé le vecteur, noté $\mathbf{e}$, dont toutes les coordonnées sont égales à 1.

Solution de l'exercice 5.2. Soit $P = (p(x,y))$ une matrice stochastique définie sur E ; pour établir que le vecteur $\mathbf{e} = (\mathbf{e}(x))$ est un vecteur propre associé à la valeur propre 1, il suffit d'établir l'égalité :

$$P\mathbf{e} = \mathbf{e}.$$

Or nous obtenons pour tout x :

$$(P\mathbf{e})(x) = \sum_{y\in E} p(x,y)\mathbf{e}(y) = \sum_{y\in E} p(x,y) = 1,$$

la matrice P étant stochastique.

Exercice 5.3. Montrer que le produit de deux matrices stochastiques indexées par le même ensemble E est encore une matrice stochastique.

Solution de l'exercice 5.3. Soit $P = (p(x,y))$ et $Q = (q(x,y))$ deux matrices stochastiques. Appelons R la matrice produit $R = PQ = (r(x,y))$. Nous voyons premièrement que les coefficients de la matrice R sont positifs comme somme de produits de coefficients eux-mêmes positifs.

De plus, en effectuant la somme sur la ligne x des coefficients de la matrice R nous obtenons, par définition du produit matriciel :

$$\begin{aligned}\sum_{y\in E} r(x,y) &= \sum_{y\in E}\Big(\sum_{z\in E} p(x,z)\,q(z,y)\Big)\\ &= \sum_{z\in E}\Big(\sum_{y\in E} p(x,z)\,q(z,y)\Big),\end{aligned}$$

en échangeant l'ordre des sommations, ce qui est possible en raison de la positivité

$$\begin{aligned}&= \sum_{z\in E} p(x,z)\Big(\sum_{y\in E} q(z,y)\Big), \text{ les coefficients } p(x,z) \text{ ne dépendant pas de } y\\ &= \sum_{z\in E} p(x,z) = 1, \ P \text{ et } Q \text{ étant deux matrices stochastiques.}\end{aligned}$$

La matrice produit R est donc une matrice stochastique.

Nous pouvons en déduire immédiatement par récurrence le corollaire suivant, où P^n est la matrice P à la puissance n :

Corollaire 5.7. *Soit P une matrice stochastique, alors pour tout $n \geq 0$, la matrice P^n est encore une matrice stochastique.*

Démonstration. La propriété est vraie à l'ordre 0 et 1, la matrice P étant une matrice stochastique. Nous la supposons vraie à l'ordre $n-1$, hypothèse de récurrence. Or nous avons par définition du produit matriciel à l'ordre n :

$$P^n = PP^{n-1},$$

ce qui indique que la matrice P^n est une matrice stochastique comme produit de deux matrices stochastiques : P et P^{n-1}. □

Le résultat précédent interviendra, lorsque que nous étudierons la dynamique des chaînes de Markov Section 6.

5.3.4 Graphe associé à une chaîne de Markov homogène

A partir de la matrice de transition d'une chaîne de Markov, nous construisons un graphe. Il donne les mêmes informations que la matrice mais, comme toute représentation graphique, a l'avantage d'être plus parlant.

Définition 5.8. Étant donnée une chaîne de Markov $(X_n)_{n\geq 0}$ définie sur un espace d'états E, le *graphe de la chaîne* est le graphe construit à partir de la matrice de transition ainsi : les sommets sont les états et les arêtes (orientées) représentent les transitions possibles d'un état vers un autre. Au dessus de chaque arête on écrit la probabilité de transition correspondante.

5.4 Exercices : Introduction aux chaînes de Markov

Exercice 5.4. [La marche aléatoire sur $\mathbb{Z}$]

On considère une suite de variables aléatoires $(Y_n)_{n\geq 1}$ à valeurs dans $\mathbb{Z}$ indépendantes et de même loi, et Y_0 une variable aléatoire à valeurs dans $\mathbb{Z}$, indépendante des $(Y_n)_{n\geq 1}$. Posons

$$\begin{cases} X_0 & = Y_0 \\ X_{n+1} & = X_n + Y_{n+1} = \sum_{i=0}^{n+1} Y_i, \quad \forall\, n \geq 0, \end{cases}$$

1. Montrer que $(X_n)_{n\geq 0}$ est une chaîne de Markov.

On suppose maintenant que Y_1 est à valeurs dans $\{-1, 1\}$ de loi,

$$\mathbb{P}[Y_1 = 1] = p, \quad \mathbb{P}[Y_1 = -1] = 1 - p,$$

où $0 < p < 1$.

2. Donner la matrice de transition et dessiner le graphe de cette chaîne de Markov.

Définition. Si $p = 1 - p = \frac{1}{2}$, $(X_n)_{n\geq 0}$ s'appelle la marche aléatoire simple *sur $\mathbb{Z}$.*

Solution de l'exercice 5.4.

1. La chaîne $(X_n)_{n\geq 0}$ est à valeurs dans $E = \mathbb{Z}$. Pour tout $x_0, \dots, x_{n+1} \in \mathbb{Z}$, on a :

$$\begin{aligned}
\mathbb{P}[X_{n+1} = x_{n+1} | X_n = x_n, \dots, X_0 = x_0] &= \frac{\mathbb{P}[X_{n+1} = x_{n+1}, X_n = x_n, \dots, X_0 = x_0]}{\mathbb{P}[X_n = x_n, \dots, X_0 = x_0]} \\
&= \frac{\mathbb{P}[Y_{n+1} = x_{n+1} - x_n, X_n = x_n, \dots, X_0 = x_0]}{\mathbb{P}[X_n = x_n, \dots, X_0 = x_0]} \\
&= \mathbb{P}[Y_{n+1} = x_{n+1} - x_n], \text{ (indépendance)} \\
&= \frac{\mathbb{P}[Y_{n+1} = x_{n+1} - x_n]\mathbb{P}[X_n = x_n]}{\mathbb{P}[X_n = x_n]} \\
&= \frac{\mathbb{P}[Y_{n+1} = x_{n+1} - x_n, X_n = x_n]}{\mathbb{P}[X_n = x_n]} \text{ (indépendance)} \\
&= \mathbb{P}[X_{n+1} = x_{n+1} | X_n = x_n].
\end{aligned}$$

   Ainsi $(X_n)_{n\geq 0}$ est une chaîne de Markov. De plus, on a montré que pour tout $(x, y) \in \mathbb{Z}^2$,

$$\mathbb{P}[X_{n+1} = y | X_n = x] = \mathbb{P}[Y_{n+1} = y - x] = \mathbb{P}[Y_1 = y - x],$$

   la suite $(Y_n)_{n\geq 0}$ étant identiquement distribuée. Cette quantité ne dépend pas de n et la chaîne est donc homogène.

2. La matrice de transition $P = (p(x, y))$ est de taille infinie. Les lignes et les colonnes sont indexées par $\mathbb{Z}$ et on a

$$\forall\, (x, y) \in \mathbb{Z}^2,\ p(x, y) = \mathbb{P}[X_{n+1} = y | X_n = x] = \mathbb{P}[Y_1 = y - x].$$

   Ainsi, pour tout $x \in \mathbb{Z}$, la x-ième ligne a pour seuls coefficients non nuls :

$$p(x, x - 1) = 1 - p \text{ et } p(x, x + 1) = p.$$

$$P = \begin{array}{c} \\ \vdots \\ -2 \\ -1 \\ 0 \\ 1 \\ 2 \\ \vdots \end{array} \begin{array}{c} \begin{array}{ccccccc} \cdots & -2 & -1 & 0 & 1 & 2 & \cdots \end{array} \\ \begin{pmatrix} \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \\ \cdots & 0 & p & 0 & 0 & 0 & \cdots \\ \cdots & 1-p & 0 & p & 0 & 0 & \cdots \\ \cdots & 0 & 1-p & 0 & p & 0 & \cdots \\ \cdots & 0 & 0 & 1-p & 0 & p & \cdots \\ \cdots & 0 & 0 & 0 & 1-p & 0 & \cdots \\ \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \end{array}.$$

Figure 5.1 – Matrice de transition de la marche aléatoire.

Le graphe associé à cette chaîne de Markov est,

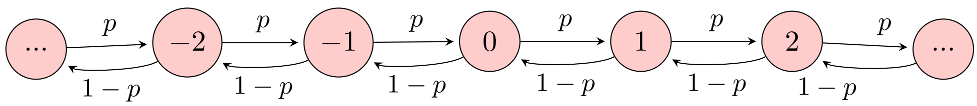


Figure 5.2 – Graphe de la marche aléatoire

Exercice 5.5. [Exemple générique]

On considère une suite de variables aléatoires $(Y_n)_{n\geq 1}$ à valeurs dans E indépendantes et de même loi, et Y_0 une variable aléatoire à valeurs dans E, indépendante des $(Y_n)_{n\geq 1}$. Posons,

$$\begin{cases} X_0 = Y_0 \\ X_{n+1} = \phi(X_n, Y_{n+1}), \end{cases}$$

où ϕ est une application de $E \times E$ à valeurs dans E. Montrer que $(X_n)_{n\geq 0}$ est une chaîne de Markov.

Solution de l'exercice 5.5. La chaîne $(X_n)_{n\geq 0}$ est à valeurs dans E. Pour tout $x_0, \dots, x_{n+1} \in E$, on a :

$$\begin{aligned} \mathbb{P}[X_{n+1} = x_{n+1} | X_n = x_n, \dots, X_0 = x_0] &= \frac{\mathbb{P}[X_{n+1} = x_{n+1}, X_n = x_n, \dots, X_0 = x_0]}{\mathbb{P}[X_n = x_n, \dots, X_0 = x_0]} \\ &= \frac{\mathbb{P}[\phi(x_n, Y_{n+1}) = x_{n+1}, X_n = x_n, \dots, X_0 = x_0]}{\mathbb{P}[X_n = x_n, \dots, X_0 = x_0]} \\ &= \mathbb{P}[\phi(x_n, Y_{n+1}) = x_{n+1}], \text{ par indépendance} \\ &= \mathbb{P}[X_{n+1} = x_{n+1} | X_n = x_n]. \end{aligned}$$

La suite $(X_n)_{n\geq 0}$ est donc une chaîne de Markov. Les variables aléatoires $(Y_n)_{n\geq 1}$ étant i.i.d. on a pour tout $(x, y) \in E^2$,

$$\mathbb{P}[X_{n+1} = y | X_n = x] = \mathbb{P}[\phi(x, Y_{n+1}) = y] = \mathbb{P}[\phi(x, Y_1) = y],$$

et la chaîne de Markov est homogène. L'équation ci-dessus permet également de trouver la matrice de transition.

Exercice 5.6. [Le fort carré / Marche aléatoire sur $\mathbb{Z}/4\mathbb{Z}$]

Considérons un fort carré pourvu d'un poste de garde à chaque coin. Une seule sentinelle est de garde ce jour-là. Elle a pour rôle de tromper l'ennemi et pour cela elle a l'ordre d'effectuer sa ronde de la manière aléatoire suivante.

Elle monte la garde 5 minutes dans un des quatre postes, puis elle tire à pile ou face une pièce équilibrée ; si elle tire pile, elle se rend au premier poste sur sa gauche et si elle tire face, elle se rend au premier poste sur sa droite. Elle y monte la garde 5 minutes et de nouveau elle tire à pile ou face le nouveau poste de garde et ainsi de suite.

Le parcours de la sentinelle peut être décrit à l'aide de la figure suivante :

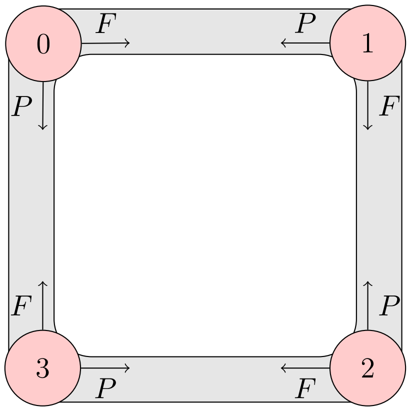


Figure 5.3 – fort carré

Appelons Y_0 le numéro du poste au départ, X_n le numéro du poste après n déplacements.

Montrer que $(X_n)_{n\geq 0}$ est une chaîne de Markov. Donner la matrice de transition et le graphe associés.

Solution de l'exercice 5.6. La suite $(X_n)_{n\geq 0}$ est une chaîne à quatre états, le numéro des postes de garde : $E = \{0,1,2,3\}$. Il est utile d'identifier E avec les entiers modulos 4 : $\mathbb{Z}/4\mathbb{Z}$. Soit $(Y_n)_{n\geq 1}$ une suite de variables aléatoires i.i.d. et indépendante de Y_0 de loi,

$$\mathbb{P}[Y_1 = -1] = \mathbb{P}[Y_1 = 1] = \frac{1}{2}.$$

La suite $(Y_n)_{n\geq 1}$ représente l'issue des jets de la pièce. Alors, la suite $(X_n)_{n\geq 0}$ peut s'écrire,

$$\begin{cases} X_0 = Y_0 \\ X_{n+1} \equiv [X_n + Y_{n+1}]\mathrm{mod}(4). \end{cases}$$

D'après l'Exercice 5.5 il s'agit d'une chaîne de Markov homogène.

La matrice de transition $P = (p(x,y))$ est de taille 4×4 et pour tout $(x,y)\in E^2$,

$$p(x,y) = \mathbb{P}[X_{n+1} = y | X_n = x] = \mathbb{P}[[x+Y_1]\mathrm{mod}\ 4 = y].$$

Il s'agit de la matrice de la marche aléatoire simple sur $\mathbb{Z}/4\mathbb{Z}$:

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

FIGURE 5.4 – Matrice de transition du fort carré

Le graphe associé à cette chaîne de Markov est,

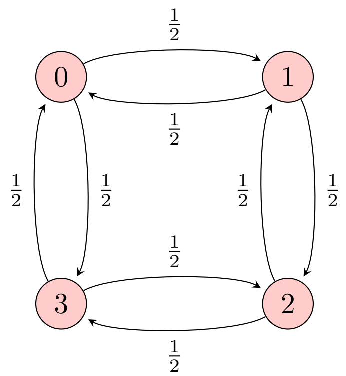


FIGURE 5.5 – Graphe du fort carré.

Exercice 5.7. [Transmission d'un bit informatique]

Un bit informatique (qu'il sera plus utile de modéliser par un -1/1 plutôt que 0/1) est transmis à travers n intermédiaires. On suppose que chaque intermédiaire transmet le message de façon correcte avec une probabilité p et le déforme en son contraire avec une probabilité $1-p$, où $0 < p < 1$. Le bit d'entrée est modélisé par une variable aléatoire Y_0 indépendante des intermédiaires.

Modéliser cette situation par un processus stochastique $(X_n)_{n\geq 0}$ et montrer qu'il s'agit d'une chaîne de Markov. Donner la matrice de transition et le graphe associé.

Solution de l'exercice 5.7. Soit $(X_n)_{n\geq 0}$ une suite de variables aléatoires telles que X_n représente le bit informatique avant la $(n+1)$-ième transmission et X_0 représente le bit initial. Cette suite est à valeurs dans $E = \{-1, 1\}$. Soit $(Y_n)_{n\geq 1}$ une suite de variables aléatoires telle que, pour tout $n \geq 1$,

$$Y_n = \begin{cases} 1 & \text{si la transmission de l'intermédiaire } n \text{ est correcte} \\ -1 & \text{si la transmission de l'intermédiaire } n \text{ est incorrecte.} \end{cases}$$

Les hypothèses nous disent que la suite $(Y_n)_{n\geq 1}$ est i.i.d. de loi,

$$\mathbb{P}[Y_n = 1] = p, \quad \mathbb{P}[Y_n = -1] = 1 - p,$$

et indépendante de Y_0. Alors on a, pour tout $n \geq 0$,

$$\begin{cases} X_0 = Y_0 \\ X_{n+1} = X_n Y_{n+1}. \end{cases}$$

D'après l'Exercice 5.5, il s'agit d'une chaîne de Markov homogène. La matrice de transition $P = (p(x,y))$ est de taille 2×2 et pour tout $(x,y) \in E^2$,

$$p(x,y) = \mathbb{P}[X_{n+1} = y | X_n = x] = \mathbb{P}[xY_1 = y] = \mathbb{P}\Big[Y_1 = \frac{y}{x}\Big].$$

Ainsi, on obtient

$$p(-1,-1) = p(1,1) = \mathbb{P}[Y_1 = 1] = p, \text{ et } p(-1,1) = p(1,-1) = \mathbb{P}[Y_1 = -1] = 1-p,$$

autrement dit,

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

Exercice 5.8. [Urne d'Ehrenfest]

L'exemple suivant, proposé par les époux Ehrenfest, sert à la modélisation de la diffusion de molécules gazeuses entre deux compartiments séparés par une cloison poreuse. Deux urnes A et B contiennent au total N billes. Les urnes correspondent aux compartiments et les billes aux molécules. À chaque instant il y a tirage au hasard d'une bille, qui est alors changée d'urne. La distribution des billes au départ suit une loi Y_0 indépendante des tirages ultérieurs.

Modéliser cette expérience par un processus stochastique. On admettra qu'il s'agit d'une chaîne de Markov. Donner sa matrice de transition et son graphe.

Solution de l'exercice 5.8. Soit $(X_n)_{n\geq 0}$ le processus défini par $X_0 = Y_0$ et tel que, pour tout $n \geq 1$, X_n représente le nombre de billes dans l'urne A à l'instant n. On admet qu'il s'agit d'une chaîne de Markov (ce qui paraît naturel au vu des conditions de l'expérience). Cette chaîne est à valeurs dans $E = \{0, \dots, N\}$. Les coefficients de la matrice de transition $P = (p(x,y))$ sont, pour tout $(x,y) \in E^2$,

$$p(x,y) = \mathbb{P}[X_{n+1} = y | X_n = x] = \begin{cases} \frac{x}{N} & \text{si } y = x-1 \text{ et } x \in \{1, \dots, N\} \\ \frac{N-x}{N} & \text{si } y = x+1 \text{ et } x \in \{0, \dots, N-1\}. \end{cases}$$

Exercice 5.9. [Ruine du joueur]

Deux joueurs A et B disposent d'une fortune initiale de a et b euros respectivement, avec $a, b > 0$. Ils jouent au jeu de hasard suivant : la mise est de 1 euro par partie ; les parties sont indépendantes et à chacune d'entre elle le joueur A a une probabilité p de gagner (et donc $1-p$ de perdre), avec $0 < p < 1$. Le jeu se déroule jusqu'à la ruine d'un des deux joueurs.

Modéliser la fortune du premier joueur par un processus stochastique. Montrer qu'il s'agit d'une chaîne de Markov. Donner sa matrice de transition et le graphe associé.

Solution de l'exercice 5.9. Soit $(X_n)_{n\geq 1}$ le processus modélisant la fortune du joueur A. Cette suite est à valeurs dans $E = \{0, \dots, N\}$, où $N = a + b$. Soit $(Y_n)_{n\geq 1}$ la suite de variables aléatoires à valeurs dans $\{-1, 1\}$ telle que Y_n représente le gain/perte du joueur A à l'instant n. D'après les conditions du jeu, cette suite est i.i.d. de loi

$$\mathbb{P}[Y_n = 1] = p, \quad \mathbb{P}[Y_n = -1] = 1 - p,$$

et indépendante de l'état initial $X_0 = a$. Alors on a, pour tout $n \geq 0$,

$$X_{n+1} = \begin{cases} X_n + Y_{n+1} & \text{si } X_n \notin \{0, N\} \\ X_n & \text{si } X_n \in \{0, N\}. \end{cases}$$

D'après l'Exercice 5.5, il s'agit d'une chaîne de Markov homogène. La matrice de transition $P = (p(x, y))$ est de taille $(N + 1) \times (N + 1)$ et pour tout $(x, y) \in E^2$,

$$p(x, y) = \mathbb{P}[X_{n+1} = y | X_n = x] = \begin{cases} p & \text{si } y = x + 1 \text{ et } x \in \{1, \dots, N - 1\} \\ 1 - p & \text{si } y = x - 1 \text{ et } x \in \{1, \dots, N - 1\} \\ 1 & \text{si } x = y \text{ et } x \in \{0, N\}. \end{cases}$$

Le graphe associé à cette chaîne de Markov est,

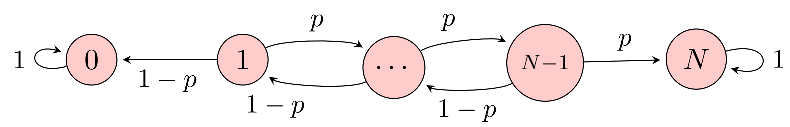


Figure 5.6 – Graphe de la ruine du joueur.

Exercice 5.10. [Le plagiste et ses voiliers]

Un plagiste possède quatre voiliers qu'il peut louer chaque jour pour la journée. Nous supposons que

- lorsqu'ils sortent, chaque voilier a, indépendamment les uns des autres, la probabilité p de subir une avarie au cours de cette sortie ;
- le plagiste ne peut faire les réparations nécessaires que le soir ;
- le plagiste ne peut réparer le soir qu'un et seul voilier, qui est en état de marche le matin ;
- pour des raisons de sécurité, il ne laisse sortir ses bateaux que s'il en a au moins deux de disponibles ;
- en cette période de l'année, il a assez de demandes pour sortir tous ses bateaux en état de marche.

Notons X_n le nombre de voiliers en état de marche le $(n+1)$-ième matin – attention au décalage de l'indice temps : le premier matin étant l'instant 0 –.

Justifier que $(X_n)_{n\geq 0}$ est une chaîne de Markov. Donner sa matrice de transition et son graphe.

Solution de l'exercice 5.10. Remarquons qu'il y a de 1 à 4 voiliers en état de marche le matin, car, si tous étaient en panne le soir, il y en a un et un seul qui est réparé le lendemain matin. L'espace des états est donc $E = \{1, 2, 3, 4\}$.

S'il n' y a qu'un seul voilier en état de marche, il ne sort pas et donc nous nous retrouvons avec deux voiliers en état de marche le lendemain.

Si les voiliers sortent, le nombre de voiliers en état de marche le lendemain ne dépendra que du nombre de voiliers en marche ce matin là – donc de l'état présent – et du nombre de voiliers tombés en panne au cours de la journée, panne indépendante du passé de chacun des voiliers considérés. Par suite le déroulement futur du processus ne dépend que du présent. Cette suite est bien une chaîne de Markov.

Étudions maintenant sa matrice de transition $P = (p(x,y))$ qui est de taille 4×4. Grâce à la première remarque, nous voyons que la première ligne est constituée d'un 1 en deuxième position – probabilité de passe de 1 à 2 voiliers – et de 0 partout ailleurs.

Lorsque x voiliers sortent, avec $x \in \{2,3,4\}$, chacun peut tomber en panne indépendamment des autres et avec la même probabilité p, par suite la loi du nombre de voiliers tombant en panne est la loi binomiale de paramètres x et p. Nous retrouvons ainsi sur les lignes 2 et 3, la loi binomiale respectivement de paramètres $(2,p)$, $(3,p)$. Pour la ligne 4, la même chose est vraie pour les coefficients $p(4,1)$, $p(4,2)$, $p(4,3)$. Par contre, pour le coefficient $p(4,4)$, sachant que le matin il y a 4 voiliers sortis, il y a deux manières d'avoir de nouveau 4 voiliers en état de marche le lendemain matin : soit 1 voilier tombe en panne et est réparé le soir, ce qui arrive avec probabilité $4p(1-p)^3$, soit aucun ne tombe en panne, ce qui arrive avec probabilité $(1-p)^4$.

$$
P = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array}
\begin{array}{c}
\begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\
\begin{pmatrix}
0 & 1 & 0 & 0 \\
p^2 & 2p(1-p) & (1-p)^2 & 0 \\
p^3 & 3p^2(1-p) & 3p(1-p)^2 & (1-p)^3 \\
p^4 & 4p^3(1-p) & 6p^2(1-p)^2 & 4p(1-p)^3 + (1-p)^4
\end{pmatrix}
\end{array}.
$$

FIGURE 5.7 – Matrice de transition du plagiste et ses voiliers.

Le graphe associé à cette chaîne de Markov est,

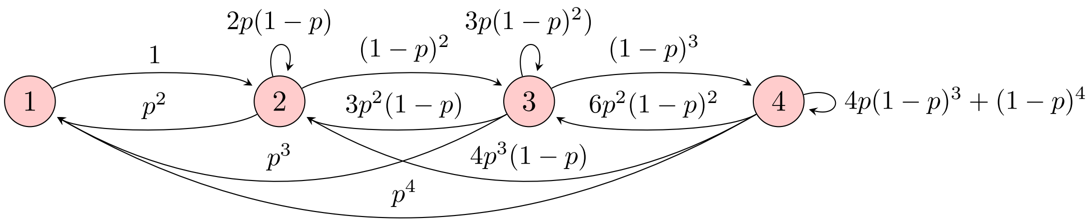


FIGURE 5.8 – Graphe du plagiste et ses voiliers.

Chapitre 6

Dynamique d'une chaîne de Markov

Dans ce chapitre, nous donnons une caractérisation d'une chaîne de Markov et étudions sont évolution dans le temps.

6.1 Caractérisation d'une chaîne de Markov

Le but de cette section est de caractériser une chaîne de Markov $(X_n)_{n\geq 0}$ à partir de sa matrice de transition et de la loi initiale.

Définition 6.1. Soit une chaîne de Markov $(X_n)_{n\geq 0}$ à valeurs dans E, nous appelons *loi initiale* la loi μ_0 de la variable aléatoire X_0 :

$$\forall x \in E, \quad \mu_0(x) = \mathbb{P}[X_0 = x].$$

Le théorème suivant montre que la donnée de la loi initiale μ_0 et de la matrice de transition P caractérise une chaîne de Markov.

Théorème 6.2. *Un processus $(X_n)_{n\geq 0}$ à valeurs dans E est une chaîne de Markov de loi initiale μ_0 et de matrice de transition P si et seulement si, pour tout $n \geq 0$ et pour tout $(x_0, x_1, \cdots, x_n) \in E^{n+1}$:*

$$\mathbb{P}[X_0 = x_0, \ldots, X_{n-1} = x_{n-1}, X_n = x_n] = \mu_0(x_0)p(x_0, x_1)\ldots p(x_{n-1}, x_n). \tag{6.1}$$

Démonstration. Supposons que $(X_n)_{n\geq 0}$ est une chaîne de Markov et montrons l'égalité (6.1).

En utilisant la formule,

$$\mathbb{P}[A_0 \cap A_1 \cap \cdots \cap A_{n-1} \cap A_n] = \mathbb{P}[A_0]\mathbb{P}[A_1|A_0]\cdots\mathbb{P}[A_n|A_0 \cap \cdots \cap A_{n-1}]$$

et la propriété de Markov. Nous avons :

$$\begin{aligned}
&\mathbb{P}[X_0 = x_0, \ldots, X_{n-1} = x_{n-1}, X_n = x_n] = \\
&\qquad = \mathbb{P}[X_0 = x_0]\mathbb{P}[X_1 = x_1|X_0 = x_0]\cdots\mathbb{P}[X_n = x_n|X_0 = x_0, \ldots, X_{n-1} = x_{n-1}] \\
&\qquad = \mathbb{P}[X_0 = x_0]\mathbb{P}[X_1 = x_1|X_0 = x_0]\cdots\mathbb{P}[X_n = x_n|X_{n-1} = x_{n-1}] \\
&\qquad = \mu_0(x_0)p(x_0, x_1)\ldots p(x_{n-1}, x_n);
\end{aligned}$$

d'où l'égalité (6.1).

Supposons maintenant que l'équation (6.1) est vraie. En l'appliquant pour $n = 0$ on obtient, pour tout $x_0 \in E$, $\mathbb{P}[X_0 = x_0] = \mu_0(x_0)$. Pour tout $n \geq 0$, nous obtenons en utilisant la définition de probabilité conditionnelle,

$$\begin{aligned}
\mathbb{P}[X_{n+1} = x_{n+1} | X_n = x_n, \dots, X_0 = x_0] &= \frac{\mathbb{P}[X_{n+1} = x_{n+1}, X_n = x_n, \dots, X_0 = x_0]}{\mathbb{P}[X_n = x_n, \dots, X_0 = x_0]} \\
&\overset{(6.1)}{=} \frac{\mu_0(x_0)p(x_0, x_1)\dots p(x_{n-1}, x_n)p(x_n, x_{n+1})}{\mu_0(x_0)p(x_0, x_1)\dots p(x_{n-1}, x_n)} \\
&= p(x_n, x_{n+1}). \qquad (6.2)
\end{aligned}$$

D'autre part, d'après la formule des probabilités totales et la définition des probabilités conditionnelles,

$$\begin{aligned}
\mathbb{P}[X_{n+1} = x_{n+1} | X_n = x_n] &= \frac{\sum_{x_0, \cdots, x_{n-1} \in E} \mathbb{P}[X_{n+1} = x_{n+1}, X_n = x_n, \cdots, X_0 = x_0]}{\sum_{x_0, \cdots, x_{n-1} \in E} \mathbb{P}[X_n = x_n, \cdots, X_0 = x_0]} \\
&= \frac{\sum_{x_0, \cdots, x_{n-1} \in E} \mathbb{P}[X_{n+1} = x_{n+1} | X_n = x_n, \cdots, X_0 = x_0]\mathbb{P}[X_n = x_n, \cdots, X_0 = x_0]}{\sum_{x_0, \cdots, x_{n-1} \in E} \mathbb{P}[X_n = x_n, \cdots, X_0 = x_0]} \\
&= \frac{\sum_{x_0, \cdots, x_{n-1} \in E} p(x_n, x_{n+1})\mathbb{P}[X_n = x_n, \cdots, X_0 = x_0]}{\sum_{x_0, \cdots, x_{n-1} \in E} \mathbb{P}[X_n = x_n, \cdots, X_0 = x_0]}, \text{ d'après (6.2),} \\
&= p(x_n, x_{n+1}).
\end{aligned}$$

Ainsi,

$$\mathbb{P}[X_{n+1} = x_{n+1} | X_n = x_n, \dots, X_0 = x_0] = \mathbb{P}[X_{n+1} = x_{n+1} | X_n = x_n] = p(x_n, x_{n+1}).$$

La propriété de Markov est vérifiée et la matrice de transition est $P = (p(x, y))$. $\square$

Du Théorème 6.2, nous déduisons le corollaire suivant.

Corollaire 6.3. *La loi d'une chaîne de Markov est invariante par translation dans le temps. Autrement dit, pour tout* $(x_0, \dots, x_{n+m}) \in E^{n+m+1}$ *tels que* $\mathbb{P}[X_0 = x_0, \dots, X_n = x_n] > 0$,

$$\begin{aligned}
\mathbb{P}[X_{n+m} = x_{n+m}, \dots, X_{n+1} = x_{n+1} | X_n = x_n, \dots, X_0 = x_0] &= \\
&= \mathbb{P}[X_m = x_{n+m}, \dots, X_1 = x_{n+1} | X_0 = x_n] \\
&= p(x_n, x_{n+1}) \dots p(x_{n+m-1}, x_{n+m}).
\end{aligned}$$

Démonstration. D'après la définition des probabilités conditionnelles et le Théorème 6.2 on a

$$\begin{aligned}\mathbb{P}[X_{n+m} = x_{n+m}, \ldots, &X_{n+1} = x_{n+1} | X_n = x_n, \ldots, X_0 = x_0] = \\ &= \frac{\mathbb{P}[X_{n+m} = x_{n+m}, \ldots, X_0 = x_0]}{\mathbb{P}[X_n = x_n, \ldots, X_0 = x_0]} \\ &= \frac{\mu_0(x_0)p(x_0, x_1) \ldots p(x_{n+m-1}, x_{n+m})}{\mu_0(x_0)p(x_0, x_1) \ldots p(x_{n-1}, x_n)}, \text{ d'après le Théorème 6.2} \\ &= p(x_n, x_{n+1}) \ldots p(x_{n+m-1}, x_{n+m}) \\ &= \frac{\mathbb{P}[X_m = x_{n+m}, \cdots, X_0 = x_n]}{\mathbb{P}[X_0 = x_n]}, \text{ d'après le Théorème 6.2} \\ &= \mathbb{P}[X_m = x_{n+m}, \ldots, X_1 = x_{n+1} | X_0 = x_n]. \end{aligned}$$

□

La question qui se pose à cette étape, est celle de l'existence d'une chaîne de Markov, c'est à dire de l'existence du processus $(X_n)_{n\geq 0}$ dont les marginales sont données par (6.1). Le théorème suivant donne la réponse. Il se démontre en utilisant le théorème d'extension de Kolmogorov.

Théorème 6.4. *Étant donné une probabilité μ_0 sur un ensemble discret E et une matrice stochastique P sur E, nous pouvons leur associer une chaîne de Markov $(X_n)_{n\geq 0}$ dont les marginales sont données par* (6.1).

6.2 Transitions d'ordre n et loi à l'instant n

Nous pouvons maintenant étudier les transitions d'ordre n d'une chaîne de Markov et en déduire sa loi à l'instant n. Auparavant rappelons quelques faits sur le produit matriciel et introduisons quelques notations.

Fonctions, vecteurs et matrices. La donnée d'une fonction f définie sur E est la donnée de $(f(x))_{x\in E}$. Si l'on écrit $E = \{x_1, x_2, \ldots\}$, on peut identifier f avec un vecteur colonne, noté f également :

$$f = \begin{pmatrix} f(x_1) \\ f(x_2) \\ \vdots \end{pmatrix}$$

Soit maintenant $A = (A(x, y))$ une matrice dont les lignes et les colonnes sont indexées par E. Le produit matriciel $f^t A$ définit un vecteur ligne dont la coordonnée y est :

$$(f^t A)(y) = \sum_{x\in E} f(x)A(x, y).$$

D'après l'identification ci-dessus, le vecteur $f^t A$ est aussi une fonction $((f^t A)(y))_{y\in E}$ définie sur E. De manière analogue, le produit matriciel Af définit un vecteur colonne dont la coordonnée x est :

$$(Af)(x) = \sum_{y\in E} A(x, y)f(y).$$

Le vecteur Af est également identifié à une fonction $((Af)(x))_{x\in E}$ définie sur E.

Notations. Soit $(X_n)_{n\geq 0}$ une chaîne de Markov définie sur E.

- Pour tout $n \geq 0$, notons μ_n la loi de $(X_n)_{n\geq 0}$ à l'instant n :

$$\forall\, x \in E, \quad \mu_n(x) = \mathbb{P}[X_n = x].$$

- Pour tout $n \geq 0$, notons $P^{(n)} = (P^{(n)}(x,y))$ la matrice des transitions d'ordre n :

$$\forall\, (x,y) \in E^2, \quad P^{(n)}(x,y) = \mathbb{P}[X_n = y | X_0 = x].$$

Proposition 6.5. *Soit $(X_n)_{n\geq 0}$ une chaîne de Markov de matrice de transition P.*
- *Pour tout $k \geq 0$, pour tout $(x,y) \in E^2$,*

$$\mathbb{P}[X_{n+k} = y | X_k = x] = P^{(n)}(x,y) = P^n(x,y).$$

- *La loi à l'instant n est donnée par*

$$\mu_n^t = \mu_0^t P^n, \quad \textit{autrement dit,} \quad \forall\, y \in E,\ \mathbb{P}[X_n = y] = \sum_{x\in E} \mathbb{P}[X_0 = x] P^n(x,y).$$

Démonstration. Démontrons le premier point. Il suffit de montrer que pour tout $k \geq 0$, pour tout $(x,y) \in E^2$,

$$\mathbb{P}[X_{n+k} = y | X_k = x] = P^n(x,y).$$

D'après la formule des probabilités totales,

$$\begin{aligned}
\mathbb{P}[X_{n+k} = y | X_k = x] &= \sum_{x_{n+k-1},\dots,x_{k+1}\in E} \mathbb{P}[X_{n+k} = y, X_{n+k-1} = x_{n+k-1}, \dots, X_{k+1} = x_{k+1} | X_k = x] \\
&= \sum_{x_{k+1},\dots,x_{n+k-1}\in E} p(x, x_{k+1}) p(x_{k+1,k+2}) \dots p(x_{n+k-1}, y), \text{ d'après le Cor. 6.3} \\
&= P^n(x,y), \text{ par définition du produit matriciel.}
\end{aligned}$$

Pour le deuxième point, en utilisant à nouveau la formule des probabilités totales, pour tout $y \in E$,

$$\begin{aligned}
\mathbb{P}[X_n = y] &= \sum_{x\in E} \mathbb{P}[X_n = y | X_0 = x] \mathbb{P}[X_0 = x] \\
&= \sum_{x\in E} \mu_0(x) P^n(x,y),
\end{aligned}$$

en utilisant les notations introduites et le premier point. □

Nous pouvons en déduire les équations importantes suivantes.

Corollaire 6.6 (Équations de Chapman-Kolmogorov)**.** *Soit $(X_n)_{n\geq 0}$ une chaîne de Markov de matrice de transition P. Alors les matrices de transitions d'ordre $n+m$ et n,m satisfont à la relation suivante :*

$$P^{(n+m)} = P^{(n)} P^{(m)}. \tag{6.3}$$

En explicitant les notations, ceci s'écrit aussi :

$$\forall\, (x,y) \in E^2, \quad \mathbb{P}[X_{n+m} = y | X_0 = x] = \sum_{z\in E} \mathbb{P}[X_n = z | X_0 = x] \mathbb{P}[X_m = y | X_0 = z].$$

Démonstration. L'égalité (6.3) découle immédiatement de la proposition précédente et de la définition du produit matriciel. □

6.3 Exercices : dynamique d'une chaîne de Markov

Exercice 6.1. [Transmission d'un bit informatique, suite] On reprend l'Exercice 5.7. Calculer la probabilité que le bit avant la transmission par le $(n+1)$-ième intermédiaire soit égal au bit initial. Que se passe-t-il lorsque $n \to \infty$?

Solution de l'exercice 6.1. On cherche $\mathbb{P}[X_n = X_0]$. On a,

$$\begin{aligned}\mathbb{P}[X_n = X_0] &= \mathbb{P}[X_n = 1, X_0 = 1] + \mathbb{P}[X_n = -1, X_0 = -1] \\ &= \mathbb{P}[X_n = 1 | X_0 = 1]\mathbb{P}[X_0 = 1] + \mathbb{P}[X_n = -1 | X_0 = -1]\mathbb{P}[X_0 = -1].\end{aligned}$$

D'après la Proposition 6.5, ceci est donné par :

$$\mathbb{P}[X_n = X_0] = P^n(1,1)\mathbb{P}[X_0 = 1] + P^n(-1,-1)\mathbb{P}[X_0 = -1].$$

Nous devons calculer la puissance n-ième de la matrice de transition P, où

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

Cette matrice est symétrique donc diagonalisable dans une base orthonormée et ses valeurs propres sont réelles. La somme des lignes étant constante égale à 1, on en déduit que 1 est valeur propre et $(1,1)^t$ est vecteur propre associé. La trace étant égale à $2p$, on déduit que la deuxième valeur propre est $2p-1$. Le vecteur propre associé étant orthogonal au premier, on déduit que $(-1,1)^t$ est vecteur propre associé à la valeur propre $2p-1$. Ainsi,

$$P = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 2p-1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix},$$

et

$$P^n = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & (2p-1)^n \end{pmatrix}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1+(2p-1)^n & 1-(2p-1)^n \\ 1-(2p-1)^n & 1+(2p-1)^n \end{pmatrix}.$$

On conclut que,

$$\mathbb{P}[X_n = X_0] = \frac{1+(2p-1)^n}{2}(\mathbb{P}[X_0 = 1] + \mathbb{P}[X_0 = -1]) = \frac{1+(2p-1)^n}{2}.$$

Comme $0 < p < 1$, $|2p-1| < 1$ et par suite, $\lim_{n\to\infty} \mathbb{P}[X_n = X_0] = \frac{1}{2}$. Asymptotiquement, il y a une chance sur deux que le bit soit transmis sans erreur, ceci indépendamment de la probabilité p.

Remarque. (*En rapport avec le chapitre 8*). La chaîne est irréductible, récurrente positive et apériodique. Donc convergence de la loi à l'instant n vers la mesure stationnaire.

Exercice 6.2. [Dépenses énergétiques, d'après le livre "Exercices de probabilité" Cottrel & al.] On dispose dans une maison individuelle de deux systèmes de chauffage, l'un de base l'autre d'appoint. On dira que l'on est dans l'état 1 si seul le chauffage de base fonctionne, dans l'état 2 si les deux systèmes fonctionnent.

Si un jour on est dans l'état 1, on estime qu'on y reste le lendemain avec probabilité $\frac{1}{2}$; par contre si l'on est dans l'état 2, le lendemain la maison est chaude et l'on passe à l'état 1 avec probabilité $\frac{3}{4}$.

Soit X_n l'état du système au jour numéro n ; on admet que $(X_n)_{n\geq 0}$ est une chaîne de Markov.

1. Déterminer la matrice de transition et son graphe.
2. On pose $p_n = \mathbb{P}[X_n = 1]$. Déterminer une relation de récurrence entre p_n et p_{n+1}, puis exprimer p_n en fonction de p_0. Que vaut $\lim_{n\to\infty} p_n$?
3. On suppose que l'on est dans l'état 1 un dimanche, trouver la probabilité d'être dans le même état le dimanche suivant.
4. Montrer que si un jour on se retrouve dans l'état 1 avec probabilité $\frac{3}{5}$ alors il en est de même les jours qui suivent.

Solution de l'exercice 6.2.

1. La chaîne de Markov $(X_n)_{n\geq 0}$ est à valeurs dans $E = \{1, 2\}$. Sa matrice de transition $P = (p(x,y))$ est de taille 2×2 et d'après les hypothèses de l'énoncé, elle est donnée par :

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

2. D'après la formule des probabilités totales et conditionnelles,

$$\begin{aligned} p_{n+1} = \mathbb{P}[X_{n+1} = 1] &= \mathbb{P}[X_{n+1} = 1, X_n = 1] + \mathbb{P}[X_{n+1} = 1, X_n = 2] \\ &= \mathbb{P}[X_{n+1} = 1 | X_n = 1]\mathbb{P}[X_n = 1] + \mathbb{P}[X_{n+1} = 1 | X_n = 2]\mathbb{P}[X_n = 2] \\ &= p(1,1)p_n + p(2,1)(1-p_n), \text{ car } \mathbb{P}[X_n = 2] = 1 - \mathbb{P}[X_n = 1] \\ &= \frac{3}{4} - \frac{1}{4}p_n. \end{aligned}$$

On peut se ramener à une suite géométrique en soustrayant de l'équation ci-dessus la solution de l'équation $x = \frac{3}{4} - \frac{1}{4}x$, soit $x = \frac{3}{5}$.

$$p_n - \frac{3}{5} = -\frac{1}{4}\Big(p_{n-1} - \frac{3}{5}\Big) = \Big(-\frac{1}{4}\Big)^n\Big(p_0 - \frac{3}{5}\Big),$$

d'où $p_n = \frac{3}{5} + \big(-\frac{1}{4}\big)^n\big(p_0 - \frac{3}{5}\big)$, et

$$\lim_{n\to\infty} p_n = \frac{3}{5}.$$

3. On cherche $\mathbb{P}[X_{n+7} = 1 | X_n = 1] = \mathbb{P}[X_7 = 1 | X_0 = 1]$. Supposer que l'on est dans l'état 1 le dimanche revient à supposer que $\mathbb{P}[X_0 = 1] = 1 = p_0$ (et donc $\mathbb{P}[X_0 = 2] = 0$). Donc,

$$\begin{aligned} p_7 &= \mathbb{P}[X_7 = 1] = \mathbb{P}[X_7 = 1 | X_0 = 1]\mathbb{P}[X_0 = 1] + \mathbb{P}[X_7 = 1 | X_0 = 2]\mathbb{P}[X_0 = 2] \\ &= \mathbb{P}[X_7 = 1 | X_0 = 1]. \end{aligned}$$

On cherche donc p_7. D'après la question précédente,

$$p_7 = \frac{3}{5} + \Big(-\frac{1}{4}\Big)^7\Big(\frac{2}{5}\Big) \simeq \frac{3}{5}.$$

4. La relation de récurrence établie sur (p_n) nous dit que si $p_{n-1} = \frac{3}{5}$ alors $p_n = \frac{3}{5}$.

Remarque. (*En rapport avec le chapitre 8*). La chaîne est irréductible, récurrente positive et apériodique. Donc convergence de la loi à l'instant n vers la mesure stationnaire. Si on part de la mesure stationnaire, on y reste.

Exercice 6.3. [Casino / Marche aléatoire sur $\mathbb{Z}/3\mathbb{Z}$] Un joueur fréquente 3 casinos numérotés 1, 2 et 3. Chaque jour il choisit l'un des deux casinos où il n'est pas allé la veille suivant une même probabilité $\frac{1}{2}$. Le premier jour, jour 0, il choisit l'un des trois casinos selon une loi μ_0 sur $E = \{1, 2, 3\}$. On note X_n la variable aléatoire égale au numéro du casino fréquenté par le joueur le jour n. On admet que $(X_n)_{n\geq 0}$ est une chaîne de Markov.

1. Déterminer la matrice de transition P et le graphe de la chaîne de Markov.
2. Calculer les puissance P^n de la matrice P, puis $\lim_{n\to\infty} P^n$.
3. Calculer $\lim_{n\to\infty} \mathbb{P}[X_n = x]$ pour $x \in \{1, 2, 3\}$.

Solution de l'exercice 6.3.

1. La matrice de transition $P = (p(x, y))$ est de taille 3×3 et d'après l'énoncé on a,

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

2. Posons

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

Alors pour tout $k \geq 1$, $M^k = 3^{k-1}M$. D'autre part, remarquons que $P = \frac{1}{2}(M - I_3)$, ainsi,

$$\begin{aligned} P^n &= \frac{1}{2^n}\Big(\sum_{k=0}^{n} \binom{n}{k}(-1)^{n-k}M^k\Big) \\ &= \frac{1}{2^n}\Big((-1)^n I_3 + \frac{1}{3}\Big(\sum_{k=1}^{n} \binom{n}{k}(-1)^{n-k}3^k\Big)M\Big) \\ &= \frac{1}{2^n}\Big((-1)^n I_3 + \frac{1}{3}(2^n - (-1)^n)M\Big). \end{aligned}$$

D'où, $P^n(x, y) = \begin{cases} \frac{1}{3\cdot 2^n}(2^n - (-1)^n) & \text{si } x \neq y \\ \frac{1}{3\cdot 2^n}(2^n + 2(-1)^n) & \text{si } x = y. \end{cases}$

On en déduit,

$$\lim_{n\to\infty} P^n = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}.$$

3. D'après le cours,
$$\mathbb{P}[X_n = y] = \sum_{x\in E} \mu_0(x) P^n(x,y).$$
Donc, $\lim_{n\to\infty} \mathbb{P}[X_n = y] = \frac{1}{3}\sum_{x\in E}\mu_0(x) = \frac{1}{3}$.

Remarque. (*En rapport avec le chapitre 8*). La chaîne est irréductible, récurrente positive et apériodique. Donc convergence de la loi à l'instant n vers la mesure stationnaire.

Exercice 6.4. [Marche aléatoire sur $\mathbb{Z}$, suite]

On reprend l'Exercice 5.4 sur la marche aléatoire.

1. Calculer la probabilité $P^{(n)}(x,x)$ que le marcheur revienne en un point x au bout de n pas, sachant qu'il est parti de ce même point x.
2. Donner une formule exacte pour $\sum_{n\geq 0} P^{(n)}(x,x)$ lorsque $p\in(0,1)$ et $p\neq \frac{1}{2}$.
3. En utilisant la formule de Stirling, donner un équivalent ($n\sim\infty$) de $P^{(n)}(x,x)$ lorsque $p=\frac{1}{2}$. En déduire que $\sum_{n\geq 0} P^{(n)}(x,x) = \infty$.

Solution de l'exercice 6.4.

1. Remarquons que si n est impair, on ne pourra pas revenir au point de départ, donc $P^{(n)}(x,x) = 0$ si n est impair. Supposons $n = 2m$ pair. Pour revenir en x après $2m$ pas, il faut et il suffit de faire m pas vers la droite et m pas vers la gauche ; il y a $\binom{2m}{m}$ telles marches. La probabilité de chacune est égale à $p^m(1-p)^m$ car les pas sont indépendants. Ainsi,
$$P^{(2m)}(x,x) = \binom{2m}{m}[p(1-p)]^m.$$
2. On a la formule $\sum_{m=0}^{\infty}\binom{2m}{m}\left(\frac{x}{4}\right)^m = (1-x)^{-\frac{1}{2}}$, pour tout x, $|x|<1$. (Généralisation de la formule du binôme pour les puissances rationnelles).
Par ailleurs, $0 < 4p(1-p) < 1$ tant que $p\in(0,1)$ et $p\neq\frac{1}{2}$ (variations d'un polynôme de degré 2). Ainsi,
$$\sum_{n=0}^{\infty} P^{(n)}(x,x) = \sum_{m=0}^{\infty} P^{(2m)}(x,x) = \Big(1-4p(1-p)\Big)^{-\frac{1}{2}} = \frac{1}{|2p-1|}.$$
3. D'après la formule de Stirling, $n! = n^n e^{-n}\sqrt{2\pi n}(1+o(1))$. Donc,
$$\binom{2m}{m} = \frac{(2m)^{2m}e^{-2m}\sqrt{2\pi 2m}}{m^{2m}e^{-2m}Z\pi m}(1+o(1)) = \frac{2^{2m}}{\sqrt{\pi m}}(1+o(1)).$$
Ainsi, lorsque $p=\frac{1}{2}$, et $m\to\infty$ $P^{(2m)}(x,x)\sim\frac{1}{\sqrt{\pi m}}$. La série de terme général $\frac{1}{\sqrt{m}}$ étant divergente, on en déduit que
$$\sum_{n=0}^{\infty} P^{(n)}(x,x) = \sum_{m=0}^{\infty} P^{(2m)}(x,x) = \infty.$$

Chapitre 7

Classification des états

L'espace des états étant souvent de grande taille, il est utile de le partitionner en des sous-ensembles dont les éléments ont les mêmes comportements au regard de propriétés importantes. Pour ce faire, nous introduisons une relation d'équivalence sur les états.

7.1 Classes de communication

On considère une chaîne de Markov $(X_n)_{n\geq 0}$ définie sur un espace d'états E, de matrice de transition P.

Définition 7.1.

- Soient deux états x et y. Nous dirons que x *conduit à* y ou que y *est accessible depuis* x, noté $x \to y$, si :
$$\exists\, n \in \mathbb{N}, \quad P^{(n)}(x,y) = \mathbb{P}[X_n = y | X_0 = x] > 0. \tag{7.1}$$
Cette relation signifie que partant de x nous avons une probabilité non nulle d'atteindre y à un certain temps n.
- Nous dirons que x *communique avec* y, noté $x \leftrightarrow y$, si chacun des états x, y est accessible depuis l'autre, c'est-à-dire $x \to y$ et $y \to x$.

Proposition 7.2. *La relation de communication $\leftrightarrow$ est une relation d'équivalence.*

Démonstration.

- Réflexivité : $x \leftrightarrow x$. Immédiat en observant que
$$P^{(0)}(x,x) = \mathbb{P}[X_0 = x | X_0 = x] = 1.$$
- Symétrie : $x \leftrightarrow y \;\Rightarrow\; y \leftrightarrow x$. Évident par définition.
- Transitivité : $x \leftrightarrow y$ et $y \leftrightarrow z \Rightarrow x \leftrightarrow z$. Pour cela il suffit de montrer la transitivité de la relation d'accessibilité. Supposons que $x \to y$ et $y \to z$. Alors il existe $m, n \in \mathbb{N}$ tels quel $P^{(n)}(x,y) > 0$ $P^{(m)}(y,z) > 0$. Or d'après les équations de Chapman-Kolmogorov, on a
$$P^{(n+m)}(x,z) = \sum_{y' \in E} P^{(n)}(x,y')P^{(m)}(y',z) \geq P^{(n)}(x,y)P^{(m)}(y,z) > 0.$$
La première inégalité provient du fait que tous les termes de la somme sont positifs ou nuls. En conséquence $x \to z$. $\square$

Définition 7.3.

- Les états E de la chaîne peuvent être partitionnés en classes d'équivalence appelées *classes irréductibles*. Si E est réduit à une seule classe, la chaîne de Markov est dite *irréductible*.
- La relation d'accessibilité s'étend aux classes : une classe d'équivalence C' est *accessible* depuis une classe C, noté $C \to C'$, si

$$\forall\,(x,x') \in C \times C',\ x \to x'.$$

A noter l'équivalence :

$$\forall\,(x,x') \in C \times C',\ x \to x' \quad \Leftrightarrow \quad \exists\,(x,x') \in C \times C',\ x \to x'.$$

La relation d'accessibilité définit une relation d'ordre partiel entre les classes d'équivalence.
- Une classe d'équivalence C est dite *fermée* si, pour tout x, y tels que ($x \in C$ et $x \to y \Rightarrow y \in C$). Autrement dit, $\forall\, x \in C$, $\forall\, n \in \mathbb{N}$, $\sum_{y \in C} P^{(n)}(x,y) = 1$, autrement dit encore C est une classe dont on ne peut pas sortir.
- Une classe fermée réduite à un point $C = \{x\}$ est appelée un *état absorbant*. Un état x est absorbant ssi $p(x,x) = 1$.

7.2 Période

Dans cette partie, nous étudions la période associée à une classe.

Définition 7.4. Étant donné un état $x \in E$, la *période* de l'état x, notée $d(x)$, est le plus grand commun diviseur des entiers n tels que $P^{(n)}(x,x)$ est strictement positif :

$$d(x) = \mathrm{PGCD}\{n \geq 1 \mid P^{(n)}(x,x) > 0\}. \tag{7.2}$$

Par convention $d(x) = 0$ si l'ensemble des $n \geq 1$ tels que $P^{(n)}(x,x) > 0$ est vide.

Nous avons le résultat suivant, qui montre que la notion de période est une notion de classe.

Théorème 7.5. *Si deux états communiquent alors ils ont même période.*

Démonstration. Soient x et y deux états qui communiquent, $x \leftrightarrow y$. Montrons que $d(y)$ divise $d(x)$. Ceci revient à prouver que $d(y)$ divise tout $n \geq 1$ tel que $P^{(n)}(x,x) > 0$; soit donc un tel n. Comme x et y communiquent, il existe deux entiers $k, \ell \geq 1$ tels que $P^{(k)}(x,y) > 0$ et $P^{(\ell)}(y,x) > 0$. De plus, d'après les équations de Chapman-Kolmogorov :

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

Ainsi, $d(y)$ divise $n + k + \ell$ et $k + \ell$; et par suite divise la différence n. On a donc montré que $d(y)$ divise $d(x)$. En intervertissant les rôles de x et y nous prouvons que $d(x)$ divise $d(y)$, ce qui entraîne que $d(x) = d(y)$. □

Définition 7.6. La *période* d'une classe est la période de chacun de ses éléments. Une classe est dite *apériodique* si sa période est 1.

7.3 Exercices : classes de communication, période

Exercice 7.1.

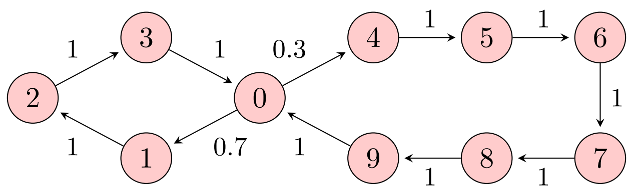
Considérons une chaîne de Markov définie sur un espace $E = \{0, \ldots, 9\}$ formé de 10 états, dont la matrice de transition est :

$$P = \begin{pmatrix} 0 & 0.7 & 0 & 0 & 0.3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

1. Tracer son graphe.
2. Déterminer les classes de la chaîne et leurs périodes.

Solution de l'exercice 7.1.

1. Traçons son graphe :



2. Nous constatons qu'il y a deux cycles donc les états de chacun des deux cycles communiquent. De plus, ces deux cycles contiennent le même état 0, ce qui implique qu'ils communiquent. Il n'y a donc qu'une seule classe ; la chaîne est irréductible.

   Afin de déterminer la période de cette chaîne de Markov, il suffit de déterminer celle d'un de ses états ; prenons $x = 0$. Partant de 0 nous sommes de nouveau en 0 au bout de 4 transitions en passant par le petit cycle et au bout de 7 transitions en passant par le grande cycle, nous avons :

$$P^{(4)}(0,0) = 0.7 \text{ et } P^{(7)}(0,0) = 0.3.$$

   Nous en déduisons que le PGCD ne peut être que 1 ; la chaîne est donc apériodique. En résumé, cette chaîne est formée d'une seule classe apériodique.

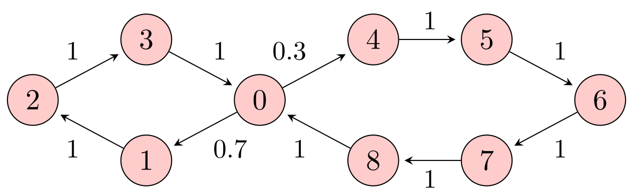
Exercice 7.2. Reprenons la chaîne précédente avec un état de moins dans le second cycle. C'est-à-dire, considérons la chaîne de Markov définie sur un espace $E = \{0, \dots, 8\}$ formé de 9 états, dont la matrice de transition est :

$$P = \begin{pmatrix} 0 & 0.7 & 0 & 0 & 0.3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

1. Tracer son graphe.
2. Déterminer les classes de la chaîne et leurs périodes.

Solution de l'exercice 7.2.

1. Traçons son graphe :



2. Comme précédemment cette chaîne de Markov ne contient qu'une seule classe et est donc irréductible. Étudions la période en étudiant celle du sommet 0. Partant de 0, nous sommes de nouveau en 0 au bout de 4 transitions en passant par le petit cycle et au bout de 6 transitions en passant par le grand cycle. De plus, pour revenir en 0 on est obligé de passer par le petit ou par le grand cycle ("ou" non exclusif). Ainsi, nous avons $P^{(n)}(0,0) > 0$ si et seulement si $n = 4k + 6\ell$ pour $(k, \ell) \neq (0,0)$. Donc

   $$d(0) = \mathrm{PGCD}\{n \geq 1 \,|\, P^{(n)}(0,0) > 0\} = \mathrm{PGCD}\{4k+6\ell \,|\, (k,\ell) \neq (0,0)\} = \mathrm{PGCD}\{4,6\} = 2.$$

   La chaîne est périodique de période 2. Il existe donc deux sous-classes,

$$0 \longrightarrow \begin{array}{c}1\\4\end{array} \longrightarrow \begin{array}{c}2\\5\end{array} \longrightarrow \begin{array}{c}3\\6\end{array} \longrightarrow \begin{array}{c}0\\7\end{array} \longrightarrow \begin{array}{c}1\\4\\8\end{array} \longrightarrow \begin{array}{c}0\\2\\5\end{array} \longrightarrow \begin{array}{c}1\\3\\4\\6\end{array} \longrightarrow \begin{array}{c}0\\2\\5\\7\end{array} \longrightarrow \begin{array}{c}1\\3\\4\\6\\8\end{array} \longrightarrow \begin{array}{c}0\\2\\5\\7\end{array} \longrightarrow \cdots$$

Figure 7.1 – sous-classes

qui sont $\{0, 2, 5, 7\}$ et $\{1, 3, 4, 6, 8\}$.

Exercice 7.3.

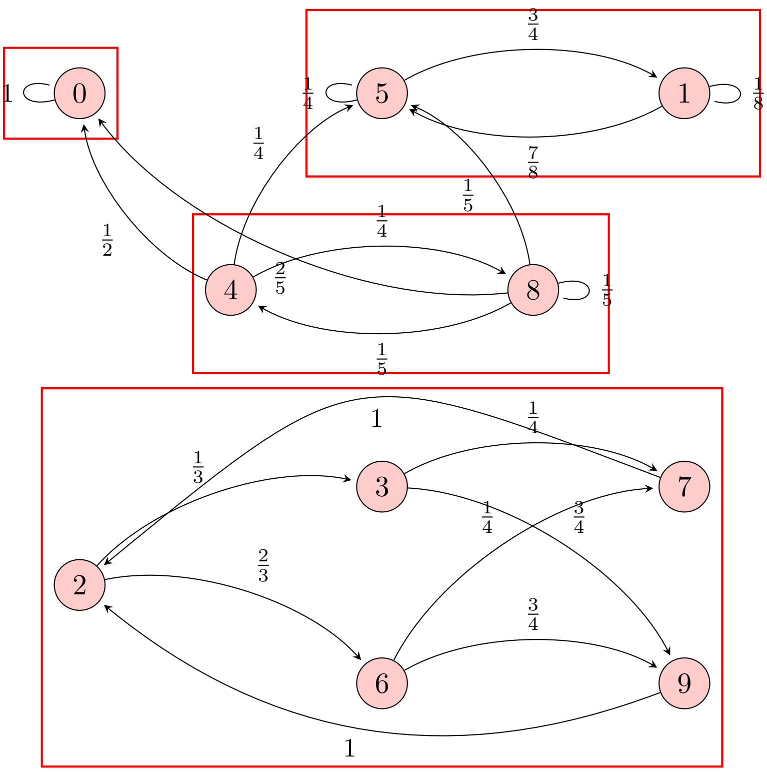
Soit la chaîne de Markov à 10 états $\{0, \dots, 9\}$ de matrice de transition :

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{8} & 0 & 0 & 0 & \frac{7}{8} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & \frac{2}{3} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{4} & 0 & \frac{3}{4} \\ \frac{1}{2} & 0 & 0 & 0 & 0 & \frac{1}{4} & 0 & 0 & \frac{1}{4} & 0 \\ 0 & \frac{3}{4} & 0 & 0 & 0 & \frac{1}{4} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{4} & 0 & \frac{3}{4} \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{2}{5} & 0 & 0 & 0 & \frac{1}{5} & \frac{1}{5} & 0 & 0 & \frac{1}{5} & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

1. Tracer son graphe.
2. Déterminer les classes de la chaîne et leurs périodes.

Solution de l'exercice 7.3.

1. Le graphe associé est :



2. Nous constatons l'existence de 4 classes :
   — la classe $\{0\}$ qui est fermée et réduite à un point, et forme donc un état absorbant 0 ;
   — la classe $\{1,5\}$ qui est fermée et apériodique en raison des boucles en 1 et 5 ;
   — la classe $\{2,3,6,7,9\}$ qui est fermée et de période 3, les 3 sous-classes étant dans l'ordre $\{2\}$, $\{3,6\}$ et $\{7,9\}$;
   — enfin la classe $\{4,8\}$, que nous pouvons quitter vers les classes $\{0\}$ et $\{1,5\}$ et qui n'est donc pas fermée.

   Nous pouvons voir que nous quittons la classe non-fermée $\{4,8\}$ vers les classes fermées $\{0\}$ et $\{1,5\}$. Il serait intéressant de savoir avec quelle probabilité nous nous retrouvons dans chacune des deux classes terminales. Or, nous voyons que nous quittons la classe $\{4,8\}$ de 4 ou de 8, et que nous avons toujours deux fois plus de chance de nous retrouver en 0 que dans la classe $\{1,5\}$. Par suite la probabilité de se retrouver en 0 est de $\frac{2}{3}$ et en $\{1,5\}$ de $\frac{1}{3}$.

Exercice 7.4. Déterminer les classes et périodes des exercices suivants du Chapitre 5 : Exercice 5.4 (marche aléatoire sur $\mathbb{Z}$), Exercice 5.6 (fort carré), Exercice 5.7 (transmission d'un bit informatique), Exercice 5.9 (ruine du joueur), Exercice 5.10 (plagiste et ses voiliers).

Solution de l'exercice 7.4.
- Exercice 5.4 (marche aléatoire sur $\mathbb{Z}$). Chaîne irréductible de période 2.
- Exercice 5.6 (fort carré). Chaîne irréductible de période 2, avec comme sous-classes : $\{0,2\}$ et $\{1,3\}$. De manière générale, la marche aléatoire sur $\mathbb{Z}/n\mathbb{Z}$ est irréductible, apériodique si n est impair, et périodique de période 2 si n est pair.
- Exercice 5.7 (transmission d'un bit informatique). Chaîne irréductible apériodique (car contient des boucles).
- Exercice 5.9 (ruine du joueur). Trois classes : $\{0\},\{N\},\{1,\dots,N-1\}$. Les classes $\{0\}$ et $\{N\}$ consistent en un seul sommet et sont fermées, il s'agit d'états absorbants. La classe $\{1,\dots,N-1\}$ n'est pas fermée car les classes $\{0\}$ et $\{N\}$ sont accessibles depuis elle. Les deux états absorbants sont de période 1 chacun. La classe non-fermée $\{1,\dots,N-1\}$ est de période 2.
- Exercice 5.10 (plagiste et ses voiliers). Chaîne irréductible apériodique (car contient des boucles).

7.4 Récurrence et transience

Nous allons étudier une seconde classification des états dépendant du type de comportement de la chaîne. Dans toute cette section, nous considérons une chaîne de Markov $(X_n)_{n\geq 0}$ à valeurs dans un espace E, de matrice de transition P.

Étant donné un état $x \in E$, tel que $\mathbb{P}[X_0 = x] > 0$, nous notons simplement $\mathbb{P}_x$ la probabilité conditionnelle $\mathbb{P}_{\{X_0=x\}}[\cdot]$ sachant l'événement $\{X_0 = x\}$ réalisé.

7.4.1 Définitions

Définition 7.7. Soit $x \in E$ un état. Le *temps d'atteinte de x*, noté T_x, est le premier instant où x est visité après le départ. Par convention, le temps d'atteinte est infini si nous n'atteignons jamais x :

$$\forall\, \omega \in \Omega, \quad T_x(\omega) = \begin{cases} \inf\{n > 0 \mid X_n(\omega) = x\} & \text{si un tel entier existe} \\ +\infty & \text{sinon.} \end{cases}$$

Si la chaîne part de l'état x lui-même, nous employons plutôt le terme de *temps de retour*.

Définition 7.8. Un état $x \in E$ est dit *récurrent* si

$$\mathbb{P}_x[T_x < +\infty] = 1.$$

L'état x est dit *transient* ou *transitoire* sinon, c'est-à-dire quand

$$\mathbb{P}_x[T_x < +\infty] < 1, \text{ autrement dit, } \mathbb{P}_x[T_x = \infty] > 0.$$

Un état est récurrent si nous sommes sûr d'y revenir, il est transient s'il existe une probabilité non nulle de ne jamais y revenir, et donc de le quitter définitivement.

7.4.2 Critères de récurrence/transience

Une manière de caractériser la récurrence/transience d'un état est d'utiliser le nombre de visites, notion que nous introduisons maintenant.

Définition 7.9. Le *nombre de visites en un état x* est la variable aléatoire N_x définie par :

$$N_x = \sum_{n=0}^{\infty} \mathbb{I}_{\{X_n=x\}}.$$

La quantité, $G(x,y) := \mathbb{E}_x[N_y]$ est *l'espérance du nombre de visites en y partant de x*. G s'appelle le *noyau de Green*, le *noyau potentiel* ou simplement la *fonction de Green*.

Lemme 7.10. *Nous avons l'égalité suivante :*

$$\forall\,(x,y) \in E^2, \quad G(x,y) = \sum_{n=0}^{\infty} P^{(n)}(x,y). \tag{7.3}$$

Démonstration. Par définition de la fonction de Green,

$$\begin{aligned}
G(x,y) &= \mathbb{E}_x[N_y] \\
&= \mathbb{E}_x\Big[\sum_{n\geq 0} \mathbf{1}_{\{X_n=y\}}\Big] && \text{(par définition du nombre de visites en } y\text{)} \\
&= \sum_{n\geq 0} \mathbb{E}_x\left[\mathbf{1}_{\{X_n=y\}}\right] && \text{(par inversion de la sommation et de l'espérance (Fubini))} \\
&= \sum_{n\geq 0} \mathbb{P}_x[X_n=y] && \text{(espérance de l'indicatrice)} \\
&= \sum_{n\geq 0} P^{(n)}(x,y). \qquad \square
\end{aligned}$$

Le théorème suivant caractérise la récurrence/transience en utilisant le nombre de visites.

Théorème 7.11. *Les conditions suivantes sont équivalentes :*

1. *L'état $x \in E$ est récurrent,* i.e., $\mathbb{P}_x[T_x < \infty] = 1$.
2. $\mathbb{P}_x[N_x = \infty] = 1$.
3. $G(x,x) = \infty$.

De manière analogue, les conditions suivantes sont équivalentes :

1. *L'état $x \in E$ est transient,* i.e., $\mathbb{P}_x[T_x < \infty] < 1$.
2. $\mathbb{P}_x[N_x = \infty] = 0$.
3. $G(x,x) < \infty$ *et* $G(x,x) = \dfrac{1}{\mathbb{P}_x[T_x = \infty]}$.

Dans ce cas, la loi conditionnelle de N_x sachant que l'on part de l'état x est une loi géométrique de paramètre $\mathbb{P}_x[T_x = \infty]$.

Avant de démontrer le Théorème 7.11, nous prouvons le résultat préliminaire suivant.

Lemme 7.12. *Pour tout $(x,y) \in E^2$,*

- *pour tout $n \in \mathbb{N}^*$,*

$$\mathbb{P}_x[N_y \geq n+1] = \mathbb{P}_x[T_y < \infty]\mathbb{P}_y[N_y \geq n];$$

  cette égalité est aussi vraie pour $n = 0$ si $x \neq y$.
- *pour la fonction de Green,*

$$G(x,y) = \delta_{\{x=y\}} + \mathbb{P}_x[T_y < \infty]G(y,y).$$

Démonstration. Démontrons la première égalité. Soit $n \in \mathbb{N}^*$; d'après la formule des probabilités totales,

$$\begin{aligned}
\mathbb{P}_x[N_y \geq n+1] &= \mathbb{P}_x[N_y \geq n+1, T_y < \infty] + \mathbb{P}_x[N_y \geq n+1, T_y = \infty] \\
&= \mathbb{P}_x[N_y \geq n+1, T_y < \infty] \\
&= \sum_{\ell=1}^{\infty} \mathbb{P}_x[N_y \geq n+1, T_y = \ell], \text{ (formule des probabilités totales)} \\
&= \sum_{\ell=1}^{\infty} \mathbb{P}_x\Big[\Big(\sum_{k=\ell+1}^{\infty} \mathbb{I}_{\{X_k=y\}}\Big) \geq n, T_y = \ell\Big] \\
&= \sum_{\ell=1}^{\infty} \mathbb{P}\Big[\Big(\sum_{k=\ell+1}^{\infty} \mathbb{I}_{\{X_k=y\}}\Big) \geq n \,\Big|\, T_y = \ell, X_0 = x\Big]\, \mathbb{P}_x[T_y = \ell], \text{ (proba. cond.)}.
\end{aligned}$$

Or,

$$\begin{aligned}
&\mathbb{P}\Big[\Big(\sum_{k=\ell+1}^{\infty} \mathbb{I}_{\{X_k=y\}}\Big) \geq n \,\Big|\, T_y = \ell, X_0 = x\Big] \\
&\qquad = \mathbb{P}\Big[\Big(\sum_{k=\ell+1}^{\infty} \mathbb{I}_{\{X_k=y\}}\Big) \geq n \,\Big|\, X_\ell = y, X_{\ell-1} \neq y, \ldots, X_1 \neq y, X_0 = x\Big] \\
&\qquad = \mathbb{P}\Big[\Big(\sum_{k=1}^{\infty} \mathbb{I}_{\{X_k=y\}}\Big) \geq n \,\Big|\, X_0 = y\Big] = \mathbb{P}_y[N_y \geq n] \text{ (d'après le Corollaire 6.3)}.
\end{aligned}$$

Ainsi on conclut,

$$\mathbb{P}_x[N_y \geq n+1] = \mathbb{P}_y[N_y \geq n] \sum_{\ell=1}^{\infty} \mathbb{P}_x[T_y = \ell] = \mathbb{P}_y[N_y \geq n]\mathbb{P}_x[T_y < \infty].$$

Démontrons l'égalité pour la fonction de Green.

$$\begin{aligned}
G(x,y) &= \sum_{n=0}^{\infty} P^{(n)}(x,y), \text{ d'après le Lemme 7.10} \\
&= \delta_{\{x=y\}} + \sum_{n=1}^{\infty} P^{(n)}(x,y) = \delta_{\{x=y\}} + \sum_{n=1}^{\infty} \mathbb{P}_x[X_n = y] \\
&= \delta_{\{x=y\}} + \sum_{n=1}^{\infty} \sum_{\ell=1}^{n} \mathbb{P}_x[X_n = y, T_y = \ell], \text{ (formule des probabilités totales)} \\
&= \delta_{\{x=y\}} + \sum_{n=1}^{\infty} \sum_{\ell=1}^{n} \mathbb{P}[X_n = y | T_y = \ell, X_0 = x]\mathbb{P}_x[T_y = \ell], \text{ (formule probas cond.)}
\end{aligned}$$

Or d'après le Corollaire 6.3, on a pour tout $\ell \in \{1, \ldots, n\}$

$$\mathbb{P}[X_n = y | T_y = \ell, X_0 = x] = \mathbb{P}[X_n = y | X_\ell = y] = \mathbb{P}_y[X_{n-\ell} = y]$$

En remplaçant et en inversant les sommes sur n et ℓ, on obtient,

$$\begin{aligned}
G(x,y) &= \delta_{\{x=y\}} + \sum_{\ell=1}^{\infty}\sum_{n=\ell}^{\infty} \mathbb{P}_y[X_{n-\ell}=y]\mathbb{P}_x[T_y=\ell] \\
&= \delta_{\{x=y\}} + \sum_{\ell=1}^{\infty} \mathbb{P}_x[T_y=\ell]\Big(\sum_{n=\ell}^{\infty} \mathbb{P}_y[X_{n-\ell}=y]\Big) \\
&= \delta_{\{x=y\}} + \mathbb{P}_x[T_y<\infty]G(y,y). \qquad \square
\end{aligned}$$

Démonstration du Théorème 7.11. Puisque $\mathbb{P}_x[N_x \geq 1] = 1$, le Lemme 7.12 et une récurrence immédiate nous disent que, pour tout $n \in \mathbb{N}^*$,

$$\mathbb{P}_x[N_x \geq n] = \mathbb{P}_x[T_x < \infty]^{n-1}.$$

La suite d'événement $(\{N_x \geq n\})_{n\geq 1}$ étant décroissante, on a aussi

$$\mathbb{P}_x[N_x = \infty] = \mathbb{P}_x[\cap_{n\geq 1}\{N_x \geq n\}] = \lim_{n\to\infty} \mathbb{P}_x[\{N_x \geq n\}] = \lim_{n\to\infty} \mathbb{P}_x[T_x < \infty]^{n-1}$$

Ceci montre l'équivalence entre les points 2 et 1. En effet, on a soit

$$\begin{aligned}
&\mathbb{P}_x[N_x = \infty] = 1 \Leftrightarrow \mathbb{P}_x[T_x < \infty] = 1, \quad \text{ou bien} \\
&\mathbb{P}_x[N_x = \infty] = 0 \Leftrightarrow \mathbb{P}_x[T_x < \infty] < 1.
\end{aligned}$$

Montrons maintenant l'équivalence entre les points 1 et 3. D'après la deuxième égalité du Lemme 7.12, la fonction de Green $G(x,x)$ satisfait à l'égalité,

$$G(x,x)(1 - \mathbb{P}_x[T_x < \infty]) = 1.$$

Ainsi,

$$\begin{aligned}
&G(x,x) = \frac{1}{1-\mathbb{P}_x[T_x<\infty]} < \infty \quad \Leftrightarrow \quad \mathbb{P}_x[T_x<\infty] < 1 \text{ ou bien} \\
&G(x,x) = \infty \quad \Leftrightarrow \quad \mathbb{P}_x[T_x<\infty] = 1.
\end{aligned}$$

et cela démontre l'équivalence entre le point 1 et 3 et termine la preuve des équivalences. Montrons encore que dans le cas transient, la loi conditionnelle de N_x sachant $\{X_0 = x\}$ est géométrique de paramètre $\mathbb{P}_x[T_x = \infty]$. Pour cela on utilise à nouveau le Lemme 7.12. Pour tout $n \geq 1$,

$$\mathbb{P}_x[N_x = n] = \mathbb{P}_x[N_x \geq n] - \mathbb{P}_x[N_x \geq n+1] = \mathbb{P}_x[T_x<\infty]^{n-1}(1 - \mathbb{P}_x[T_x<\infty]). \qquad \square$$

7.4.3 Classes récurrentes/transientes

Nous allons établir que la notion de récurrence/transience est une propriété de classes.

Proposition 7.13. *Soient deux états x et y qui communiquent, alors x et y sont soit tous les deux récurrents soit tous les deux transients.*

Démonstration. D'après le Théorème 7.11 il suffit de montrer que $G(x,x)$ et $G(y,y)$ sont de même nature (soit finies, soit infinies). D'après le Lemme 7.10, pour tout $x \in E$, $G(x,x) = \sum_{n=0}^{\infty} P^{(n)}(x,x)$, donc cela revient à montrer que les séries de terme général $P^{(n)}(x,x)$ et $P^{(n)}(y,y)$ sont de même nature. Comme x et y communiquent, il existe k,ℓ tels que $P^{(k)}(x,y) > 0$ et $P^{(\ell)}(y,x) > 0$. Ainsi, pour tout $n \geq k+\ell$,

$$P^{(n)}(x,x) \geq P^{(k)}(x,y)P_{n-k-\ell}(y,y)P^{(\ell)}(y,x) \text{ et } P^{(n)}(y,y) \geq P^{(\ell)}(y,x)P_{n-k-\ell}(x,x)P^{(k)}(x,y),$$

ce qui démontre que les deux séries convergent et divergent en même temps. □

Définition 7.14. Une classe d'équivalence est dite *récurrente*, resp. *transiente*, si un de ses sommets est récurrent, resp. transient.

Allons un peu plus loin dans la description des classes.

Proposition 7.15. *Une classe récurrente est fermée, autrement dit, la probabilité de sortir d'une classe récurrente est nulle.*

Démonstration. Soit x un état quelconque appartenant à la classe de récurrence C. Supposons qu'il existe $y \notin C$ tel que $x \to y$ et montrons que l'on a une contradiction. Remarquons d'abord que y ne conduit à aucun sommet de C, car sinon on aurait $y \to x$ et donc $x \leftrightarrow y$ et $y \in C$. De plus, on a :

$$x \to y \quad \Leftrightarrow \quad \mathbb{P}_x[T_y < \infty] > 0.$$

Or, la probabilité $\mathbb{P}_x[T_x = \infty]$ de ne pas revenir en x est bornée inférieurement par la probabilité d'aller en y en temps fini (vu que y ne conduit à aucun état de C). Ainsi,

$$\mathbb{P}_x[T_x = \infty] \geq \mathbb{P}_x[T_y < \infty] > 0,$$

d'où $\mathbb{P}_x[T_x < \infty] < 1$, ce qui est une contradiction avec x récurrent. □

Nous voyons qu'une classe récurrente est fermée, mais la réciproque est fausse en général (voir l'exemple de la marche aléatoire non-symétrique, Exercice 7.7, qui est fermée et transiente), malgré tout elle est vérifiée si cette classe est de cardinal fini.

Proposition 7.16. *Toute classe fermée et de cardinal fini est récurrente.*

Démonstration. Soit C cette classe fermée, et soit $x \in C$, alors,

$$\begin{aligned}\sum_{y\in C} G(x,y) &= \sum_{y\in C}\sum_{n\geq 0} P^{(n)}(x,y)\\ &= \sum_{n\geq 0}\sum_{y\in C} P^{(n)}(x,y) = \infty,\end{aligned}$$

car pour tout $n \in \mathbb{N}$, $\sum_{y\in C} P^{(n)}(x,y) = \mathbb{P}_x[X_n \in C] = 1$; puisque la classe C est fermée nous ne la quittons pas.

Supposons que pour tout $y \in C$, y est transient, alors d'après l'Exercice 7.5, $G(x,y) < \infty$. Par hypothèse la classe C est de cardinal fini, ce qui implique que $\sum_{y\in C} G(x,y) < \infty$ et on obtient une contradiction. En conséquence, il existe un état $y \in C$ récurrent et la classe C est donc récurrente. □

Comme corollaire, nous déduisons,

Corollaire 7.17. *Une chaîne de Markov définie sur un espace d'états fini admet au moins un état récurrent.*

Démonstration. Ceci découle du fait que si l'espace d'état est fini, il doit y avoir au moins une classe fermée. □

7.5 Exercices : récurrence/transience

Exercice 7.5. [Fonction de Green]

Allons un peu plus loin dans l'étude de la fonction de Green. Soit $x, y \in E$ deux états. Montrer que :

1. Si $x \nrightarrow y$, alors $G(x,y) = 0$.
2. Si $x \to y$ et
   - y est transient, alors $G(x,y) < \infty$
   - y est récurrent, alors $G(x,y) = \infty$
3. La fonction de Green satisfait à l'équation,

$$G = PG + I \quad \Leftrightarrow \quad \forall\,(x,y) \in E^2,\ G(x,y) = \sum_{z \in E} p(x,z)G(z,y) + \delta_{\{x=y\}}.$$

Solution de l'exercice 7.5.

1. Si $x \nrightarrow y$, alors pour tout $n \geq 0$, $P^{(n)}(x,y) = 0$. Ainsi, d'après le Lemme 7.10, $G(x,y) = 0$.
2. On a,
$$x \to y \quad \Leftrightarrow \quad \mathbb{P}_x[T_y < \infty] > 0,$$
(Preuve dans l'Exercice 9.6. de "Exercices de Probabilités", Cottrell & al.)
De plus, si y est transient, $G(y,y) < \infty$ et si y est récurrent $G(y,y) = \infty$. En utilisant la deuxième égalité du Lemme 7.12, ceci démontre le point 2.
3. D'après le Lemme 7.10,

$$\begin{aligned}
G(x,y) &= \sum_{n=0}^{\infty} P^{(n)}(x,y) = P^{(0)}(x,y) + \sum_{n=1}^{\infty} P^{(n)}(x,y) \\
&= \delta_{\{x=y\}} + \sum_{n=1}^{\infty} \sum_{z \in E} p(x,z) P^{(n-1)}(z,y), \text{ d'après Chapman-Kolmogorov} \\
&= \delta_{\{x=y\}} + \sum_{z \in E} p(x,z) \sum_{n=1}^{\infty} P^{(n-1)}(z,y) = \delta_{\{x=y\}} + \sum_{z \in E} p(x,z) \sum_{n=0}^{\infty} P^{(n)}(z,y) \\
&= \delta_{\{x=y\}} + \sum_{z \in E} p(x,z) G(z,y).
\end{aligned}$$

Exercice 7.6. On considère une chaîne de Markov $(X_n)_{n \geq 0}$ d'espace d'état $E = \{1,2,3\}$ et de matrice de transition

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

1. Dessiner le graphe associé à cette chaîne de Markov et classifier les états (récurrents/transients).
2. Calculer la fonction de Green pour toutes les paires de points $(x,y) \in E^2$. En déduire la probabilité de ne jamais revenir en 3 sachant que l'on est parti de 3.

Solution de l'exercice 7.6.

1. Graphe à faire. Les classes d'équivalence sont $\{1, 2\}$ et $\{3\}$. La classe $\{1, 2\}$ est fermée et finie, donc récurrente. La classe $\{3\}$ permet d'atteindre la classe $\{1, 2\}$; elle n'est pas fermée, donc transiente.
2. D'après les points 1 et 2 de l'Exercice 7.5, on a que la matrice des coefficients de la fonction de Green est de la forme,

$$G = \begin{pmatrix} \infty & \infty & 0 \\ \infty & \infty & 0 \\ \infty & \infty & G(3,3) \end{pmatrix}$$

Pour déterminer $G(3, 3)$ on utilise le Point 3 de l'Exercice 7.5 :

$$G(3,3) = 1 + \sum_{z=1}^{3} p(z,3)G(z,3) = 1 + \frac{1}{3}G(3,3).$$

Ainsi, $\frac{2}{3}G(3, 3) = 1$, autrement dit $G(3, 3) = \frac{3}{2}$. On en déduit la probabilité de ne jamais revenir en 3 sachant que l'on est parti de 3 :

$$\mathbb{P}_3[T_3 = \infty] = \frac{1}{G(3,3)} = \frac{2}{3}.$$

Exercice 7.7. Classifier les états (récurrent / transient) des exercices du Chapitre 5.3 : Exercice 5.4 (marche aléatoire sur $\mathbb{Z}$), Exercice 5.6 (fort carré), Exercice 5.7 (transmission d'un bit informatique), Exercice 5.9 (ruine du joueur), Exercice 5.10 (plagiste et ses voiliers).

Solution de l'exercice 7.7.

- Exercice 5.4 (marche aléatoire sur $\mathbb{Z}$). Chaîne irréductible. Il suffit donc de considérer un état. Dans l'Exercice 6.4 on a montré que

$$\text{Si } p \neq \frac{1}{2}, \quad G(x,x) = \sum_{n=0}^{\infty} P^{(n)}(x,x) < \infty$$
$$\text{Si } p = \frac{1}{2}, \quad G(x,x) = \sum_{n=0}^{\infty} P^{(n)}(x,x) = \infty.$$

  Ainsi, la marche aléatoire est récurrente dans le cas symétrique et transiente sinon. Remarquer que dans le cas non-symétrique, on a un exemple de classe fermée (infinie) et non-récurrente.
- Exercice 5.6 (fort carré) ou marche aléatoire sur $\mathbb{Z}/n\mathbb{Z}$. Chaîne irréductible. Elle est récurrente car finie.
- Exercice 5.7 (transmission d'un bit informatique). Chaîne irréductible. Elle est récurrente car finie.
- Exercice 5.9 (ruine du joueur). Trois classes : $\{0\}, \{N\}, \{1, \dots, N-1\}$. Les classes $\{0\}, \{N\}$ consistent en un seul sommet et sont fermées, il s'agit d'états absorbants. La classe $\{1, \dots, N-1\}$ n'est pas fermée car les classes $\{0\}, \{N\}$ sont accessibles depuis elle. Ainsi les états absorbants sont récurrents et la classe $\{1, \dots, N-1\}$ est transiente.

- Exercice 5.10 (plagiste et ses voiliers). Chaîne irréductible. Elle est récurrente car finie.

Exercice 7.8. [Qui perd, perd tout]

Nous considérons un joueur dont la probabilité de donner une bonne réponse est p quelle que soit la question posée et la probabilité de donner une réponse fausse est $1-p$, avec $0 < p < 1$. Nous supposons que les réponses sont indépendantes les unes des autres et que s'il donne une bonne réponse, il gagne 1 et s'il donne une mauvaise réponse, il perd tout. La richesse de départ est $X_0 = 0$ et, pour tout $n \geq 1$, on note X_n sa fortune à l'instant n.

1. Montrer qu'il s'agit d'une chaîne de Markov. Donner son graphe et sa matrice de transition.
2. Montrer que cette chaîne de Markov est irréductible.
3. Soit $T_0 = \inf\{n > 0 | X_n = 0\}$ le temps d'atteinte de 0. Pour tout $n \in \mathbb{N}^*$, calculer $\mathbb{P}_0[T_0 = n]$.
4. En déduire que la chaîne est récurrente.

Solution de l'exercice 7.8.

1. Le processus $(X_n)_{n\geq 0}$ est à valeurs dans $E = \mathbb{N}$. Soit Y_n la variable aléatoire valant 1 si le joueur répond juste à la n-ième question, et 0 sinon. Alors $(Y_n)_{n\geq 1}$ est une suite de variables aléatoires i.i.d. de loi de Bernoulli de paramètre p, indépendante de X_0. Pour tout $n \geq 0$, la fortune du joueur à l'instant $n+1$ est :

$$\begin{aligned} X_{n+1} &= \begin{cases} X_n + 1 & \text{si } Y_{n+1} = 1 \\ 0 & \text{si } Y_{n+1} = 0 \end{cases} \\ &= (X_n + 1)\mathbb{I}_{\{Y_{n+1}=1\}}. \end{aligned}$$

Ainsi d'après l'Exercice 5.5, $(X_n)_{n\geq 0}$ est une chaîne de Markov.
La matrice de transition $P = (p(x,y))$ a pour coefficients non nuls,

$$\forall\, x \in \mathbb{N}, \quad p(x, x+1) = p \quad p(x, 0) = 1 - p.$$

$$P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ \vdots \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & \cdots \end{array} \\ \begin{pmatrix} 1-p & p & 0 & 0 & \cdots \\ 1-p & 0 & p & 0 & \cdots \\ 1-p & 0 & 0 & p & \cdots \\ 1-p & 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \end{array}.$$

Figure 7.2 – Matrice de transition du jeu.

Le graphe associé à cette chaîne de Markov est,

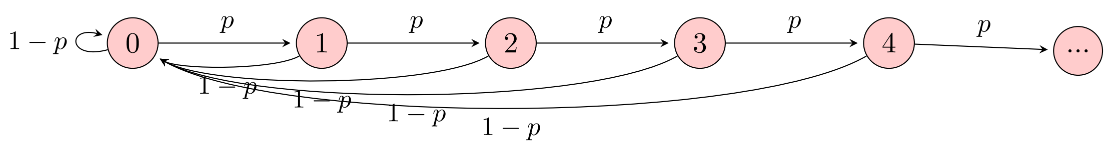


Figure 7.3 – Graphe de la ruine du joueur.

2. Nous constatons que tous les points communiquent entre eux. En effet, soient deux états x et y :
   - Si $y > x$, alors partant de x nous pouvons atteindre y en $y - x$ étapes en gagnant successivement $y - x$ fois : $P^{(x-y)}(x, y) = p^{y-x} > 0$.
   - Si $y < x$, alors partant de x nous pouvons perdre immédiatement et gagner y fois pour atteindre la fortune y : $P^{(1+y)}(x, y) = (1 - p)p^y > 0$.

   Il n'y a donc qu'une seule classe et la chaîne est irréductible.
3. Soit $n \in \mathbb{N}^*$. Alors,

$$\begin{aligned}\mathbb{P}_0[T_0 = n] &= \mathbb{P}[\{\text{le joueur répond juste aux } (n-1) \text{ premières questions et faux à la } n\text{-ième}\}] \\ &= \mathbb{P}_0[\{Y_1 = 1, \cdots, Y_{n-1} = 1, Y_n = 0\}] \\ &= p^{n-1}(1-p), \text{ par indépendance.}\end{aligned}$$

   Ainsi la loi conditionnelle de T_0 sachant $\{X_0 = 0\}$ est une loi géométrique de paramètre $(1 - p)$ sur $\mathbb{N}^*$.
4. On a,

$$\mathbb{P}_0[T_0 < \infty] = \sum_{n=1}^{\infty} \mathbb{P}_0[T_0 = n] = (1-p) \sum_{n=1}^{\infty} p = \frac{1-p}{1-p} = 1,$$

   et la chaîne est donc récurrente.

Exercice 7.9. [Ruine du joueur, suite]

On reprend l'Exercice 5.9 de la ruine du joueur : deux joueurs A et B disposent d'une fortune initiale de a et b euros respectivement, avec $a, b > 0$. Ils jouent au jeu de hasard suivant : la mise est de 1 euro par partie ; les parties sont indépendantes et à chacune d'entre elle le joueur A a une probabilité p de gagner (et donc $1 - p$ de perdre), avec $0 < p < 1$. Le jeu se déroule jusqu'à la ruine d'un des deux joueurs. Notons $N = a + b$, alors le durée du jeu est $T = \inf\{n > 0 | X_n \in \{0, N\}\}$. Calculer la probabilité que le joueur A gagne.

Solution de l'exercice 7.9. Notons $u(x) = \mathbb{P}_x[X_T = N]$, c'est-à-dire la probabilité que le joueur gagne sachant que sa fortune initiale est x, avec $x \in \{0, \ldots, N\}$. Finalement, on sera intéressé par la probabilité $u(a)$ mais on verra qu'il est utile d'élargir le problème pour le résoudre.

Montrons que pour tout $x \in \{1, \ldots, N-1\}$, en écrivant $q = 1 - p$, $u(x)$ satisfait à l'équation :

$$u(x) = pu(x+1) + qu(x-1).$$

En effet, comme $a, b > 0$, on sait que $T \geq 1$. En décomposant sur toutes les valeurs possibles de la chaîne à l'instant 1, on obtient,

$$\begin{aligned}
u(x) = \mathbb{P}_x[X_T = N] &= \sum_{y\in E} \mathbb{P}_x[X_T = N | X_1 = y]\mathbb{P}_x[X_1 = y] \\
&= \sum_{y\in E} \mathbb{P}_y[X_T = N]\mathbb{P}_x[X_1 = y], \text{ (en utilisant la propriété de Markov)} \\
&= pu(x+1) + (1-p)u(x-1) \text{ (car les autres termes } \mathbb{P}_x[X_1 = y] = p(x,y) \text{ sont nuls).}
\end{aligned}$$

On a les conditions au bord suivantes,

$$u(0) = 0, \quad u(N) = 1.$$

Il s'agit de résoudre une récurrence linéaire d'ordre 2. L'équation caractéristique est :

$$pz^2 - z + q = 0.$$

Ce polynôme a deux racines $z_{1,2} = 1, \frac{q}{p}$ si $p \neq q$ et une racine double $z_1 = 1$ si $p = q = \frac{1}{2}$. Ainsi,

- Si $p \neq q$, la solution générale est de la forme

$$u(x) = \lambda z_1^x + \mu z_2^x = \lambda + \mu \Big(\frac{q}{p}\Big)^x.$$

En tenant compte des conditions initiales, on résout pour λ, μ et on trouve,

$$u(x) = \frac{1 - \big(\frac{q}{p}\big)^x}{1 - \big(\frac{q}{p}\big)^N}.$$

- Si $p = q = \frac{1}{2}$, la solution générale est de la forme

$$u(x) = (\lambda + \mu x)z_1^x = (\lambda + \mu x).$$

En tenant compte des conditions initiales, on résout pour λ, μ et on trouve,

$$u(x) = \frac{x}{N}.$$

Chapitre 8

Mesures stationnaires

Notre but est ici d'étudier des mesures particulières pour les chaînes de Markov : les mesures stationnaires. Elles interviennent dans l'étude de l'équilibre et du comportement en temps long des chaînes.

Dans tout ce chapitre, nous considérons une chaîne de Markov $(X_n)_{n\geq 0}$, définie sur un espace des états E et de matrice de transition $P = (p(x,y))$.

8.1 Mesures stationnaires et réversibles

Définition 8.1. Soit μ une mesure positive sur E, telle que $\mu(x) < \infty$ pour tout $x \in E$ et $\mu \not\equiv 0$. On dit que μ est une mesure *invariante* ou *stationnaire* si :

$$\mu^t = \mu^t P, \quad \text{autrement dit,} \quad \forall\, y \in E,\ \mu(y) = \sum_{x\in E} \mu(x)p(x,y).$$

Remarque 8.2.

- Si μ est de *masse* finie, *i.e.*, $\mu(E) = \sum_{x\in E} \mu(x) < \infty$, on peut associer à μ une unique probabilité π telle que μ et π soient proportionnelles :

$$\forall\, x \in E, \quad \pi(x) = \frac{\mu(x)}{\mu(E)}.$$

- Si μ est une mesure stationnaire alors, pour tout $n \geq 0$,

$$\mu^t = \mu^t P^n.$$

  En particulier, en utilisant la Proposition 6.5, si la loi initiale π_0 est stationnaire, alors la loi à l'instant n, π_n, est égale à la loi initiale :

$$\pi_0^t = \pi_0^t P^n = \pi_n^t, \quad \text{autrement dit,} \quad \forall\, y \in E, \quad \mathbb{P}[X_0 = y] = \mathbb{P}[X_n = y].$$

Proposition 8.3. *Si la loi initiale est stationnaire, alors la chaîne de Markov est stationnaire, c'est-à-dire que, pour tout* $(x_0,\dots,x_k) \in E^{k+1}$ *et pour tout* $n \geq 0$,

$$\mathbb{P}[X_n = x_0,\dots,X_{n+k} = x_k] = \mathbb{P}[X_0 = x_0,\dots,X_k = x_k].$$

Démonstration. En utilisant un argument similaire à celui de la preuve du Théorème 6.2 on obtient,

$$\begin{aligned}\mathbb{P}[X_n = x_0, \dots, X_{n+k} = x_k] &= \mathbb{P}[X_n = x_0]p(x_0, x_1) \dots p(x_{k-1}, x_k)\\ &= \mathbb{P}[X_0 = x_0]p(x_0, x_1) \dots p(x_{k-1}, x_k)\\ &= \mathbb{P}[X_0 = x_0, \dots, X_k = x_k],\end{aligned}$$

où dans la deuxième égalité on a utilisé la Remarque 8.2. □

Certaines mesures satisfont à une propriété plus forte, la réversibilité, qui implique la stationnarité. Cette condition a l'avantage d'être en général plus simple à vérifier ; elle est adaptée aux chaînes de Markov qui sont réversibles dans le temps. Cependant attention, si une chaîne de Markov n'a pas de mesure réversible, cela ne veut pas dire qu'elle n'a pas de mesure stationnaire.

Définition 8.4. Soit μ une mesure positive sur E, telle que $\mu(x) < \infty$ pour tout $x \in E$ et $\mu \not\equiv 0$. On dit que μ est une mesure *réversible* si, pour tout $(x, y) \in E^2$,

$$\mu(x)p(x, y) = \mu(y)p(y, x).$$

Proposition 8.5. *Toute mesure réversible est stationnaire.*

Démonstration. Pour tout $y \in E$,

$$\sum_{x\in E} \mu(x)p(x, y) = \sum_{x\in E} \mu(y)p(y, x) = \mu(y) \sum_{x\in E} p(y, x) = \mu(y),$$

où dans la dernière égalité on a utilisé que la matrice P est stochastique. Donc la mesure μ est stationnaire. □

Proposition 8.6. *Si la loi initiale π_0 est réversible, alors pour tout $n \geq 0$, pour tout $(x, y) \in E^2$,*

$$\mathbb{P}[X_n = y|X_{n+1} = x] = \mathbb{P}[X_{n+1} = y|X_n = x].$$

Démonstration. Puisque la loi initiale est réversible elle est stationnaire et on a, pour tout $n \geq 0$, pour tout $x \in E$, $\mathbb{P}[X_n = x] = \mathbb{P}[X_0 = x] = \pi_0(x)$. De plus, d'après la formule de Bayes,

$$\begin{aligned}\mathbb{P}[X_n = y|X_{n+1} = x] &= \frac{\mathbb{P}[X_{n+1} = x|X_n = y]\mathbb{P}[X_n = y]}{\mathbb{P}[X_{n+1} = x]} = \frac{p(y, x)\pi_0(y)}{\pi_0(x)}\\ &= \frac{p(x, y)\pi_0(x)}{\pi_0(x)} \text{ car la mesure est réversible}\\ &= p(x, y) = \mathbb{P}[X_{n+1} = y|X_n = x].\end{aligned}$$

□

Remarque. Puisque qu'une chaîne de Markov est entièrement déterminée par sa loi initiale et sa matrice de transition, ceci signifie que la chaîne de Markov et celle retournée dans le temps ont la même loi.

8.2 Existence

Nous allons établir l'existence d'une mesure invariante dans le cas d'une chaîne irréductible et récurrente. Si la chaîne n'est pas irréductible, des résultats analogues s'appliquent à la chaîne restreinte à chacune des classes récurrentes.

Pour tout $x \in E$, définissons une mesure positive μ_x sur E telle que, pour tout $y \in E$, $\mu_x(y)$ est l'espérance du nombre de visites en y partant de x jusqu'au premier retour en x :

$$\mu_x(y) = \mathbb{E}_x\Big(\sum_{n=1}^{T_x} \mathbb{I}_{\{X_n=y\}}\Big).$$

Nous allons montrer que la mesure μ_x est stationnaire et strictement positive.

Théorème 8.7. *Soit une chaîne de Markov irréductible et récurrente, alors la mesure μ_x est stationnaire et strictement positive.*

Démonstration. Pour montrer la stationnarité, nous devons établir l'égalité suivante,

$$\forall\, y \in E, \quad \sum_{z\in E} \mu_x(z)p(z,y) = \mu_x(y).$$

Nous avons :

$$\begin{aligned}
\mu_x(y) &= \mathbb{E}_x\Big(\sum_{n=1}^{T_x} \mathbb{I}_{\{X_n=y\}}\Big) = \mathbb{E}_x\Big(\sum_{n=1}^{\infty} \mathbb{I}_{\{T_x\geq n\}}\mathbb{I}_{\{X_n=y\}}\Big) \\
&= \sum_{n=1}^{\infty} \mathbb{E}_x\Big(\mathbb{I}_{\{T_x\geq n\}\cap\{X_n=y\}}\Big) = \sum_{n=1}^{\infty} \mathbb{P}_x[T_x \geq n, X_n = y]
\end{aligned}$$

Or si $n = 1$, $\mathbb{P}_x[T_x \geq n, X_n = y] = \mathbb{P}_x[T_x \geq 1, X_1 = y] = p(x,y)$, et si $n \geq 2$,

$$\begin{aligned}
\mathbb{P}_x[T_x \geq n, X_n = y] &= \mathbb{P}_x[X_n = y, X_{n-1} \neq x, \cdots, X_1 \neq x] \\
&= \sum_{z\in E\setminus\{x\}} \mathbb{P}_x[X_n = y, X_{n-1} = z, \cdots, X_1 \neq x], \text{ formule des probabilités totales} \\
&= \sum_{z\in E\setminus\{x\}} \mathbb{P}_x[X_n = y | X_{n-1} = z, \cdots, X_1 \neq x]\mathbb{P}_x[X_{n-1} = z, \cdots, X_1 \neq x] \\
&= \sum_{z\in E\setminus\{x\}} p(z,y)\mathbb{P}_x[X_{n-1} = z, \cdots, X_1 \neq x], \text{ propriété de Markov et homogénéité} \\
&= \sum_{z\in E\setminus\{x\}} p(z,y)\mathbb{P}[T_x \geq n-1, X_{n-1} = z].
\end{aligned}$$

Ainsi, en mettant ceci dans la définition de $\mu_x(y)$, on obtient :

$$\mu_x(y) = p(x,y) + \sum_{n=2}^{\infty} \sum_{z\in E\setminus\{x\}} p(z,y)\mathbb{P}[T_x \geq n-1, X_{n-1} = z]$$

Nous avons,

$$\sum_{n=2}^{\infty} \sum_{z\in E\setminus\{x\}} p(z,y)\mathbb{P}[T_x \geq n-1, X_{n-1}=z] = \sum_{z\in E\setminus\{x\}} p(z,y) \sum_{n=1}^{\infty} \mathbb{P}[T_x \geq n, X_n = z]$$
$$= \sum_{z\in E\setminus\{x\}} p(z,y)\mu_x(z).$$

De plus,

$$\mu_x(x) = \mathbb{E}_x\Big(\sum_{n=1}^{T_x} \mathbb{I}_{\{X_n=x\}}\Big) = \mathbb{P}_x[X_0 = x] = 1.$$

Ainsi, on conclut que

$$\mu_x(y) = p(x,y)\cdot 1 + \sum_{z\in E\setminus\{x\}} p(z,y)\mu_x(z) = \sum_{z\in E} p(z,y)\mu_x(z).$$

Établissons maintenant que, pour tout $y \in E$, $0 < \mu_x(y) < \infty$. Puisque la chaîne de Markov est irréductible il existe $m, n \geq 0$ tels que $P^{(m)}(x,y) = P^m(x,y) > 0$ et $P^{(n)}(y,x) = P^n(y,x) > 0$. De plus, la mesure μ_x étant stationnaire on a, pour tout $n \geq 0$, $\mu_x^t P^n = \mu_x^t$. Ainsi,

$$1 = \mu_x(x) = \sum_{y'\in E} \mu_x(y')P(y',x) \geq \mu_x(y)P^n(y,x) \quad \Rightarrow \quad \mu_x(y) < \infty$$
$$\mu_x(y) = \sum_{z\in E} \mu_x(z)P(z,y) \geq \mu_x(x)P^m(x,y) = P^m(x,y) > 0 \quad \Rightarrow \quad \mu_x(y) > 0. \qquad \square$$

8.3 Unicité

Montrons maintenant que cette mesure stationnaire est unique à une constante multiplicative près.

Théorème 8.8. *Soit une chaîne de Markov irréductible et récurrente. Alors, la mesure stationnaire est unique à une constante multiplicative près.*

Afin de prouver ce théorème, nous démontrons d'abord quelques lemmes.

Lemme 8.9. *Soit une chaîne de Markov irréductible et récurrente. Si une mesure positive satisfait,*

$$\forall\, x \in E, \quad \mu(x) \geq (\mu^t P)(x),$$

alors μ est une mesure stationnaire.

Démonstration. Soit $y \in E$. Pour tout $n \geq 0$, on définit $u_n(y) = \sum_{k=0}^n (\mu^t - \mu^t P)P^k(y)$. Montrons que la suite $(u_n(y))_{n\geq 0}$ est croissante et bornée supérieurement, donc convergente.

$$u_{n+1}(y) - u_n(y) = (\mu^t - \mu^t P)P^{n+1}(y) = \sum_{x\in E} (\mu(x) - (\mu^t P)(x))P^{n+1}(x,y) \geq 0, \text{ par hypothèse}$$
$$u_n(y) = \sum_{k=0}^{n} (\mu^t - \mu^t P)P^k(y) = \sum_{k=0}^{n} [(\mu^t P^k)(y) - (\mu^t P^{k+1})(y)] = \mu(y) - (\mu^t P^{n+1})(y) \leq \mu(y),$$

où dans la troisième égalité de la deuxième ligne on remarque que les termes s'annulent par télescopage et dans la dernière égalité on utilise que la mesure est positive. Ainsi $(u_n(y))_{n\geq 0}$ converge vers,

$$\sum_{k=0}^{\infty}(\mu^t - \mu^t P)P^k(y) = \sum_{k=0}^{\infty}\left(\sum_{x\in E}(\mu - \mu^t P)(x)P^k(x,y)\right) = \sum_{x\in E}\left((\mu - \mu^t P)(x)\sum_{k=0}^{\infty}P^k(x,y)\right).$$

La chaîne étant irréductible récurrente, $\sum_{k=0}^{\infty} P^k(x,y) = G(x,y) = \infty$. De plus, tous les termes de la somme sont positifs par hypothèse. Ainsi, pour que cette suite converge il faut que, pour tout $x \in E$, $(\mu - \mu^t P)(x) = 0$, c'est-à-dire que la mesure μ soit stationnaire. □

Corollaire 8.10. *Soit une chaîne de Markov et μ, ν deux mesures satisfaisant la condition de stationnarité, alors le minimum $\mu \wedge \nu$ des deux mesures :*

$$\forall\, x \in E, \quad \mu \wedge \nu(x) = \min(\mu(x), \nu(x)),$$

est encore une mesure stationnaire.

Démonstration. En utilisant le lemme précédent, il suffit de montrer que, pour tout $x \in E$, $(\mu \wedge \nu)^t P(x) \leq (\mu \wedge \nu)(x)$. On a,

$$(\mu \wedge \nu)^t P(x) = \sum_{y\in E}(\mu \wedge \nu)(y)P(y,x) \leq \min\Big(\sum_{y\in E}\mu(y)P(y,x), \sum_{y\in E}\nu(y)P(y,x)\Big) = \min(\mu(x), \nu(x)),$$

les mesures μ, ν étant invariantes. □

Lemme 8.11. *Soit une chaîne de Markov irréductible et récurrente et μ une mesure stationnaire non identiquement nulle. Alors μ est une mesure strictement positive.*

Démonstration. Par définition une mesure stationnaire est positive ; nous devons donc montrer que μ est strictement positive. Si μ est non identiquement nulle, il existe $x \in E$ tel que $\mu(x) > 0$. Supposons qu'il existe $y \in E$ tel que $\mu(y) = 0$. Comme la chaîne est irréductible récurrente, il existe n tel que $P^n(x,y) > 0$. Ainsi, utilisant la stationnarité,

$$0 = \mu(y) = \sum_{x'\in E}\mu(x')P^n(x',y) \geq \mu(x)P^n(x,y) > 0,$$

ce qui est une contradiction. On conclut que la mesure μ est strictement positive. □

Démonstration du Théorème 8.8. Soit ν une mesure stationnaire (non-identiquement nulle). Montrons que ν est égale, à une constante multiplicative près, à la mesure stationnaire μ_x. D'après le lemme ci-dessus, les mesures ν et μ_x sont strictement positives. On définit la mesure $\tilde{\nu} = C\nu$ où $C = \frac{\mu_x(y)}{\nu(y)}$ pour un état $y \in E$. Alors, $\tilde{\nu}(y) = \mu_x(y)$. On considère la mesure $\tilde{\nu} \wedge \mu_x$; elle est stationnaire d'après le corollaire ci-dessus et il en est de même pour la mesure $\tilde{\nu} - \tilde{\nu} \wedge \mu_x$ (comme différence de deux mesures stationnaires). La mesure $\tilde{\nu} - \tilde{\nu} \wedge \mu_x$ est aussi positive et nulle en y ; ainsi d'après le lemme ci-dessus elle est identiquement nulle, d'où $\tilde{\nu} \leq \mu_x$. Par un argument symétrique, on montre que $\mu_x \leq \tilde{\nu}$ et donc que $C\nu = \tilde{\nu} = \mu_x$. □

Corollaire 8.12. *Soit une chaîne de Markov irréductible et récurrente. Alors,*

- *Soit, pour tout $x \in E$, $\mathbb{E}_x(T_x) < \infty$ et il existe une unique probabilité stationnaire π donnée par,*
$$\forall\, x \in E, \quad \pi(x) = \frac{1}{\mathbb{E}_x(T_x)}.$$
- *Soit, pour tout $x \in E$, $\mathbb{E}_x(T_x) = \infty$ et toute mesure stationnaire a une masse totale infinie.*

Démonstration. Si la chaîne de Markov est irréductible et récurrente, d'après le Théorème 8.8 toutes les mesures stationnaires sont proportionnelles : elles ont toutes soit masse finie, soit masse infinie. Fixons donc $x \in E$ et considérons la mesure stationnaire μ_x. Alors,

$$\begin{aligned}\mu_x(E) &= \sum_{y\in E} \mu_x(y) = \sum_{y\in E} \mathbb{E}_x\Big(\sum_{n=1}^{T_x} \mathbb{I}_{\{X_n=y\}}\Big)\\ &= \mathbb{E}_x\Big(\sum_{n=1}^{T_x}\sum_{y\in E} \mathbb{I}_{\{X_n=y\}}\Big) = \mathbb{E}_x(T_x).\end{aligned}$$

Ainsi, si $\mathbb{E}_x(T_x) < \infty$, on est dans le premier cas et l'unique probabilité stationnaire π satisfait, pour tout $x \in E$, pour tout $y \in E$, $\pi(y) = \frac{\mu_x(y)}{\mathbb{E}_x(T_x)}$. En prenant $x = y$, on obtient pour tout $x \in E$,
$$\pi(x) = \frac{\mu_x(x)}{\mathbb{E}_x(T_x)} = \frac{1}{\mathbb{E}_x(T_x)}.$$
Si $\mathbb{E}_x(T_x) = \infty$, on est dans le deuxième cas. □

Définition 8.13. Soit une chaîne de Markov et x un état récurrent. Alors x est dit *récurrent positif* si $\mathbb{E}_x(T_x) < \infty$, il est dit *récurrent nul* si $\mathbb{E}_x(T_x) = \infty$.

Proposition 8.14. *Soient x et y deux états récurrents qui communiquent. Alors ils sont tous les deux soit récurrents positifs, soit récurrents nuls.*

Démonstration. Si la chaîne de Markov est irréductible, cela découle du Corollaire 8.12. Si ce n'est pas le cas, on applique le même raisonnement à chacune des classes récurrentes. □

Définition 8.15. Une classe d'équivalence récurrente est dite *récurrente positive* si un de ses états est récurrent positif ; elle est dite *récurrente nulle* sinon.

8.4 Caractérisation des chaînes de Markov récurrentes positives

La proposition suivante permet de caractériser les chaînes de Markov irréductibles et récurrentes positives.

Proposition 8.16. *Une chaîne de Markov irréductible est récurrente positive si et seulement si elle admet une probabilité stationnaire. En particulier, toute chaîne de Markov irréductible définie sur un espace d'états de cardinal fini est récurrente positive.*

Démonstration. Démontrons la première partie. Le sens direct provient du Corollaire 8.12. Pour la réciproque, supposons qu'il existe une probabilité stationnaire π et soit $x \in E$ tel que $\pi(x) > 0$. D'après le Lemme 7.12, on a pour tout $x, y \in E$, $G(x,y) \leq G(y,y)$, autrement dit,

$$\sum_{n \geq 0} P^n(x,y) \leq G(y,y).$$

En multipliant par $\pi(x)$, en sommant sur x, en utilisant le théorème de Fubini et le fait que la probabilité π soit stationnaire, on obtient pour le membre de gauche,

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

En multipliant par $\pi(x)$ et en sommant sur x, le membre de droite devient,

$$\pi(E) G(y,y) = G(y,y).$$

Ainsi, d'après l'inégalité, $G(y,y) = \infty$ et on déduit que l'état y est récurrent et donc récurrent positif puisqu'il existe une probabilité stationnaire.

La deuxième partie découle du fait que toute chaîne de Markov irréductible défini sur un espace d'état fini est récurrente, ce qui implique l'existence d'une mesure stationnaire ; et que puisque l'espace des états est fini, toute mesure stationnaire a forcément une masse finie. $\square$

Remarque. L'existence d'une mesure invariante infinie ne permet pas de conclure quand à la récurrence de la marche. Voir par exemple l'Exercice 8.5 où l'on montre que la marche aléatoire sur $\mathbb{Z}$ non-symétrique admet une mesure invariante (de masse infinie) et pourtant cette marche est irréductible et *transiente*.

8.5 Exercices : mesures stationnaires et invariantes

Exercice 8.1. On considère une chaîne de Markov $(X_n)_{n\geq 0}$ à valeurs dans $E = \{1, 2\}$, dont la matrice de transition P est donnée par

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix},$$

où $0 < p, q < 1$.

1. Classifier les états de cette chaîne.
2. Déterminer la probabilité stationnaire.
3. En déduire l'espérance du temps de retour dans l'état 1 sachant qu'elle est partie de l'état 1.

Solution de l'exercice 8.1.

1. Cette chaîne possède une seule classe, elle est donc irréductible. L'espace des états étant fini, on déduit en utilisant un résultat du cours que la chaîne est récurrente positive ; elle admet donc une unique probabilité stationnaire.
2. La probabilité stationnaire satisfait aux équations $\pi^t = \pi^t P$ et $\sum_{x\in E} \pi(x) = 1$. Autrement dit, en notant $\pi^t = (\pi(1), \pi(2))$, on a

$$\begin{cases} \pi(1)(1-p) + \pi(2) = \pi(1) \\ \pi(1)q + \pi(2)(1-q) = \pi(2) \end{cases} \quad \Leftrightarrow \quad p\pi(1) = q\pi(2) \quad \Leftrightarrow \quad \pi(1) = \frac{q}{p}\pi(2)$$

$$\pi(1) + \pi(2) = 1, \text{ d'où } \left(1 + \frac{p}{q}\right)\pi(1) = 1$$

   Ainsi, la probabilité stationnaire est $\pi^t = \left(\frac{q}{p+q}, \frac{p}{p+q}\right)$.
3. D'après un résultat du cours, $\pi(1) = \frac{1}{\mathbb{E}_1(T_1)}$, ainsi l'espérance du temps de retour en 1 sachant que l'on est parti de 1 est $\frac{p+q}{q}$.

Exercice 8.2. [Urne d'Ehrenfest, suite] On reprend l'Exercice 5.8.

1. Classifier les états de la chaîne.
2. Déterminer une probabilité réversible pour cette chaîne. En déduire la probabilité stationnaire.
3. En supposant que $X_0 = \frac{N}{2}$, calculer l'espérance du temps pour revenir dans cet état. Même question si $X_0 = N$, c'est-à-dire si toutes les billes sont dans l'urne A. Application : $N = 12$.

Solution de l'exercice 8.2.

1. Tous les états de la chaîne communiquent, elle est donc irréductible. L'espace des états étant fini, on en déduit d'après un résultat du cours que la chaîne est récurrente positive ; elle admet donc une unique probabilité stationnaire.

2. Une probabilité réversible π satisfait aux équations,

$$\forall (x,y) \in E^2,\ \pi(x)p(x,y) = \pi(y)p(y,x) \text{ et } \sum_{x\in E} \pi(x) = 1.$$

Dans le cas de l'urne d'Ehrenfest, la condition de réversibilité donne,

$$\begin{aligned}\forall x \in \{0,\dots,N-1\}, \quad & \pi(x)p(x,x+1) = \pi(x+1)p(x+1,x)\\ &\Leftrightarrow \pi(x)\frac{N-x}{N} = \pi(x+1)\frac{x+1}{N}\\ &\Leftrightarrow \pi(x+1) = \frac{N-x}{x+1}\pi(x) = \cdots = \frac{(N-x)\dots N}{(x+1)\dots 1}\pi(0) = \binom{N}{x+1}\pi(0).\end{aligned}$$

La condition d'être une probabilité donne,

$$\pi(0)\sum_{x=0}^{N}\binom{N}{x} = 1 \quad \Leftrightarrow \quad \pi(0)2^N = 1.$$

Ainsi, la probabilité réversible est égale à,

$$\forall x \in \{0,\dots,N\}, \quad \pi(x) = \frac{\binom{N}{x}}{2^N}.$$

D'après le cours, il s'agit aussi de la probabilité stationnaire qui suit donc la loi binomiale de paramètres $(N, \frac{1}{2})$.

3. D'après le cours, pour tout $x \in \{0,\dots,N\}$,

$$\mathbb{E}_x(T_x) = \frac{1}{\pi(x)} = \frac{2^N}{\binom{N}{x}}.$$

Ainsi, on a $\mathbb{E}_{N/2}(T_{N/2}) = \frac{2^N}{\binom{N}{N/2}} \sim \sqrt{\frac{\pi N}{2}}$, et $\mathbb{E}_N(T_N) = 2^N$.

Application $N = 12$: $\mathbb{E}_{N/2}(T_{N/2}) = \frac{4096}{924} \simeq 4.43$, et $\mathbb{E}_N(T_N) = 4096$.

Exercice 8.3. [Qui perd, perd tout, suite]

On reprend la chaîne de Markov de l'Exercice 7.8.

1. Trouver une probabilité stationnaire pour cette chaîne de Markov.
2. En déduire que cette chaîne est récurrente positive et donner l'espérance du temps de retour en 0 sachant qu'elle est partie de 0.
3. Retrouver ces résultats en utilisant les réponses aux questions de l'Exercice 7.8.

Solution de l'exercice 8.3.

1. Rappelons que l'espace des états est $\mathbb{N}$. Une probabilité stationnaire π satisfait à $\pi^t = \pi^t P$ et $\sum_{x=0}^{\infty}\pi(x) = 1$. Ceci donne :

$$\begin{cases}\forall\, y \geq 1,\ \pi(y-1)p = \pi(y)\\ \left(\sum_{x=0}^{\infty}\pi(x)\right)(1-p) = \pi(0)\end{cases} \quad \Leftrightarrow \quad \begin{cases}\forall\, y \geq 1,\ \pi(y) = \pi(0)p^y\\ (1-p) = \pi(0),\ \text{car } \pi \text{ est une probabilité.}\end{cases}$$

Ainsi, pour tout $y \geq 0$, $\pi(y) = (1-p)p^y$. On vérifie qu'il s'agit bien d'une probabilité (loi géométrique de paramètre $1-p$ sur $\mathbb{N}$).

2. Cette chaîne de Markov est irréductible. Comme elle admet une probabilité stationnaire, d'après la Proposition 8.16, on en déduit qu'elle est récurrente positive. D'après le Corollaire 8.12, on déduit aussi que :
$$\mathbb{E}_0(T_0) = \frac{1}{\pi(0)} = \frac{1}{1-p}.$$
3. Dans l'Exercice 7.8, on a montré que cette chaîne de Markov est irréductible et récurrente. D'après le Corollaire 8.12, pour montrer qu'elle est récurrente nulle il suffit de voir que $\mathbb{E}_0(T_0) < \infty$. Or nous avons montré que la loi conditionnelle de T_0 sachant que l'on part de l'état 0 est géométrique sur $\mathbb{N}^*$ de paramètre $1-p$. Ainsi, on a $\mathbb{E}_0(T_0) = \frac{1}{1-p} < \infty$.

Exercice 8.4. [Marche aléatoire sur un graphe]

On considère un graphe $G = (V, E)$ fini, simple et connexe. A partir du graphe G, on définit un graphe orienté G' où les sommets de G' sont ceux de G, et chaque arête est remplacée dans G' par deux arêtes orientées. A chaque arête orientée (x, y) de G', on associe une probabilité de transition $p(x, y) = \frac{1}{d_x}$, où d_x est le degré du sommet x dans G. Le graphe G' pondéré est alors celui d'une chaîne de Markov $(X_n)_{n\geq 0}$ définie sur l'espace des états V, de matrice de transition $P = (p(x, y))$. Il s'agit de la *marche aléatoire sur le graphe G*.

1. Classifier les états de cette chaîne.
2. Montrer que $\sum_{x\in V} d_x = 2\,\mathrm{card}(E)$.
3. Montrer que la mesure μ définie par, pour tout $x \in V$, $\mu(x) = d_x$ est une mesure réversible. En déduire la probabilité stationnaire de cette chaîne.

Solution de l'exercice 8.4.

1. Le graphe étant connexe, la chaîne de Markov est irréductible. L'espace des états est l'ensemble V des sommets du graphe, il donc de cardinal fini. On en déduit, d'après un résultat du cours, que la chaîne est récurrente positive.
2. Pour tout sommet x, le degré d_x du sommet x représente le nombre d'arêtes incidentes à x. En remarquant que chaque arête est formée de deux sommets, on a qu'en sommant d_x sur tous les sommets, on compte chaque arête deux fois exactement, d'où l'égalité.
3. La mesure μ est clairement positive et $\mu(x) < \infty$ pour tout sommet x. Elle est réversible :
$$\forall\,(x, y) \in V^2,\ x \sim y, \quad \mu(x)p(x, y) = \mu(y)p(y, x), \quad \text{car} \quad d_x\frac{1}{d_x} = d_y\frac{1}{d_y}.$$
La mesure μ étant réversible, elle est stationnaire. Il faut la normaliser à 1 pour en faire une probabilité (qui existe et est unique car la chaîne est récurrente positive). Pour cela, on divise μ par sa masse. En utilisant, la question 2, on déduit que la probabilité stationnaire π est donnée par,
$$\forall\, x \in V, \quad \pi(x) = \frac{\mu(x)}{\sum_{x\in V}\mu(x)} = \frac{d_x}{2\,\mathrm{card}(E)}.$$

Exercice 8.5. [Marche aléatoire sur $\mathbb{Z}$, suite]

On reprend l'Exercice 5.4. On se rappelle que la chaîne est irréductible. Dans les questions 1-4, on n'utilise pas les résultats montrés dans l'Exercice 7.7.

1. Montrer que la mesure μ définie par, pour tout $x \in \mathbb{Z}$, $\mu(x) = \left(\frac{p}{1-p}\right)^x$ est réversible. En déduire une mesure stationnaire.
2. Montrer que la mesure $\nu \equiv 1$ est stationnaire.
3. Montrer que les mesures μ et ν sont égales à une constante multiplicative près si et seulement si $p = \frac{1}{2}$.
4. En déduire que si $p \neq \frac{1}{2}$, la marche aléatoire sur $\mathbb{Z}$ est transiente. On retrouve le résultat montré dans l'Exercice 7.7.
5. En utilisant le fait, démontré dans l'Exercice 7.7, que la marche aléatoire est récurrente si $p = \frac{1}{2}$, montrer qu'elle est récurrente nulle.

Solution de l'exercice 8.5.

1. La mesure μ est clairement positive et $\mu(x) < \infty$ pour tout sommet x. Elle est réversible :
$$\forall\, x \in \mathbb{Z}, \quad \mu(x)p(x, x+1) = \mu(x+1)p(x+1, x) \quad \text{car} \quad \left(\frac{p}{1-p}\right)^x p = \left(\frac{p}{1-p}\right)^{x+1}(1-p).$$
D'après un résultat du cours, cette mesure étant réversible elle est stationnaire.
2. La mesure ν est positive et finie pour tout sommet. Pour montrer la stationnarité de ν, nous devons voir que $\nu^t = \nu^t P$ autrement dit que $\nu = P^t \nu$. Comme $\nu \equiv 1$, ceci revient à voir que la matrice P est bi-stochastique, autrement dit que ses colonnes somment aussi à 1. Chaque colonne a comme seuls termes non-nuls p et $1-p$ donc c'est bien le cas.
3. Les mesures μ et ν sont égales à une constante multiplicative C près ssi
$$\forall\, x \in \mathbb{Z}, \quad \left(\frac{p}{1-p}\right)^x = C.$$
Ceci est vrai ssi $p = 1-p$ et alors $C = 1$, autrement dit ssi $p = \frac{1}{2}$.
4. Si $p \neq \frac{1}{2}$, il y a deux mesures stationnaires qui ne sont pas égales à une constante près. Ceci implique que la marche aléatoire est transiente, sinon on aurait une contradiction avec le Théorème 8.8.
5. Comme la marche aléatoire est irréductible et récurrente, d'après le Théorème 8.8, elle a une unique mesure stationnaire à constante près. Pour montrer que la marche aléatoire est récurrente nulle, il suffit de voir qu'une mesure stationnaire a masse totale infinie. Ceci est bien le cas pour la mesure $\nu \equiv 1$ ci-dessus (égale à la mesure μ dans ce cas) car le graphe $\mathbb{Z}$ est infini.

8.6 Convergence vers l'équilibre

On considère une chaîne de Markov $(X_n)_{n\geq 0}$ irréductible et récurrente positive, elle possède donc une unique probabilité invariante π. On a vu que si la loi initiale de la chaîne est la probabilité stationnaire π, alors la loi reste la même à chaque instant n. La chaîne est en régime d'équilibre, autrement dit en *régime stationnaire*. Si maintenant la loi initiale n'est pas la probabilité stationnaire, que se passe-t-il ? Le théorème suivant montre que si la chaîne est de plus *apériodique*, alors asymptotiquement la loi de la chaîne de Markov est la loi stationnaire.

Théorème 8.17. *Soit $(X_n)_{n\geq 0}$ une chaîne de Markov irréductible, récurrente positive et* apériodique. *Alors pour tout état $x \in E$, et toute loi initiale μ_0,*

$$\lim_{n\to\infty} \mathbb{P}[X_n = x] = \pi(x),$$

où π est la probabilité stationnaire de la chaîne.

Afin de démontrer ce théorème, nous montrons d'abord un lemme.

Lemme 8.18. *Soit $(X_n)_{n\geq 0}$ une chaîne de Markov irréductible et apériodique. Alors, pour tout $(x,y) \in E^2$, il existe $N \in \mathbb{N}$ tel que $\forall\, n \geq N$, $P^{(n)}(x,y) > 0$.*

Démonstration. Étant donné que la chaîne est irréductible, il suffit de montrer que, pour tout $x \in E$, il existe $N \in \mathbb{N}$ tel que $\forall\, n \geq N$, $P^{(n)}(x,x) > 0$.

La chaîne étant apériodique, il existe $\ell \geq 1$ et $n_1, \dots, n_\ell \geq 1$, tels que $\mathrm{pgcd}(n_1, \dots, n_\ell) = 1$ et $P^{(n_1)}(x,x) > 0, \dots, P^{(n_\ell)}(x,x) > 0$.

D'après le Théorème de Bezout, si $\mathrm{pgcd}(n_1, \dots, n_\ell) = d$, pour tout $n \geq n_1 \dots n_\ell$, il existe $a_1, \dots, a_\ell \in \mathbb{N}$ tels que

$$a_1 n_1 + \dots a_\ell n_\ell = nd.$$

Dans notre cas on a $d = 1$. Posons $N = n_1 \dots n_\ell$, alors pour tout $n \geq N$, il existe $a_1, \dots, a_\ell \in \mathbb{N}$ tel que $a_1 n_1 + \dots a_\ell n_\ell = n$. On conclut en utilisant les équations de Chapman Kolmogorov qui impliquent que :

$$P^{(n)}(x,x) \geq [P^{(n_1)}(x,x)]^{a_1} \dots [P^{(n_\ell)}(x,x)]^{a_\ell} > 0. \qquad \square$$

Nous pouvons maintenant démontrer le théorème.

Démonstration du Théorème 8.17. Supposons démontré que

$$\forall\, (x,y) \in E^2, \quad \lim_{n\to\infty} \mathbb{P}_x[X_n = y] = \pi(y). \tag{8.1}$$

Alors, pour tout $y \in E$, et pour toute loi initiale μ_0,

$$\lim_{n\to\infty} \mathbb{P}[X_n = y] = \sum_{x\in E} \mu_0(x) \lim_{n\to\infty} \mathbb{P}_x[X_n = y] = \Big(\sum_{x\in E} \mu_0(x)\Big)\pi(y) = \pi(y).$$

Montrons donc l'Équation (8.1). Pour cela nous utilisons un argument de couplage. On introduit une chaîne de Markov $(Y_n = (X_n, X'_n))_{n\geq 0}$ définie sur l'espace des états E^2, formée de deux copies indépendantes X_n, X'_n de X_n. Montrons que la chaîne $(Y_n)_{n\geq 0}$ est irréductible et récurrente positive. Étant donné $(x,x') \in E^2$, notons $\mathbb{P}_{(x,x')}$ la probabilité conditionnelle de Y_n sachant que $\{(X_0, X'_0) = (x,x')\}$.

<u>$(Y_n)_{n\geq 0}$ est irréductible.</u> En utilisant l'indépendance de X_n et X'_n, les transitions de la chaîne $(Y_n)_{n\geq 0}$ sont données par, pour tout (x,x'), $(y,y') \in E^2$,

$$\mathbb{P}_{(x,x')}(Y_n = (y,y')) = \mathbb{P}_x[X_n = y]\mathbb{P}_{x'}[X'_n = y'] = P^{(n)}(x,y)P^{(n)}(x',y').$$

Ainsi, en utilisant le Lemme 8.18 pour chacune des chaînes X_n, X'_n, on a $\mathbb{P}_{(x,x')}(Y_n = (y,y')) > 0$ pour tout $n \geq \max\{N, N'\}$. Ainsi, la chaîne $(Y_n)_{n\geq 0}$ est bien irréductible.

$(Y_n)_{n\geq 0}$ est récurrente positive. D'après la Proposition 8.16, il suffit de voir que la chaîne possède une probabilité stationnaire. La chaîne $(X_n)_{n\geq 0}$ étant récurrente positive, elle a une probabilité stationnaire, notée π. Définissons la probabilité produit $\bar{\pi}$ sur E^2 par :

$$\forall\,(x,x') \in E^2, \quad \bar{\pi}(x,x') = \pi(x)\pi(x').$$

Alors, par indépendance de X_n et X'_n, et par stationnarité de π pour $(X_n)_{n\geq 0}$,

$$\begin{aligned}\sum_{(x,x')\in E^2} \bar{\pi}(x,x')\mathbb{P}_{(x,x')}[Y_1 = (y,y')] &= \Big(\sum_{x\in E} \pi(x)\mathbb{P}_x[X_n = y]\Big)\Big(\sum_{x'\in E} \pi(x')\mathbb{P}_{x'}[X_n = y']\Big)\\ &= \pi(y)\pi(y') = \bar{\pi}(y,y').\end{aligned}$$

Comme la chaîne $(Y_n)_{n\geq 0}$ est irréductible et récurrente positive on a, pour tout $(x,x'),(y,y') \in E^2$, $\mathbb{P}_{(x,x')}[T_{(y,y')} < \infty] = 1$, d'où $\lim_{n\to\infty} \mathbb{P}_{(x,x')}[T_{(y,y')} > n] = 0$.
Argument de couplage. On fixe $c \in E$ et on considère,

$$T := T_{(c,c)} = \inf\{n > 0 : Y_n = (c,c)\} = \inf\{n > 0 : X_n = X'_n = c\}.$$

Remarquons que, pour tout $n \geq 0$, les deux chaînes $(X_{T+n})_{n\geq 0}$ et $(X'_{T+n})_{n\geq 0}$ ont même loi car elles ont le même état initial c et la même matrice de transitions. On peut donc écrire, pour tout $x, x', y \in E$,

$$\begin{aligned}\mathbb{P}_x[X_n = y] &= \mathbb{P}_{(x,x')}[X_n = y]\\ &= \mathbb{P}_{(x,x')}[X_n = y, T \leq n] + \mathbb{P}_{(x,x')}[X_n = y, T > n]\\ &= \mathbb{P}_{(x,x')}[X'_n = y, T \leq n] + \mathbb{P}_{(x,x')}[X_n = y, T > n]\\ &\leq \mathbb{P}_{(x,x')}[X'_n = y] + \mathbb{P}_{(x,x')}[T > n]\\ &= \mathbb{P}_{x'}[X'_n = y] + \mathbb{P}_{(x,x')}[T > n].\end{aligned}$$

Ainsi,

$$|\mathbb{P}_x[X_n = y] - \mathbb{P}_{x'}[X'_n = y]| \leq \mathbb{P}_{(x,x')}[T > n] \overset{n\to\infty}{\longrightarrow} 0.$$

On en déduit que, pour tout $(x,y) \in E^2$,

$$\begin{aligned}|\pi(y) - \mathbb{P}_x[X_n = y]| &= \left|\sum_{x'\in E} \pi(x')(\mathbb{P}_{x'}[X'_n = y] - \mathbb{P}_x[X_n = y])\right|\\ &\leq \sum_{x'\in E} \pi(x')\left|\mathbb{P}_{x'}[X'_n = y] - \mathbb{P}_x[X_n = y]\right|\\ &\leq \sum_{x'\in E} \pi(x')\mathbb{P}_{(x,x')}[T > n] \overset{n\to\infty}{\longrightarrow} 0.\end{aligned}$$

□

8.7 Exercices : convergence vers l'équilibre

Exercice 8.6. On reprend les Exercices 6.2 (dépenses énergétiques) et 6.3 (Casino/marche aléatoire sur $\mathbb{Z}/3\mathbb{Z}$). Sans calculer les puissances n-ième de la matrice de transition, déterminer la limite de la loi à l'instant n de la chaîne de Markov associée à chacun de ces deux exemples.

Solution de l'exercice 8.6. Les chaînes de Markov de ces deux exercices sont irréductibles et récurrentes positives ; elles ont donc chacune une unique probabilité stationnaire π. Elles sont de plus apériodiques, d'où d'après le Théorème 8.17, on a

$$\forall\, x, \quad \lim_{n\to\infty} \mathbb{P}[X_n = x] = \pi(x).$$

- (Dépenses énergétiques). On est dans le cas particulier de l'Exercice 8.1 avec $p = \frac{1}{2}$ et $q = \frac{3}{4}$. Ainsi,
$$\pi(1) = \frac{q}{p+q} = \frac{3}{5}, \quad \pi(2) = \frac{p}{p+q} = \frac{2}{5}.$$
On a donc, $\lim_{n\to\infty} \mathbb{P}[X_n = 1] = \pi(1) = \frac{3}{5}$ et on retrouve le Point 2. de l'Exercice 6.2 (sans calculer P^n).
- (Casino/marche aléatoire). On est dans le cas particulier de l'Exercice 8.4 où le graphe $G = (V, E)$ est un triangle. Ainsi, pour tout $x \in V$, $d_x = 2$, $\mathrm{card}(E) = 3$, et
$$\forall\, x \in V, \quad \pi(x) = \frac{d_x}{2\,\mathrm{card}(E)} = \frac{1}{3}.$$
On retrouve le Point 3. de l'Exercice 6.3 (sans calculer P^n).

Bibliographie

[CGCCT16] M. Cottrell, V. Genon-Catalot, Duhamel C., and Meyre T. *Exercices de probabilités.* Cassini, 2016.

[Del10] Jean-François Delmas. *Introduction au calcul des probabilités et à la statistique.* Les presses de l'ENSTA, `http://cermics.enpc.fr/~delmas/Enseig/ensta_cours.pdf`, 2010.

[Fel68] William Feller. *An introduction to probability theory and its applications. Vol. I.* Third edition. John Wiley & Sons Inc., New York, 1968.

[Fra13] Jacques Franchi. *Processus aléatoires à temps discret.* Ellipses, 2013.

[Lel09] Marc Lelarge. *Notes de cours ENS : chaînes de Markov.* http ://www.di.ens.fr/ lelarge/proba09/ENSmarkov.pdf, 2009.

[LG06] Jean-François Le Gall. *Cours de l'ENS : Intégration, Probabilités et Processus Aléatoires.* https ://www.math.u-psud.fr/ jflegall/IPPA2.pdf, 2006.

[Mar16] Bernard Marchal. *Notes de cours ESIAG : Processus stochastiques.* 2016.

[PR12] Thérèse Phan and Jean-Pierre Rowenczyk. *Exercices et problèmes. Statistique et Probabilités.* Sciences sup. Dunod, 2012.

[Pri05] Pierre Priouret. Polycopié du cours de Probabilités de L3. `http://www.proba.jussieu.fr/cours/proba_L_priouret.pdf`, 2004-2005.

[Shi96] A. N. Shiryaev. *Probability*, volume 95 of *Graduate Texts in Mathematics.* Springer-Verlag, New York, second edition, 1996. Translated from the first (1980) Russian edition by R. P. Boas.

[Vel14] Yvan Velenik. Polycopié du cours de Probabilités et Statistique. `http://www.unige.ch/math/folks/velenik/Cours/2013-2014/ProbaStat/probastat.pdf`, 2014.