

MARTINGALES EN TEMPS DISCRET ET CHAINES DE MARKOV

Nizar TOUZI

Ecole Polytechnique Paris
Département de Mathématiques Appliquées

nizar.touzi@polytechnique.edu

Septembre 2009

Table des matières

1	Préliminaires de la théorie des mesures	11
1.1	Espaces mesurables et mesures	11
1.1.1	Algèbres, σ -algèbres	11
1.1.2	Mesures	12
1.1.3	Propriétés élémentaires des mesures	13
1.2	L'intégrale de Lebesgue	15
1.2.1	Fonction mesurable	15
1.2.2	Intégration des fonctions positives	16
1.2.3	Intégration des fonctions réelles	19
1.2.4	De la convergence p.p. à la convergence \mathcal{L}^1	19
1.2.5	Intégrale de Lebesgue et intégrale de Riemann	21
1.3	Transformées de mesures	22
1.3.1	Mesure image	22
1.3.2	Mesures définies par des densités	22
1.4	Inégalités remarquables	23
1.5	Espaces produits	24
1.5.1	Construction et intégration	24
1.5.2	Mesure image et changement de variable	26
1.6	Annexe du chapitre 1	27
1.6.1	π -système, d -système et unicité des mesures	27
1.6.2	Mesure extérieure et extension des mesures	28
1.6.3	Démonstration du théorème des classes monotones	31
2	Préliminaires de la théorie des probabilités	33
2.1	Variables aléatoires	33
2.1.1	σ -algèbre engendrée par une v.a.	33
2.1.2	Distribution d'une v.a.	34
2.2	Espérance de variables aléatoires	35
2.2.1	Variables aléatoires à densité	35
2.2.2	Inégalité de Jensen	36
2.2.3	Fonction caractéristique	37
2.3	Espaces \mathcal{L}^p et convergences fonctionnelles des variables aléatoires	39
2.3.1	Géométrie de l'espace \mathcal{L}^2	39

2.3.2	Espaces \mathcal{L}^p et \mathbb{L}^p	40
2.3.3	Espaces \mathcal{L}^0 et \mathbb{L}^0	41
2.3.4	Lien entre les convergences \mathbb{L}^p , en proba et p.s.	42
2.4	Convergence en loi	45
2.4.1	Définitions	45
2.4.2	Caractérisation de la convergence en loi par les fonctions de répartition	46
2.4.3	Convergence des fonctions de répartition	47
2.4.4	Convergence en loi et fonctions caractéristiques	48
2.5	Indépendance	49
2.5.1	σ -algèbres indépendantes	49
2.5.2	Variables aléatoires indépendantes	50
2.5.3	Asymptotique des suites d'événements indépendants	51
2.5.4	Moyennes de variables indépendantes	52
3	Espérance conditionnelle	55
3.1	Premières intuitions	55
3.1.1	Espérance conditionnelle en espace d'états fini	55
3.1.2	Cas des variables à densités	56
3.2	Définition et premières propriétés	57
3.3	Propriétés de l'espérance conditionnelle	59
3.4	Application au filtre de Kalman-Bucy	61
3.4.1	Lois conditionnelles pour les vecteurs gaussiens	62
3.4.2	Filtre de Kalman-Bucy	63
4	Processus aléatoires et structure d'information	67
4.1	Processus aléatoire	67
4.2	Temps d'arrêt	68
4.3	Information à un temps d'arrêt et conditionnement	69
5	Chaînes de Markov : premières définitions	71
5.1	Premières définitions	71
5.1.1	Dynamique markovienne	71
5.1.2	Distributions marginales, espérance	72
5.1.3	Propriété de Markov forte	74
5.1.4	Chaînes de Markov homogènes	75
5.2	Exemples	75
5.3	Problème de Dirichlet et chaînes de Markov homogènes	81
5.3.1	Problème de Dirichlet	81
5.3.2	Temps d'atteinte d'un ensemble par une chaîne de Markov	81
5.3.3	Représentation stochastique de la solution du problème de Dirichlet	82
5.3.4	Application au problème de ruine du joueur	84
5.3.5	Monte Carlo pour le problème de Dirichlet	86
5.4	Temps de retours et excursions	88

6	Lois invariantes et classification des états	91
6.1	Loi invariante	91
6.2	Loi invariante en espace d'état fini	92
6.3	Loi invariante pour les chaînes de Markov réversibles	94
6.4	Existence en espace d'état dénombrable	95
6.5	Classes fermées irréductibles	98
6.6	Récurrence positive et loi invariante	100
6.7	Application aux marches aléatoires	102
6.7.1	Marche aléatoire au plus proche voisin sur \mathbb{Z}	102
6.7.2	Marche aléatoire unidimensionnelle générale	103
6.7.3	Marche aléatoire symétrique sur \mathbb{Z}^d	104
6.8	Autres exemples	106
7	Théorèmes ergodiques pour les chaînes de Markov	111
7.1	Théorèmes ergodiques	111
7.2	Convergences des lois marginales et apériodicité	116
7.2.1	Apériodicité	116
7.2.2	Convergence et vitesse de convergence des lois marginales	118
7.2.3	Application à l'algorithme PageRank de Google	120
7.3	Algorithme de Hastings-Metropolis	123
7.3.1	Description de l'algorithme	123
7.3.2	Simulation de la loi de Gibbs	125
7.4	Recuit simulé	127
7.4.1	Description de la méthode	127
7.4.2	Un résultat de convergence	127
7.4.3	Application au voyageur de commerce	130
8	Martingales en temps discret	133
8.1	Martingales et temps d'arrêt	133
8.2	Martingales fermées	136
8.3	Inégalités de martingales	137
8.4	Décomposition des surmartingales	138
8.5	Martingales locales	139
9	Convergence des martingales	143
9.1	Convergence des martingales de carré intégrable	143
9.2	Loi des Grands Nombres	145
9.3	Convergence des sous-martingales	146
9.4	Théorème central limite martingale	150
9.5	Application : l'algorithme de Robbins-Monro	151
9.5.1	Méthode itérative de recherche de racine	151
9.5.2	L'algorithme de Robins-Monro	152

10 Introduction aux processus de branchement	157
10.1 Processus de Galton-Watson	157
10.2 Propriétés de la fonction génératrice	158
10.3 Loi de reproduction géométrique	159
10.4 Probabilité d'extinction et classification	160
10.5 Comportement asymptotique	163
11 Arrêt optimal	165
11.1 Arrêt optimal en horizon fini	165
11.2 Exemples en horizon fini	167
11.2.1 Le problème du parking	167
11.2.2 Problème de la secrétaire	168
11.3 Arrêt optimal en horizon infini	170
11.4 Application aux options américaines perpétuelles	172
12 Introduction aux mathématiques financières	177
12.1 Contrats d'options et principe de non domination	178
12.1.1 Options européennes et américaines	178
12.1.2 Principe de non domination et premières propriétés	179
12.1.3 Parité call-put	181
12.1.4 Bornes sur les prix des calls et exercice anticipé	181
12.1.5 Quelques exemples populaires de produits dérivés	182
12.2 Du modèle binomial à la formule de Black-Scholes	183
12.2.1 Modèle binomial à une période	183
12.2.2 Le modèle de Cox-Ross-Rubinstein	185
12.2.3 Limite en temps continu	188
12.3 Evaluation et couverture dans un modèle général en temps discret	190
12.3.1 Formulation du modèle	190
12.3.2 Probabilités neutres au risque	191
12.3.3 Evaluation et couverture	192

Introduction, guide de lecture

Ces notes de cours sont à l'intention des élèves de deuxième année de l'Ecole Polytechnique. Le cours introduit aux modèles dynamiques aléatoires en temps discret et leurs diverses utilisations dans de nombreux domaines d'application : la physique, la dynamique des populations, la génétique, les télécommunications, l'économie et la finance. La description suivante du contenu constitue un guide de lecture.

La structure générale du cours suit celle introduite par Michel Benaïm et Nicole El Karoui qui ont conçu cet enseignement à l'Ecole Polytechnique. Ainsi leur livre [4] est une référence importante dont s'inspire l'ensemble de ces notes.

1 Les deux premiers chapitres récapitulent les éléments essentiels de la théorie de la mesure et des probabilités, dont certains ont été vus dans les cours de mathématiques et de mathématiques appliquées de première année. Ils sont abordés très rapidement en cours. Je suis conscient que ce résumé trop rapide peut paraître assez aride. L'ensemble des résultats qui y sont regroupés peuvent servir de référence tout le long des chapitres suivants, donnant un caractère auto-suffisant à ces notes de cours.

Les élèves sont ainsi appelés à étudier ces deux premiers chapitres, tout en veillant à ne pas s'y perdre, car ils ne constituent qu'un préliminaire au cours. Voici les résultats qu'on devrait absolument retenir :

- les théorèmes de convergence monotone, dominée, et le lemme de Fatou,
- les inégalités de Markov, Chebysev, Cauchy-Schwarz, Hölder, Minkowsky, et de Jensen,
- la complétude des espaces \mathbb{L}^p , $p \geq 1$,
- le théorème de Fubini,
- la notion d'indépendance, la loi des grands nombres et le théorème central limite.

Les références essentielles pour ces deux premiers chapitres sont les excellents livres de David Williams [15] et de Jean Jacod et Philip Protter [9] qui contiennent plus de résultats et d'exemples pour ceux qui souhaitent approfondir le sujet.

2 Les chapitres 3 et 4 complètent les deux premiers chapitres par les notions essentielles afin d'aborder les modèles dynamiques aléatoires : l'espérance conditionnelle, les filtrations et les temps d'arrêt. On devrait retenir essentiellement que tous les résultats énoncés dans les deux premiers chapitres sont valables sous la forme conditionnée, et que les résultats valables avec un temps déterministe s'étendent aux temps aléatoires qui sont des temps d'arrêt, avec quelques précautions...

Une première application aborde le problème de filtrage dans le cadre gaussien. Il s'agit du filtre de Kalman-Bucy qui est utilisé en traitement du signal et en statistique pour de nombreuses applications. Ce problème est par ailleurs une excellente opportunité pour réviser les connaissances concernant les vecteurs gaussiens.

La référence essentielle pour ces deux chapitres est le livre de Williams [15].

3 Les chapitres 5, 6 et 7 constituent la première partie du cours, après les chapitres de préliminaires précédents. Il s'agit de l'étude des chaînes de Markov à espace d'état au plus dénombrable. Ce cadre permet d'ores et déjà d'aborder plusieurs exemples intéressants dans diverses applications. Remarquons que cette partie aurait pu être présentée sans référence à la notion de conditionnement, et en utilisant au minimum les outils probabilistes. Cependant, j'espère que la manipulation de ces outils dans le cadre simple des chaînes de Markov permettra aux élèves d'en acquérir une intuition plus forte et de s'en familiariser.

Le chapitre 5 donne les définitions essentielles ainsi que les premières propriétés. La représentation du problème de Dirichlet à l'aide des chaînes de Markov introduit la première utilisation des méthodes de simulation pour l'approximation de la solution. Puis, nous abordons la notion importante de loi invariante dans le chapitre 6. L'existence est toujours satisfaite en dimension finie, et fait appel aux notions de transience, de récurrence nulle, et de récurrence positive dans le cadre dénombrable. L'unicité requiert une classification préalable des états de la chaîne et introduit la notion importante d'irréductibilité.

La référence pour l'ensemble de ces résultats, de nature algébrique, est le livre de Carl Graham [8], qui contient beaucoup plus de résultats, d'exemples et d'exercices (corrigés). Les résultats extraits de cet ouvrage sont exprimés en utilisant le langage probabiliste.

Les propriétés asymptotiques des chaînes de Markov sont étroitement liées aux propriétés des excursions, i.e. l'observation de la chaîne entre deux points de passage par le même point. L'indépendance de ces excursions et leur identité en loi ouvre la porte à l'application de la loi des grands nombres et du théorème central limite, et permet d'obtenir des résultats asymptotiques généraux. L'approche de ces résultats est tirée du livre de Jean-François Delmas et Benjamin Jourdain [6] qui contient des applications plus avancées des chaînes de Markov.

Enfin, la convergence des lois marginales nécessite l'introduction de la propriété supplémentaire d'apériodicité. Des vitesses de convergence exponentielles sont obtenues sous la condition de Doeblin, qui est automatiquement vérifiée en dimension finie. Une étude plus détaillée conduirait à la notion de gap spec-

tral et de formes de Dirichlet que nous n'aborderons pas dans ces notes, nous renvoyons les élèves intéressés à [8].

Cette partie contient de nombreuses applications à divers domaines d'ingénierie. La description de l'algorithme PageRank de Google est peut être la plus frappante par sa simplicité, son caractère actuel, et son utilisation universelle.

4 La deuxième partie de ces notes de cours, contenue dans les chapitres 8 et 9, aborde la théorie des martingales en temps discret. Contrairement à la partie précédente, l'outil probabiliste et la notion d'espérance conditionnelle jouent maintenant un rôle fondamental, et ne peuvent être contournés. Dans cette partie, la notion de dynamique de l'information devient essentielle, et sa modélisation mathématique est mieux précisée. Les résultats essentiels à retenir concernent les théorèmes de convergence des martingales, qui permettent en particulier d'obtenir une démonstration simple de la loi forte des grands nombres pour des variables indépendantes identiquement distribuées intégrables. Enfin cette partie permet d'initier les élèves au cadre des modèles stochastiques en temps continu qui seront abordés dans certains cours de troisième année en vue des applications à la finance et à la biologie.

Les références essentielles pour ces deux chapitres sont les livres de David Williams [15] et de Jean Jacod et Philip Protter [9].

Les trois derniers chapitres présentent des applications à des domaines où la théorie des martingales et les chaînes de Markov jouent un rôle central :

- Le chapitre 10 est une introduction très simpliste aux modèles de dynamique des populations. Il présente les premiers résultats pour les processus de Galton-Watson. La référence essentielle pour ce chapitre est le livre de Athreya et Ney [1] et celui de Michel Benaïm et Nicole El Karoui [4].

- Le chapitre 11 donne les éléments essentiels de la théorie de l'arrêt optimal en horizon fini et infini. On y reporte l'exemple classique du problème de la secrétaire, ainsi qu'une application au problème d'évaluation des options américaines en finance. La référence principale pour ce chapitre est le livre de Jacques Neveu [13].

- Enfin, le chapitre 12 fournit une introduction aux mathématiques financières. Ce domaine nécessite un coût d'entrée non négligeable afin de comprendre la nature des problèmes qui y sont posés et de se familiariser avec le vocabulaire qui y est pratiqué. Une grande partie du chapitre est dédiée à ces aspects, et est essentiellement tirée de l'article de Robert Merton [12]. Nous développons ensuite la théorie de l'évaluation et de couverture dans le modèle le plus simple de l'arbre binomial qui permet, par passage à la limite temps continu, d'obtenir la formule de Fisher Black et Myron Scholes [2].

Pour finir, un Grand Merci à Arnaud Guillin, Jean-François Delmas et Thierry Bodineau pour leurs commentaires et corrections de la première version de ces notes de cours.

Bonne lecture !

Chapitre 1

Préliminaires de la théorie des mesures

1.1 Espaces mesurables et mesures

Dans toute cette section, Ω désigne un ensemble quelconque, et $\mathcal{P}(\Omega)$ est l'ensemble des toutes ses parties.

1.1.1 Algèbres, σ -algèbres

Définition 1.1. Soient $\mathcal{A}_0, \mathcal{A} \subset \mathcal{P}(\Omega)$. On dit que

- (i) \mathcal{A}_0 est une algèbre sur Ω si \mathcal{A}_0 contient Ω et est stable par passage au complémentaire et par réunion.
- (ii) \mathcal{A} est une σ -algèbre si c'est une algèbre stable par union dénombrable. On dit alors que (Ω, \mathcal{A}) est un espace mesurable.

Notons qu'une algèbre doit aussi contenir \emptyset , et est stable par intersection et par différence symétrique, i.e.

$$A \cap B \text{ et } A \Delta B := (A \cup B) \setminus (A \cap B) \in \mathcal{A} \quad \text{pour tous } A, B \in \mathcal{A}_0,$$

et qu'une σ -algèbre est stable par intersection dénombrable. $\mathcal{P}(\Omega)$ est la plus grande σ -algèbre sur Ω . Il s'avère cependant que, si Ω n'est pas dénombrable, cette σ -algèbre est souvent trop grande pour qu'on puisse y développer les outils mathématiques nécessaires.

En dehors des cas très simples, il est souvent impossible de lister les éléments d'une algèbre ou d'une σ -algèbre. Il est alors commode de les caractériser par des sous-ensembles "assez riches".

Ainsi, on définit pour tout $\mathcal{C} \subset \mathcal{P}(\Omega)$ la σ -algèbre $\sigma(\mathcal{C})$ engendrée par \mathcal{C} . C'est la plus petite σ -algèbre sur Ω contenant \mathcal{C} , définie comme intersection de toutes les σ -algèbres sur Ω contenant \mathcal{C} .

Exemple 1.2. Si Ω est un espace topologique, la σ -algèbre Borelienne, notée par \mathcal{B}_Ω , est la σ -algèbre engendrée par les ouverts de Ω . Pour la droite réelle, on peut même simplifier la compréhension de $\mathcal{B}_\mathbb{R}$:

$$\mathcal{B}_\mathbb{R} = \sigma(\pi(\mathbb{R})) \quad \text{où} \quad \pi(\mathbb{R}) := \{]-\infty, x] : x \in \mathbb{R} \}$$

(Exercice !)

L'exemple précédent se généralise par la notion suivante :

Définition 1.3. Soit $\mathcal{I} \subset \mathcal{P}(\Omega)$. On dit que \mathcal{I} est un π -système s'il est stable par intersection finie.

Ainsi l'ensemble $\pi(\mathbb{R})$ de l'exemple ci-dessus est un π -système. L'importance de cette notion apparaîtra dans la proposition 1.5 ci-dessous ainsi que dans le théorème des classes monotones 1.18 de la section 1.2.

1.1.2 Mesures

Définition 1.4. Soit \mathcal{A}_0 une algèbre sur Ω , et $\mu_0 : \mathcal{A}_0 \rightarrow \mathbb{R}_+$ une fonction positive.

(i) μ_0 est dite additive si $\mu_0(\emptyset) = 0$ et pour tous $A, B \in \mathcal{A}_0$:

$$\mu_0(A \cup B) = \mu_0(A) + \mu_0(B) \quad \text{dès que} \quad A \cap B = \emptyset.$$

(ii) μ_0 est dite σ -additive si $\mu_0(\emptyset) = 0$ et pour toute suite $(A_n)_{n \geq 0} \subset \mathcal{A}_0$:

$$A = \cup_{n \geq 0} A_n \in \mathcal{A}_0 \text{ et les } A_n \text{ disjoints} \implies \mu_0(A) = \sum_{n \geq 0} \mu_0(A_n).$$

(iii) Une fonction σ -additive $\mu : \mathcal{A} \rightarrow \mathbb{R}_+$ sur un espace mesurable (Ω, \mathcal{A}) est appelée mesure, et on dit que $(\Omega, \mathcal{A}, \mu)$ est un espace mesuré.

(iv) Un espace mesuré $(\Omega, \mathcal{A}, \mu)$ est dit fini si $\mu(\Omega) < \infty$, et σ -fini s'il existe une suite $(\Omega_n)_{n \geq 0} \subset \mathcal{A}$ telle que $\mu(\Omega_n) < \infty$ et $\cup_{n \geq 0} \Omega_n = \Omega$.

Proposition 1.5. Soient \mathcal{I} un π -système, et μ, ν deux mesures finies sur l'espace mesurable $(\Omega, \sigma(\mathcal{I}))$. Si $\mu = \nu$ sur \mathcal{I} alors $\mu = \nu$ sur $\sigma(\mathcal{I})$.

La démonstration est reportée, à titre de complément, dans l'annexe de ce chapitre. Le résultat suivant est essentiel pour construire des mesures "intéressantes".

Théorème 1.6. (*extension de Carathéodory, théorème*) Soient \mathcal{A}_0 une algèbre sur Ω , et $\mu_0 : \mathcal{A}_0 \rightarrow \mathbb{R}_+$ une fonction σ -additive. Alors il existe une mesure μ sur $\mathcal{A} := \sigma(\mathcal{A}_0)$ telle que $\mu = \mu_0$ sur \mathcal{A}_0 . Si de plus $\mu_0(\Omega) < \infty$, alors une telle extension μ est unique.

La démonstration est reportée, à titre de complément, dans l'annexe de ce chapitre. Avec ce résultat, on peut maintenant construire une mesure importante sur l'espace mesurable $(]0, 1], \mathcal{B}_{]0, 1]})$.

Exemple 1.7. (*Mesure de Lebesgue*) Nous allons définir une mesure sur $\mathcal{B}_{]0,1]}$ qui mesure les longueurs.

1- On remarque tout d'abord que \mathcal{A}_0 constitué des parties $A \subset]0,1]$ de la forme

$$A = \cup_{1 \leq i \leq n}]a_i, b_i] \quad \text{pour } n \in \mathbb{N} \text{ et } 0 \leq a_1 \leq b_1 \leq \dots \leq a_r \leq b_r \leq 1, \quad (1.1)$$

est une algèbre telle que $\mathcal{B}_{]0,1]} = \sigma(\mathcal{A}_0)$. Pour tout $A \in \mathcal{A}_0$ de la forme (1.1), on définit

$$\lambda_0(A) := \sum_{i=1}^n (b_i - a_i).$$

2- Alors $\lambda_0 : \mathcal{A}_0 \rightarrow \mathbb{R}_+$ est une application bien définie et est évidemment additive. On peut montrer qu'elle est σ -additive (c'est moins évident, nous renonçons à le justifier ici pour alléger ces notes, et nous renvoyons au livre de Williams [15]). Comme $\lambda_0(]0,1]) < \infty$, on déduit du théorème de Carathéodory l'existence d'une unique extension λ définie sur $\mathcal{B}_{]0,1]}$.

Cette mesure finie λ est appelée *mesure de Lebesgue* sur $]0,1]$. La mesure de Lebesgue sur $[0,1]$ est obtenue par une modification triviale puisque le singleton $\{0\}$ est de mesure de Lebesgue nulle.

3- Par le même raisonnement, on peut construire la mesure de Lebesgue sur $\mathcal{B}_{\mathbb{R}}$ comme extension d'une application d'ensembles sur l'algèbre des unions finies d'intervalles semi-ouverts disjoints. Dans ce cas, la mesure de Lebesgue est seulement σ -finie.

Définition 1.8. (i) *Sur un espace mesuré $(\Omega, \mathcal{A}, \mu)$, un ensemble $N \in \mathcal{A}$ est dit négligeable si $\mu(N) = 0$.*

(ii) *Soit $P(\omega)$ une propriété qui ne dépend que d'un élément $\omega \in \Omega$. On dit que P est vraie μ -presque partout, et on note μ -p.p., si l'ensemble $\{\omega \in \Omega : P(\omega) \text{ n'est pas vraie}\}$ est inclus dans un ensemble négligeable.*

Remarque 1.9. D'après la propriété de σ -additivité de la mesure, on voit aisément que toute union dénombrable de négligeables est négligeable.

1.1.3 Propriétés élémentaires des mesures

Nous commençons par des propriétés mettant en jeu un nombre fini d'ensembles.

Proposition 1.10. *Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré, et $(A_i)_{i \leq n} \subset \mathcal{A}$. Alors :*

- (i) $\mu(\cup_{i \leq n} A_i) \leq \sum_{i \leq n} \mu(A_i)$,
- (ii) *Si de plus $\mu(\Omega) < \infty$, on a*

$$\mu(\cup_{i \leq n} A_i) = \sum_{k \leq n} (-1)^{k-1} \sum_{i_1 < \dots < i_k \leq n} \mu(A_{i_1} \cap \dots \cap A_{i_k}).$$

La preuve de ce résultat est une conséquence immédiate de la définition de mesure. La partie (ii), spécifique aux mesures finies, donne une formule pour

la mesure de l'union finie d'ensemble qui alterne entre sur-estimation et sous estimation. Pour $n = 2$ cette formule n'est autre que la propriété bien connue $\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$ pour $A, B \in \mathcal{A}$.

Le résultat (simple) suivant est fondamental en théorie de la mesure. Pour une suite d'ensembles $(A_n)_n$, nous notons simplement $A_n \uparrow A$ pour indiquer que la suite est croissante ($A_n \subset A_{n+1}$) et $\cup_n A_n = A$. La notation $A_n \downarrow A$ a un sens similaire dans le cas où la suite est décroissante.

Proposition 1.11. *Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré, et $(A_n)_n$ une suite de \mathcal{A} . Alors*

- (i) $A_n \uparrow A \implies \mu(A_n) \uparrow \mu(A)$,
- (ii) $A_n \downarrow A$ et $\mu(A_k) < \infty$ pour un certain entier $k \implies \mu(A_n) \downarrow \mu(A)$,

La démonstration simple de ce résultat est laissée comme exercice. Faisons juste deux remarques :

- La proposition 1.11 (i) implique que l'union dénombrable d'ensembles de mesure nulle est de mesure nulle (Remarque 1.9).
- l'exemple $A_n =]n, \infty[$ dans l'espace mesuré $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \lambda)$, λ étant la mesure de Lebesgue sur \mathbb{R} , montre que la condition supplémentaire dans (ii) est nécessaire.

Ces résultats permettent de montrer les outils importants pour l'analyse de la convergence des mesures des ensembles. On rappelle les notions de liminf et limsup pour une suite réelle $(x_n)_{n \geq 1} \subset \mathbb{R}$

$$\limsup_{n \rightarrow \infty} x_n := \inf_{p \geq 1} \sup_{n \geq p} x_n \quad \text{et} \quad \liminf_{n \rightarrow \infty} x_n := \sup_{p \geq 1} \inf_{n \geq p} x_n$$

et pour une suite d'ensembles $(A_n)_n$:

$$\begin{aligned} \limsup A_n &:= \cap_n \cup_{k \geq n} A_k = \{\omega \in \Omega : \omega \in A_n \text{ pour une infinité de } n\}, \\ \liminf A_n &:= \cup_n \cap_{k \geq n} A_k = \{\omega \in \Omega : \omega \in A_n \text{ à partir d'un rang } n_0(\omega)\}. \end{aligned}$$

Le résultat suivant est très utile.

Lemme 1.12. *(de Fatou pour les ensembles) Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré, et $(A_n)_n$ une suite dans \mathcal{A} . Alors*

$$\mu[\liminf A_n] \leq \liminf \mu[A_n].$$

Démonstration. Par définition, nous avons $B_n := \cap_{k \geq n} A_k \uparrow B := \liminf A_n$, et on déduit de la proposition 1.11 (i) que $\mu[B] = \lim \uparrow \mu[B_n]$. Pour conclure, il suffit de remarquer que $B_n \subset A_k$ pour tout $k \geq n$, et par suite $\mu[B_n] \leq \inf_{k \geq n} \mu[A_k]$, impliquant que $\lim \uparrow \mu[B_n] \leq \liminf \mu[A_n]$. \diamond

Si la mesure est finie, le résultat suivant montre que l'inégalité inverse dans le lemme de Fatou pour les ensembles a lieu en échangeant liminf et limsup. Nous verrons plus tard que la situation est plus compliquée pour les fonctions...

Lemme 1.13. (*inverse Fatou pour les ensembles*) Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré fini, et $(A_n)_n$ une suite dans \mathcal{A} . Alors

$$\mu[\limsup A_n] \geq \limsup \mu[A_n].$$

Démonstration. Par définition, nous avons $C_n := \cup_{k \geq n} A_k \downarrow C := \limsup A_n$. La proposition 1.11 (ii), qui requiert que la mesure soit finie, donne $\mu[C] = \lim \downarrow \mu[C_n]$. Pour conclure, il suffit de remarquer que $C_n \supset A_k$ pour tout $k \geq n$, et par suite $\mu[C_n] \geq \inf_{k \geq n} \mu[A_k]$, impliquant que $\lim \downarrow \mu[C_n] \geq \limsup \mu[A_n]$. \diamond

Enfin, nous énonçons le résultat suivant qui sera utilisé à plusieurs reprises. Notons que cet énoncé sera complété dans la suite quand nous aurons abordé les notions d'indépendance.

Lemme 1.14. (*Premier lemme de Borel-Cantelli*) Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré, et $(A_n)_n \subset \mathcal{A}$. Alors

$$\sum_n \mu[A_n] < \infty \implies \mu[\limsup A_n] = 0.$$

Démonstration. Comme $\limsup A_n \subset C_p := \cup_{k \geq p} A_k$ pour tout $p \geq 1$, on déduit que $\mu(\limsup A_n) \leq \mu(C_p) \leq \sum_{k \geq p} \mu(A_k)$. Le résultat est obtenu en envoyant p vers l'infini. \diamond

1.2 L'intégrale de Lebesgue

Dans cette section, on considère un espace mesuré $(\Omega, \mathcal{A}, \mu)$, et nous développons la théorie d'intégration d'une fonction par rapport à la mesure μ . Si Ω est dénombrable, $\mathcal{A} = \mathcal{P}(\Omega)$, et $\mu(\{\omega\}) = 1$ pour tout $\omega \in \Omega$, une fonction est identifiée à une suite $(a_n)_n$, et elle est intégrable si et seulement si $\sum_n |a_n| < \infty$, et l'intégrale est donnée par la valeur de la série $\sum_n a_n$. La réelle difficulté est donc pour les espaces non dénombrables.

1.2.1 Fonction mesurable

L'objet central en topologie est la structure des ouverts, et les fonctions continues sont caractérisées par la propriété que les images réciproques des ouverts de l'ensemble d'arrivée sont des ouverts de l'ensemble de départ. Dans la théorie de la mesure, les ouverts sont remplacés par les ensembles mesurables, et les fonctions mesurables remplacent les fonctions continues.

Définition 1.15. On dit qu'une fonction $f : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ est mesurable si l'image réciproque de tout ensemble borélien est dans \mathcal{A} . On note par $\mathcal{L}^0(\mathcal{A})$ l'ensemble des fonctions mesurables. Les sous-ensembles des fonctions mesurables positives (resp. bornées) seront notés $\mathcal{L}_+^0(\mathcal{A})$ (resp. $\mathcal{L}^\infty(\mathcal{A})$).

De manière équivalente $f \in \mathcal{L}^0(\mathcal{A})$ si et seulement l'inverse f^{-1} est bien définie comme une application de $\mathcal{B}_{\mathbb{R}}$ dans \mathcal{A} , i.e. $f^{-1} : \mathcal{B}_{\mathbb{R}} \rightarrow \mathcal{A}$. Si $\mathcal{C} \subset \mathcal{B}_{\mathbb{R}}$ est tel que $\sigma(\mathcal{C}) = \mathcal{B}_{\mathbb{R}}$, alors il suffit de vérifier $f^{-1} : \mathcal{C} \rightarrow \mathcal{A}$.

Remarque 1.16. (i) En prenant $\mathcal{C} = \pi(\mathbb{R})$ le π -système des intervalles de la forme $] -\infty, c]$, $c \in \mathbb{R}$, on voit que

$$f \in \mathcal{L}^0(\mathcal{A}) \quad \text{ssi} \quad \{f \leq c\} \in \mathcal{A} \text{ pour tout } c \in \mathbb{R}.$$

(ii) Supposons que Ω est un espace topologique, et que $f : \Omega \rightarrow \mathbb{R}$ est continue. Alors f est \mathcal{B}_{Ω} -mesurable. En effet, avec $\mathcal{C} = \{\text{ouverts de } \mathbb{R}\}$, la continuité s'écrit $f^{-1} : \mathcal{B}_{\mathbb{R}} \rightarrow \mathcal{A}$. On dit que f est une fonction borelienne.

(iii) Soit X une application de Ω dans un ensemble dénombrable $X(\Omega) = \{x_n, n \in \mathbb{N}\}$. On munit $X(\Omega)$ de la plus grande σ -algèbre $\mathcal{P}(X(\Omega))$ et on remarque que $\mathcal{P}(X(\Omega)) = \sigma(\{\{\omega\} : \omega \in \Omega\})$. Ceci permet de conclure que X est mesurable si et seulement si $\{X = x_n\} \in \mathcal{A}$ pour tout $n \in \mathbb{N}$.

La mesurabilité est conservée par les opérations usuelles pour les fonctions.

Proposition 1.17. (i) Pour $f, g \in \mathcal{L}^0(\mathcal{A})$, $h \in \mathcal{L}^0(\mathcal{B}_{\mathbb{R}})$, et $\lambda \in \mathbb{R}$, on a $f + g$, λf , fg , $f \circ h$ et $\lambda f \in \mathcal{L}^0(\mathcal{A})$.

(ii) Pour une suite $(f_n)_n \subset \mathcal{L}^0(\mathcal{A})$, on a $\inf h_n$, $\liminf h_n$, $\sup h_n$ et $\limsup h_n \in \mathcal{L}^0(\mathcal{A})$.

La preuve est simple et est laissée en exercice. Avant d'aborder l'objet central de ce chapitre, à savoir la construction de l'intégrale de Lebesgue, nous reportons une version simple du théorème des classes monotones, qui ne sera utilisé que plus tard dans la construction d'espaces mesurés produits.

Théorème 1.18. (classes monotones) Soit \mathcal{H} une classes de fonctions réelles bornées sur Ω vérifiant les conditions suivantes :

(H1) \mathcal{H} est un espace vectoriel contenant la fonction constante $\mathbf{1}$,

(H2) pour toute suite croissante $(f_n)_n \subset \mathcal{H}$ de fonctions positives dont la limite $f := \lim \uparrow f_n$ est bornée, on a $f \in \mathcal{H}$.

Soit \mathcal{I} un π -système tel que $\{\mathbf{1}_A : A \in \mathcal{I}\} \subset \mathcal{H}$. Alors $\mathcal{L}^\infty(\sigma(\mathcal{I})) \subset \mathcal{H}$.

La démonstration est reportée à titre de complément dans l'annexe de ce chapitre.

1.2.2 Intégration des fonctions positives

Le but de ce paragraphe est de définir pour toute fonction mesurable positive f une notion d'intégrale par rapport à la mesure μ :

$$\int f d\mu \quad \text{que l'on note aussi} \quad \mu(f),$$

qui est un abus de notation communément accepté ($\mu : \mathcal{A} \rightarrow \mathbb{R}!$) du fait que notre définition doit vérifier

$$\int \mathbf{1}_A = \mu(A) \quad \text{pour tout} \quad A \in \mathcal{A}.$$

Plus généralement, soit \mathcal{S}^+ l'ensemble des fonctions de Ω dans \mathbb{R}_+ de la forme

$$g = \sum_{i=1}^n a_i \mathbf{1}_{A_i}, \quad (1.2)$$

pour un certain entier $n \geq 1$, des ensembles $A_i \in \mathcal{A}$, et des scalaires $a_i \in [0, \infty]$, $1 \leq i \leq n$. Ici, il est commode d'autoriser la valeur $+\infty$, et on utilisera les règles de calcul $0 \times \infty = \infty \times 0 = 0$. l'intégrale sur \mathcal{S}^+ est définie par :

$$\mu(g) = \sum_{i=1}^n a_i \mu(A_i). \quad (1.3)$$

Il est clair que $\mu(g)$ est bien défini, i.e. deux représentations différentes (1.2) d'un élément $f \in \mathcal{S}^+$ donnent la même valeur. Nous étendons à présent la définition de μ à l'ensemble $\mathcal{L}_+^0(\mathcal{A})$ des fonctions \mathcal{A} -mesurables positives.

Définition 1.19. Pour $f \in \mathcal{L}_+^0(\mathcal{A})$, l'intégrale de f par rapport à μ est définie par

$$\mu(f) := \sup \{ \mu(g) : g \in \mathcal{S}^+ \text{ et } g \leq f \}.$$

L'ensemble $\{g \in \mathcal{S}^+ : g \leq f\}$, dont la borne supérieure définit l'intégrale, contient la fonction nulle. On peut aussi construire des éléments non triviaux en introduisant la fonction

$$\alpha_n(x) := n \wedge \sum_{i \geq 1} (i-1)2^{-n} \mathbf{1}_{B_i^n}(x), \quad B_i^n :=](i-1)2^{-n}, i2^{-n}].$$

En effet, pour tout $f \in \mathcal{L}^0(\mathcal{A})$:

$$(\alpha_n \circ f)_n \subset \mathcal{S}^+ \quad \text{est une suite croissante qui converge vers } f. \quad (1.4)$$

La définition de l'intégrale implique immédiatement que

$$\mu(cf) = c\mu(f) \quad \text{pour tous } c \in \mathbb{R}_+ \text{ et } f \in \mathcal{L}_+^0(\mathcal{A}), \quad (1.5)$$

ainsi que la propriété de monotonie suivante.

Lemme 1.20. Pour $f_1, f_2 \in \mathcal{L}_+^0(\mathcal{A})$. Si $f_1 \leq f_2$, alors $0 \leq \mu(f_1) \leq \mu(f_2)$. De plus $\mu(f_1) = 0$ si et seulement si $f_1 = 0$, μ -p.p.

Démonstration. Pour la première partie, il suffit de remarquer que $\{g \in \mathcal{S}^+ : g \leq f_1\} \subset \{g \in \mathcal{S}^+ : g \leq f_2\}$. Pour la deuxième partie de l'énoncé, rappelons que $\mu(\{f > 0\}) = \lim \uparrow \mu(\{f > n^{-1}\})$ d'après la proposition 1.11. Si $\mu(\{f > 0\}) > 0$, on a $\mu(\{f > n^{-1}\}) > 0$ pour n assez grand. Alors $f \geq g := n^{-1} \mathbf{1}_{\{f > n^{-1}\}} \in \mathcal{S}^+$, et on déduit de la définition de l'intégrale que $\mu(f) \geq \mu(g) = n^{-1} \mu(\{f > n^{-1}\}) > 0$. \diamond

Le résultat à la base de la théorie de l'intégration est l'extension suivante de la propriété de convergence monotone des mesures d'ensembles énoncée dans la proposition 1.11 (i).

Théorème 1.21. (convergence monotone) Soit $(f_n)_n \subset \mathcal{L}_+^0(\mathcal{A})$ une suite croissante μ -p.p., i.e. pour tout $n \geq 1$, $f_n \leq f_{n+1}$ μ -p.p. Alors

$$\mu(\lim \uparrow f_n) = \lim \uparrow \mu(f_n).$$

Démonstration. On procède en trois étapes.

Etape 1 On commence par supposer que $f_n \leq f_{n+1}$ sur Ω . On note $f := \lim \uparrow f_n$. D'après le lemme 1.20, la suite des intégrales $(\mu(f_n))_n$ hérite la croissance de la suite $(f_n)_n$ et est majorée par $\mu(f)$. Ceci montre l'inégalité $\lim \uparrow \mu(f_n) \leq \mu(\lim \uparrow f_n)$.

Pour établir l'inégalité inverse, nous devons montrer que $\lim \uparrow \mu(f_n) \geq \mu(g)$ pour tout $g = \sum_{i=1}^k a_i \mathbf{1}_{A_i} \in \mathcal{S}^+$ vérifiant $g \leq f$. Pour tout $c \in [0, 1[$, on déduit du lemme 1.20 et de (1.5) que :

$$\mu(f_n) \geq \mu(f_n \mathbf{1}_{\{f_n \geq cg\}}) \geq c\mu(g \mathbf{1}_{\{f_n \geq cg\}}) = c \sum_{i=1}^k a_i \mu(A_i \cap \{f_n \geq ca_i\}).$$

En utilisant la propriété de convergence monotone des mesures d'ensembles énoncée dans la proposition 1.11 (i), on obtient alors :

$$\lim \uparrow \mu(f_n) \geq c \sum_{i=1}^k a_i \mu(A_i) = c\mu(g) \longrightarrow \mu(g) \text{ quand } c \rightarrow 1.$$

Etape 2 Dans le reste de la preuve, on veut passer de la monotonie de la suite $(f_n)_n$ sur Ω à la monotonie μ -p.p. Pour cela, introduisons $\Omega_0 = \{\omega \in \Omega : (f_n(\omega))_n \text{ croissante}\}$ et la suite croissante (sur Ω) $\tilde{f}_n := f_n \mathbf{1}_{\Omega_0}$. La première étape de cette preuve s'applique à la suite $(\tilde{f}_n)_n$ alors il suffit de montrer que $\mu(\tilde{f}_n) = \mu(f_n)$. Comme $\tilde{f}_n \leq f_n$, l'inégalité $\mu(\tilde{f}_n) \leq \mu(f_n)$ découle du lemme 1.20. Pour tout $\varepsilon > 0$, il existe $g_n^\varepsilon \in \mathcal{S}^+$ tel que $g_n^\varepsilon \leq \tilde{f}_n$ et $\mu(g_n^\varepsilon) \geq \mu(\tilde{f}_n) - \varepsilon$. Remarquons que $\tilde{g}_n^\varepsilon := g_n^\varepsilon \mathbf{1}_{\Omega_0} \in \mathcal{S}^+$ et vérifie $\tilde{g}_n^\varepsilon \leq f_n$. Alors, $\mu(\tilde{g}_n^\varepsilon) \leq \mu(f_n)$. Comme $\mu(\tilde{g}_n^\varepsilon) = \mu(g_n^\varepsilon)$, on déduit que $\mu(f_n) \leq \mu(g_n^\varepsilon) + \varepsilon \leq \mu(\tilde{f}_n) + \varepsilon \searrow \mu(\tilde{f}_n)$ pour $\varepsilon \searrow 0$. \diamond

Remarque 1.22. Par le même argument que l'étape 2 ci-dessus (approximation par les fonctions simples (1.4) et utilisation du théorème de convergence monotone), on montre facilement que :

- (i) Pour $f_1, f_2 \in \mathcal{L}_+^0(\mathcal{A})$ telles que $f_1 = f_2$ μ -p.p., on a $\mu(f_1) = \mu(f_2)$.
- (ii) Pour $f_1, f_2 \in \mathcal{L}_+^0(\mathcal{A})$, on a $\mu(f_1 + f_2) = \mu(f_1) + \mu(f_2)$.

Voici une conséquence simple et très utile du théorème de convergence monotone.

Lemme 1.23. (Fatou) Pour une suite de fonctions $(f_n)_n$ de $\mathcal{L}_+^0(\mathcal{A})$, on a

$$\mu(\lim \inf f_n) \leq \lim \inf \mu(f_n).$$

Démonstration. D'après la monotonie de l'intégrale, $\inf_{k \geq n} \mu(f_k) \geq \mu(\inf_{k \geq n} f_k)$ pour tout $n \geq 1$. Comme la suite $(\inf_{k \geq n} f_k)_{n \geq 1}$ est croissante μ -p.p., on obtient le résultat par application du théorème de convergence monotone. \diamond

1.2.3 Intégration des fonctions réelles

Pour une fonction $f \in \mathcal{L}^0(\mathcal{A})$, on note $f^+ := \max\{f, 0\}$ et $f^- := \max\{-f, 0\}$ si bien que $|f| = f^+ + f^-$. Ces fonctions héritent la \mathcal{A} -mesurabilité de f .

Définition 1.24. Une fonction $f \in \mathcal{L}^0(\mathcal{A})$ est dite μ -intégrable si $\mu(|f|) = \mu(f^+) + \mu(f^-) < \infty$, et son intégrale est définie par

$$\mu(f) := \mu(f^+) - \mu(f^-).$$

On note par $\mathcal{L}^1(\mathcal{A}, \mu)$ l'ensemble des fonctions μ -intégrables.

On voit immédiatement que $\mathcal{L}^1(\mathcal{A}, \mu)$ est un espace vectoriel dont on donnera d'autres propriétés topologiques dans la suite.

Avant de continuer, levons tout de suite une source d'ambiguïté concernant l'intégration d'une fonction $f \in \mathcal{L}^1(\mathcal{A}, \mu)$ sur une partie $A \in \mathcal{A}$. En effet celle-ci peut se faire soit en intégrant la fonction intégrable $f\mathbf{1}_A$, soit en intégrant la restriction $f|_A$ par rapport à la restriction μ_A de μ à l'espace mesurable (A, \mathcal{A}_A) , où \mathcal{A}_A est la σ -algèbre définie par $\mathcal{A}_A := \mathcal{P}(A) \cap \mathcal{A}$.

Proposition 1.25. Pour tout $f \in \mathcal{L}^1(\mathcal{A}, \mu)$ et $A \in \mathcal{A}$, on a $\mu(f\mathbf{1}_A) = \mu_A(f|_A)$.

Démonstration. Tout d'abord, cette propriété est vraie pour les fonctions $f = \mathbf{1}_B$, $B \in \mathcal{A}$, puisque dans ce cas $\mu(\mathbf{1}_B\mathbf{1}_A) = \mu(A \cap B) = \mu_A(\mathbf{1}_B|_A)$. Par linéarité, cette égalité reste vraie pour les fonctions simples, puis par convergence monotone pour les fonctions mesurables positives. Enfin, pour $f \in \mathcal{L}^1(\mathcal{A}, \mu)$, on décompose $f = f^+ - f^-$, et on obtient le résultat voulu en appliquant l'égalité à f^+ et f^- . \diamond

Voici un résultat qui rappelle une propriété classique sur les intégrales de Riemann éventuellement impropres.

Lemme 1.26. Soit $f \in \mathcal{L}^1(\mathcal{A}, \mu)$ et $\varepsilon > 0$. Alors, il existe $\delta > 0$ tel que pour tout $A \in \mathcal{A}$ vérifiant $\mu(A) < \delta$, on a $\mu(|f|\mathbf{1}_A) < \varepsilon$.

Démonstration. Supposons, au contraire, qu'il existe ε_0 et une suite $(A_n)_n \subset \mathcal{A}$ tels que $\mu(A_n) < 2^{-n}$ et $\mu(|f|\mathbf{1}_{A_n}) \geq \varepsilon_0$. D'après le premier lemme de Borel-Cantelli, lemme 1.14, on déduit que $A := \limsup A_n$ est négligeable. En particulier $\mu(|f|\mathbf{1}_A) = 0$, et on obtient une contradiction en remarquant que $\mu(|f|\mathbf{1}_A) = \mu(|f|) - \mu(|f|\mathbf{1}_{A^c}) \geq \mu(|f|) - \liminf \mu(|f|\mathbf{1}_{A_n^c}) = \limsup \mu(|f|\mathbf{1}_{A_n}) \geq \varepsilon_0$, où on a utilisé le lemme de Fatou. \diamond

1.2.4 De la convergence p.p. à la convergence \mathcal{L}^1

Théorème 1.27. (convergence dominée) Soient $(f_n)_n \subset \mathcal{L}^0(\mathcal{A})$ une suite telle que $f_n \rightarrow f$ μ -p.p. pour une certaine fonction $f \in \mathcal{L}^0(\mathcal{A})$. Si $\sup_n |f_n| \in \mathcal{L}^1(\mathcal{A}, \mu)$, alors

$$f_n \rightarrow f \text{ dans } \mathcal{L}^1(\mathcal{A}, \mu) \quad \text{i.e.} \quad \mu(|f_n - f|) \rightarrow 0.$$

En particulier, $\mu(f_n) \rightarrow \mu(f)$.

Démonstration. On note $g_n := |f_n - f|$ et $h := \sup_n g_n$. Par le lemme de Fatou, $\mu(|f|) \leq \liminf_n \mu(|f_n|) \leq \mu(\sup_n |f_n|) < \infty$. Alors $f \in \mathcal{L}^1(\mathcal{A}, \mu)$ et $|h| \leq |f| + \sup_n |f_n| \in \mathcal{L}^1(\mathcal{A}, \mu)$. Comme la fonction $h - g_n$ est positive et que $g_n \rightarrow 0$, μ -p.p., on obtient par le lemme de Fatou que $\liminf \mu(h - g_n) \geq \mu(h)$. Simplifiant par $\mu(h)$, ceci implique que $0 \geq \limsup_n \mu(g_n) \geq \liminf_n \mu(g_n) \geq 0$ du fait de la positivité de g_n . \diamond

Le résultat suivant donne une condition nécessaire et suffisante pour qu'une suite convergente μ -p.p. soit convergente dans $\mathcal{L}^1(\mathcal{A})$.

Lemme 1.28. (Scheffé) Soit $(f_n)_n \subset \mathcal{L}^1(\mathcal{A}, \mu)$ telle que $f_n \rightarrow f$ μ -p.p. pour une certaine fonction $f \in \mathcal{L}^1(\mathcal{A}, \mu)$. Alors :

$$f_n \rightarrow f \text{ dans } \mathcal{L}^1(\mathcal{A}, \mu) \quad \text{ssi} \quad \mu(|f_n|) \rightarrow \mu(|f|).$$

Démonstration. L'implication " \implies " est triviale. Pour l'inégalité inverse, on procède en deux étapes.

Étape 1 Supposons que $f_n, f \geq 0$, μ -p.p. Alors $(f_n - f)^- \leq f \in \mathcal{L}^1(\mathcal{A})$, et on déduit du théorème de convergence dominée que $\mu((f_n - f)^-) \rightarrow 0$. Pour conclure, on écrit que $\mu(|f_n - f|) = \mu(f_n) - \mu(f) + 2\mu((f_n - f)^-) \rightarrow 0$.

Étape 2 Pour f_n et f de signe quelconque, on utilise le lemme de Fatou pour obtenir $\mu(|f|) = \lim\{\mu(f_n^+) + \mu(f_n^-)\} \geq \mu(f^+) + \mu(f^-) = \mu(|f|)$ et par suite toutes les inégalités sont des égalités, i.e. $\lim \mu(f_n^+) = \mu(f^+)$ et $\lim \mu(f_n^-) = \mu(f^-)$. On est alors ramené au contexte de l'étape 1, qui permet d'obtenir $f_n^+ \rightarrow f^+$ et $f_n^- \rightarrow f^-$ dans $\mathcal{L}^1(\mathcal{A})$, et on conclut en écrivant $|f_n - f| \leq |f_n^+ - f^+| + |f_n^- - f^-|$ et en utilisant la monotonie de l'intégrale. \diamond

Exercice 1.29. Soient $(\Omega, \mathcal{A}, \mu)$ un espace mesuré, I un intervalle ouvert de \mathbb{R} , et $f : I \times \Omega \rightarrow \mathbb{R}$ une fonction telle que $f(x, \cdot) \in \mathcal{L}^0(\mathcal{A})$ pour tout $x \in I$.

1. On suppose qu'il existe une fonction $g \in \mathcal{L}_+^1(\mathcal{A}, \mu)$ telle que $|f(x, \cdot)| \leq g$, μ -p.p. Montrer alors que, si $f(\cdot, \omega)$ est continue en un point $x_0 \in I$, μ -p.p., la fonction $\phi : I \rightarrow \mathbb{R}$ définie par

$$\phi(x) := \int f(x, \omega) d\mu(\omega); \quad x \in I,$$

est bien définie, et qu'elle est continue au point x_0 .

2. On suppose que la dérivée partielle $f_x := (\partial f / \partial x)$ existe pour tout $x \in I$, μ -p.p. et qu'il existe une fonction $h \in \mathcal{L}_+^1(\mathcal{A}, \mu)$ telle que $|f_x(x, \cdot)| \leq h$, μ -p.p. Montrer alors que ϕ est dérivable sur I , et

$$\phi'(x) = \int \frac{\partial f}{\partial x}(x, \omega) d\mu(\omega); \quad x \in I.$$

3. Donner des conditions qui assurent que ϕ soit continuellement dérivable sur I .

1.2.5 Intégrale de Lebesgue et intégrale de Riemann

Dans ce paragraphe, nous donnons quelques éléments qui expliquent l'avantage de l'intégrale de Lebesgue par rapport à celle de Riemann. Pour être plus concret, on considère le problème d'intégration sur \mathbb{R} .

(a) *L'intégrale de Riemann est construite sur un intervalle $[a, b]$ compact de \mathbb{R} . Il y a bien une extension par les intégrales impropres, mais cela conduit à un cadre assez restrictif.*

(b) *L'intégrale de Riemann est construite en approximant la fonction par des fonctions en escalier, i.e. constantes sur des sous-intervalles de $[a, b]$ de longueur petite. Sur un dessin, il s'agit d'une approximation verticale. Par contre, l'intégrale de Lebesgue est construite en découpant l'intervalle image et en approximant f sur les images réciproques de ces intervalles. Il s'agit dans ce cas d'une approximation horizontale de la fonction à intégrer.*

(c) *Les fonctions Riemann intégrables sont Lebesgue intégrables. Montrons ceci dans $[0, 1]$. Soit f une fonction Riemann intégrable bornée sur $\Omega = [0, 1]$ d'intégrale (au sens de Riemann) $\int_0^1 f(x)dx$. Alors f est Lebesgue intégrable d'intégrale $\lambda(f) = \int_0^1 f(x)dx$. Si f est une fonction en escalier, ce résultat est trivial. Pour une fonction Riemann intégrable f arbitraire, on peut trouver deux suites de fonctions en escalier $(g_n)_n$ et $(h_n)_n$ respectivement croissante et décroissante telles que $g_n \leq f \leq h_n$ et*

$$\inf_n \int_0^1 (g_n - h_n)(x)dx = \lim_{n \rightarrow \infty} \int_0^1 (g_n - h_n)(x)dx = 0.$$

Sans perte de généralité, on peut supposer $h_n \leq 2\|f\|_\infty$. Les fonctions $f_* := \sup_n g_n$ et $f^* := \inf_n h_n$ sont boreliennes, et on a $f_* \leq f \leq f^*$. D'après la monotonie de l'intégrale :

$$0 \leq \mu(f^* - f_*) = \mu(\inf_n (h_n - g_n)) \leq \inf_n \mu(h_n - g_n) = 0,$$

et par suite $f = f^* = f_*$. Enfin :

$$\mu(f_*) = \lim \uparrow \mu(g_n) = \lim \uparrow \int_0^1 g_n(x)dx = \int_0^1 f(x)dx$$

La réciproque n'est pas vraie. Par exemple, la fonction $f = \mathbf{1}_{\mathbb{Q} \cap [0, 1]}$ est Lebesgue-intégrable, mais n'est pas Riemann-intégrable.

(d) *Le théorème de convergence dominée n'a pas son équivalent dans le cadre de l'intégrale de Riemann, et permet d'obtenir un espace de fonctions intégrables complet (on verra ce résultat plus tard). Par contre, on peut construire des exemples de suites de Cauchy de fonctions Riemann intégrables dont la limite n'est pas Riemann intégrable.*

(e) Pour les fonctions définies par des intégrales, les résultats de continuité et de dérivabilité sont simplement obtenus grâce au théorème de convergence dominée. Leur analogue dans le cadre des intégrales de Riemann conduit à des résultats assez restrictifs.

(f) L'intégrale de Lebesgue se définit naturellement dans \mathbb{R}^n , comme on le verra dans la section 1.5. En particulier, le théorème de Fubini est d'une grande simplicité dans le cadre de l'intégrale de Lebesgue. La situation est un peu plus compliquée pour l'intégrale de Riemann.

1.3 Transformées de mesures

1.3.1 Mesure image

Soit $(\Omega_1, \mathcal{A}_1, \mu_1)$ un espace mesuré, $(\Omega_2, \mathcal{A}_2)$ un espace mesurable et $f : \Omega_1 \rightarrow \Omega_2$ une fonction mesurable, i.e. $f^{-1} : \mathcal{A}_2 \rightarrow \mathcal{A}_1$. On vérifie immédiatement que l'application :

$$\mu_2(A_2) := \mu_1(f^{-1}(A_2)) \quad \text{pour tout } A_2 \in \mathcal{A}_2,$$

définit une mesure sur $(\Omega_2, \mathcal{A}_2)$.

Définition 1.30. μ_2 est appelée mesure image de μ_1 par f , et est notée $\mu_1 f^{-1}$.

Théorème 1.31. (transfert) Soient $\mu_2 := \mu_1 f^{-1}$, la mesure image de μ_1 par f , et $h \in \mathcal{L}^0(\mathcal{A}_2)$. Alors $h \in \mathcal{L}^1(\mathcal{A}_2, \mu_2)$ si et seulement si $h \circ f \in \mathcal{L}^1(\mathcal{A}_1, \mu_1)$. Dans ces conditions, on a

$$\int_{\Omega_2} h d(\mu_1 f^{-1}) = \int_{\Omega_1} (h \circ f) d\mu_1. \quad (1.6)$$

Démonstration. On commence par vérifier la formule de transfert (1.6) pour les fonctions positives. La formule est vraie pour les fonctions $\mathbf{1}_{A_2}$, $A_2 \in \mathcal{A}_2$, puis, par linéarité, pour les fonctions simples positives, et on conclut par le biais du théorème de convergence monotone. Pour h de signe arbitraire intégrable, on applique le résultat précédent à h^+ et h^- . Enfin, la formule de transfert montre que $h \in \mathcal{L}^1(\mathcal{A}_2, \mu_2)$ ssi $h^+ \circ f$ et $h^- \circ f \in \mathcal{L}^1(\mathcal{A}_1, \mu_1)$, et l'équivalence découle du fait que $h^+ \circ f = (h \circ f)^+$ et $h^- \circ f = (h \circ f)^-$. \diamond

1.3.2 Mesures définies par des densités

Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré, et soit $f \in \mathcal{L}_+^0(\mathcal{A})$ une fonction mesurable positive finie. On définit

$$\nu(A) := \mu(f\mathbf{1}_A) = \int_A f d\mu \quad \text{pour tout } A \in \mathcal{A}.$$

Exercice 1.32. Vérifier que ν est une mesure sur (Ω, \mathcal{A}) .

Définition 1.33. (i) La mesure ν est appelée mesure de densité f par rapport à μ , et on note $\nu = f \cdot \mu$.

(ii) Soient μ_1, μ_2 deux mesures sur un espace mesurable (Ω, \mathcal{A}) . On dit que μ_2 est absolument continue par rapport à μ_1 , et on note $\mu_2 \prec \mu_1$, si pour tout $A \in \mathcal{A}$:

$$\mu_2(A) = 0 \implies \mu_1(A) = 0.$$

Si on dit que μ_2 est étrangère à μ_1 .

(iii) Si $\mu_2 \prec \mu_1$ et $\mu_1 \prec \mu_2$, on dit que μ_1 et μ_2 sont équivalentes, et on note $\mu_1 \sim \mu_2$. Si $\mu_2 \not\prec \mu_1$ et $\mu_1 \not\prec \mu_2$, on dit que μ_1 et μ_2 sont singulières.

Ainsi, la mesure $f \cdot \mu$ est absolument continue par rapport à μ .

Théorème 1.34. (i) Pour $g : \Omega \rightarrow [0, \infty]$ \mathcal{A} -mesurable positive, on a $(f \cdot \mu)(g) = \mu(fg)$.

(ii) Pour $g \in \mathcal{L}_+^0(\mathcal{A})$, on a $g \in \mathcal{L}^1(\mathcal{A}, f \cdot \mu)$ ssi $fg \in \mathcal{L}^1(\mathcal{A}, \mu)$, et alors $(f \cdot \mu)(g) = \mu(fg)$.

Exercice 1.35. Prouver le théorème 1.34 (considérer d'abord les fonctions simple, puis passer aux fonctions positives par convergence monotone, enfin les fonctions intégrable en décomposant $f = f^+ - f^-$).

1.4 Inégalités remarquables

Dans ce paragraphe, nous énonçons trois inégalités qui sont très utiles. Afin d'habituer le lecteur à la manipulation des mesures et de l'intégration, nous formulons les résultats sous forme d'exercices.

Exercice 1.36. (Inégalité de Markov) Soit f une fonction \mathcal{A} -mesurable, et $g : \mathbb{R} \rightarrow \mathbb{R}_+$ une fonction borelienne croissante positive.

1. Justifier que $g \circ f$ est une fonction mesurable, et montrer l'inégalité de Markov :

$$\mu(g \circ f) \geq g(c)\mu(\{f \geq c\}) \quad \text{pour tout } c \in \mathbb{R}. \quad (1.7)$$

2. Montrer que

$$\begin{aligned} c\mu(\{f \geq c\}) &\leq \mu(f) \quad \text{pour tout } f \in \mathcal{L}_+^0(\mathcal{A}) \text{ et } c > 0, \\ c\mu(\{|f| \geq c\}) &\leq \mu(|f|) \quad \text{pour tout } f \in \mathcal{L}^1(\mathcal{A}, \mu) \text{ et } c > 0. \end{aligned}$$

3. Montrer l'inégalité de Chebyshev :

$$c^2\mu(\{|f| \geq c\}) \leq \mu(f^2) \quad \text{pour tout } f \text{ tel que } f^2 \in \mathcal{L}^1(\mathcal{A}, \mu) \text{ et } c > 0.$$

4. Montrer que

$$\mu(\{f \geq c\}) \leq \inf_{\tau > 0} e^{-\tau c} \mu(e^{\tau f}) \quad \text{pour tout } f \in \mathcal{L}^0(\mathcal{A}), \text{ et } c \in \mathbb{R}.$$

Exercice 1.37. (*Inégalité de Schwarz*) Soient $(\Omega, \mathcal{A}, \mu)$ un espace mesuré, et $f, g : \mathcal{A} \rightarrow \mathbb{R}_+$ deux fonctions mesurables positives telle que $\mu(f^2) + \mu(g^2) < \infty$.

1. Montrer que $\mu(fg) < \infty$.
2. Montrer que $\mu(fg)^2 \leq \mu(f^2)\mu(g^2)$ (Indication : considérer la fonction $xf + g, x \in \mathbb{R}$).
3. Montrer que l'inégalité de Schwarz dans la question 2 est valable sans la condition de positivité de f et g .

Exercice 1.38. (*Inégalité de Hölder, inégalité de Minkowski*) On admet l'inégalité de Jensen, valable pour une mesure positive ν sur $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ telle que $\nu(\mathbb{R}) = 1$:

$$\nu(c(f)) \geq c(\nu(f)) \quad \text{pour } f, c(f) \in \mathcal{L}^1(\mathcal{B}_{\mathbb{R}}, \nu) \text{ et } c(\cdot) \text{ convexe,}$$

qui sera démontrée dans le chapitre 2, théorème 2.6.

Soient $(\Omega, \mathcal{A}, \mu)$ un espace mesuré et $f, g : \Omega \rightarrow \mathbb{R}$ deux fonctions mesurables avec

$$\mu(|f|^p) < \infty \text{ et } \mu(|g|^q) < \infty \quad \text{où } p > 1, \frac{1}{p} + \frac{1}{q} = 1. \quad (1.8)$$

1. On suppose $f, g \geq 0$ et $\mu(f^p) > 0$. Montrer l'inégalité de Hölder :

$$\mu(|fg|) \leq \mu(|f|^p)^{1/p} \mu(|g|^q)^{1/q}.$$

(Indication : introduire la mesure $\nu := \frac{f^p}{\mu(f^p)} \cdot \mu$.)

2. Montrer que l'inégalité de Hölder de la question 1 est valable sous les conditions (1.8) sans les conditions supplémentaires de la question précédente.
3. En déduire l'inégalité de Minkowski :

$$\mu(|f + g|^p)^{1/p} \leq \mu(|f|^p)^{1/p} + \mu(|g|^p)^{1/p}.$$

(Indication : décomposer $|f + g|^p = (f + g)|f + g|^{p-1}$.)

1.5 Espaces produits

1.5.1 Construction et intégration

Dans ce paragraphe, nous faisons la construction de la mesure produit sur le produit de deux espaces mesurés.

Soient $(\Omega_1, \mathcal{A}_1, \mu_1), (\Omega_2, \mathcal{A}_2, \mu_2)$ deux espaces mesurés. Sur l'espace produit $\Omega_1 \times \Omega_2$, on vérifie immédiatement que $\mathcal{A}_1 \times \mathcal{A}_2$ est un π -système. On définit alors la σ -algèbre qu'il engendre

$$\mathcal{A}_1 \otimes \mathcal{A}_2 := \sigma(\mathcal{A}_1 \times \mathcal{A}_2).$$

Sur cette structure d'espace mesurable $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$, on veut définir une mesure μ telle que

$$\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2) \quad \text{pour tous } (A_1, A_2) \in \mathcal{A}_1 \times \mathcal{A}_2, \quad (1.9)$$

puis définir l'intégrale d'une fonction $f : \Omega_1 \times \Omega_2 \rightarrow \mathbb{R}$ intégrable :

$$\int_{\Omega_1 \times \Omega_2} f d\mu.$$

Une question importante est de relier cette quantité aux intégrales doubles

$$\int_{\Omega_2} \left(\int_{\Omega_1} f d\mu_1 \right) d\mu_2 \quad \text{et} \quad \int_{\Omega_1} \left(\int_{\Omega_2} f d\mu_2 \right) d\mu_1,$$

qui pose tout d'abord les questions de

(1a) la μ_1 -intégrabilité de la fonction $f_2^{\omega_2} : \omega_1 \mapsto f(\omega_1, \omega_2)$,

(2a) la μ_2 -intégrabilité de la fonction $f_1^{\omega_1} : \omega_2 \mapsto f(\omega_1, \omega_2)$,

puis, une fois ces questions réglées,

(1b) la μ_1 -intégrabilité de la fonction $I_1^f : \omega_1 \mapsto \int f(\omega_1, \omega_2) d\mu_2(\omega_2)$,

(2b) la μ_2 -intégrabilité de la fonction $I_2^f : \omega_2 \mapsto \int f(\omega_1, \omega_2) d\mu_1(\omega_1)$.

Ces deux problèmes sont résolus aisément grâce au théorème des classes monotones :

Lemme 1.39. (a) Soit $f \in \mathcal{L}^\infty(\mathcal{A}_1 \otimes \mathcal{A}_2)$. Alors, pour tous $\omega_1 \in \Omega_1, \omega_2 \in \Omega_2$:

$$f_1^{\omega_1} \in \mathcal{L}^\infty(\mathcal{A}_2) \quad \text{et} \quad f_2^{\omega_2} \in \mathcal{L}^\infty(\mathcal{A}_1).$$

(b) Supposons de plus que μ_1 et μ_2 soient finies. Alors $I_i^f \in \mathcal{L}^1(\mathcal{A}_i, \mu_i)$ pour $i = 1, 2$ et

$$\int_{\Omega_1} I_1^f d\mu_1 = \int_{\Omega_2} I_2^f d\mu_2.$$

Démonstration. (a) Soit $\mathcal{H} := \{f \in \mathcal{L}^\infty(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2) : f_1^{\omega_1} \in \mathcal{L}^0(\Omega_2, \mathcal{A}_2) \text{ et } f_2^{\omega_2} \in \mathcal{L}^0(\Omega_1, \mathcal{A}_1)\}$. Les conditions H1 et H2, du théorème 1.18 des classes monotones, sont trivialement satisfaites par \mathcal{H} . De plus, rappelons que $\mathcal{A}_1 \times \mathcal{A}_2$ est un π -système engendrant $\mathcal{A}_1 \otimes \mathcal{A}_2$, par définition. Il est clair que $\mathcal{H} \supset \{\mathbf{1}_A : A \in \mathcal{A}_1 \times \mathcal{A}_2\}$. Le théorème des classes monotones permet de conclure que $\mathcal{H} = \mathcal{L}^\infty(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$.

Pour une fonction $f(\omega_1, \omega_2)$ non bornée, l'argument précédent montre que $f_n := (-n) \wedge f \wedge n \in \mathcal{H}$, et par passage à la limite, on obtient $f_1^{\omega_1} \in \mathcal{L}^0(\Omega_2, \mathcal{A}_2)$ et $f_2^{\omega_2} \in \mathcal{L}^0(\Omega_1, \mathcal{A}_1)$.

(b) Il suffit de refaire le même type d'argument que pour (a). \diamond

Grâce au dernier résultat, nous pouvons maintenant définir un candidat pour la mesure sur l'espace produit $\Omega_1 \times \Omega_2$ par :

$$\mu(A) := \int \left(\int \mathbf{1}_A d\mu_1 \right) d\mu_2 = \int \left(\int \mathbf{1}_A d\mu_2 \right) d\mu_1 \quad \text{pour tout } A \in \mathcal{A}_1 \otimes \mathcal{A}_2.$$

Théorème 1.40. (Fubini) L'application μ est une mesure sur $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$, appelée mesure produit de μ_1 et μ_2 , et notée $\mu_1 \otimes \mu_2$. C'est l'unique mesure sur $\Omega_1 \times \Omega_2$ vérifiant (1.9). De plus, pour tout $f \in \mathcal{L}_+^0(\mathcal{A}_1 \otimes \mathcal{A}_2)$,

$$\int f d\mu_1 \otimes \mu_2 = \int \left(\int f d\mu_1 \right) d\mu_2 = \int \left(\int f d\mu_2 \right) d\mu_1 \in [0, \infty] \quad (1.10)$$

Enfin, si $f \in \mathcal{L}^1(\mathcal{A}_1 \otimes \mathcal{A}_2, \mu_1 \otimes \mu_2)$, les égalités (1.10) sont valides.

Démonstration. On vérifie que $\mu_1 \otimes \mu_2$ est une mesure grâce aux propriétés élémentaires de l'intégrale de Lebesgue. L'unicité est une conséquence immédiate de la proposition 1.5. Les égalités (1.10) ont déjà été établies dans le lemme 1.39 (b) pour f bornée et des mesures finies. Pour généraliser à des fonctions f mesurables positives, on introduit des approximations croissantes, et on utilise le théorème de convergence monotone. Enfin, pour des fonctions $f \in \mathcal{L}^1(\mathcal{A}_1 \otimes \mathcal{A}_2, \mu_1 \otimes \mu_2)$, on applique le résultat précédent à f^+ et f^- . \diamond

Remarque 1.41. (i) La construction de ce paragraphe, ainsi que les résultats d'intégration ci-dessous, s'étendent sans difficulté pour la construction du produit de n espaces mesurés au prix de notations plus encombrantes.

(ii) Soit maintenant $(\Omega_i, \mathcal{A}_i)_{i \geq 1}$ une famille dénombrable d'espaces mesurés, et $\Omega := \prod_{i \geq 1} \Omega_i$. Pour tout sous-ensemble fini $I \subset \mathbb{N}$, et pour tous $A_i \in \mathcal{A}_i, i \in I$, on définit le *cylindre*

$$\mathcal{C}(A_i, i \in I) := \{\omega \in \Omega : \omega_i \in A_i \text{ pour } i \in I\}.$$

La σ -algèbre produit est alors définie par

$$\mathcal{A} := \otimes_{n \geq 1} \mathcal{A}_i := \sigma(\mathcal{C}(A_i, i \in I) : I \subset \mathbb{N}, \text{card}(I) < \infty).$$

1.5.2 Mesure image et changement de variable

Soit $\mathcal{O} = \mathbb{R}^n$, ou un sous-ensemble d'un espace de dimension n . Les outils développés dans les paragraphes précédents permettent de définir la mesure de Lebesgue sur \mathbb{R}^n à partir de notre construction de la mesure de Lebesgue sur \mathbb{R} .

Dans ce paragraphe, on considère une fonction

$$g : \Omega_1 \longrightarrow \Omega_2 \quad \text{où } \Omega_1, \Omega_2 \text{ ouverts de } \mathbb{R}^n.$$

On note $g = (g_1, \dots, g_n)$. Si g est différentiable en un point $x \in \Omega_1$, on note par

$$Dg(x) := \left(\frac{\partial g_i}{\partial x_j} \right)_{1 \leq i, j \leq n} \quad \text{et} \quad \det[Dg(x)]$$

la matrice jacobienne de f en x et son déterminant. Rappelons enfin que g est un C^1 -difféomorphisme si g est une bijection telle que g et g^{-1} sont de classe C^1 , et que dans ce cas

$$\det[Dg^{-1}(y)] = \frac{1}{\det[Dg \circ g^{-1}(y)]}.$$

Théorème 1.42. Soit μ_1 une mesure sur $(\Omega_1, \mathcal{B}_{\Omega_1})$ de densité par rapport à la mesure de Lebesgue $f_1 \in \mathcal{L}_+^0(\mathcal{B}_{\Omega_1})$, i.e. $\mu_1(dx) = \mathbf{1}_{\Omega_1} f_1(x) \cdot dx$. Si g est un C^1 -difféomorphisme, la mesure image $\mu_2 := \mu_1 \circ g^{-1}$ est absolument continue par rapport à la mesure de Lebesgue de densité

$$f_2(y) = \mathbf{1}_{\Omega_2}(y) f(g^{-1}(y)) |\det[Dg^{-1}(y)]| \text{ et } \int_{\Omega_1} h \circ g(x) f_1(x) dx = \int_{\Omega_2} h(y) f_2(y) dy$$

pour toute fonction $h : \Omega_2 \rightarrow \mathbb{R}$ positive ou μ_2 -intégrable.

Pour la démonstration, on renvoie au cours de première année.

1.6 Annexe du chapitre 1

1.6.1 π -système, d -système et unicité des mesures

Le but de ce paragraphe est de démontrer la proposition 1.5 dont nous rappelons l'énoncé.

Proposition 1.5 Soient \mathcal{I} un π -système, et μ, ν deux mesures finies sur l'espace mesurable $(\Omega, \sigma(\mathcal{I}))$. Si $\mu = \nu$ sur \mathcal{I} alors $\mu = \nu$ sur $\sigma(\mathcal{I})$.

Commençons par introduire une notion supplémentaire de classes d'ensembles.

Définition 1.43. Une classe $\mathcal{D} \subset \mathcal{P}(\Omega)$ est appelée d -système si $\Omega \in \mathcal{D}$, $B \setminus A \in \mathcal{D}$ pour tous $A, B \in \mathcal{D}$ avec $A \subset B$, et $\cup_n A_n \in \mathcal{D}$ pour toute suite croissante $(A_n)_n \subset \mathcal{D}$.

Lemme 1.44. Une classe $\mathcal{C} \subset \mathcal{P}(\Omega)$ est une σ -algèbre si et seulement si \mathcal{C} est un π -système et un d -système.

La preuve facile de ce résultat est laissée en exercice. Pour toute classe \mathcal{C} , on définit l'ensemble

$$d(\mathcal{C}) := \cap \{ \mathcal{D} \supset \mathcal{C} : \mathcal{D} \text{ est un } d\text{-système} \},$$

qui est le plus petit d -système contenant \mathcal{C} . L'inclusion $d(\mathcal{C}) \subset \sigma(\mathcal{C})$ est évidente.

Lemme 1.45. Pour un π -système \mathcal{I} , on a $d(\mathcal{I}) = \sigma(\mathcal{I})$.

Démonstration. D'après le lemme 1.44, il suffit de montrer que $d(\mathcal{I})$ est un π -système, i.e. que $d(\mathcal{I})$ est stable par intersection finie. On définit l'ensemble $\mathcal{D}' := \{A \in d(\mathcal{I}) : A \cap B \in d(\mathcal{I}) \text{ pour tout } B \in d(\mathcal{I})\}$, et on va montrer que $\mathcal{D}' = d(\mathcal{I})$ ce qui termine la démonstration.

1- On commence par montrer que l'ensemble $\mathcal{D}_0 := \{B \in d(\mathcal{I}) : B \cap C \in d(\mathcal{I}) \text{ pour tout } C \in \mathcal{I}\}$ est un d -système. En effet :

- $\Omega \in \mathcal{D}_0$;

- soient $A, B \in \mathcal{D}_0$ tels que $A \subset B$, et $C \in \mathcal{I}$; comme $A, B \in \mathcal{D}_0$, on a $(A \cap C)$ et $(B \cap C) \in d(\mathcal{I})$, et du fait que $d(\mathcal{I})$ est un d -système, on voit que $(B \setminus A) \cap C = (B \cap C) \setminus (A \cap C) \in d(\mathcal{I})$;
 - enfin, si $\mathcal{D}_0 \ni A_n \uparrow A$ et $C \in \mathcal{I}$, on a $A_n \cap C \in d(\mathcal{I})$ et donc $\lim \uparrow (A_n \cap C) = A \cap C \in d(\mathcal{I})$ du fait que $d(\mathcal{I})$ est un d -système;
 2- par définition $\mathcal{D}_0 \subset d(\mathcal{I})$, et comme on vient de montrer que c'est un d -système contenant \mathcal{I} , on voit qu'on a en fait $\mathcal{D}_0 = d(\mathcal{I})$; on vérifie maintenant que ceci implique que $\mathcal{I} \in \mathcal{D}'$;
 3- enfin, en procédant comme dans les étapes précédentes, on voit que \mathcal{D}' est un d -système. \diamond

Preuve de la proposition 1.5 On vérifie aisément que l'ensemble $\mathcal{D} := \{A \in \sigma(\mathcal{I}) : \mu(A) = \nu(A)\}$ est un d -système (c'est à ce niveau qu'on utilise que les mesures sont finies afin d'éviter des formes indéterminées du type $\infty - \infty$). Or, par hypothèse, \mathcal{D} contient le π -système \mathcal{I} . On déduit alors du lemme 1.45 que \mathcal{D} contient $\sigma(\mathcal{I})$ et par suite $\mathcal{D} = \sigma(\mathcal{I})$. \diamond

1.6.2 Mesure extérieure et extension des mesures

Le but de ce paragraphe est de démontrer du théorème de Carathéodory 1.6 dont nous rappelons l'énoncé.

Théorème 1.6 Soient \mathcal{A}_0 une algèbre sur Ω , et $\mu_0 : \mathcal{A}_0 \rightarrow \mathbb{R}_+$ une fonction σ -additive. Alors il existe une mesure μ sur $\mathcal{A} := \sigma(\mathcal{A}_0)$ telle que $\mu = \mu_0$ sur \mathcal{A}_0 . Si de plus $\mu_0(\Omega) < \infty$, alors une telle extension μ est unique.

Pour préparer la démonstration, nous considérons une σ -algèbre $\mathcal{A}' \subset \mathcal{P}(\Omega)$, et une application $\lambda : \mathcal{A}' \rightarrow [0, \infty]$ vérifiant $\lambda(\emptyset) = 0$.

Définition 1.46. On dit que λ est une mesure extérieure sur (Ω, \mathcal{A}') si

- (i) $\lambda(\emptyset) = 0$,
- (ii) λ est croissante : pour $A_1, A_2 \in \mathcal{A}'$, $\lambda(A_1) \leq \lambda(A_2)$ dès que $A_1 \subset A_2$,
- (iii) λ est σ -sous-additive : pour $(A_n)_n \subset \mathcal{A}'$, on a $\lambda(\cup_n A_n) \leq \sum_n \lambda(A_n)$.

Définition 1.47. On dit qu'un élément $A \in \mathcal{A}'$ est un λ -ensemble si

$$\lambda(A \cap B) + \lambda(A^c \cap B) = \lambda(B) \quad \text{pour tout } B \in \mathcal{A}_0,$$

(en particulier, $\lambda(\emptyset) = 0$). On note par \mathcal{A}'_λ l'ensemble de tous les λ -ensembles de \mathcal{A}' .

Le résultat suivant utilise uniquement le fait que \mathcal{A}' est une algèbre.

Lemme 1.48. L'ensemble \mathcal{A}'_λ est une algèbre, et la restriction de λ à \mathcal{A}'_λ est additive et vérifie pour tout $B \in \mathcal{A}'$:

$$\lambda(\cup_{i=1}^n (A_i \cap B)) = \sum_{i=1}^n \lambda(A_i \cap B) \quad \text{dès que } A_1, \dots, A_n \in \mathcal{A}'_\lambda \text{ sont disjoints.}$$

Ce lemme, dont la démonstration (facile) est reportée pour la fin du paragraphe, permet de montrer le résultat suivant :

Lemme 1.49. (*Carathéodory*) Soit λ une mesure extérieure sur (Ω, \mathcal{A}') . Alors \mathcal{A}'_λ est une σ -algèbre, et la restriction de λ à \mathcal{A}'_λ est σ -additive, et par suite λ est une mesure sur $(\Omega, \mathcal{A}'_\lambda)$.

Démonstration. En vue du lemme 1.48, il reste à montrer que pour une suite d'ensembles disjoints $(A_n)_n \subset \mathcal{A}'_\lambda$, on a

$$\cup_n A_n \in \mathcal{A}'_0(\lambda) \quad \text{et} \quad \lambda(\cup_n A_n) = \sum_n \lambda(A_n). \quad (1.11)$$

Notons $\bar{A}_n := \cup_{i \leq n} A_i$, $\bar{A} := \cup_n A_n$, et remarquons que $\bar{A}^c \subset \bar{A}_n^c$. D'après le lemme 1.48, $\bar{A}_n \in \mathcal{A}'_\lambda$ et pour tout $B \in \mathcal{A}'$:

$$\lambda(B) = \lambda(\bar{A}_n^c \cap B) + \lambda(\bar{A}_n \cap B) \geq \lambda(\bar{A}^c \cap B) + \lambda(\bar{A}_n \cap B) = \lambda(\bar{A}^c \cap B) + \sum_{i \leq n} \lambda(A_i \cap B).$$

On continue en faisant tendre n vers l'infini, et en utilisant (deux fois) la sous-additivité de λ :

$$\lambda(B) \geq \lambda(\bar{A}^c \cap B) + \sum_n \lambda(A_n \cap B) \geq \lambda(\bar{A}^c \cap B) + \lambda(A \cap B) \geq \lambda(B).$$

On déduit que toutes les inégalités sont des égalités, prouvant que $\bar{A} \in \mathcal{A}'_\lambda$, et pour $B = \bar{A}$ on obtient la propriété de sous-additivité de λ , finissant la preuve de (1.11). \diamond

Nous avons maintenant tous les ingrédients pour montrer le théorème d'extension de Carathéodory.

Preuve du théorème 1.6 On considère la σ -algèbre $\mathcal{A}' := \mathcal{P}(\Omega)$, et on définit l'application sur Ω :

$$\lambda(A) := \inf \left\{ \sum_n \mu_0(B_n) : (B_n)_n \subset \mathcal{A}_0, B_n \text{ disjoints et } A \subset \cup_n B_n \right\}.$$

Etape 1 Montrons que λ est une mesure extérieure sur (Ω, \mathcal{P}) , ce qui implique par le lemme 1.49 que

$$\lambda \text{ est une mesure sur } (\Omega, \mathcal{A}'_\lambda). \quad (1.12)$$

Il est clair que $\lambda(\emptyset) = 0$, et que λ est croissante, il reste donc à vérifier que λ est σ -sous-additive. Soit une suite $(A_n)_n \subset \mathcal{P}$ telle que $\lambda(A_n) < \infty$ pour tout n , et soit $A := \cup_n A_n$. Pour tout $\varepsilon > 0$ et $n \geq 1$, on considère une suite ε -optimale $(B_i^{n,\varepsilon})_i \subset \mathcal{A}_0$ du problème de minimisation $\lambda(A_n)$, i.e. $B_i^{n,\varepsilon} \cap B_j^{n,\varepsilon} = \emptyset$,

$$A_n \subset \cup_k B_k^{n,\varepsilon} \quad \text{et} \quad \lambda(A_n) > \sum_k \mu_0(B_k^{n,\varepsilon}) - \varepsilon 2^{-n}.$$

Alors, $\lambda(A) \leq \sum_{n,k} \mu_0(B_k^{n,\varepsilon}) < \varepsilon + \sum_n \lambda(A_n) \rightarrow \sum_n \lambda(A_n)$ quand $\varepsilon \rightarrow 0$.

Etape 2 Rappelons que $\sigma(\mathcal{A}_0) \subset \mathcal{A}'_\lambda$. Alors, pour finir la démonstration de l'existence d'une extension, il nous reste à montrer que

$$\mathcal{A}_0 \subset \mathcal{A}'_\lambda \quad \text{et} \quad \lambda = \mu_0 \text{ sur } \mathcal{A}_0, \quad (1.13)$$

pour ainsi définir μ comme la restriction de λ à $\sigma(\mathcal{A}_0)$.

1- Commençons par montrer que $\lambda = \mu_0$ sur \mathcal{A}_0 . L'inégalité $\lambda \leq \mu_0$ sur \mathcal{A}_0 est triviale. Pour l'inégalité inverse, on considère $A \in \mathcal{A}_0$ et une suite $(B_n)_n \subset \mathcal{A}_0$ d'éléments disjoints telle $A \subset \cup_n B_n$. Alors, en utilisant la σ -additivité de μ_0 sur \mathcal{A}_0 :

$$\mu_0(A) = \mu_0(\cup_n (A \cap B_n)) = \sum_n \mu_0(A \cap B_n) \leq \sum_n \mu_0(B_n) = \lambda(A).$$

2- Montrons maintenant que $\mathcal{A}_0 \subset \mathcal{A}'_\lambda$. Soient $A \in \mathcal{A}'$, $\varepsilon > 0$ et $(B_n)_n \subset \mathcal{A}_0$ une suite ε -optimale pour le problème de minimsation $\lambda(A)$. Alors, pour tout $A_0 \in \mathcal{A}_0$, on a

$$\begin{aligned} \lambda(A) + \varepsilon &\geq \sum_n \mu_0(B_n) = \sum_n \mu_0(A_0 \cap B_n) + \sum_n \mu_0(A_0^c \cap B_n) \\ &\geq \lambda((A_0 \cap A) + \lambda((A_0^c \cap A) \\ &\geq \lambda(A), \end{aligned}$$

où les deux dernières inégalités découlent respectivement de la monotonie et la sous-linéarité de λ . Comme $\varepsilon > 0$ est arbitraire, ceci montre que A_0 est un λ -ensemble, i.e. $A_0 \in \mathcal{A}'_\lambda$. \diamond

Preuve du lemme 1.48 1- Commençons par montrer que \mathcal{A}'_λ est une algèbre. Il est clair que $\Omega \in \mathcal{A}'_\lambda$ et que \mathcal{A}'_λ est stable par passage au complémentaire. Il reste à montrer que $A = A_1 \cap A_2 \in \mathcal{A}_0(\lambda)$ pour tous $A_1, A_2 \in \mathcal{A}_0(\lambda)$. En utilisant successivement le fait que $A_2 \in \mathcal{A}'_\lambda$ et que $A_2 \cap A^c = A_1^c \cap A_2$, $A_2^c \cap A^c = A_2^c$, on calcule directement :

$$\lambda(A^c \cap B) = \lambda(A_2 \cap A^c \cap B) + \lambda(A_2^c \cap A^c \cap B) = \lambda(A_1^c \cap A_2 \cap B) + \lambda(A_2^c \cap B).$$

On continue en utilisant le fait que $A_1, A_2 \in \mathcal{A}'_\lambda$:

$$\lambda(A^c \cap B) = \lambda(A_2 \cap B) - \lambda(A \cap B) + \lambda(A_2^c \cap B) = \lambda(B) - \lambda(A \cap B).$$

2- Pour des ensembles disjoints $A_1, A_2 \in \mathcal{A}'_\lambda$, on a $(A_1 \cup A_2) \cap A_1 = A_1$ et $(A_1 \cup A_2) \cap A_1^c = A_2$, et on utilise le fait que $A_1 \in \mathcal{A}'_\lambda$ pour voir que $\lambda((A_1 \cup A_2) \cap B) = \lambda(A_1 \cap B) + \lambda(A_2 \cap B)$, ce qui est l'égalité annoncée pour $n = 2$. L'extension pour un n plus grand est triviale, et la σ -additivité de λ en est une conséquence immédiate. \diamond

1.6.3 Démonstration du théorème des classes monotones

Rappelons l'énoncé.

Théorème 1.18 *Soit \mathcal{H} une classes de fonctions réelles bornées sur Ω vérifiant les conditions suivantes :*

- (H1) *\mathcal{H} est un espace vectoriel contenant la fonction constante $\mathbf{1}$,*
- (H2) *pour toute suite croissante $(f_n)_n \subset \mathcal{H}$ de fonctions positives telle que $f := \lim \uparrow f_n$ est bornée, on a $f \in \mathcal{H}$.*

Soit \mathcal{I} un π -système tel que $\{\mathbf{1}_A : A \in \mathcal{I}\} \subset \mathcal{H}$. Alors $\mathcal{L}^\infty(\sigma(\mathcal{I})) \subset \mathcal{H}$.

Démonstration. D'après les conditions H1 et H2, on voit immédiatement que l'ensemble $\mathcal{D} := \{F \subset \Omega : \mathbf{1}_F \in \mathcal{H}\}$ est un d -système. De plus, comme \mathcal{D} contient le π -système \mathcal{I} , on déduit du lemme 1.44 que $\sigma(\mathcal{I}) \subset \mathcal{D}$. Soit maintenant $f \in \mathcal{L}^\infty(\sigma(\mathcal{I}))$ bornée par $M > 0$, et

$$\phi_n(\omega) := \sum_{i=0}^{M2^{-n}} i2^{-n} \mathbf{1}_{A_i^n}(\omega), \text{ où } A_i^n := \{\omega \in \Omega : i2^{-n} \leq f^+(\omega) < (i+1)2^{-n}\}.$$

Comme $A_i^n \in \sigma(\mathcal{I})$, on déduit de la structure d'espace vectoriel (condition H1) de \mathcal{H} que $\phi_n \in \mathcal{H}$. De plus $(\phi_n)_n$ étant une suite croissante de fonctions positives convergeant vers la fonction bornée f^+ , la condition H2 assure que $f^+ \in \mathcal{H}$. On montre de même que $f^- \in \mathcal{H}$ et, par suite, $f = f^+ - f^- \in \mathcal{H}$ d'après H1. \diamond

Chapitre 2

Préliminaires de la théorie des probabilités

Dans ce chapitre, on spécialise l'analyse aux cas d'une mesures de probabilité, i.e. une mesure $\mathbb{P} : \mathcal{A} \rightarrow \mathbb{R}_+$ telle que $\mathbb{P}[\Omega] = 1$. On dit alors que $(\Omega, \mathcal{A}, \mathbb{P})$ est un espace probabilisé.

Bien évidemment, tous les résultats du chapitre précédent sont valables dans le cas présent. En plus de ces résultats, nous allons exploiter l'intuition probabiliste pour introduire de nouveaux concepts et obtenir de nouveaux résultats.

Ainsi, l'ensemble Ω s'interprète comme l'ensemble de tous les événements élémentaires, et tout point $\omega \in \Omega$ est un événement élémentaire. La σ -algèbre \mathcal{A} est l'ensemble de tous les événements réalisables.

On remplacera systématiquement la terminologie \mathbb{P} -p.p. par \mathbb{P} -*presque sur-ement*, notée \mathbb{P} -p.s. ou plus simplement p.s. s'il n'y a pas de risque de confusion.

Les fonctions \mathbb{P} -mesurables sont appelées variables aléatoires (on écrira v.a.), et sont le plus souvent notées avec des lettres majuscules, typiquement X . La loi image $\mathbb{P}X^{-1}$ est appelée distribution de la v.a. X , et sera notée \mathbb{P}^X s'il n'y a pas besoin de rappeler la probabilité \mathbb{P} .

2.1 Variables aléatoires

2.1.1 σ -algèbre engendrée par une v.a.

Nous commençons par donner un sens précis à l'information révélée par une famille de variables aléatoires.

Définition 2.1. Soient \mathbb{T} un ensemble, et $\{X_\tau, \tau \in \mathbb{T}\}$ une famille quelconque de v.a. La σ -algèbre engendrée par cette famille $\mathcal{X} := \sigma(X_\tau : \tau \in \mathbb{T})$ est la plus petite σ -algèbre sur Ω telle que X_τ est \mathcal{X} -mesurable pour tout $\tau \in \mathbb{T}$, i.e.

$$\sigma(X_\tau : \tau \in \mathbb{T}) = \sigma(\{X_\tau^{-1}(A) : \tau \in \mathbb{T} \text{ et } A \in \mathcal{B}_{\mathbb{R}}\}). \quad (2.1)$$

Il est clair que si les X_τ sont \mathcal{A} -mesurables, alors $\sigma(X_\tau : \tau \in \mathbb{T}) \subset \mathcal{A}$.

Lemme 2.2. *Soient X et Y deux v.a. sur $(\Omega, \mathcal{A}, \mathbb{P})$ prenant valeurs respectivement dans \mathbb{R} et dans \mathbb{R}^n . Alors X est $\sigma(Y)$ -mesurable si et seulement si il existe une fonction borélienne $f : \mathbb{R}^n \rightarrow \mathbb{R}$ telle que $X = f(Y)$.*

Démonstration. Seule la condition nécessaire est non triviale. Par ailleurs quitte à transformer X par une fonction bijective bornée, on peut se limiter au cas où X est bornée. On définit

$$\mathcal{H} := \{X \in \mathcal{L}^\infty(\sigma(Y)) : \exists f \in \mathcal{L}^\infty(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n}), X = f(Y) : \},$$

et on remarque que $\{1_A : A \in \sigma(Y)\} \subset \mathcal{H}$: d'après (2.1), pour tout $A \in \sigma(Y)$, il existe $B \in \mathcal{A}$ tel que $A = Y^{-1}(B)$, et par suite $\mathbf{1}_A = \mathbf{1}_B(Y)$.

Pour conclure, il nous suffit de montrer que \mathcal{H} vérifie les conditions du théorème des classes monotones. Il est clair que \mathcal{H} est un espace vectoriel contenant la v.a. constante 1. Soient $X \in \mathcal{L}_+^\infty(\mathcal{A}, \mathbb{P})$ et $(f_n(Y))_n$ une suite croissante de \mathcal{H} telle que $f_n(Y) \uparrow X$. Alors $X = f(Y)$, où $f = \limsup f_n$ est $\mathcal{B}_{\mathbb{R}^n}$ -mesurable bornée (puisque X l'est). \diamond

2.1.2 Distribution d'une v.a.

La distribution, ou la loi, d'une v.a. X sur $(\Omega, \mathcal{A}, \mathbb{P})$ est définie par la mesure image $\mathbb{P}^X := \mathbb{P}X^{-1}$. En utilisant le π -système $\pi(\mathbb{R}) = \{] - \infty, c] : c \in \mathbb{R}\}$, on déduit de la proposition 1.5 que la loi \mathbb{P}^X est caractérisée par la fonction

$$F_X(c) := \mathbb{P}^X(] - \infty, c]) = \mathbb{P}[X \leq c], \quad c \in \mathbb{R}. \quad (2.2)$$

La fonction F_X est appelée fonction de répartition.

Proposition 2.3. (i) *La fonction F_X est croissante continue à droite, et $F_X(-\infty) = 0$, $F_X(\infty) = 1$,*
(ii) *Soit F une fonction croissante continue à droite, et $F(-\infty) = 0$, $F(\infty) = 1$. Alors il existe une variable aléatoire \tilde{X} sur un espace de probabilité $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{\mathbb{P}})$ telle que $F = F_{\tilde{X}}$.*

Démonstration. (i) est triviale. Pour (ii), une première approche consiste à construire une loi $\tilde{\mathcal{L}}$ en suivant le schémas de construction de la mesure de Lebesgue dans l'exemple 1.7 qui utilise le théorème d'extension de Carathéodory ; on prend alors $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{\mathbb{P}}) = (\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \tilde{\mathcal{L}})$ et $X(\omega) = \omega$. La remarque suivante donne une approche alternative. \diamond

Remarque 2.4. Etant donnée une fonction de répartition, ou une loi, voici une construction explicite d'une v.a. lui correspondant. Cette construction est utile, par exemple, pour la simulation de v.a. Sur l'espace de probabilité $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{\mathbb{P}}) := ([0, 1], \mathcal{B}_{[0,1]}, \lambda)$, λ étant la mesure de Lebesgue, on définit

$$\bar{X}(\omega) := \inf\{u : F(u) > \omega\} \quad \text{et} \quad \underline{X}(\omega) := \inf\{u : F(u) \geq \omega\}$$

1- $F_{\underline{X}} = F$: nous allons montrer que

$$\omega \leq F(c) \iff \underline{X}(\omega) \leq c, \quad (2.3)$$

et par suite $\mathbb{P}[\underline{X} \leq c] = F(c)$.

L'implication \implies découle de la définition. Pour l'implication inverse, on observe que $F(\underline{X}(\omega)) \geq \omega$. En effet, si ce n'était pas le cas, on déduirait de la continuité à droite de F que $F(\underline{X}(\omega) + \varepsilon) < \omega$ pour $\varepsilon > 0$ assez petit, impliquant l'absurdité $\underline{X}(\omega) + \varepsilon \leq \underline{X}(\omega)$!

Avec cette observation et la croissance de F , on voit que $\underline{X}(\omega) \leq c$ implique $\omega \leq F(\underline{X}(\omega)) \leq F(c)$ implique $\omega \leq F(c)$.

2- $F_{\overline{X}} = F$: par définition de \overline{X} , on a $\omega < F(c)$ implique $\overline{X}(\omega) \leq c$. Mais $\overline{X}(\omega) \leq c$ implique $\underline{X}(\omega) \leq c$ puisque $\underline{X} \leq \overline{X}$. On en déduit que $F(c) \leq \mathbb{P}[\overline{X} \leq c] \leq \mathbb{P}[\underline{X} \leq c] = F(c)$.

2.2 Espérance de variables aléatoires

Pour une v.a. $X \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})$, l'espérance dans le vocabulaire probabiliste est l'intégrale de X par rapport à \mathbb{P} :

$$\mathbb{E}[X] := \mathbb{P}(X) = \int_{\Omega} X d\mathbb{P}.$$

Pour une v.a. positive, $\mathbb{E}[X] \in [0, \infty]$ est toujours bien définie. Bien sûr, toutes les propriétés du chapitre 1 sont valides. Nous allons en obtenir d'autres comme conséquence de $\mathbb{P}[\Omega] = 1$.

2.2.1 Variables aléatoires à densité

Revenons à présente à la loi \mathbb{P}^X sur $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ d'une v.a. X sur $(\Omega, \mathcal{A}, \mathbb{P})$. Par définition, on a :

$$\mathbb{P}^X(B) = \mathbb{P}[X \in B] \quad \text{pour tout } B \in \mathcal{B}_{\mathbb{R}}.$$

Par linéarité de l'intégrale (par rapport à \mathbb{P}^X), on obtient $\mathbb{E}[g(X)] = \mathbb{P}^X(g) = \int_{\mathbb{R}} g d\mathbb{P}^X$ pour toute fonction simple $g \in \mathcal{S}^+$. On étend alors cette relation aux fonction g mesurables positives, par le théorème de convergence monotone, puis à \mathcal{L}^1 en décomposant $g = g^+ - g^-$. Ceci montre que $g(X) \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})$ ssi $g \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \mathbb{P}^X)$ et

$$\mathbb{E}[g(X)] = \mathbb{P}^X(g) = \int_{\mathbb{R}} g d\mathbb{P}^X. \quad (2.4)$$

Définition 2.5. On dit que X a une densité de probabilité f_X si \mathbb{P}^X est absolument continue par rapport à la mesure de Lebesgue sur \mathbb{R} et :

$$\mathbb{P}[X \in B] = \int_B f_X(x) dx \quad \text{pour tout } B \in \mathcal{B}_{\mathbb{R}}.$$

Le lien entre la densité de probabilité, si elle existe, et la fonction de répartition (qui existe toujours) est facilement établi en considérant $B =] - \infty, c]$:

$$F_X(c) = \int_{]-\infty, c]} f_X(x) dx \quad \text{pour tout } c \in \mathbb{R}.$$

qui exprime que “ f_X est la dérivée de F_X ” aux points de continuité de f . Enfin, pour une v.a. X à densité f_X , on peut reformuler (2.4) sous la forme :

$$g(X) \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P}) \quad \text{ssi} \quad \int_{\mathbb{R}} |g(x)| f_X(x) dx < \infty$$

et on peut ré-écrire l’espérance sous la forme

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x) f_X(x) dx.$$

2.2.2 Inégalité de Jensen

Une fonction convexe $g : \mathbb{R}^n \rightarrow \mathbb{R}$ est au dessus de son hyperplan tangent en tout point de l’intérieur du domaine. Si on admet ce résultat, alors, on peut écrire pour une v.a. intégrable X que

$$g(X) \geq g(\mathbb{E}[X]) + \langle p_{\mathbb{E}[X]}, X - \mathbb{E}[X] \rangle,$$

où $p_{\mathbb{E}[X]}$ est le gradient de g au point $\mathbb{E}[X]$, si g est dérivable en ce point. si g n’est pas dérivable ce résultat est encore valable en remplaçant le gradient par la notion de sous-gradient... Dans la démonstration qui va suivre, nous allons éviter de passer par cette notion d’analyse convexe, et utiliser un argument d’approximation. En prenant l’espérance dans la dernière inégalité, on obtient l’inégalité de Jensen :

Théorème 2.6. Soient $X \in \mathcal{L}^1(\mathcal{A}, \mathbb{P})$ et $g : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{\infty\}$ une fonction convexe telle que $\mathbb{E}[|g(X)|] < \infty$. Alors $\mathbb{E}[g(X)] \geq g(\mathbb{E}[X])$.

Démonstration. Si g est dérivable sur l’intérieur du domaine, le résultat découle de la discussion qui précède l’énoncé. Dans le cas général, on commence par montrer le résultat pour une fonction g bornée, puis on étend aux fonctions g vérifiant les hypothèses du théorème.

1- Supposons d’abord g bornée. Soit ϕ un noyau de convolution ($\phi \geq 0$ et $\int \phi = 1$) de classe C^1 à support compact, et $\varphi^n(x) := n^d \phi(nx)$. Alors la fonction $g_n := g * \varphi_n$ est bornée, de classe C^1 , et converge vers g . De plus, du fait que $\varphi \geq 0$, on voit que g_n hérite la convexité de g . Alors, l’inégalité de Jensen est alors vérifiée pour g_n :

$$\mathbb{E}[g_n(X)] \geq g_n(\mathbb{E}[X]),$$

et on la déduit pour g par passage à la limite en utilisant le théorème de convergence dominée.

2- Si g n’est pas bornée, on note pour tout $n \geq 1$:

$$D_n := \{x \in \mathbb{R}^d : |g(x)| \leq n\} \quad \text{et} \quad X_n := X \mathbf{1}_{D_n}(X) + n \mathbf{1}_{D_n^c}(X).$$

D'après l'étape précédente, on a :

$$\mathbb{E}[g(X_n)] \geq g(\mathbb{E}[X_n]).$$

Remarquons maintenant que $|X_n| \leq |X| + |g(X)| \in \mathcal{L}^1(\mathcal{A}, \mathbb{P})$ et $|g(X_n)| \leq |g(X)| \in \mathcal{L}^1(\mathcal{A}, \mathbb{P})$. On obtient alors le résultat souhaité par passage à la limite en utilisant le théorème de convergence dominée. \diamond

2.2.3 Fonction caractéristique

Dans tout ce paragraphe X désigne un vecteur aléatoire sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, à valeurs dans \mathbb{R}^d .

Définition 2.7. On appelle fonction caractéristique de X la fonction $\Phi_X : \mathbb{R}^d \rightarrow \mathbb{C}$ définie par

$$\Phi_X(u) := \mathbb{E} \left[e^{i\langle u, X \rangle} \right] \quad \text{pour tout } u \in \mathbb{R}^d.$$

La fonction caractéristique dépend uniquement de la loi de X :

$$\Phi_X(u) = \int_{\mathbb{R}^d} e^{i\langle u, x \rangle} d\mathbb{P}^X(x),$$

et n'est rien d'autre que la transformée de Fourier de \mathbb{P}^X au point $-u/2\pi$. L'intégrale de Lebesgue d'une fonction à valeurs complexes est définie de manière naturelle en séparant partie réelle et partie imaginaire. La fonction caractéristique est bien définie pour tout $u \in \mathbb{R}^d$ comme intégrale d'une fonction de module 1. Enfin, pour deux v.a. X et Y , on a

$$\Phi_X(u) = \overline{\Phi_{-X}(u)} \text{ et } \Phi_{aX+b}(u) = e^{ib} \Phi_X(au) \quad \text{pour tous } u \in \mathbb{R}^d, a, b \in \mathbb{R}.$$

Les propriétés suivantes des fonctions caractéristiques peuvent être démontrées facilement grâce au théorème de convergence dominée.

Lemme 2.8. Soit Φ_X la fonction caractéristique d'une v.a. X . Alors $\Phi_X(0) = 1$, et Φ_X est continue bornée (par 1) sur \mathbb{R}^d .

Démonstration. $\Phi_X(0) = 1$ et $|\Phi_X| \leq 1$ sont des propriétés évidentes, la continuité est une conséquence immédiate du théorème de convergence dominée.

\diamond

Exercice 2.9. 1. Pour un vecteur gaussien X de moyenne b et de matrice de variance V , montrer que

$$\Phi_X(u) = e^{\langle u, b \rangle - \frac{1}{2} \langle u, V u \rangle}.$$

(Il s'agit d'une formule utile à retenir.)

2. Si \mathbb{P}^X est symétrique par rapport à l'origine, i.e. $\mathbb{P}^X = \mathbb{P}^{-X}$, montrer que Φ_X est à valeurs réelles.

3. Pour une v.a. réelle, supposons que $\mathbb{E}[|X|^p] < \infty$ pour un certain entier $p \geq 1$. Montrer que Φ_X est p fois dérivable et

$$\Phi_X^{(k)}(0) = i^k \mathbb{E}[X^k] \quad \text{pour } k = 1, \dots, p.$$

Le but de ce paragraphe est de montrer que la fonction caractéristique permet, comme son nom l'indique, de caractériser la loi \mathbb{P}^X de X . Ceci donne un moyen alternatif d'aborder les vecteurs aléatoires pour lesquels la fonction de répartition est difficile à manipuler. Cependant, l'intérêt de cette notion ne se limite pas à la dimension $d > 1$. Par exemple, la manipulation de sommes de v.a. est souvent plus simple par le biais des fonctions caractéristiques.

Dans ces notes, nous nous limitons à montrer ce résultat dans le cas unidimensionnel.

Théorème 2.10. *Pour une v.a. réelle, la fonction Φ_X caractérise la loi \mathbb{P}^X . Plus précisément*

$$\frac{1}{2} \mathbb{P}^X(\{a\}) + \frac{1}{2} \mathbb{P}^X(\{b\}) + \mathbb{P}^X(]a, b[) = \frac{1}{2\pi} \lim_{T \rightarrow \infty} \int_{-T}^T \Phi_X(u) \frac{e^{-iua} - e^{-iub}}{iu} du$$

pour tous $a < b$. De plus, si Φ_X est intégrable, \mathbb{P}^X est absolument continue par rapport à la mesure de Lebesgue, de densité

$$f_X(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-iux} \Phi_X(u) du, \quad x \in \mathbb{R}.$$

Démonstration. Pour $a < b$, on vérifie sans peine que la condition d'application du théorème de Fubini est satisfaite, et on calcule que :

$$\begin{aligned} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-iua} - e^{-iub}}{iu} \Phi_X(u) du &= \frac{1}{2\pi} \int_{-T}^T \frac{e^{-iua} - e^{-iub}}{iu} \left(\int_{\mathbb{R}} e^{iuv} d\mathbb{P}^X(v) \right) du \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} \left(\int_{-T}^T \frac{e^{iu(v-a)} - e^{iu(v-b)}}{iu} du \right) d\mathbb{P}^X(v). \end{aligned}$$

Puis, on calcule directement que

$$\frac{1}{2\pi} \int_{-T}^T \frac{e^{iu(v-a)} - e^{iu(v-b)}}{iu} du = \frac{S((v-a)T) - S((v-b)T)}{\pi T}, \quad (2.5)$$

où $S(x) := \operatorname{sgn}(x) \int_0^{|x|} \frac{\sin t}{t} dt$, $t > 0$, et $\operatorname{sgn}(x) = \mathbf{1}_{\{x>0\}} - \mathbf{1}_{\{x<0\}}$. On peut vérifier que $\lim_{x \rightarrow \infty} S(x) = \frac{\pi}{2}$, que l'expression (2.5) est uniformément bornée en v et T , et qu'elle converge vers

$$0 \text{ si } x \notin [a, b], \quad \frac{1}{2} \text{ si } x \in \{a, b\}, \quad \text{et } 1 \text{ si } x \notin]a, b[.$$

On obtient alors le résultat annoncé par le théorème de convergence dominée.

Supposons de plus que $\int_{\mathbb{R}} |\phi_X(u)| du < \infty$. Alors, en prenant la limite $T \rightarrow \infty$ dans l'expression du théorème, et en supposant dans un premier temps que \mathbb{P}^X est absolument continue par rapport à la mesure de Lebesgue, on obtient :

$$\mathbb{P}^X([a, b]) = F_X(b) - F_X(a) = \frac{1}{2\pi} \int_{\mathbb{R}} \frac{e^{-iua} - e^{-iub}}{iu} \Phi_X(u) du$$

par le théorème de convergence dominée. On réalise alors que le membre de droite est continu en a et b et, par suite, \mathbb{P}^X n'a pas d'atomes et l'expression ci-dessus est vraie. Pour trouver l'expression de la densité f_X , il suffit de prendre la limite $b \rightarrow a$ après normalisation par $b - a$, et d'utiliser le théorème de convergence dominée. \diamond

2.3 Espaces \mathcal{L}^p et convergences fonctionnelles des variables aléatoires

2.3.1 Géométrie de l'espace \mathcal{L}^2

On désigne par $\mathcal{L}^2 = \mathcal{L}^2(\mathcal{A}, \mathbb{P})$ l'espace vectoriel des variables aléatoires réelles de carré \mathbb{P} -intégrable. Une application simple de l'inégalité de Jensen montre que $\mathcal{L}^2 \subset \mathcal{L}^1 = \mathcal{L}^1(\mathcal{A}, \mathbb{P})$.

L'application $(X, Y) \mapsto \mathbb{E}[XY]$ définit un produit scalaire sur \mathcal{L}^2 si on identifie les v.a. égales p.s. On note la norme correspondante par $\|X\|_2 := \mathbb{E}[X^2]^{1/2}$. En particulier, ceci garantit l'inégalité de Schwarz (valable pour les mesures, voir exercice 1.37) :

$$|\mathbb{E}[XY]| \leq \mathbb{E}[|XY|] \leq \|X\|_2 \|Y\|_2 \quad \text{pour tous } X, Y \in \mathcal{L}^2,$$

ainsi que l'inégalité triangulaire

$$\|X + Y\|_2 \leq \|X\|_2 + \|Y\|_2 \quad \text{pour tous } X, Y \in \mathcal{L}^2.$$

En probabilité, l'espérance quantifie la moyenne de la v.a. Il est aussi important, au moins intuitivement, d'avoir une mesure de la dispersion de la loi. ceci est quantifié par les notions de variance et de covariance :

$$\mathbb{V}[X] := \mathbb{E}[(X - \mathbb{E}X)^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

et

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

Si X est à valeurs dans \mathbb{R}^d , ces notions sont étendues de manière naturelle. Dans ce cadre $\mathbb{V}[X]$ est une matrice symétrique positive de taille d .

Enfin, la corrélation entre les v.a. X et Y est définie par

$$\text{Cor}[X, Y] := \frac{\text{Cov}[X, Y]}{\|X\|_2 \|Y\|_2} = \frac{\langle X, Y \rangle_2}{\|X\|_2 \|Y\|_2},$$

i.e. le cosinus de l'angle formé par les vecteurs X et Y . L'inégalité de Schwarz garantit que la corrélation est un réel dans l'intervalle $[-1, 1]$. Le théorème de Pythagore s'écrit

$$\mathbb{E}[(X + Y)^2] = \mathbb{E}[X^2] + \mathbb{E}[Y^2] \quad \text{dès que} \quad \mathbb{E}[XY] = 0,$$

ou, en termes de variances,

$$\mathbb{V}[X + Y] = \mathbb{V}[X] + \mathbb{V}[Y] \quad \text{dès que} \quad \text{Cov}[X, Y] = 0.$$

Attention, la variance n'est pas un opérateur linéaire, la formule ci-dessus est uniquement valable si $\text{Cov}[X, Y] = 0$. Enfin, la loi du parallélogramme s'écrit

$$\|X + Y\|_2^2 + \|X - Y\|_2^2 = 2\|X\|_2^2 + 2\|Y\|_2^2 \quad \text{pour tous} \quad X, Y \in \mathcal{L}^2.$$

2.3.2 Espaces \mathcal{L}^p et \mathbb{L}^p

Pour $p \in [1, \infty[$, on note par $\mathcal{L}^p := \mathcal{L}^p(\mathcal{A}, \mathbb{P})$ l'espace vectoriel des variables aléatoires X telles que $\mathbb{E}[|X|^p] < \infty$. On note $\|X\|_p := (\mathbb{E}[|X|^p])^{1/p}$. Remarquons que $\|X\|_p = 0$ implique seulement que $X = 0$ p.s. donc $\|\cdot\|_p$ ne définit pas une norme sur \mathcal{L}^p .

Définition 2.11. *L'espace \mathbb{L}^p est l'ensemble des classes d'équivalence de \mathcal{L}^p pour la relation définie par l'égalité p.s.*

Ainsi l'espace \mathbb{L}^p identifie les variables aléatoires égales p.s. et $\|\cdot\|$ définit bien une norme sur \mathbb{L}^p .

Nous continuerons tout de même à travailler sur l'espace \mathcal{L}^p et nous ne passerons à \mathbb{L}^p que si nécessaire.

Par une application directe de l'inégalité de Jensen, on voit que

$$\|X\|_p \leq \|X\|_r \quad \text{si} \quad 1 \leq p \leq r < \infty \quad \text{pour tout} \quad X \in \mathcal{L}^r, \quad (2.6)$$

en particulier, $X \in \mathcal{L}^p$. Ceci montre que $\mathcal{L}^p \supset \mathcal{L}^r$ dès que $1 \leq p \leq r < \infty$.

Nous allons montrer que l'espace \mathcal{L}^p peut être transformé (toujours par quotientnement par la classe des v.a. nulles p.s.) en un espace de Banach.

Théorème 2.12. *Pour $p \geq 1$, l'espace \mathbb{L}^p est un espace de Banach, et \mathbb{L}^2 est espace de Hilbert. Plus précisément, soit $(X_n)_n$ une suite de Cauchy dans \mathcal{L}^p , i.e. $\|X_n - X_m\|_p \rightarrow 0$ pour $n, m \rightarrow \infty$. Alors il existe une v.a. $X \in \mathcal{L}^p$ telle que $\|X_n - X\|_p \rightarrow 0$.*

Démonstration. Si $(X_n)_n$ est une suite de Cauchy, on peut trouver une suite croissante $(k_n)_n \subset \mathbb{N}$, $k_n \uparrow \infty$, telle que

$$\|X_{m_1} - X_{m_2}\|_p \leq 2^{-n} \quad \text{pour tous} \quad m_1, m_2 \geq k_n. \quad (2.7)$$

Alors, on déduit de l'inégalité (2.6) que

$$\mathbb{E}[|X_{k_{n+1}} - X_{k_n}|] \leq \|X_{k_{n+1}} - X_{k_n}\|_p \leq 2^{-n},$$

et que $\mathbb{E}[\sum_n |X_{k_{n+1}} - X_{k_n}|] < \infty$. Alors la série $\sum_n (X_{k_{n+1}} - X_{k_n})$ est absolument convergente p.s. Comme il s'agit d'une série télescopique, ceci montre que :

$$\lim_n X_{k_n} = X \text{ p.s.} \quad \text{où} \quad X := \limsup_n X_{k_n}.$$

Revenant à (2.7), on voit que pour $m_1 \geq k_n$ et $m \geq n$, on a $\mathbb{E}[|X_{m_1} - X_{k_m}|^p] = \|X_n - X_{k_m}\|_p^p \leq 2^{-np}$. Pour $m \rightarrow \infty$, on déduit du lemme de Fatou que $\mathbb{E}[|X_{m_1} - X|^p] \leq 2^{-np}$. \diamond

2.3.3 Espaces \mathcal{L}^0 et \mathbb{L}^0

On note $\mathcal{L}^0 := \mathcal{L}^0(\mathcal{A})$ l'espace vectoriel des variables aléatoires \mathcal{A} -mesurables sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, et on introduit l'espace quotient \mathbb{L}^0 constitué des classes d'équivalence de \mathcal{L}^0 pour la relation définie par l'égalité p.s.

Définition 2.13. (*Convergence en probabilité*) Soient $(X_n)_n$ et X des v.a. dans \mathcal{L}^0 . On dit que $(X_n)_n$ converge en probabilité vers X si

$$\lim_{n \rightarrow \infty} \mathbb{P}[|X_n - X| \geq \varepsilon] = 0 \quad \text{pour tout} \quad \varepsilon > 0.$$

Cette notion de convergence est plus faible que la convergence p.s. et que la convergence dans \mathbb{L}^p dans le sens suivant.

Lemme 2.14. (i) *La convergence p.s. implique la convergence en probabilité.*
(ii) *Soit $p \geq 1$. La convergence en norme dans \mathbb{L}^p implique la convergence en probabilité.*

Démonstration. (i) découle d'une application immédiate du théorème de convergence dominée. Pour (ii), il suffit d'utiliser l'inégalité de Markov de l'exercice 1.36. \diamond

Le but de ce paragraphe est de montrer que la convergence en probabilité est métrisable et qu'elle confère à \mathbb{L}^0 une structure d'espace métrique complet. Pour cela, on introduit la fonction $D : \mathcal{L}^0 \times \mathcal{L}^0 \rightarrow \mathbb{R}_+$ définie par :

$$D(X, Y) = \mathbb{E}[|X - Y| \wedge 1] \quad \text{pour tous} \quad X, Y \in \mathcal{L}^0. \quad (2.8)$$

On vérifie immédiatement que D est une distance sur \mathbb{L}^0 , mais ne l'est pas sur \mathcal{L}^0 , pour les mêmes raisons que celles du paragraphe précédent.

Lemme 2.15. *La convergence en probabilité est équivalente à la convergence au sens de la distance D .*

Démonstration. Pour $X \in \mathcal{L}^0$, on obtient par l'inégalité de Markov de l'exercice 1.36 :

$$\mathbb{P}[|X| \geq \varepsilon] = \mathbb{P}[|X| \wedge 1 \geq \varepsilon] \leq \frac{\mathbb{E}[|X| \wedge 1]}{\varepsilon},$$

qui permet de déduire que la convergence au sens de D implique la convergence en probabilité. Pour l'implication inverse, on estime :

$$\mathbb{E}[|X| \wedge 1] = \mathbb{E}[(|X| \wedge 1)\mathbf{1}_{|X| \geq \varepsilon}] + \mathbb{E}[(|X| \wedge 1)\mathbf{1}_{|X| < \varepsilon}] \leq \mathbb{P}[|X| \geq \varepsilon] + \varepsilon,$$

d'où on tire que la convergence en probabilité implique la convergence au sens de D . \diamond

Théorème 2.16. (\mathbb{L}^0, D) est un espace métrique complet.

Démonstration. Soit $(X_n)_n$ une suite de Cauchy pour D . Alors c'est une suite de Cauchy pour la convergence en probabilité d'après le lemme 2.15, et on peut construire une suite $(n_k)_k \uparrow \infty$ telle que

$$\mathbb{P}[|X_{n_{k+1}} - X_{n_k}| \geq 2^{-k}] \leq 2^{-k} \quad \text{pour tout } k \geq 1,$$

et par suite $\sum_k \mathbb{P}[|X_{n_{k+1}} - X_{n_k}| \geq 2^{-k}] < \infty$. Le premier lemme de Borel-Cantelli (lemme 1.14) implique alors que $\mathbb{P}[\cup_n \cap_{m \geq n} \{|X_{n_{k+1}} - X_{n_k}| \geq 2^{-k}\}] = 1$ et, par suite, pour presque tout $\omega \in \Omega$, $(X_{n_k}(\omega))_n$ est une suite de Cauchy dans \mathbb{R} . Ainsi, la v.a. $X := \limsup_k X_{n_k}$ vérifie $X_{n_k} \rightarrow X$ p.s. donc en probabilité, et on termine comme dans la démonstration du théorème 2.12. \diamond

2.3.4 Lien entre les convergences \mathbb{L}^p , en proba et p.s.

Nous avons vu que la convergence en probabilité est plus faible que la convergence p.s. Le résultat suivant établit un lien précis entre ces deux notions de convergence.

Théorème 2.17. Soient $\{X_n, n \geq 1\}$ et X des v.a. dans \mathcal{L}^0 .

- (i) $X_n \rightarrow X$ p.s. *ssi* $\sup_{m \geq n} |X_m - X| \rightarrow 0$ en probabilité.
- (ii) $X_n \rightarrow X$ en probabilité *ssi* de toute suite croissante d'entiers $(n_k)_k$, on peut extraire une sous-suite $(n_{k_j})_j$ telle que $X_{n_{k_j}} \rightarrow X$ p.s.

La démonstration est reportée à la fin de ce paragraphe. On continue par une conséquence immédiate du théorème 2.17 (ii).

Corollaire 2.18. (Slutsky) Soient $(X_n)_n$ une suite à valeur dans \mathbb{R}^d , et $\phi : \mathbb{R}^d \rightarrow \mathbb{R}^p$ une fonction continue. Si $X_n \rightarrow X$ en probabilité, alors $\phi(X_n) \rightarrow \phi(X)$ en probabilité.

En particulier, ce corollaire montre que la convergence en probabilité est stable pour les opérations usuelles d'addition, de multiplication, de min, de max, etc...

Avant de démontrer le théorème 2.17, énonçons le résultat établissant le lien précis entre la convergence en probabilité et la convergence dans \mathbb{L}^1 .

Définition 2.19. Une famille \mathcal{C} de v.a. est dite uniformément intégrable, et on note U.I. si

$$\lim_{c \rightarrow \infty} \sup_{X \in \mathcal{C}} \mathbb{E}[|X| \mathbf{1}_{\{|X| \geq c\}}] = 0.$$

Théorème 2.20. Soient $\{X_n, n \geq 1\}$ et X des v.a. dans \mathcal{L}^1 . Alors $X_n \rightarrow X$ dans \mathbb{L}^1 si et seulement si

- (a) $X_n \rightarrow X$ en probabilité,
- (b) $(X_n)_n$ est U.I.

La démonstration de ce résultat est aussi reportée à la fin de ce paragraphe. L'exercice suivant regroupe les résultats essentiels qui concernent l'uniforme intégrabilité.

Exercice 2.21. Soit $(X_n)_n$ une suite de v.a. à valeurs réelles.

1. Supposons que $(X_n)_n$ est U.I.
 - (a) Montrer que $(X_n)_n$ est bornée dans \mathcal{L}^1 , i.e. $\sup_n \mathbb{E}[|X_n|] < \infty$.
 - (b) Sur l'espace probabilisé $([0, 1], \mathcal{B}_{[0,1]}, \lambda)$, λ étant la mesure de Lebesgue, on considère la suite $Y_n := n\mathbf{1}_{[0, 1/n]}$. Montrer que $(Y_n)_n$ est bornée dans \mathcal{L}^1 , mais n'est pas U.I.
2. Supposons que $\mathbb{E}[\sup_n |X_n|] < \infty$. Montrer que (X_n) est U.I. (Indication : utiliser la croissance de la fonction $x \mapsto x\mathbf{1}_{\{x \geq c\}}$ \mathbb{R}_+).
3. Supposons qu'il existe $p > 1$ tel que $(X_n)_n$ est bornée dans \mathcal{L}^p .
 - (a) Montrer que $\mathbb{E}[|X_n|\mathbf{1}_{\{|X_n| \geq c\}}] \leq \|X_n\|_p \mathbb{P}[|X_n| \geq c]^{1-1/p}$
 - (b) En déduire que (X_n) est U.I.

Nous allons maintenant passer aux démonstrations des théorèmes de ce paragraphe.

Preuve du théorème 2.17 (i) Remarquons que

$$C := \{X_n \rightarrow X\} = \bigcap_k \bigcup_n \bigcap_{m \geq n} \{|X_m - X| \leq k^{-1}\} = \lim_k \downarrow \bigcup_n A_n^k,$$

où $A_n^k := \bigcap_{m \geq n} \{|X_m - X| \leq k^{-1}\}$. La convergence p.s. de X_n vers X s'écrit $\mathbb{P}[C] = 1$, et est équivalente à $\mathbb{P}[\bigcup_n A_n^k] = 1$ pour tout $k \geq 1$. Comme la suite $(A_n^k)_n$ est croissante, ceci est équivalent à $\lim_n \uparrow \mathbb{P}[A_n^k] = 1$ pour tout $k \geq 1$, ce qui exprime exactement la convergence en probabilité de $\sup_{m \geq n} |X_m - X|$ vers 0.

(ii) Supposons d'abord que $X_n \rightarrow X$ en probabilité. Soit (n_k) une suite croissante d'indices, et $\bar{X}_k := X_{n_k}$. On définit

$$k_j := \inf \{i : \mathbb{P}[|\bar{X}_i - X| \geq 2^{-j}] \leq 2^{-j}\}.$$

Alors, $\sum_j \mathbb{P}[|\bar{X}_{k_j} - X| \geq 2^{-j}] < \infty$, et on déduit du premier lemme de Borel-Cantelli, lemme 1.14, que $|\bar{X}_{k_j} - X| < 2^{-j}$ pour j assez grand, p.s. En particulier, ceci montre que $\bar{X}_{k_j} \rightarrow X$, p.s.

Pour la condition suffisante, supposons au contraire que $X_n \not\rightarrow X$ en probabilité. Alors, d'après le lemme 2.15, il existe une sous-suite (n_k) croissante et $\varepsilon > 0$ tels que $D(X_{n_k}, X) \geq \varepsilon$. On arrive à une contradiction en extrayant une sous-suite $(X_{n_{k_j}})_j$ qui converge p.s. vers X , et en évoquant le théorème de convergence dominée pour le passage à la limite. \diamond

Preuve du théorème 2.20 Supposons d'abord que les conditions (a) et (b) sont satisfaites. La fonction $\varphi_c(x) := -c \vee x \wedge c$, $x \in \mathbb{R}$ est lipschitzienne, et vérifie $|\varphi_c(x) - x| \leq |x| \mathbf{1}_{|x| \geq c}$. On déduit alors
- de l'U.I. de $(X_n)_n$ et l'intégrabilité de X que, quand $c \rightarrow \infty$:

$$\mathbb{E}[|\varphi_c(X_n) - X_n|] \longrightarrow 0 \quad \text{pour tout } n \text{ et } \quad \mathbb{E}[|\varphi_c(X) - X|] \longrightarrow 0,$$

- de la convergence en probabilité de X_n vers X , et du corollaire 2.18, que

$$\varphi_c(X_n) \longrightarrow \varphi_c(X) \quad \text{en probabilité.}$$

On peut maintenant conclure que $X_n \longrightarrow X$ dans \mathbb{L}^1 en décomposant

$$\mathbb{E}[|X_n - X|] \leq \mathbb{E}[|X_n - \varphi_c(X_n)|] + \mathbb{E}[|\varphi_c(X_n) - \varphi_c(X)|] + \mathbb{E}[|\varphi_c(X) - X|].$$

Réciproquement, supposons que $X_n \longrightarrow X$ dans \mathbb{L}^1 , alors la convergence en probabilité (a) est une conséquence immédiate de l'inégalité de Markov (exercice 1.7). Pour montrer (b), on se donne $\varepsilon > 0$. La convergence \mathcal{L}^1 de $(X_n)_n$ montre l'existence d'un rang N à partir duquel

$$\mathbb{E}|X_n - X| < \varepsilon \quad \text{pour tout } n > N. \quad (2.9)$$

Par ailleurs, d'après le lemme 1.26, il existe $\delta > 0$ tel que pour tout $A \in \mathcal{A}$:

$$\sup_{n \leq N} \mathbb{E}[|X_n| \mathbf{1}_A] < \varepsilon \text{ et } \mathbb{E}[|X| \mathbf{1}_A] < \varepsilon \quad \text{dès que } \mathbb{P}[A] < \delta. \quad (2.10)$$

Nous allons utiliser cette inégalité avec les ensembles $A_n := \{|X_n| > c\}$ qui vérifient bien

$$\sup_n \mathbb{P}[A_n] \leq c^{-1} \sup_n \mathbb{E}[|X_n|] < \delta \quad \text{pour } c \text{ assez grand,} \quad (2.11)$$

où nous avons utilisé l'inégalité de Markov, exercice 1.7, ainsi que la bornitude dans \mathcal{L}^1 de la suite $(X_n)_n$ du fait de sa convergence dans \mathbb{L}^1 . Ainsi, on déduit de (2.10) et (2.11) que

$$\begin{aligned} \sup_n \mathbb{E}[|X_n| \mathbf{1}_{\{|X_n| > c\}}] &= \max \left\{ \sup_{n \leq N} \mathbb{E}[|X_n| \mathbf{1}_{\{|X_n| > c\}}], \sup_{n > N} \mathbb{E}[|X_n| \mathbf{1}_{\{|X_n| > c\}}] \right\} \\ &\leq \max \left\{ \varepsilon, \sup_{n > N} \mathbb{E}[|X_n| \mathbf{1}_{\{|X_n| > c\}}] \right\} \\ &\leq \max \left\{ \varepsilon, \sup_{n > N} \mathbb{E}[|X| \mathbf{1}_{\{|X_n| > c\}}] + \mathbb{E}[|X - X_n| \mathbf{1}_{\{|X_n| > c\}}] \right\} \\ &\leq \max \left\{ \varepsilon, \sup_{n > N} \mathbb{E}[|X| \mathbf{1}_{A_n}] + \mathbb{E}[|X - X_n|] \right\} < 2\varepsilon, \end{aligned}$$

où la dernière inégalité est due à (2.9), (2.10) et (2.11). \diamond

2.4 Convergence en loi

Dans ce paragraphe, nous nous intéressons à la convergence des lois. Remarquons immédiatement qu'il ne peut s'agir que d'un sens de convergence plus faible que les convergences fonctionnelles étudiées dans le paragraphe précédent puis qu'on ne pourra en général rien dire sur les variables aléatoires sous-jacentes. A titre d'exemple, si X est une v.a. de loi gaussienne centrée, alors $-X$ a la même loi que X (on écrit $X \stackrel{L}{=} -X$). Pire encore, on peut avoir deux v.a. réelles X et Y sur des espaces probabilisés différents $(\Omega_2, \mathcal{A}_1, \mathbb{P}_1)$ et $(\Omega_2, \mathcal{A}_2, \mathbb{P}_2)$ qui ont la même distribution.

Dans ce paragraphe, on désignera par $C_b(\mathbb{R})$ l'ensemble des fonctions continues bornées sur \mathbb{R} , et $\Sigma(\mathbb{R})$ l'ensemble des mesures de probabilité sur \mathbb{R} .

2.4.1 Définitions

Soient μ et $\mu_n \in \Sigma(\mathbb{R})$, $n \in \mathbb{N}$. On dit que $(\mu_n)_n$ converge faiblement, ou étroitement, vers μ si

$$\mu_n(f) \longrightarrow \mu(f) \quad \text{pour toute fonction } f \in C_b(\mathbb{R}).$$

Soient X et $X_n, n \in \mathbb{N}$ des v.a. dans $\mathcal{L}^0(\mathcal{A})$. On dit que $(X_n)_n$ converge en loi vers X si $(\mathbb{P}^{X_n})_n$ converge faiblement vers \mathbb{P}^X , i.e.

$$\mathbb{E}[f(X_n)] \longrightarrow \mathbb{E}[f(X)] \quad \text{pour tout } f \in C_b(\mathbb{R}).$$

Dans la dernière définition, il n'est pas nécessaire que les v.a. $X, X_n, n \in \mathbb{N}$ soient définies sur le même espace probabilisé. Montrons maintenant que les convergences introduites dans les chapitres précédents sont plus fortes que la convergence en loi.

Proposition 2.22. *La convergence en probabilité implique la convergence en loi.*

Démonstration. Supposons que $X_n \longrightarrow X$ en probabilité, et soient $g \in C_b(\mathbb{R})$. La suite réelle $u_n := \mathbb{E}[g(X_n)], n \in \mathbb{N}$, est bornée. Pour montrer la convergence en loi, il suffit de vérifier que toute sous-suite convergente $(u_{n_k})_k$ converge vers $\mathbb{E}[g(X)]$. Pour cela, il suffit d'utiliser le lemme 2.17 et le théorème de convergence dominée. \diamond

Comme la convergence en probabilité est plus faible que la convergence \mathbb{L}^1 et la convergence p.s. on a le schémas suivant expliquant les liens entre les différents types de convergence rencontrés :

$$\begin{array}{ccccc} & & \text{====} & & \\ & & \text{====} & & \\ \mathbb{L}^p & \implies & \mathbb{L}^1 & & \\ & & \Downarrow & & \\ \text{p.s.} & \implies & \mathbb{P} & \implies & \text{Loi} \\ & & \text{====} & & \\ & & \text{====} & & \end{array}$$

2.4.2 Caractérisation de la convergence en loi par les fonctions de répartition

Toute loi $\mu \in \Sigma(\mathbb{R})$ est caractérisée par la fonction de répartition correspondante $F(x) := \mu(]-\infty, x])$. Ainsi, si $F, F_n, n \in \mathbb{N}$, sont des fonctions de répartition sur \mathbb{R} , on dira que $(F_n)_n$ converge en loi vers F si la convergence en loi a lieu pour les mesures correspondantes.

Dans ce paragraphe, nous allons exprimer la définition de la convergence faible de manière équivalente en terme des fonctions de répartition.

Remarque 2.23. Voici un exemple qui montre que les points de discontinuité de F , s'il y en a, jouent un rôle particulier. Sur $([0, 1], \mathcal{B}_{[0,1]}, \lambda)$, soit $\mu_n := \delta_{1/n}$ la masse de Dirac au point $1/n$ (c'est la loi de la v.a. déterministe $X_n = 1/n$). Alors (μ_n) converge en loi vers δ_0 , la masse de Dirac au point 0. Mais pour tout $n \geq 1$, $F_n(0) = 0 \not\rightarrow F_{\delta_0}(0)$.

Théorème 2.24. Soient $F, F_n, n \in \mathbb{N}$, des fonctions de répartition sur \mathbb{R} . Alors, (F_n) converge en loi vers F si et seulement si

$$\text{Pour tout } x \in \mathbb{R}, \quad F(x-) = F(x) \implies F_n(x) \longrightarrow F(x).$$

Démonstration. 1- Pour $\eta > 0$ et $x \in \mathbb{R}$, on définit les fonctions

$$g_1(y) := \mathbf{1}_{]-\infty, x+\eta]}(y) - \frac{y-x}{\eta} \mathbf{1}_{]x, x+\eta]} \quad \text{et} \quad g_2(y) := g_1(y+\eta), \quad y \in \mathbb{R},$$

et on observe que $\mathbf{1}_{]-\infty, x]} \leq g_1 \leq \mathbf{1}_{]-\infty, x+\eta]}$, $\mathbf{1}_{]-\infty, x-\eta]} \leq g_2 \leq \mathbf{1}_{]-\infty, x]}$ et, par suite,

$$F_n(x) \leq \mu_n(g_1), \quad \mu(g_1) \leq F(x+\eta), \quad \text{et} \quad F_n(x) \geq \mu_n(g_2), \quad \mu(g_2) \geq F(x-\eta)$$

Comme $g_1, g_2 \in C_b(\mathbb{R})$, on déduit de la convergence faible de $(F_n)_n$ vers F que $\mu_n(g_1) \longrightarrow \mu(g_1)$, $\mu_n(g_2) \longrightarrow \mu(g_2)$, et

$$F(x-\eta) \leq \liminf_n F_n(x) \leq \limsup_n F_n(x) \leq F(x+\eta) \quad \text{pour tout } \eta > 0,$$

qui implique bien que $F_n(x) \longrightarrow F(x)$ si x est un point de continuité de F .

2- Pour la condition suffisante, on définit comme dans la remarque 2.4 les v.a. $\bar{X}, \underline{X}, \bar{X}_n, \underline{X}_n$ qui ont pour fonction de répartition F et F_n . Par définition de \bar{X} , pour tout $x > \bar{X}(\omega)$ on a $F(x) > \omega$. Si x est un point de continuité de F , ceci implique que $F_n(x) > \omega$ pour n assez grand et, par suite, $x \geq \bar{X}_n(\omega)$. Comme F est croissante, l'ensemble de ses points de discontinuité est au plus dénombrable. On peut donc faire tendre x vers $\bar{X}(\omega)$ le long de points de continuité de F , et on tire l'inégalité $\bar{X}(\omega) \geq \bar{X}_n(\omega)$ pour n assez grand. On obtient le résultat symétrique en raisonnant sur \underline{X} et \underline{X}_n . D'où :

$$\bar{X}(\omega) \geq \bar{X}_n(\omega) \geq \underline{X}_n(\omega) \geq \underline{X}(\omega) \quad \text{pour } n \text{ assez grand.}$$

Comme $\mathbb{P}[\bar{X} = \underline{X}] = 1$, ceci montre que $\bar{X}_n \longrightarrow \bar{X}$ p.s. et donc en loi. \diamond

2.4.3 Convergence des fonctions de répartition

L'importance de la convergence en loi provient de la facilité d'obtenir des théorèmes limites. En effet, les suites de mesures convergent en loi "à peu de frais", le long d'une sous-suite, vers une limite qui n'est cependant pas nécessairement une loi. Si la limite n'est pas une loi, on dit qu'il y a perte de masse.

Avant d'énoncer un résultat précis, expliquons les idées qu'il y a derrière ces résultats profonds. Les fonctions de répartition ont une structure très spéciale : on regardant le graphe d'une fonction de répartition dans les coordonnées $(x + y, -x + y)$ (obtenu par rotation des coordonnées initiale de 45°), celui-ci correspond à une fonction dont la valeur absolue de la pente est majorée par 1 : les pentes -1 et 1 correspondent respectivement aux "plats" et aux sauts de la fonction de répartition. Ainsi dans ce système de coordonnées le graphe perd la propriété de croissance, mais devient 1-Lipschitzien. Par conséquent, pour toute suite de fonctions de répartition, le théorème d'Ascoli nous garantit alors l'existence d'une sous-suite convergente. La démonstration ci-dessous utilise un argument encore plus élémentaire.

Lemme 2.25. *Soit $(F_n)_n$ une suite de fonctions de répartition sur \mathbb{R} . Alors, il existe une fonction croissante continue à droite $F : \mathbb{R} \rightarrow [0, 1]$, et une sous-suite (n_k) telles que $F_{n_k} \rightarrow F$ simplement en tout point de continuité de F .*

Démonstration. On dénombre les éléments de l'ensemble des rationnels $\mathbb{Q} = \{q_i, i \in \mathbb{N}\}$. La suite $(F_n(q_1))_n$ est bornée, donc converge le long d'une sous-suite $F_{n_k^1}(q_1) \rightarrow G(q_1)$ quand $k \rightarrow \infty$. De même la suite $(F_{n_k^1}(q_2))_n$ est bornée, donc converge le long d'une sous-suite $F_{n_k^2}(q_2) \rightarrow G(q_2)$ quand $k \rightarrow \infty$, etc... Alors, en posant $k_j := n_{k_j}^j$, on obtient

$$F_{k_j}(q) \rightarrow G(q) \quad \text{pour tout } q \in \mathbb{Q}.$$

Il est clair que G est croissante sur \mathbb{Q} et à valeurs dans $[0, 1]$. On définit alors la fonction F par

$$F(x) := \lim_{\mathbb{Q} \ni q \searrow x} G(q) \quad \text{pour tout } x \in \mathbb{R},$$

qui vérifie les propriétés annoncées dans le lemme. \diamond

Afin d'éviter la perte de masse à la limite, on introduit une nouvelle notion.

Définition 2.26. *Une suite $(F_n)_{n \geq 1}$ de fonctions de répartition sur \mathbb{R} est dite tendue si pour tout $\varepsilon > 0$, il existe $K > 0$ tel que*

$$\mu_n([-K, K]) := F_n(K) - F_n(-K) > 1 - \varepsilon \quad \text{pour tout } n \geq 1.$$

Le résultat suivant est une conséquence directe du lemme précédent.

Lemme 2.27. Soit $(F_n)_n$ une suite de fonctions de répartition sur \mathbb{R} .

- (i) Si $F_n \rightarrow F$ en loi, alors $(F_n)_n$ est tendue.
- (ii) Si $(F_n)_n$ est tendue, alors il existe une fonction de répartition F sur \mathbb{R} , et une sous-suite (n_k) telles que $F_{n_k} \rightarrow F$ en loi.

2.4.4 Convergence en loi et fonctions caractéristiques

La fonction caractéristique caractérise une loi de distribution tout aussi bien que la fonction de répartition. Le résultat suivant donne la caractérisation de la convergence en loi en termes de fonctions caractéristiques.

Théorème 2.28. (convergence de Lévy) Soit $(F_n)_{n \geq 1}$ une suite de fonctions de répartition sur \mathbb{R} , et $\phi_n := \Phi_{F_n}$, $n \geq 1$, la suite correspondante de fonctions caractéristiques. Supposons qu'il existe une fonction ϕ sur \mathbb{R} telle que

$$\phi_n \rightarrow \phi \quad \text{simplement sur } \mathbb{R} \text{ et } \phi \text{ continue en } 0.$$

Alors ϕ est une fonction caractéristique correspondant à une fonction de répartition F , et $F_n \rightarrow F$ en loi.

Démonstration. 1- Montrons d'abord que

$$(F_n)_n \text{ est tendue.} \tag{2.12}$$

Soit $\varepsilon > 0$. D'après la continuité de ϕ en 0, il existe $\alpha > 0$ tel que $|1 - \phi| < \varepsilon$ sur $[-\alpha, \alpha]$. Il est clair que $2 - \phi_n(u) - \phi_n(-u) \in \mathbb{R}_+$ et que cette propriété est héritée par ϕ à la limite. Alors $0 \leq \int_0^\alpha [2 - \phi(u) - \phi(-u)] du \leq 2\varepsilon\alpha$, et on déduit de la convergence de ϕ_n vers ϕ et du théorème de convergence dominée qu'à partir d'un certain rang $n \geq N$:

$$\begin{aligned} 4\varepsilon &\geq \frac{1}{\alpha} \int_0^\alpha [2 - \phi_n(u) - \phi_n(-u)] du \\ &= \frac{1}{\alpha} \int_{-\alpha}^\alpha \int_{\mathbb{R}} (1 - e^{iut}) dF_n(t) du \\ &= \frac{1}{\alpha} \int_{\mathbb{R}} \int_{-\alpha}^\alpha (1 - e^{iut}) dudF_n(t) = 2 \int_{\mathbb{R}} \left(1 - \frac{\sin(\alpha t)}{\alpha t}\right) dF_n(t) \end{aligned}$$

par le théorème de Fubini. Comme $\sin x \leq x$ pour tout $x \in \mathbb{R}$, on déduit alors que pour tout $\varepsilon > 0$, il existe $\alpha > 0$:

$$4\varepsilon \geq 2 \int_{|t| \geq 2\alpha^{-1}} \left(1 - \frac{\sin(\alpha t)}{\alpha t}\right) dF_n(t) \geq \int_{|t| \geq 2\alpha^{-1}} dF_n(t),$$

prouvant (2.12).

2- Comme $(F_n)_n$ est tendue, on déduit du lemme 2.27 que $F_{n_k} \rightarrow F$ en loi le long d'une sous-suite $(n_k)_k$, où F est une fonction de répartition. D'après la définition de la convergence en loi, on a aussi convergence des fonctions caractéristiques correspondantes $\phi_{n_k} \rightarrow \Phi_F$. Alors $\phi = \Phi_F$.

3- Il reste à monter que $F_n \rightarrow F$ en loi. Supposons au contraire qu'il existe un point de continuité x tel que $F_n(x) \not\rightarrow F(x)$. Alors, il existe une sous-suite $(n_k)_k$ telle que

$$F(x-) = F(x) \text{ et } |F_{n_k}(x) - F(x)| \geq \varepsilon \text{ pour tout } k. \quad (2.13)$$

Comme $(F_{n_k})_k$ est tendue d'après l'étape 1, on a $F_{n_{k_j}} \rightarrow \tilde{F}$ en loi le long d'une sous-suite $(n_{k_j})_j$, où \tilde{F} est une fonction de répartition. Raisonnant comme dans l'étape précédente, on voit que $\phi_{n_{k_j}} \rightarrow \Phi_{\tilde{F}} = \phi = \Phi_F$, et on déduit que $\tilde{F} = F$ par injectivité. Ainsi $F_{n_{k_j}} \rightarrow F$ en loi, contredisant (2.13). \diamond

2.5 Indépendance

2.5.1 σ -algèbres indépendantes

Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, et $(\mathcal{A}_n)_n \subset \mathcal{A}$ une suite de σ -algèbres. On dit que les $(\mathcal{A}_n)_n$ sont indépendantes (sous \mathbb{P}) si pour tous entiers $n \geq 1$ et $1 \leq i_1 < \dots < i_n$:

$$\mathbb{P}[\cap_{k=1}^n A_{i_k}] = \prod_{k=1}^n \mathbb{P}[A_{i_k}] \text{ pour tous } A_{i_k} \in \mathcal{A}_{i_k}, 1 \leq k \leq n. \quad (2.14)$$

Remarquons que le théorème de convergence monotone permet d'affirmer que (2.14) est aussi valide pour $n = \infty$, i.e.

$$\mathbb{P}[\cap_{k \geq 1} A_{i_k}] = \prod_{k \geq 1} \mathbb{P}[A_{i_k}] \text{ pour tous } A_{i_k} \in \mathcal{A}_{i_k}, k \geq 1. \quad (2.15)$$

A partir de cette définition générale pour les σ -algèbres, on étend l'indépendance à des sous-familles arbitraires de \mathcal{A} et aux v.a.

Définition 2.29. *On dit que les événements $(A_n)_n \subset \mathcal{A}$ sont indépendants si $(\sigma(A_n))_n$ sont indépendantes.*

Dans la définition précédente, il est inutile de vérifier (2.14) pour tous les choix possibles dans les σ -algèbres $\sigma(A_n) = \{\Omega, \emptyset, A_n, A_n^c\}$. En effet, il suffit de vérifier que

$$\mathbb{P}[\cap_{k=1}^n A_{i_k}] = \prod_{k=1}^n \mathbb{P}[A_{i_k}] \text{ pour } n \geq 1 \text{ et } 1 \leq i_1 < \dots < i_n.$$

Voici une formulation plus générale de ce résultat.

Lemme 2.30. *Soit $(\mathcal{I}_n)_n \subset \mathcal{A}$ une suite de π -systèmes. Alors les sous- σ -algèbres $(\sigma(\mathcal{I}_n))_n$ sont indépendantes si et seulement si (2.14) est vraie pour les événements des \mathcal{I}_n , i.e. si pour tous entiers $n \geq 1$ et $1 \leq i_1 < \dots < i_n$, on a :*

$$\mathbb{P}[\cap_{k=1}^n I_{i_k}] = \prod_{k=1}^n \mathbb{P}[I_{i_k}] \text{ pour tous } I_{i_k} \in \mathcal{I}_{i_k}, 1 \leq k \leq n.$$

Démonstration. il suffit de vérifier le résultat pour deux π -systèmes $\mathcal{I}_1, \mathcal{I}_2$. Fixons un événement $I_1 \in \mathcal{I}_1$, et introduisons les applications de $\sigma(\mathcal{I}_2)$ dans $[0, \mathbb{P}[I_1]]$ définies par $\mu(I_2) := \mathbb{P}(I_1 \cap I_2)$ et $\nu(I_2) := \mathbb{P}(I_1)\mathbb{P}(I_2)$. Il est clair que μ et ν sont des mesures sur $\sigma(\mathcal{I}_2)$ égales sur le π -système \mathcal{I}_2 . Alors elles sont égales sur $\sigma(\mathcal{I}_2)$ d'après la proposition 1.5. Il suffit maintenant d'évoquer le rôle arbitraire de $I_1 \in \mathcal{I}_1$, et de répéter exactement le même argument en inversant \mathcal{I}_1 et \mathcal{I}_2 . \diamond

2.5.2 Variables aléatoires indépendantes

Définition 2.31. On dit que des v.a. $(X_n)_n$ sont indépendantes si les sous- σ -algèbres correspondantes $(\sigma(X_n))_n$ sont indépendantes.

Une application directe du lemme 2.30 et du théorème de Fubini permet d'établir le critère suivant d'indépendance de v.a.

Proposition 2.32. Les v.a. $(X_n)_n$ sont indépendantes si et seulement si pour tous $n \geq 1$ et $1 \leq i_1 < \dots < i_n$, l'une des assertions suivantes est vérifiée :

- (a) $\mathbb{P}[X_{i_k} \leq x_k \text{ pour } 1 \leq k \leq n] = \prod_{k=1}^n \mathbb{P}[X_{i_k} \leq x_k]$ pour tous $x_1, \dots, x_k \in \mathbb{R}$,
- (b) $\mathbb{E}[\prod_{k=1}^n f_{i_k}(X_{i_k})] = \prod_{k=1}^n \mathbb{E}[f_{i_k}(X_{i_k})]$ pour toutes $f_{i_k} : \mathbb{R} \rightarrow \mathbb{R}$, $1 \leq k \leq n$, mesurables bornées.
- (c) $\mathbb{P}^{(X_{i_1}, \dots, X_{i_n})} = \mathbb{P}^{X_{i_1}} \otimes \dots \otimes \mathbb{P}^{X_{i_n}}$

Exercice 2.33. Montrer la proposition 2.32.

Remarque 2.34. Si X, Y sont deux v.a. réelles indépendantes, la proposition 2.32 implique que la fonction caractéristique du couple se factorise comme :

$$\Phi_{(X,Y)}(u, v) = \Phi_X(u)\Phi_Y(v) \quad \text{pour tous } u, v \in \mathbb{R}.$$

Remarque 2.35. Soient X, Y deux v.a. réelles indépendantes intégrables, alors d'après la proposition 2.32, on a

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y], \quad \text{Cov}[X, Y] = 0 \quad \text{et} \quad \mathbb{V}[X + Y] = \mathbb{V}[X] + \mathbb{V}[Y].$$

Observons que la nullité de la covariance n'implique pas l'indépendance, en général. Dans le cas très particulier où le couple (X, Y) est un vecteur gaussien, on a cependant équivalence entre l'indépendance et la nullité de la covariance.

Si les $(X_n)_n$ sont des v.a. indépendantes à densité, alors on déduit de l'assertion (a) ci-dessus que le vecteur aléatoire $(X_{i_1}, \dots, X_{i_n})$ est absolument continu par rapport à la mesure de Lebesgue sur \mathbb{R}^n de densité

$$f_{(X_{i_1}, \dots, X_{i_n})}(x_1, \dots, x_n) := f_{X_{i_1}}(x_1) \dots f_{X_{i_n}}(x_n). \quad (2.16)$$

Réciproquement si le vecteur aléatoire $(X_{i_1}, \dots, X_{i_n})$ est absolument continu par rapport à la mesure de Lebesgue sur \mathbb{R}^n de densité séparable, comme dans (2.16) $f_{(X_{i_1}, \dots, X_{i_n})}(x_1, \dots, x_n) = \varphi_1(x_1) \dots \varphi_n(x_n)$ alors, les v.a. X_{i_k} sont indépendantes à densité $f_{X_{i_k}} = \varphi_k$.

2.5.3 Asymptotique des suites d'événements indépendants

Le résultat suivant joue un rôle central en probabilités. Remarquons tout de suite que la partie (i) reprend le résultat établi plus généralement pour les mesures dans le lemme 1.14.

Lemme 2.36. (Borel-Cantelli) Soit $(A_n)_n$ une suite d'événements d'un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$.

- (i) Si $\sum_n \mathbb{P}[A_n] < \infty$, alors $\mathbb{P}[\limsup_n A_n] = 0$,
- (ii) Si $\sum_n \mathbb{P}[A_n] = \infty$ et $(A_n)_n$ sont indépendants, alors $\mathbb{P}[\limsup_n A_n] = 1$.
- (iii) Si $(A_n)_n$ sont indépendants, alors soit $\limsup_n A_n$ est négligeable, soit $(\limsup_n A_n)^c$ est négligeable.

Démonstration. Il reste à montrer (ii). Par définition de l'indépendance et (2.15), on a

$$\mathbb{P}[\cap_{m \geq n} A_m^c] = \prod_{m \geq n} (1 - \mathbb{P}[A_m]) \leq \prod_{m \geq n} e^{-\mathbb{P}[A_m]} = e^{-\sum_{m \geq n} \mathbb{P}[A_m]} = 0.$$

Ainsi, pour tout $n \geq 1$, l'événement $\cap_{m \geq n} A_m^c$ est négligeable, et l'union dénombrable de négligeables $(\limsup_n A_n)^c = \cup_{n \geq 1} \cap_{m \geq n} A_m^c$ est alors négligeable. \diamond

Le résultat suivant est assez frappant, et est une conséquence du Lemme de Borel-Cantelli.

Théorème 2.37. (Loi du zéro-un) Soient $(X_n)_{n \geq 1}$ une suite de v.a. indépendantes, et $\mathcal{T} := \cap_n \sigma(X_m, m > n)$ la σ -algèbre de queue associée. Alors \mathcal{T} est triviale, c'est à dire :

- (i) Pour tout événement $A \in \mathcal{T}$, on a $\mathbb{P}[A]\mathbb{P}[A^c] = 0$,
- (ii) Toute v.a. \mathcal{T} -mesurable est déterministe p.s.

Démonstration. (i) De indépendance des $(X_n)_n$, on déduit que pour tout $n \geq 1$, les σ -algèbres $\mathcal{A}_n := \sigma(X_1, \dots, X_n)$ et $\mathcal{T}_n := \sigma(X_m, m > n)$ sont indépendantes. Comme $\mathcal{T} \subset \mathcal{T}_n$, on voit que \mathcal{A}_n et \mathcal{T} sont indépendantes, et par suite $\cup_n \mathcal{A}_n$ et \mathcal{T} sont indépendantes. En observant que $\cup_n \mathcal{A}_n$ est un π -système, on déduit du lemme 2.30 que $\mathcal{A}_\infty := \sigma(\cup_n \mathcal{A}_n)$ et \mathcal{T} sont indépendants.

Or, $\mathcal{T} \subset \mathcal{A}_\infty$, donc l'indépendance entre \mathcal{T} et \mathcal{A}_∞ implique que \mathcal{T} est indépendant de lui même, et pour tout $A \in \mathcal{T}$, $\mathbb{P}[A] = \mathbb{P}[A \cap A] = \mathbb{P}[A]^2$.

(ii) Soit ξ une v.a. \mathcal{T} -mesurable. Pour tout $x \in \mathbb{R}$, l'événement $\mathbb{P}[\xi \leq x] \in \{0, 1\}$ d'après (i). Soit $c := \sup\{x : \mathbb{P}[\xi \leq x] = 0\}$. Si $c = -\infty$, ou $c = +\infty$, on voit immédiatement que $\xi = c$ (déterministe), p.s. Si $|c| < \infty$, la définition de c implique que $\mathbb{P}[\xi \leq c - \varepsilon] = \mathbb{P}[\xi > c + \varepsilon] = 0$ pour tout $\varepsilon > 0$. Alors $1 \geq \mathbb{E}[\mathbf{1}_{]c-\varepsilon, c+\varepsilon]}(\xi)] = \mathbb{P}[c - \varepsilon < \xi \leq c + \varepsilon] = 1$, i.e. $\mathbf{1}_{]c-\varepsilon, c+\varepsilon]}(\xi) = 1$ p.s. et on termine la preuve en envoyant ε vers 0. \diamond

La σ -algèbre de queue introduite dans le théorème 2.37 contient de nombreux événements intéressants comme par exemple

$$\{\lim_n X_n \text{ existe}\}, \quad \{\lim_n n^{-1} \sum_{i=1}^n X_i \text{ existe}\}, \quad \{\sum_n X_n \text{ converge}\}.$$

Des exemples de v.a. \mathcal{T} -mesurable sont donnés par

$$\limsup \sum_n X_n, \liminf \frac{1}{n} \sum_{i=1}^n X_i, \dots$$

2.5.4 Moyennes de variables indépendantes

Dans ce paragraphe, nous manipulerons des suites de v.a. indépendantes et identiquement distribuées, on écrira plus simplement iid.

On commencera par énoncer la loi des grands nombres pour les suites de v.a. iid intégrables, sans en reporter la démonstration, et nous en déduirons une version pour les suites de v.a. positives, sans hypothèse supplémentaire d'intégrabilité. Rappelons qu'une loi forte des grands nombres a été établie dans le cours de 1^{ère} année pour des v.a. iid seulement de carré intégrable. La démonstration de la version plus forte suivante sera vue dans le chapitre 9 en utilisant l'approche des martingales. Enfin, nous montrerons le théorème central limite.

Théorème 2.38. (*Loi forte des grands nombres*) Soit $(X_n)_n$ une suite de v.a. iid intégrables. Alors

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

Démonstration. Voir théorème 9.3. ◇

Pour l'énoncé suivant, on rappelle que l'espérance d'une v.a. positive est toujours définie dans $[0, \infty]$.

Corollaire 2.39. Soit $(X_n)_n$ une suite de v.a. iid à valeurs dans $[0, \infty]$. Alors

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

Démonstration. Si $\mathbb{E}[X_1] = \infty$, la loi des grands nombres du théorème 2.38 appliquée à la suite $(X_n \wedge K)_n$ pour une certaine constante $K > 0$ donne

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i \geq \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n (X_i \wedge K) \longrightarrow \mathbb{E}[X_1 \wedge K] \text{ pour tout } K \text{ p.s.}$$

On prend la limite lorsque $K \rightarrow \infty$ en évoquant le théorème de convergence monotone, et on obtient le résultat voulu. ◇

Enfin, si les v.a. iid sont de carré intégrable, le théorème central limite donne une information précise sur le taux de convergence de la moyenne empirique vers l'espérance, ou la moyenne théorique.

Théorème 2.40. *Soit $(X_n)_n$ une suite de v.a. iid de carré intégrable. Alors*

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n X_i - \mathbb{E}[X_1] \right) \longrightarrow \mathcal{N}(0, \mathbb{V}[X_1]) \text{ en loi,}$$

où $\mathcal{N}(0, \mathbb{V}[X_1])$ désigne la loi normale centrée de variance $\mathbb{V}[X_1]$.

Démonstration. On note $\bar{X}_i = X_i - \mathbb{E}[X_1]$ et $G_n := \sqrt{n} \frac{1}{n} \sum_{i=1}^n \bar{X}_i$. En utilisant les propriétés de la fonction caractéristique du lemme 2.8, l'indépendance des X_i , la remarque 2.34 et l'identité des distributions des X_i , on obtient :

$$\Phi_{G_n}(u) = \Phi_{\sum_{i=1}^n \frac{\bar{X}_i}{\sqrt{n}}}(u) = \prod_{i=1}^n \Phi_{\frac{\bar{X}_i}{\sqrt{n}}}(u) = \left(\Phi_{\bar{X}_1} \left(\frac{u}{\sqrt{n}} \right) \right)^n.$$

D'après la question 3 de l'exercice 2.9 et le fait que $\mathbb{E}[\bar{X}_1] = 0$ et $\mathbb{E}[\bar{X}_1^2] = \mathbb{V}[X_1] < \infty$, on peut écrire le développement au second ordre suivant :

$$\Phi_{G_n}(u) = \left(1 - \frac{1}{2} \frac{u^2}{n} \mathbb{V}[X_1] + o\left(\frac{1}{n}\right) \right)^n \longrightarrow \phi(u) := e^{-\frac{u^2}{2} \mathbb{V}[X_1]}.$$

On reconnaît alors que $\phi = \Phi_{\mathcal{N}(0, \mathbb{V}[X_1])}$, voir question 1 de l'exercice 2.9, et on conclut grâce au théorème 2.28 de convergence de Lévy. \diamond

Chapitre 3

Espérance conditionnelle

3.1 Premières intuitions

3.1.1 Espérance conditionnelle en espace d'états fini

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et X, Y deux v.a. réelles. Dans ce paragraphe, on suppose que X et Y prennent un nombre fini de valeurs respectivement dans les ensembles $\{x_i, 1 \leq i \leq n\}$ et $\{y_j, 1 \leq j \leq m\}$. Il est alors naturel de définir la distribution conditionnelle de X sachant $Y = y_j$ par

$$\mathbb{P}[X = x_i | Y = y_j] := \frac{\mathbb{P}[(X, Y) = (x_i, y_j)]}{\mathbb{P}[Y = y_j]}, \quad 1 \leq i \leq n, \quad 1 \leq j \leq m, \quad (3.1)$$

i.e. parmi tous les événements où la modalité y_j de Y s'est réalisée, la quantité $\mathbb{P}[X = x_i | Y = y_j]$ exprime la fréquence de réalisation de la modalité x_i de X . En notant $\mathbb{P}^{Y=y_j} := \mathbb{P}[X = x_i | Y = y_j]$, on vérifie immédiatement que pour tout j , $\mathbb{P}^{Y=y_j}$ définit une mesure de probabilité sur Ω . L'espérance conditionnelle est alors naturellement définie par :

$$\mathbb{E}[X | Y = y_j] = \sum_{i=1}^n x_i \mathbb{P}[X = x_i | Y = y_j], \quad 1 \leq j \leq m.$$

Nous pouvons ainsi définir une v.a. $\xi := \mathbb{E}[X | Y]$ par $\xi(\omega) = \mathbb{E}[X | Y = Y(\omega)]$, appelée espérance conditionnelle de X sachant Y . Notons que $\mathbb{E}[X | Y]$ est complètement déterminée par la réalisation de Y . On retrouve la notion de mesurabilité puisque ceci peut s'écrire mathématiquement ξ est $\sigma(Y)$ -mesurable, ou de manière équivalente $\xi = \varphi(Y)$ pour une certaine fonction déterministe φ , voir le lemme 2.2.

Par ailleurs, un calcul direct montre que pour toute fonction déterministe φ :

$$\begin{aligned}\mathbb{E}[\mathbb{E}[X|Y]\varphi(Y)] &= \sum_{j=1}^m \mathbb{P}[Y = y_j] \varphi(y_j) \mathbb{E}[X|Y = y_j] \\ &= \sum_{j=1}^m \mathbb{P}[Y = y_j] \varphi(y_j) \sum_{i=1}^n x_i \frac{\mathbb{P}[X = x_i, Y = y_j]}{\mathbb{P}[Y = y_j]} \\ &= \sum_{j=1}^m \sum_{i=1}^n \mathbb{P}[X = x_i, Y = y_j] \varphi(y_j) x_i = \mathbb{E}[X\varphi(Y)].\end{aligned}$$

Notons $\langle \cdot, \cdot \rangle_2$ le produit scalaire dans \mathbb{L}^2 , et écrivons ce dernier résultat sous la forme

$$\langle X - \mathbb{E}[X|Y], \varphi(Y) \rangle_2 = 0 \quad \text{pour tout fonction } \varphi : \mathbb{R} \rightarrow \mathbb{R}.$$

Il s'agit de la condition d'orthogonalité de $X - \mathbb{E}[X|Y]$ à l'espace vectoriel de toutes les fonctions de Y ou, de manière équivalente, l'espace vectoriel de toutes les v.a. $\sigma(Y)$ -mesurable. Ainsi, $\mathbb{E}[X|Y]$ s'interprète géométriquement comme la projection orthogonale, au sens de \mathbb{L}^2 , de X sur l'e.v. des v.a. $\sigma(Y)$ -mesurable, et est la solution du problème variationnel :

$$\min_f \|X - f(Y)\|^2.$$

Cette interprétation géométrique montre que $\mathbb{E}[X|Y]$ est la meilleure approximation, au sens de \mathbb{L}^2 , de X par une fonction de Y .

3.1.2 Cas des variables à densités

Supposons maintenant que le couple (X, Y) est à valeurs dans \mathbb{R}^2 et admet une distribution absolument continue par rapport à la mesure de Lebesgue dans \mathbb{R}^2 , de densité

$$f_{(X,Y)}(x, y) dx dy$$

La loi marginale de Y est obtenue par intégration par rapport à la variable x :

$$f_Y(y) = \int f_{(X,Y)}(x, y) dx.$$

Dans ce cas, la probabilité conditionnelle est naturellement définie par :

$$f_{X|Y=y}(x) = \frac{f_{(X,Y)}(x, y)}{f_Y(y)} = \frac{f_{(X,Y)}(x, y)}{\int f_{(X,Y)}(x, y) dx},$$

qui est l'analogue de (3.1) dans le contexte présent. Il est clair que pour tout y fixé, la fonction $f_{X|Y=y}(x)$ définit une densité sur \mathbb{R} , et qu'on peut lui associer l'opérateur d'espérance :

$$\mathbb{E}[X|Y = y] = \int x f_{X|Y=y}(x) dx$$

qui définit encore une variable aléatoire $\mathbb{E}[X|Y]$, appelée espérance conditionnelle de X sachant Y . Comme dans le cas fini, $\mathbb{E}[X|Y]$ est une fonction déterministe de Y , elle est donc $\sigma(Y)$ -mesurable. La condition d'orthogonalité de la section précédente se vérifie aussi par un calcul direct : pour toute fonction $\phi : \mathbb{R} \rightarrow \mathbb{R}$ bornée

$$\begin{aligned} \langle \mathbb{E}[X|Y], \phi(Y) \rangle_{\mathbb{L}^2} &= \mathbb{E}[\mathbb{E}[X|Y]\phi(Y)] \\ &= \int \phi(y) f_Y(y) \int x f_{X|Y=y}(x) dx dy \\ &= \iint x \phi(y) f_{(X,Y)}(x, y) dx dy = \mathbb{E}[X\phi(Y)]. \end{aligned}$$

On retrouve alors l'interprétation géométrique de l'espérance conditionnelle comme projection orthogonale, au sens de \mathbb{L}^2 , de la v.a. X sur l'e.v. des v.a. $\sigma(Y)$ -mesurables.

3.2 Définition et premières propriétés

Considérons maintenant le cadre général d'un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, et soit \mathcal{F} une sous- σ -algèbre de \mathcal{A} .

Les arguments intuitifs du paragraphe précédent suggèrent d'introduire la notion d'espérance conditionnelle par la projection orthogonale au sens du produit scalaire de \mathbb{L}^2

$$P_{\mathcal{F}}(X) := \operatorname{Argmin} \{ \|X - Y\|_2 : Y \in \mathcal{L}^2(\mathcal{F}, \mathbb{P}) \}, \quad X \in \mathcal{L}^2(\mathcal{A}, \mathbb{P}).$$

Ceci est en effet rendu possible grâce à la structure d'espace de Hilbert de l'espace quotient \mathbb{L}^2 muni de la norme $\|\cdot\|_2$.

Lemme 3.1. *L'opérateur de projection orthogonale $P_{\mathcal{F}}$ est bien défini sur $\mathbb{L}^2(\mathcal{A}, \mathbb{P})$, et vérifie*

$$\mathbb{E}[X\mathbf{1}_F] = \mathbb{E}[P_{\mathcal{F}}(X)\mathbf{1}_F] \quad \text{pour tout } F \in \mathcal{F} \text{ et } X \in \mathbb{L}^2(\mathcal{A}, \mathbb{P}).$$

De plus, on a les propriétés suivantes :

- (i) $X \geq 0$ p.s. $\implies P_{\mathcal{F}}(X) \geq 0$ p.s.
- (ii) $\mathbb{E}[P_{\mathcal{F}}(X)] = \mathbb{E}[X]$.

Démonstration. On travaille avec l'espace quotient $\mathbb{L}^2(\mathcal{A}, \mathbb{P})$ identifiant ainsi les v.a. égales p.s. La projection orthogonale $P_{\mathcal{F}}$ est bien définie car l'e.v. $\mathbb{L}^2(\mathcal{A}, \mathbb{P})$ et le s.e.v. $\mathbb{L}^2(\mathcal{F}, \mathbb{P})$ sont complets. Alors, on sait que pour tout X , il existe une (unique) v.a. $Z := P_{\mathcal{F}}(X) \in \mathbb{L}^2(\mathcal{F}, \mathbb{P})$ vérifiant les conditions d'orthogonalité

$$\mathbb{E}[(X - Z)Y] = 0 \quad \text{pour tout } Y \in \mathbb{L}^2(\mathcal{F}, \mathbb{P}).$$

En particulier, pour tout $F \in \mathcal{F}$, la v.a. $Y = \mathbf{1}_F \in \mathbb{L}^2(\mathcal{F}, \mathbb{P})$ induit la condition d'orthogonalité $\mathbb{E}[X\mathbf{1}_F] = \mathbb{E}[Z\mathbf{1}_F]$.

Supposons maintenant que $X \geq 0$ p.s., notons $Z := P_{\mathcal{F}}(X)$, et prenons $F := \{Z \leq 0\} \in \mathcal{F}$. Alors $0 \leq \mathbb{E}[X\mathbf{1}_F] = \mathbb{E}[Z\mathbf{1}_F] = -\mathbb{E}[Z^-] \leq 0$, et $Z^- = 0$ p.s. montrant la propriété (i).

Pour la propriété (ii), il suffit de remarquer que $F = \Omega \in \mathcal{F}$ du fait que \mathcal{F} est une σ -algèbre. Alors (i) donne le résultat voulu. \diamond

Théorème 3.2. *Pour toute v.a. $X \in \mathcal{L}^1(\mathcal{A}, \mathbb{P})$, il existe une v.a. Z telles que*

(a) Z est \mathcal{F} -mesurable,

(b) $\mathbb{E}[|Z|] < \infty$,

(c) *Pour tout événement $F \in \mathcal{F}$, on a $\mathbb{E}[X\mathbf{1}_F] = \mathbb{E}[Z\mathbf{1}_F]$.*

De plus, si \tilde{Z} est une autre v.a. vérifiant (a,b,c), alors $Z = \tilde{Z}$ p.s.

Définition 3.3. *Une v.a. vérifiant les propriétés (a)-(b)-(c) est appelée version de l'espérance conditionnelle de X sachant \mathcal{F} , notée $\mathbb{E}[X|\mathcal{F}]$, et est unique à l'égalité p.s. près.*

Si $\mathcal{F} = \sigma(Y_1, \dots, Y_n)$, on écrit simplement $\mathbb{E}[X|Y_1, \dots, Y_n]$.

Preuve du théorème 3.2 Commençons par montrer l'unicité. Si Z et \tilde{Z} vérifient (a,b,c), alors $\mathbb{E}[(Z - \tilde{Z})\mathbf{1}_F] = 0$ pour tout $F \in \mathcal{F}$. Mais Z et \tilde{Z} étant \mathcal{F} -mesurable, on peut choisir $F = \{Z - \tilde{Z} \geq 0\} \in \mathcal{F}$, et l'égalité précédente implique que $(Z - \tilde{Z})^+ = 0$ p.s. Le choix $F = \{Z - \tilde{Z} \leq 0\} \in \mathcal{F}$ conduit à $(Z - \tilde{Z})^- = 0$ p.s. et par suite $Z = \tilde{Z}$ p.s.

Pour l'existence, il suffit de traiter le cas $X \geq 0$ et d'utiliser la décomposition $X = X^+ - X^-$ pour conclure le cas où X a un signe arbitraire (ou plutôt n'a pas de signe!). La v.a. $X_n := X \wedge n$ est bornée, donc dans $\mathcal{L}^2(\mathcal{A}, \mathbb{P})$. La v.a. $Z_n := P_{\mathcal{F}}(X_n)$ est alors bien définie d'après le lemme 3.1 et vérifie par définition les conditions (a,b,c). Observons que la suite $(Z_n)_n$ est croissante, comme conséquence de la propriété (i) du lemme 3.1 et de la linéarité de la projection $P_{\mathcal{F}}$. On introduit alors la v.a.

$$Z := \lim \uparrow_n Z_n.$$

Il est clair que Z hérite la \mathcal{F} -mesurabilité des Z_n , et que $\mathbb{E}[Z\mathbf{1}_F] = \mathbb{E}[X\mathbf{1}_F]$ pour tout $F \in \mathcal{F}$ par le théorème de convergence monotone. Pour la condition (b), remarquons par que $\mathbb{E}[Z] = \lim_{n \rightarrow \infty} \uparrow \lim \inf_n \mathbb{E}[Z_n] = \mathbb{E}[X \wedge n] \leq \mathbb{E}[X]$, où on a utilisé les propriétés (i) et (ii) du lemme 3.1. \diamond

Remarque 3.4. Regardons deux cas extrêmes pour la σ -algèbre \mathcal{F} .

- Soit $\mathcal{F} = \{\emptyset, \Omega\}$ la plus petite σ -algèbre correspondant à l'absence totale d'information. Alors la condition (a) dit que Z est déterministe, i.e. $Z = \mathbb{E}[Z]$, et la condition d'orthogonalité (c) permet d'identifier cette constante $\mathbb{E}[X] = \mathbb{E}[Z] = Z$. Ainsi l'espérance conditionnelle dans ce cas se confond avec l'espérance.

- Soit $\mathcal{F} = \sigma(X)$. Alors la condition d'orthogonalité (c) avec $F^+ := \{X - Z \geq 0\} \in \sigma(X)$ donne $\mathbb{E}[(X - Z)\mathbf{1}_{\{X - Z \geq 0\}}] = 0$, soit $(X - Z)^+ = 0$ p.s., et avec $F^- := \{X - Z \leq 0\} \in \sigma(X)$ donne $(X - Z)^- = 0$ p.s. Ainsi $\mathbb{E}[X|\sigma(X)] = X$ qui vérifie bien les autres conditions (a) et (b).

3.3 Propriétés de l'espérance conditionnelle

Commençons par les propriétés déjà esquissées dans le paragraphe précédent.

Proposition 3.5. *L'opérateur d'espérance conditionnelle $\mathbb{E}[\cdot|\mathcal{F}]$ est linéaire, et pour tout $X \in \mathcal{L}^1(\mathcal{A}, \mathbb{P})$, on a :*

- (i) $\mathbb{E}[\mathbb{E}[X|\mathcal{F}]] = \mathbb{E}[X]$,
- (ii) si X est \mathcal{F} -mesurable, $\mathbb{E}[X|\mathcal{F}] = X$ p.s.
- (iii) si $X \geq 0$, $\mathbb{E}[X|\mathcal{F}] \geq 0$ p.s.

Exercice 3.6. *Prouver la proposition 3.5.*

Nous montrons maintenant que l'espérance conditionnelle jouit des mêmes propriétés de passage à la limite que l'espérance.

Proposition 3.7. *Pour $X, X_n \in \mathcal{L}^1(\mathcal{A}, \mathbb{P})$, $n \in \mathbb{N}$, on a :*

- (1- Convergence monotone) si $0 \leq X_n \uparrow X$, alors $\mathbb{E}[X_n|\mathcal{F}] \uparrow \mathbb{E}[X|\mathcal{F}]$,
- (2- Fatou) si $X_n \geq 0$, alors $\mathbb{E}[\liminf_n X_n|\mathcal{F}] \leq \liminf_n \mathbb{E}[X_n|\mathcal{F}]$,
- (3- Convergence dominée) si $\sup_n |X_n| \in \mathcal{L}^1(\mathcal{A}, \mathbb{P})$ et $X_n \rightarrow X$, p.s. alors $\mathbb{E}[X_n|\mathcal{F}] \rightarrow \mathbb{E}[X|\mathcal{F}]$ p.s.

Démonstration. 1- la suite $Z_n := \mathbb{E}[X_n|\mathcal{F}]$ est croissante d'après la proposition 3.5 (iii). On définit alors la variable $Z = \lim \uparrow Z_n$ qui est par définition \mathcal{F} -mesurable positive et, par Fatou, $\mathbb{E}[Z] \leq \liminf_n \mathbb{E}[Z_n] = \liminf_n \mathbb{E}[X_n] \leq \mathbb{E}[X] < \infty$. Enfin, pour tout $F \in \mathcal{F}$, on a $\mathbb{E}[X_n \mathbf{1}_F] = \mathbb{E}[Z_n \mathbf{1}_F]$ et on déduit du théorème de convergence monotone que $\mathbb{E}[X \mathbf{1}_F] = \mathbb{E}[Z \mathbf{1}_F]$. Ainsi Z vérifie les propriétés (a,b,c) du théorème 3.2 et $Z = \mathbb{E}[X|\mathcal{F}]$, p.s.

2- D'après la monotonie de l'opérateur d'espérance conditionnelle due à (iii) de la proposition 3.5, on a

$$\inf_{k \geq n} \mathbb{E}[X_k|\mathcal{F}] \geq \mathbb{E} \left[\inf_{k \geq n} X_k \middle| \mathcal{F} \right] \quad \text{pour tout } n \geq 1,$$

et on conclut en utilisant le résultat de convergence monotone démontré en première partie de cette preuve.

3- Avec $Y_n := |X_n - X|$ et $\bar{Y} := \sup_n Y_n$, on vérifie que $\bar{Y} \in \mathcal{L}^1(\mathcal{A}, \mathbb{P})$, et on applique le lemme de Fatou conditionnel, qu'on vient de démontrer, à la v.a. $\bar{Y} - Y_n$. Le résultat s'en déduit immédiatement. \diamond

L'inégalité de Jensen s'étend aussi aux espérances conditionnelles :

Proposition 3.8. *Soit $X \in \mathcal{L}^1(\mathcal{A}, \mathbb{P})$, et $g : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$ une fonction convexe telle que $\mathbb{E}[|g(X)|] < \infty$. Alors $\mathbb{E}[g(X)|\mathcal{F}] \geq g(\mathbb{E}[X|\mathcal{F}])$.*

Démonstration. Il suffit de répéter les arguments de la preuve de l'inégalité de Jensen dans le cas non conditionnel, théorème 2.6... \diamond

La propriété suivante est très utile, et est une conséquence de la propriété des projections itérées en algèbre linéaire.

Proposition 3.9. (*Projections itérées*) Pour $X \in \mathcal{L}^1(\mathcal{A}, \mathbb{P})$ et \mathcal{F}, \mathcal{G} des sous- σ -algèbre de \mathcal{A} :

$$\mathcal{F} \subset \mathcal{G} \implies \mathbb{E}[\mathbb{E}\{X|\mathcal{G}\}|\mathcal{F}] = \mathbb{E}[X|\mathcal{F}].$$

Démonstration. 1-On observe que $\mathcal{L}^2(\mathcal{F}, \mathbb{P}) \subset \mathcal{L}^2(\mathcal{G}, \mathbb{P})$, et que par suite le résultat dans le cas $X \in \mathcal{L}^2(\mathcal{A}, \mathbb{P})$ est une conséquence immédiate du théorème de projections itérées en algèbre linéaire. Puis, le théorème de convergence monotone permet de l'étendre aux variables $X \in \mathcal{L}^1(\mathcal{A}, \mathbb{P})$. \diamond

La propriété suivante généralise celle de la proposition 3.5 (i).

Proposition 3.10. Soient \mathcal{F} une sous- σ -algèbre de \mathcal{A} , $X \in \mathcal{L}^0(\mathcal{A})$ et $Y \in \mathcal{L}^0(\mathcal{F})$. On suppose $\mathbb{E}[|X|] < \infty$ et $\mathbb{E}[|XY|] < \infty$. Alors

$$\mathbb{E}[XY|\mathcal{F}] = Y\mathbb{E}[X|\mathcal{F}].$$

Démonstration. On commence par le cas $Y = \mathbf{1}_A$, $A \in \mathcal{F}$. Alors pour tout $F \in \mathcal{F}$, on a $\mathbb{E}[Y\mathbb{E}[X|\mathcal{F}]\mathbf{1}_F] = \mathbb{E}[\mathbb{E}[X|\mathcal{F}]\mathbf{1}_{A \cap F}] = \mathbb{E}[X\mathbf{1}_{A \cap F}] = \mathbb{E}[XY\mathbf{1}_F]$ d'après la définition de $\mathbb{E}[X|\mathcal{F}]$ et du fait que $A \cap F \in \mathcal{F}$. Ainsi, la proposition est vraie pour les indicatrices d'événements de \mathcal{F} .

Si X est une v.a. positive, la propriété précédente s'étend par linéarité à \mathcal{S}^+ , l'ensemble des v.a. simples positives, et par le théorème de convergence monotone à l'ensemble des v.a. positives telles que $\mathbb{E}[|XY|] < \infty$ et $\mathbb{E}[|X|] < \infty$ (pour que l'espérance conditionnelle ait un sens).

Enfin, pour des variables X, Y générales, on décompose $X = X^+ - X^-$, $Y = Y^+ - Y^-$, et on applique le résultat établi pour les v.a. positives. \diamond

Les deux dernières propriétés donnent des résultats utiles sur l'espérance conditionnelle en présence d'indépendance.

Proposition 3.11. Soient $X \in \mathcal{L}^1(\mathcal{A}, \mathbb{P})$ et \mathcal{F}, \mathcal{G} des sous- σ -algèbres de \mathcal{A} telles que \mathcal{G} est indépendante de $\sigma(\sigma(X), \mathcal{F})$. Alors

$$\mathbb{E}[X|\sigma(\mathcal{F}, \mathcal{G})] = \mathbb{E}[X|\mathcal{F}].$$

Démonstration. Il suffit de vérifier pour $X \in \mathcal{L}_+^1(\mathcal{A}, \mathbb{P})$ que

$$\mathbb{E}[X\mathbf{1}_A] = \mathbb{E}[\mathbb{E}[X|\mathcal{F}]\mathbf{1}_A] \quad \text{pour tout } A \in \sigma(\mathcal{F}, \mathcal{G}).$$

En remarquant que $A \mapsto \mathbb{E}[X\mathbf{1}_A]$ et $A \mapsto \mathbb{E}[\mathbb{E}[X|\mathcal{F}]\mathbf{1}_A]$ sont des mesures sur Ω , on déduit de la proposition 1.5 qu'il suffit de vérifier l'égalité ci-dessus pour les événements A dans le π -système $\mathcal{F} \cap \mathcal{G}$. Soient alors $F \in \mathcal{F}$ et $G \in \mathcal{G}$. En utilisant l'indépendance entre \mathcal{G} et $\sigma(\sigma(X), \mathcal{F})$, la proposition 2.32, et la définition de $\mathbb{E}[X|\mathcal{F}]$, on voit que :

$$\mathbb{E}[\mathbb{E}[X|\mathcal{F}]\mathbf{1}_{F \cap G}] = \mathbb{E}[\mathbb{E}[X|\mathcal{F}]\mathbf{1}_F]\mathbb{E}[\mathbf{1}_G] = \mathbb{E}[X\mathbf{1}_F]\mathbb{E}[\mathbf{1}_G] = \mathbb{E}[X\mathbf{1}_{F \cap G}].$$

\diamond

Proposition 3.12. *Soient (X, Y) deux v.a. à valeurs dans \mathbb{R}^n et \mathbb{R}^m , respectivement, et $g : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ une fonction telle que $\mathbb{E}[|g(X, Y)|] < \infty$. Si X et Y sont indépendantes, alors*

$$\mathbb{E}[g(X, Y)|X] = G(X) \quad \text{où} \quad G(x) := \mathbb{E}[g(x, Y)] \text{ pour tout } x \in \mathbb{R}^n.$$

Démonstration. Pour tout $A \in \sigma(X)$, on doit vérifier que $\mathbb{E}[g(X, Y)\mathbf{1}_A] = \mathbb{E}[G(X)\mathbf{1}_A]$. Comme X et Y sont indépendantes, la loi du couple (X, Y) est la loi produit $\mathbb{P}^X \otimes \mathbb{P}^Y$, et on obtient immédiatement par le théorème de Fubini que

$$\begin{aligned} \mathbb{E}[g(X, Y)\mathbf{1}_A] &= \int g(x, y)\mathbf{1}_A(x)\mathbb{P}^X \otimes \mathbb{P}^Y(dx, dy) \\ &= \int \left(\int g(x, y)\mathbb{P}^Y(dy) \right) \mathbf{1}_A(x)\mathbb{P}^X(dx) = \int G(x)\mathbf{1}_A(x)\mathbb{P}^X(dx), \end{aligned}$$

ce qui est exactement le résultat recherché. \diamond

3.4 Application au filtre de Kalman-Bucy

On considère une récurrence aléatoire définie par

$$X_k = F_k X_{k-1} + f_k + \varepsilon_k \quad (3.2)$$

$$Y_k = H_k X_k + h_k + \eta_k, \quad (3.3)$$

où $F_k \in \mathcal{M}_{\mathbb{R}}(n, n)$, $f_k \in \mathbb{R}^n$, $H_k \in \mathcal{M}_{\mathbb{R}}(m, n)$, $h_k \in \mathbb{R}^m$ sont les paramètres du système, connus par l'utilisateur, et $(\varepsilon_k)_k$, $(\eta_k)_k$ sont des suites de v.a. que l'on supposera indépendantes et même gaussiennes...

Les v.a. $(X_k)_k$ à valeurs dans \mathbb{R}^n constituent l'objet d'intérêt, mais **ne sont pas observables**. Les v.a. $(Y_k)_k$ à valeurs dans \mathbb{R}^m sont observées par l'utilisateur. On dit que (3.2) est l'équation d'état, et (3.3) est l'équation d'observation. Le problème de filtrage consiste à *chercher, à partir des variables observables $(Y_k)_k$, la meilleure approximation des variables inobservables $(X_k)_k$* .

L'indice k joue un rôle important. Il représente la date de l'observation, et induit une structure de l'information naturelle et de sa progression. Ceci représente une composante importante du problème : *à chaque date k , on cherche la meilleure approximation de X_k sur la base des observations $\{Y_j, j \leq k\}$* . D'après la définition de l'espérance conditionnelle dans \mathcal{L}^2 , comme projection orthogonale sur l'espace vectoriel engendré par les v.a. de \mathcal{L}^2 mesurables par rapport au conditionnement, la meilleure approximation est donnée par

$$\mathbb{E}[X_k | Y_0, \dots, Y_k],$$

qui est ainsi la quantité que l'utilisateur cherche à calculer à chaque date k .

Enfin, en vue d'une application réelle, l'aspect de l'implémentation est une composante essentielle du problème. Quand la taille des observations croît, l'effort numérique pour calculer l'espérance conditionnelle ci-dessus peut devenir

rapidement gigantesque. Des méthodes de type **mise à jour** sont très désirables pour l'utilisateur : *utiliser l'observation la plus récente pour mettre à jour l'approximation de la date précédente.*

3.4.1 Lois conditionnelles pour les vecteurs gaussiens

Dans ce paragraphe, nous isolons un résultat sur les vecteurs gaussiens qui sera crucial pour la résolution du problème de filtrage de Kalman-Bucy. Rappelons qu'un vecteur aléatoire Z , à valeurs dans \mathbb{R}^n , est gaussien si $a \cdot Z = \sum_{i=1}^n a_i Z_i$ est gaussien sur \mathbb{R} pour tout $a \in \mathbb{R}^n$.

Proposition 3.13. *Soit (X, Y) un vecteur gaussien à valeurs dans $\mathbb{R}^n \times \mathbb{R}^m$ de moyenne et de matrice de variances-covariances*

$$\mu = \begin{pmatrix} \mu^X \\ \mu^Y \end{pmatrix} \quad \text{et} \quad V = \begin{pmatrix} V^X & (V^{XY})^T \\ V^{XY} & V^Y \end{pmatrix}.$$

Supposons que la matrice $\mathbb{V}[Y] = V^Y$ est inversible. Alors, la loi conditionnelle de X sachant $Y = y$ est gaussienne de moyenne et variance :

$$\mathbb{E}[X|Y = y] = \mu^X + V^{XY}(V^Y)^{-1}(y - \mu^Y), \quad \text{et} \quad \mathbb{V}[X|Y = y] = V^X - V^{XY}(V^Y)^{-1}(V^{XY})^T$$

Démonstration. 1- On note $Z = (X, Y)$, et on suppose d'abord que V est inversible. Alors, pour $z = (x, y) \in \mathbb{R}^n \times \mathbb{R}^m$, on calcule la densité de la loi de X conditionnellement à $\{Y = y\}$ au point x par :

$$\begin{aligned} f_{X|Y=y}(x) &= \frac{f_Z(z)}{f_Y(y)} \\ &= (2\pi)^{-n/2} \sqrt{\frac{\det(V^Y)}{\det(V)}} e^{[-\frac{1}{2}(z-\mu)^T V^{-1}(z-\mu) + \frac{1}{2}(y-\mu^Y)^T (V^Y)^{-1}(y-\mu^Y)]}. \end{aligned} \quad (3.4)$$

On note $W := V^X - V^{XY}(V^Y)^{-1}(V^{XY})^T$, et on vérifie par un calcul direct que

$$\begin{pmatrix} I & -V^{XY}(V^Y)^{-1} \\ 0 & I \end{pmatrix} V \begin{pmatrix} I & 0 \\ -(V^Y)^{-1}(V^{XY})^T & I \end{pmatrix} = \begin{pmatrix} W & 0 \\ 0 & V^Y \end{pmatrix} \quad (3.5)$$

dont on tire que

$$\frac{\det(V)}{\det(V^Y)} = \det(W). \quad (3.6)$$

On déduit aussi de (3.5) que

$$V^{-1} = \begin{pmatrix} I & 0 \\ -(V^Y)^{-1}(V^{XY})^T & I \end{pmatrix} \begin{pmatrix} W^{-1} & 0 \\ 0 & (V^Y)^{-1} \end{pmatrix} \begin{pmatrix} I & -V^{XY}(V^Y)^{-1} \\ 0 & I \end{pmatrix}$$

qui implique que :

$$\begin{aligned} (z - \mu)^T V^{-1}(z - \mu) &= (x - M(x))^T W^{-1}(x - M(x)) \\ &\quad + (y - \mu^Y)^T (V^Y)^{-1}(y - \mu^Y), \end{aligned} \quad (3.7)$$

où $M(x) := \mu^X + V^{XY}(V^Y)^{-1}(y - \mu^Y)$. en injectant (3.6) et (3.7) dans (3.4), on obtient une expression de la densité conditionnelle $p_{X|Y=y}$ qui correspond à la loi annoncée dans la proposition.

2- Pour une matrice V générale, on considère l'approximation $Z_n := Z + \frac{1}{n}G$, où G est une v.a. indépendante de Z de loi gaussienne centrée réduite dans \mathbb{R}^{n+m} . Alors, on peut appliquer le résultat de la première étape et obtenir l'expression de la densité $f_{X_n|Y_n=y}(y)$ qui converge vers la densité annoncée dans la proposition, et par convergence dominées il y a convergence des fonctions de répartition correspondantes en tout point. Ainsi, pour finir la démonstration, il ne reste plus qu'à vérifier que $X_n|Y_n = y$ converge en loi vers $X|Y = y$, par exemple en montrant que $\Phi_{X_n|Y_n=y} \rightarrow \Phi_{X|Y=y}$ simplement. Pour cela, on calcule directement :

$$\begin{aligned}\Phi_{X_n|Y_n=y}(u) &= \mathbb{E} [e^{iu \cdot X_n} | Y_n = y] \\ &= \mathbb{E} [e^{iu \cdot (X + G_1^n)} | Y + G_2^n = y],\end{aligned}$$

où $G^n = (G_1^n, G_2^n) := \frac{1}{n}G$. En remarquant que G_1^n est indépendant de (X, Y, G_2^n) , puis que G_2^n est indépendant de (X, Y) , on obtient alors :

$$\begin{aligned}\Phi_{X_n|Y_n=y}(u) &= \Phi_{G_1^n}(u) \mathbb{E} \left[\mathbb{E} [e^{iu \cdot X} | Y = y - g]_{g=G_2^n} \right] \\ &= \Phi_{G_1^n}(u) \mathbb{E} [\{\Phi_{X|Y=y-g}(u)\}_{g=G_2^n}] \rightarrow \Phi_{X|Y=y}(u)\end{aligned}$$

par utilisation du théorème de convergence dominée. \diamond

3.4.2 Filtre de Kalman-Bucy

Précisons à présent les hypothèses sur les composantes aléatoires du système d'état-observation (3.2)-(3.3) :

- l'état initial X_0 est gaussien de moyenne $\mu_0^X := \mathbb{E}[X_0]$ et de variance $V_0^X := \mathbb{V}[X_0]$,
- l'aléa générant l'état $(\varepsilon_k)_k$ est une suite de v.a. indépendantes gaussiennes centrées de variances V_k^ε ,
- le bruit d'observation $(\eta_k)_k$ est une suite de v.a. indépendantes gaussiennes centrées de variances V_k^η ,
- les bruits $(\varepsilon_k)_k$, $(\eta_k)_k$ et l'état initial X_0 sont mutuellement indépendants.

On modélise l'information en introduisant les σ -algèbres $\mathcal{F}_k := \sigma(Y_0, \dots, Y_k)$. Pour tout N , le vecteur aléatoire $(X_k, Y_k)_{0 \leq k \leq N}$ est gaussien. La proposition 3.13 garantit que la loi conditionnelle de X_k sachant \mathcal{F}_k est une gaussienne dont il suffit de calculer la moyenne et la variance pour la caractériser :

$$\hat{X}_k := \mathbb{E}[X_k | \mathcal{F}_k], \quad V_k := \mathbb{V}[X_k | \mathcal{F}_k].$$

Notons que, d'après la proposition 3.13, la matrice de covariances conditionnelle V_k est indépendante du conditionnement :

$$V_k := \mathbb{E} \left[(X_k - \hat{X}_k)(X_k - \hat{X}_k)^\top \right].$$

Pour un calcul efficace des caractéristiques de la loi du filtre \hat{X}_k et V_k , on procède en deux étapes :

1. Étape de prédiction : étant donnés (\hat{X}_{k-1}, V_{k-1}) , on voit de l'équation d'état (3.2) que la loi de X_k sachant \mathcal{F}_{k-1} est gaussienne de caractéristiques :

$$\hat{X}_k^{\text{pr}} := \mathbb{E}[X_k | \mathcal{F}_{k-1}], \quad V_k^{\text{pr}} := \mathbb{V}[X_k | \mathcal{F}_{k-1}] = \mathbb{E} \left[(X_k - \hat{X}_k^{\text{pr}})(X_k - \hat{X}_k^{\text{pr}})^{\text{T}} \right].$$

où on a encore utilisé que la matrice de variance est indépendante du conditionnement, comme conséquence de la proposition 3.13.

2. Étape de correction : on utilise l'information supplémentaire Y_k de la date k ou, plus précisément, *l'innovation*

$$I_k := Y_k - \mathbb{E}[Y_k | \mathcal{F}_{k-1}] = H_k(X_k - \hat{X}_k^{\text{pr}}) + \eta_k,$$

où on a utilisé l'équation d'observation (3.3) et le fait que η_k est indépendant de \mathcal{F}_{k-1} . En particulier, on voit que

$$\begin{aligned} (I_k)_k \text{ est gaussien et } \mathbb{E}[I_k] &= \mathbb{E}[I_k | \mathcal{F}_{k-1}] = 0, \\ \mathbb{V}[I_k] &= \mathbb{V}[I_k | \mathcal{F}_{k-1}] = H_k V_k^{\text{pr}} H_k^{\text{T}} + V_k^{\eta}. \end{aligned} \quad (3.8)$$

Théorème 3.14. (*Kalman-Bucy*) *Supposons que la matrice de variance du bruit d'observation V_k^{η} est inversible pour tout $k \geq 0$. Alors, les caractéristiques de l'étape de prédiction sont données par $\hat{X}_0^{\text{pr}} = \mu_0^X = \mathbb{E}[X_0]$, $V_0^{\text{pr}} = V_0^X = \mathbb{V}[X_0]$,*

$$\hat{X}_k^{\text{pr}} = F_k \hat{X}_{k-1} + f_k, \quad V_k^{\text{pr}} = F_k V_{k-1} F_k^{\text{T}} + V_k^{\eta}, \quad k \geq 1, \quad (3.9)$$

et celles de l'étape de correction :

$$\hat{X}_k = \hat{X}_k^{\text{pr}} + K_k [Y_k - (H_k \hat{X}_k^{\text{pr}} + h_k)], \quad V_k = (I - K_k H_k) V_k^{\text{pr}}, \quad k \geq 0, \quad (3.10)$$

où $K_k := V_k^{\text{pr}} H_k^{\text{T}} (H_k V_k^{\text{pr}} H_k^{\text{T}} + V_k^{\eta})^{-1}$ est appelée *matrice de gain de Kalman*.

Démonstration. On décompose en trois étapes :

1- Initialisation de la prédiction. Le vecteur aléatoire (X_0, Y_0) est gaussien de moyenne et variance

$$\begin{pmatrix} \mu_0^X \\ H_0 \mu_0^X \end{pmatrix}, \quad \begin{pmatrix} V_0^X & V_0^X H_0^{\text{T}} \\ H_0 V_0^X & H_0 V_0^X H_0^{\text{T}} + V_0^{\eta} \end{pmatrix}.$$

On déduit de la proposition 3.13 que la loi de X_0 conditionnellement à Y_0 est gaussienne de caractéristiques

$$\begin{aligned} \hat{X}_0 &= \mu_0^X + V_0^X H_0^{\text{T}} (H_0 V_0^X H_0^{\text{T}} + V_0^{\eta})^{-1} [Y_0 - (H_0 \mu_0^X + h_0)], \\ V_0 &= V_0^X - V_0^X H_0^{\text{T}} (H_0 V_0^X H_0^{\text{T}} + V_0^{\eta})^{-1} H_0 V_0^X. \end{aligned}$$

2- Prédiction. D'après la proposition 3.13, la loi de X_k sachant \mathcal{F}_{k-1} est gaussienne dont on calcule la moyenne et la variance à partir de l'équation d'état :

$$\begin{aligned}\hat{X}_k^{\text{pr}} &= F_k \hat{X}_{k-1} + f_k \\ V_k^{\text{pr}} &= \mathbb{E} \left[(X_k - \hat{X}_k^{\text{pr}})(X_k - \hat{X}_k^{\text{pr}})^{\text{T}} \right] \\ &= \mathbb{E} \left[(F_k(X_{k-1} - \hat{X}_{k-1}) + \varepsilon_k)(F_k(X_{k-1} - \hat{X}_{k-1}) + \varepsilon_k)^{\text{T}} \right] \\ &= F_k V_{k-1} F_k^{\text{T}} + V_k^{\eta},\end{aligned}$$

d'après nos hypothèses sur le bruit η_k .

3- Correction. D'après la proposition 3.13, la loi de X_k sachant \mathcal{F}_k est gaussienne. D'après (3.8) et la proposition 3.11, on calcule :

$$\hat{X}_k = \hat{X}_k^{\text{pr}} + \mathbb{E} \left[X_k - \hat{X}_k^{\text{pr}} | \mathcal{F}_{k-1}, I_k \right] = \hat{X}_k^{\text{pr}} + \mathbb{E} \left[X_k - \hat{X}_k^{\text{pr}} | I_k \right]. \quad (3.11)$$

Par suite, $X_k - \hat{X}_k = (X_k - \hat{X}_k^{\text{pr}}) - \mathbb{E} \left[X_k - \hat{X}_k^{\text{pr}} | I_k \right]$, et

$$V_k = \mathbb{E} \left[(X_k - \hat{X}_k)(X_k - \hat{X}_k)^{\text{T}} \right] = \mathbb{E} \left[\mathbb{V}[X_k - \hat{X}_k^{\text{pr}} | I_k] \right]. \quad (3.12)$$

Pour calculer (3.11) et (3.12), on observe que le vecteur aléatoire $(X_k - \hat{X}_k^{\text{pr}}, I_k)$ est gaussien centré de matrice de variance

$$\begin{pmatrix} V_k^{\text{pr}} & V_k^{\text{pr}} H_k^{\text{T}} \\ H_k V_k^{\text{pr}} & H_k V_k^{\text{pr}} H_k^{\text{T}} + V_k^{\eta} \end{pmatrix}.$$

La matrice $H_k V_k^{\text{pr}} H_k^{\text{T}} + V_k^{\eta}$ est inversible, du fait de l'hypothèse d'inversibilité de V_k^{η} , on déduit alors de la proposition 3.11 que

$$\begin{aligned}\hat{X}_k &= \hat{X}_k^{\text{pr}} + V_k^{\text{pr}} H_k^{\text{T}} (H_k V_k^{\text{pr}} H_k^{\text{T}} + V_k^{\eta})^{-1} I_k, \\ V_k &= V_k^{\text{pr}} - V_k^{\text{pr}} H_k^{\text{T}} (H_k V_k^{\text{pr}} H_k^{\text{T}} + V_k^{\eta})^{-1} H_k V_k^{\text{pr}}.\end{aligned}$$

◇

Chapitre 4

Processus aléatoires et structure d'information

Dans tout ce chapitre $(\Omega, \mathcal{A}, \mathbb{P})$ désigne un espace probabilisé.

4.1 Processus aléatoire

Soit $X = (X_n)_{n \geq 0}$ une suite de v.a. sur (Ω, \mathcal{A}) à valeurs dans un ensemble mesuré (E, \mathcal{E}) . On dit que X est un *processus aléatoire*. L'indice n indique la date à laquelle la v.a. X_n est observée. Afin d'introduire le déroulement du temps et la structure de l'information qui en découle, on introduit la notion importante suivante.

Définition 4.1. (*Filtration*) Une *filtration* de \mathcal{A} est une suite croissante $\mathbb{F} = (\mathcal{F}_n)_{n \geq 0}$ de sous- σ -algèbres de \mathcal{A} . On dit que $(\Omega, \mathcal{A}, \mathbb{F})$ est un *espace probabilisable filtré* et $(\Omega, \mathcal{A}, \mathbb{F}, \mathbb{P})$ un *espace probabilisé filtré*.

Pour chaque $n \in \mathbb{N}$, la sous- σ -algèbre \mathcal{F}_n représente l'information disponible à la date n . La croissance de la suite $(\mathcal{F}_n)_n$ traduit l'idée que l'information ne peut que s'accumuler au fil du temps, et qu'il n'y a pas de possibilité d'oublier des informations passées.

Exemple 4.2. Soit $X = (X_n)_n$ un processus aléatoire de (Ω, \mathcal{A}) dans (E, \mathcal{E}) . Alors la suite

$$\mathcal{F}_n^X := \sigma(X_i, i \leq n), \quad n \geq 0,$$

est une filtration de \mathcal{A} appelée filtration naturelle de X .

La notion de filtration permet de décrire la structure de l'information et de sa dynamique dans le temps de manière précise. De la même façon qu'une v.a. fait référence à une σ -algèbre, un processus aléatoire se doit d'être décrit en relation avec une filtration.

Définition 4.3. Soient $X = (X_n)_n$ un processus aléatoire, et $\mathbb{F} = (\mathcal{F}_n)_n$ une filtration de \mathcal{A} . On dit que

- (i) X est \mathbb{F} -adapté si X_n est \mathcal{F}_n -mesurable pour tout $n \geq 0$,
- (ii) X est \mathbb{F} -prévisible si X_n est \mathcal{F}_{n-1} -mesurable pour tout $n \geq 0$, où $\mathcal{F}_{-1} := \{\emptyset, \Omega\}$.

4.2 Temps d'arrêt

Définition 4.4. Un temps d'arrêt ν est une variable aléatoire à valeurs dans $\mathbb{N} \cup \{\infty\}$ telle que

$$\{\nu = n\} \in \mathcal{F}_n \quad \text{pour tout } n \geq 0.$$

On note par \mathcal{T} l'ensemble des temps d'arrêt.

Remarque 4.5. On pourra vérifier en exercice que ν est un temps d'arrêt si et seulement si

$$\{\nu \leq n\} \in \mathcal{F}_n \quad \text{pour tout } n \geq 0.$$

C'est cette définition équivalente qui s'étend au temps continu.

La notion de temps d'arrêt joue un rôle central dans l'analyse des processus aléatoires. C'est la vraie notion de temps aussi bien pour les développements mathématiques que pour la modélisation. Les prises de décision en présence d'une structure d'information doivent prendre en considération l'information disponible à la date courante, et les seuls temps aléatoires qui sont perceptibles par un agent, dont la structure d'information est décrite par une filtration \mathbb{F} , sont les temps d'arrêt.

Le résultat suivant fournit une grande classe de temps d'arrêt.

Proposition 4.6. Soit $(X_n)_{n \geq 0}$ un processus \mathbb{F} -adapté à valeurs dans (E, \mathcal{E}) . Pour tout $A \in \mathcal{E}$, on définit le premier temps d'atteinte :

$$T_A := \inf \{n \geq 0 : X_n \in A\},$$

avec la convention $\inf \emptyset = \infty$. Alors T_A est un temps d'arrêt.

Démonstration. Pour tout $n \geq 0$, $k \leq n$, on a $(X_k)^{-1}(A) \in \mathcal{F}_k \subset \mathcal{F}_n$ du fait que X est \mathbb{F} -adapté. On déduit alors que $\{T_A \leq n\} = \cup_{k \leq n} \{X_k \in A\} = \cup_{k \leq n} (X_k)^{-1}(A) \in \mathcal{F}_n$. \diamond

Donnons maintenant les propriétés de stabilité élémentaires des temps d'arrêt.

Proposition 4.7. (i) Soient τ et θ deux temps d'arrêt. Alors $\tau \wedge \theta$, $\tau \vee \theta$, $\tau + \theta$ sont des temps d'arrêt.

(ii) Soient τ un temps d'arrêt, et $c \geq 0$ une constante. Alors $\tau + c$ et $(1 + c)\tau$ sont des temps d'arrêt.

(iii) Soit $(\tau_n)_{n \geq 0}$ une suite de temps d'arrêt. Alors $\liminf_n \tau_n$ et $\limsup_n \tau_n$ sont des temps d'arrêt.

Démonstration. Exercice! \diamond

Proposition 4.8. Soient $\{X_n, n \geq 0\}$ un processus aléatoire à valeurs dans un espace mesuré (E, \mathcal{E}) , et τ un temps d'arrêt. Alors, la fonction aléatoire

$$X_\tau : \omega \in \Omega \longmapsto X_{\tau(\omega)}(\omega)$$

est une variable aléatoire.

Démonstration. Pour voir que X_τ est \mathcal{A} -mesurable, on écrit que pour tout $A \in \mathcal{E}$,

$$(X_\tau)^{-1}(A) = \bigcup_{k \geq 0} [\{\tau = k\} \cap (X_k)^{-1}(A)],$$

et on remarque que pour tout $k \geq 0$, les événements $\{\tau = k\}, (X_k)^{-1}(A)$ sont dans $\mathcal{F}_k \subset \mathcal{A}$, leur intersection est donc dans \mathcal{A} . Par suite $(X_\tau)^{-1}(A) \in \mathcal{A}$ comme union dénombrable d'éléments de la σ -algèbre \mathcal{A} . \diamond

4.3 Information à un temps d'arrêt et conditionnement

On définit à présente l'information disponible à un temps d'arrêt. Ceci conduit naturellement à la notion suivante :

$$\mathcal{F}_\tau := \{A \in \mathcal{A} : A \cap \{\tau = n\} \in \mathcal{F}_n \text{ pour tout } n \geq 0\},$$

i.e. l'ensemble des événements qui, si le temps d'arrêt vaut n , sont dans l'ensemble d'information de la date n . Il est clair que si $\tau = n \in \mathbb{N}$ est déterministe, alors $\mathcal{F}_\tau = \mathcal{F}_n$ définit bien une σ -algèbre. Ceci reste vrai grâce à la notion de temps d'arrêt.

Proposition 4.9. Pour tout temps d'arrêt $\tau \in \mathcal{T}$, $\mathcal{F}_\tau \subset \mathcal{A}$ est une sous- σ -algèbre de \mathcal{A} . Si X est un processus aléatoire \mathbb{F} -adapté, X_τ est \mathcal{F}_τ -mesurable.

Démonstration. 1- Il est clair que $\emptyset \in \mathcal{F}_\tau$. Si $A \in \mathcal{F}_\tau$ alors $A^c \cap \{\tau = n\} = \{\tau = n\} \setminus (A \cap \{\tau = n\}) = \{\tau = n\} \cap (A \cap \{\tau = n\})^c \in \mathcal{F}_n$ pour tout $n \geq 0$, i.e. $A^c \in \mathcal{F}_\tau$. Si $(A_i)_i \subset \mathcal{F}_\tau$, alors $(\cup_i A_i) \cap \{\tau = n\} = (\cup_i (A_i \cap \{\tau = n\})) \in \mathcal{F}_n$ pour tout $n \geq 0$, i.e. $\cup_i A_i \in \mathcal{F}_\tau$. Ainsi \mathcal{F}_τ est une σ -algèbre.

2- On cacule directement que $\{X_\tau \in A\} \cap \{\tau = n\} = \{X_n \in A\} \cap \{\tau = n\} \in \mathcal{F}_n$. \diamond

Exercice 4.10. Pour un temps d'arrêt τ et un événement $A \in \mathcal{F}_\tau$, on définit

$$\tau_A := \tau \mathbf{1}_A + \infty \mathbf{1}_{A^c}.$$

Montrer que τ_A est un temps d'arrêt.

Pour un temps d'arrêt τ , nous pouvons évidemment définir la notion d'espérance conditionnelle par rapport à \mathcal{F}_τ qui obéit aux conditions habituelles. Le résultat suivant généralise la propriété des projections itérées de l'espérance conditionnelle.

Proposition 4.11. *Soient τ et θ deux temps d'arrêt. Alors, $\{\tau \leq \theta\}$, $\{\tau \geq \theta\}$ et $\{\tau = \theta\} \in \mathcal{F}_\tau \cap \mathcal{F}_\theta$, et pour toute v.a. X intégrable, on a*

$$\mathbb{E} \{ \mathbb{E}[X|\mathcal{F}_\tau] | \mathcal{F}_\theta \} = \mathbb{E} \{ \mathbb{E}[X|\mathcal{F}_\theta] | \mathcal{F}_\tau \} = \mathbb{E}[X|\mathcal{F}_{\tau \wedge \theta}].$$

Démonstration. Pour tout $n \geq 0$, on a $\{\tau \leq \theta\} \cap \{\tau = n\} = \{\theta \geq n\} \cap \{\tau = n\} = \{\theta < n\}^c \cap \{\tau = n\} \in \mathcal{F}_n$, prouvant que $\{\tau \leq \theta\} \in \mathcal{F}_\tau$. Par un argument similaire, on montre aussi que $\{\tau \leq \theta\} \in \mathcal{F}_\theta$.

Pour la propriété d'espérance conditionnelle, il suffit de montrer pour une v.a. positive X que $\mathbb{E} \{ \mathbb{E}[X|\mathcal{F}_\tau] | \mathcal{F}_\theta \} = \mathbb{E}[X|\mathcal{F}_{\tau \wedge \theta}] =: \xi$. En effet, pour tout $A \in \mathcal{F}_\theta$:

$$\begin{aligned} \mathbb{E}[\xi \mathbf{1}_A] &= \mathbb{E}[\mathbb{E}[X|\mathcal{F}_{\tau \wedge \theta}] \mathbf{1}_A \mathbf{1}_{\tau \leq \theta}] + \mathbb{E}[\mathbb{E}[X|\mathcal{F}_{\tau \wedge \theta}] \mathbf{1}_A \mathbf{1}_{\tau > \theta}] \\ &= \mathbb{E}[\mathbb{E}[X|\mathcal{F}_{\tau \wedge \theta}] \mathbf{1}_A \mathbf{1}_{\tau \leq \theta}] + \mathbb{E}[\mathbb{E}[X \mathbf{1}_A \mathbf{1}_{\tau > \theta} | \mathcal{F}_\theta]], \end{aligned}$$

car $\{\tau > \theta\} \in \mathcal{F}_\theta$ d'après la première partie de cette démonstration. En utilisant le théorème des projections itérées, on obtient alors :

$$\begin{aligned} \mathbb{E}[\xi \mathbf{1}_A] &= \mathbb{E}[\mathbb{E}[X|\mathcal{F}_{\tau \wedge \theta}] \mathbf{1}_A \mathbf{1}_{\tau \leq \theta}] + \mathbb{E}[\mathbb{E}[X \mathbf{1}_A \mathbf{1}_{\tau > \theta} | \mathcal{F}_\tau]] \\ &= \mathbb{E}[\mathbb{E}[X|\mathcal{F}_\tau] \mathbf{1}_A \mathbf{1}_{\tau \leq \theta}] + \sum_{i \geq 0} \sum_{j < i} \mathbb{E}[\mathbb{E}[X \mathbf{1}_A \mathbf{1}_{\{\tau=i\}} \mathbf{1}_{\{\theta=j\}} | \mathcal{F}_i]] \\ &= \mathbb{E}[\mathbb{E}[X|\mathcal{F}_\tau] \mathbf{1}_A \mathbf{1}_{\tau \leq \theta}] + \sum_{i \geq 0} \sum_{j < i} \mathbb{E}[\mathbb{E}[X|\mathcal{F}_i] \mathbf{1}_A \mathbf{1}_{\{\tau=i\}} \mathbf{1}_{\{\theta=j\}}] \\ &= \mathbb{E}[\mathbb{E}[X|\mathcal{F}_\tau] \mathbf{1}_A \mathbf{1}_{\tau \leq \theta}] + \mathbb{E}[\mathbb{E}[X|\mathcal{F}_\tau] \mathbf{1}_A \mathbf{1}_{\tau > \theta}] \\ &= \mathbb{E}[\mathbb{E}[X|\mathcal{F}_\tau] \mathbf{1}_A]. \end{aligned}$$

◇

Chapitre 5

Chaînes de Markov : premières définitions

5.1 Premières définitions

5.1.1 Dynamique markovienne

Dans tout ce chapitre, nous considérons un processus stochastique $X = (X_n)_{n \geq 0}$, défini sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, et prenant ses valeurs dans un espace d'états discret E , fini ou dénombrable. Nous noterons π_n la distribution marginale de X_n :

$$\pi_n(x) := \mathbb{P}[X_n = x] \quad \text{pour tout } x \in E,$$

et \mathcal{F}^X la σ -algèbre des événements passés jusqu'à la date n sur la base de l'observation du processus X :

$$\mathcal{F}_n^X := \sigma\{X_0, \dots, X_n\}.$$

Définition 5.1. *On dit que $(X_n)_{n \geq 0}$ est une chaîne de Markov si*

$$\mathbb{P}[X_n \in A | \mathcal{F}_{n-1}^X] = \mathbb{P}[X_n \in A | X_{n-1}] \quad \text{pour tous } A \subset E \text{ et } n \geq 1.$$

En d'autres termes, une chaîne de Markov est un processus stochastique dont la dépendance au passé est résumée par la seule observation présente. Les *probabilités de transition* introduites dans la définition précédente peuvent être représentées par les *matrices de transition* $P_n = (P_n(x, y))_{x, y \in E}$ définies par

$$P_n(x, y) := \mathbb{P}[X_n = y | X_{n-1} = x] \quad \text{pour } n \geq 1, x, y \in E.$$

Comme l'espace d'état E est au plus dénombrable, il s'agit de matrices de taille au plus dénombrable. En termes d'algèbre linéaire, ces matrices de transition

sont dites *stochastiques*, c'est à dire qu'elles sont à composantes positives et que leurs lignes somment à l'unité :

$$P_n(x, y) \geq 0 \text{ et } \sum_{y \in E} P_n(x, y) = 1 \text{ pour tous } x, y \in E.$$

Ceci traduit le fait que $P_n(x, \cdot)$ définit une mesure de probabilité sur l'espace d'état discret E .

Nous avons ainsi construit pour toute chaîne de Markov une suite $(P_n)_{n \geq 0}$ de matrices de transition. Le résultat suivant donne la réciproque.

Proposition 5.2. *Soit $(U_n)_{n \geq 0}$ une suite de v.a. iid de distribution uniforme sur $[0, 1]$, et soit $(P_n)_{n \geq 1}$ une suite de matrices stochastiques sur un espace d'état E . Alors, pour toute distribution initiale π_0 , il existe une chaîne de Markov de loi initiale π_0 et de matrices de transition $(P_n)_{n \geq 1}$.*

Démonstration. Avec $E = \{x_i, i \in \mathbb{N}\}$, on pose $\Pi_0(i) := \sum_{j \leq i} \pi_0(x_j)$. On définit la variable initiale

$$X_0 = \sum_{i \geq 0} x_i \mathbf{1}_{\Pi_0(i-1), \Pi_0(i)}(U_0),$$

qui est bien distribuée suivant la loi π_0 . Puis, étant donnée la v.a. X_{n-1} , on pose $\Pi_n^{X_{n-1}}(i) := \sum_{j \leq i} P_{n-1}(X_{n-1}, x_j)$, et on définit

$$X_n = \sum_{i \geq 0} x_i \mathbf{1}_{\Pi_n^{X_{n-1}}(i-1), \Pi_n^{X_{n-1}}(i)}(U_n),$$

dont la loi conditionnelle à X_{n-1} est bien $P_{n-1}(X_{n-1}, \cdot)$. ◇

5.1.2 Distributions marginales, espérance

Par conditionnements successifs, on voit immédiatement que la distribution du n -uplet (X_0, \dots, X_n) extrait d'une chaîne de Markov X se décompose en :

$$\begin{aligned} \mathbb{P}[(X_0, \dots, X_n) = (x_0, \dots, x_n)] &= \mathbb{P}[X_0 = x_0] P_1(x_0, x_1) \cdots P_n(x_{n-1}, x_n) \\ &= \pi_0(x_0) P_1(x_0, x_1) \cdots P_n(x_{n-1}, x_n), \end{aligned}$$

pour tous $x_0, \dots, x_n \in E$. En particulier, les probabilités marginales π_n se déduisent de π_0 et des matrices de transition par :

$$\pi_n(x) = \sum_{x_0, \dots, x_{n-1} \in E} \pi_0(x_0) P_1(x_0, x_1) \cdots P_n(x_{n-1}, x) \text{ pour tout } x \in E.$$

Ces formules se simplifient considérablement en utilisant les notations matricielles :

$$\pi_n = \pi_0 P_1 \cdots P_n \text{ pour tout } n \geq 1,$$

où la distribution π_0 est à comprendre dans ce contexte comme un vecteur ligne de taille $\text{card}(E)$ multipliant **à gauche** le produit de matrices $P_1 \dots P_n$.

Un argument de conditionnement similaire conduit à la *formule de Chapman-Kolmogorov* :

$$\mathbb{P}[X_n = y | X_0 = x] = \sum_{z \in E} \mathbb{P}[X_n = y | X_k = z] \mathbb{P}[X_k = z | X_0 = x]; \quad x, y \in E, \quad (5.1)$$

pour tous $0 \leq k \leq n$. Cette formule exprime le fait que la probabilité d'aller de x en y entre la date 0 et la date n se décompose comme la somme des probabilités d'aller de x en y en passant par un état z arbitraire à une date intermédiaire k .

Pour une fonction $f : E \rightarrow \mathbb{R}$ positive ou intégrable, l'espérance conditionnelle $\mathbb{E}[f(X_n) | \mathcal{F}_{n-1}^X]$ se calcule aisément à l'aide de la matrice de transition P_n :

$$\begin{aligned} \mathbb{E}[f(X_n) | \mathcal{F}_{n-1}^X] &= \mathbb{E}[f(X_n) | X_{n-1}] \\ &=: (P_n f)(X_{n-1}) = \sum_{y \in E} P_n(X_{n-1}, y) f(y). \end{aligned}$$

En représentant f par un vecteur colonne $(f(x), x \in E)$ de taille $\text{card}(E)$, la formule précédente peut être comprise comme le produit matriciel **à droite** $P_n f$:

$$\mathbb{E}[f(X_n) | X_{n-1} = x] = (P_n f)(x). \quad (5.2)$$

Enfin, l'espérance non conditionnelle de la variable aléatoire $f(X_n)$ se calcule par le produit matriciel

$$\mathbb{E}[f(X_n)] = \pi_0 P_1 \dots P_n f \quad \text{pour tout } n \geq 1. \quad (5.3)$$

Insistons sur le fait que la distribution π_0 est représentée par un vecteur ligne, alors que la fonction f est représentée par un vecteur colonne.

Enfin, pour tout $x \in E$, nous noterons par \mathbb{P}_x la probabilité conditionnelle

$$\begin{aligned} \mathbb{P}_x[(X_1, \dots, X_n) = (x_1, \dots, x_n)] &:= \mathbb{P}[(X_1, \dots, X_n) = (x_1, \dots, x_n) | X_0 = x] \\ &= P_1(x, x_1) \dots P_n(x_{n-1}, x_n), \end{aligned}$$

et par \mathbb{E}_x l'opérateur d'espérance conditionnelle correspondant :

$$\mathbb{E}_x[f(X_n)] := \mathbb{E}[f(X_n) | X_0 = x] = \delta_x P_1 \dots P_n f,$$

où $\delta_x(\cdot) = \mathbf{1}_x(\cdot)$ désigne la mesure de Dirac au point x , et est représentée par un vecteur ligne de taille $\text{card}(E)$. Si la distribution de X_0 est le vecteur ligne π_0 , on notera parfois l'espérance par :

$$\mathbb{E}_{\pi_0}[f(X_n)] := \mathbb{E}[f(X_n)] = \pi_0 P_1 \dots P_n f$$

afin de rappeler explicitement la distribution initiale.

Exercice 5.3. Soient X et Y deux chaînes de Markov.

1. Si X et Y sont indépendantes, montrer que (X, Y) est une chaîne de Markov (appelée chaîne produit). Donner un exemple de chaînes non indépendantes pour lesquelles le couple (X, Y) n'est pas une chaîne de Markov.
2. La somme $X + Y$ n'est en général pas une chaîne de Markov (donner un contre exemple).
3. On note $X_n^* := \max_{k \leq n} X_k$ le maximum courant de la chaîne ou le processus record. Montrer que $\{X_n^*, n \geq 0\}$ n'est pas une chaîne de Markov, en général, et que le couple (X, X^*) en est une.

5.1.3 Propriété de Markov forte

Si $(X_n)_{n \geq 0}$ est une chaîne de Markov caractérisée par la distribution initiale π_0 et les matrices de transition $(P_n)_{n \geq 0}$, le processus stochastique $(X_{k+n})_{n \geq 0}$ est une chaîne de Markov caractérisée par la distribution initiale π_k et les matrices de transition $(P_{k+n})_{n \geq 0}$. Le résultat suivant montre que cette propriété s'étend aux temps d'arrêt.

Théorème 5.4 (Propriété de Markov forte). Soient $X = (X_n)_{n \geq 0}$ une chaîne de Markov et τ un temps d'arrêt à valeurs dans \mathbb{N} . Alors pour tout $A \subset E$ et $n \geq 1$:

$$\mathbb{P}[X_{\tau+n} \in A | \mathcal{F}_{\tau+n-1}^X] = \mathbb{P}[X_{\tau+n} \in A | X_{\tau+n-1}].$$

Démonstration. Soient $A \subset E$, $B \in \mathcal{F}_{\tau+n-1}^X$, et $n \geq 1$ fixés. On calcule directement que :

$$\begin{aligned} \mathbb{E}\{\mathbb{P}[X_{\tau+n} \in A | X_{\tau+n-1}] \mathbf{1}_B\} &= \mathbb{E}\{\mathbb{E}[\mathbf{1}_A(X_{\tau+n}) | X_{\tau+n-1}] \mathbf{1}_B\} \\ &= \mathbb{E}\left\{ \sum_{k \geq 0} \mathbb{E}[\mathbf{1}_A(X_{k+n}) | X_{k+n-1}] \mathbf{1}_{B \cap \{\tau=k\}} \right\} \\ &= \mathbb{E}\left\{ \sum_{k \geq 0} \mathbb{E}[\mathbf{1}_A(X_{k+n}) | \mathcal{F}_{k+n-1}^X] \mathbf{1}_{B \cap \{\tau=k\}} \right\}, \end{aligned}$$

où la dernière égalité découle du fait que X est une chaîne de Markov. Comme τ est un temps d'arrêt, $B \cap \{\tau = k\} = B \cap \{\tau + n - 1 = k + n - 1\} \in \mathcal{F}_{k+n-1}^X$, et par suite

$$\mathbb{E}\{\mathbb{P}[X_{\tau+n} \in A | X_{\tau+n-1}] \mathbf{1}_B\} = \mathbb{E}\left\{ \sum_{k \geq 0} \mathbb{E}[\mathbf{1}_A(X_{k+n}) \mathbf{1}_{B \cap \{\tau=k\}} | \mathcal{F}_{k+n-1}^X] \right\}.$$

On conclut alors en utilisant le théorème de convergence monotone ainsi que la propriété des projections itérées des espérances conditionnelles :

$$\mathbb{E}\{\mathbb{P}[X_{\tau+n} \in A | X_{\tau+n-1}] \mathbf{1}_B\} = \mathbb{E}\{\mathbf{1}_A(X_{\tau+n}) \mathbf{1}_B\}.$$

Comme $B \in \mathcal{F}_{\tau+n-1}$ est arbitraire, ceci montre bien le résultat voulu d'après la définition de l'espérance conditionnelle. \diamond

5.1.4 Chaînes de Markov homogènes

Dans le cadre de ce cours, nous allons nous intéresser essentiellement aux chaînes de Markov dont les transitions sont indépendantes de la variable temporelle.

Définition 5.5. *Une chaîne de Markov X est dite homogène si sa matrice de transition P_n est indépendante de n .*

La distribution d'une chaîne de Markov homogène est caractérisée par la distribution π_0 de la variable aléatoire initiale X_0 et la matrice de transition P . Les distributions marginales se déduisent simplement des puissances de la matrice de transition :

$$\pi_n = \pi_0 P^n \quad \text{pour tout } n \geq 1.$$

La formule de Chapman-Kolmogorov (5.1) est tout simplement la traduction probabiliste de la décomposition algébrique de la puissance P^n en produits $P^k P^{n-k}$ pour tout $k \leq n$.

Enfin, si $(X_n)_{n \geq 0}$ est une chaîne de Markov homogène caractérisée par (π_0, P) , le processus stochastique $(X_{k+n})_{n \geq 0}$ est une chaîne de Markov homogène caractérisée par la distribution initiale π_k et la matrice de transition P . Ce résultat s'étend aux temps d'arrêt grâce à la propriété de Markov forte énoncée dans le théorème 5.4.

5.2 Exemples

Dans ce paragraphe, nous décrivons quelques problèmes, qui se modélisent naturellement par des chaînes de Markov, issus d'applications aussi diverses que la physique, l'économie, la dynamique des populations, la génétique, les télécommunications... D'autres problèmes seront développés dans les chapitres ultérieurs.

Exemple 5.6. (*Marche aléatoire*) Nous commençons par un exemple général qui intervient dans le cadre de plusieurs applications. On suppose que E est un réseau, i.e. un sous-groupe additif discret. L'exemple typique dans nos applications est \mathbb{Z}^d . Soit $(\xi_n)_{n \geq 1}$ une suite de v.a. iid à valeurs dans E . Etant donné une initialisation par une v.a. X_0 à valeurs dans E , la marche aléatoire sur le réseau E est définie par la suite de v.a.

$$X_n = X_{n-1} + \xi_n \quad \text{pour tout } n \geq 1.$$

Ceci définit clairement une chaîne de Markov homogène de matrice de transition

$$P(x, y) = \mathbb{P}[\xi_1 = y - x] = : p(y - x) \quad \text{pour tous } x, y \in E.$$

Comme $X_n = X_0 + \sum_{1 \leq i \leq n} \xi_i$, on déduit de la loi des grands nombres que $X_n = n\mathbb{E}[\xi_1] + o(n)$ p.s. En particulier, si $\mathbb{E}[\xi_1] \neq 0$, la suite part à l'infini le long de cette direction. Si $\mathbb{E}[\xi_1] = 0$ et $\mathbb{E}[\xi_1^2] < \infty$, le théorème central limite montre que $n^{-1/2}X_n$ converge en loi vers une gaussienne centrée de variance $\mathbb{V}[\xi_1]$.

Pour $E = \mathbb{Z}^d$, la marche aléatoire est dite *au plus proche voisin* si $P(x, y) = 0$ dès que $|x - y|_1 > 1$, et *symétrique* si de plus

$$P(x, y) = \frac{1}{2d} \mathbf{1}_{\{|x-y|_1=1\}}.$$

Pour des réseaux plus généraux, on peut définir des notions analogues de voisins directs et étendre de manière naturelle la notion de marche aléatoire au plus proche voisin symétrique.

Exemple 5.7. (*Urne d'Ehrenfest*) Il s'agit d'un modèle simple de mécanique statistique proposé par Tatiana et Paul Ehrenfest. Une urne contient un nombre N de particules d'un gaz réparties dans deux compartiments communicants numérotés 0 et 1. N est de l'ordre du nombre d'Avogadro $6,02 \times 10^{23}$. Les particules sont initialement réparties selon une distribution donnée. La répartition entre les deux compartiments varie de manière aléatoire comme suit. A chaque date de changement, une molécule choisie au hasard (uniformément parmi toutes les molécules) change de compartiment. Ce problème peut être modélisé de deux points de vue différents. La modélisation microscopique s'intéresse à la dynamique individuelle dans le temps de chacune des particules du système. La modélisation macroscopique s'intéresse au nombre de molécules se trouvant dans chacun des compartiments.

Modélisation microscopique. Pour des entiers $i \geq 0$ et $1 \leq j \leq N$, on désigne par X_i^j le numéro du compartiment contenant la $j^{\text{ème}}$ particule à la date i . Ainsi X_i^j prend ses valeurs dans $\{0, 1\}$, et l'état du système à la date i est décrit par le vecteur aléatoire $X_i = (X_i^j, 1 \leq j \leq N)$ à valeurs dans $E^{\text{micro}} := \{0, 1\}^N$. Rappelons que, N étant très grand, le vecteur X_i a une taille très grande, et tous les calculs numériques demandent rapidement un effort gigantesque.

Pour modéliser la dynamique du système, on introduit une suite $(\xi_i, i \geq 1)$ de variables aléatoires indépendantes et identiquement distribuées selon la loi uniforme sur l'ensemble $\{1, 2, \dots, N\}$, et indépendantes de X_0 . A chaque date i , la réalisation de la variable aléatoire ξ_i indique la particule qui change de compartiment, i.e.

$$X_{i+1}^j = X_i^j \text{ pour } j \neq \xi_i \quad \text{et} \quad X_{i+1}^{\xi_i} = 1 - X_i^{\xi_i}.$$

La caractère iid de la suite de v.a. $(\xi_i, i \geq 1)$ implique que la suite de vecteurs aléatoires $\{X_i, i \geq 0\}$ forme une chaîne de Markov homogène sur E^{micro} de matrice de transition

$$P^{\text{micro}}(x, y) = \frac{1}{N} \mathbf{1}_{\{|x-y|_1=1\}} \quad \text{pour tous } x, y \in E^{\text{micro}} = \{0, 1\}^N,$$

où $|x|_1 = \sum_{1 \leq j \leq N} |x^j|$ désigne la norme 1 dans \mathbb{R}^N . Il s'agit de la marche aléatoire au plus proche voisin, pour l'addition modulo 2, symétrique sur $\{0, 1\}^N$.

Modélisation macroscopique. Une représentation simplifiée du système, contournant la grande dimension, consiste à se suffire de décrire le nombre de particules dans chaque compartiment. Avec les notations précédentes, le nombre de particules dans le compartiment 1 à la date i est donné par

$$S_i := \sum_{j=1}^N X_i^j.$$

Pour tout $i \geq 0$, S_i est un vecteur aléatoire à valeurs dans $E^{\text{macro}} := \{0, 1, \dots, N\}$. La distribution de la variable aléatoire initiale S_0 est directement déduite de celle de X_0 . Pour décrire la dynamique de la suite $\{S_i, i \geq 1\}$ de manière plus simple, nous introduisons pour tout $x \in \{0, 1\}^N$ les notations suivantes :

- Soit $U^x := \{i : 1 \leq i \leq N \text{ et } x_i = 1\}$, que l'on représente par $U^x := \{n_1^x, \dots, n_k^x\}$ où $k = k^x = \text{card}(U^x)$ et $n_1^x < \dots < n_k^x$. De même, on représente l'ensemble $Z^x := \{1, \dots, N\} \setminus U^x$ par $Z^x := \{n_{k+1}^x, \dots, n_N^x\}$ en arrangeant ses éléments par ordre croissant.
- Pour tout $i \in \{1, \dots, N\}$, on pose $\pi^x(i) = n_i^x$.

Ceci définit pour tout $x \in E^{\text{micro}}$ une permutation π^x de l'ensemble $\{1, \dots, N\}$, et permet de décrire la dynamique de la suite de variables aléatoires $\{S_k, k \geq 0\}$ par :

$$S_k = S_{k-1} + \mathbf{1}_{\{\xi'_k > S_{k-1}\}} - \mathbf{1}_{\{\xi'_k \leq S_{k-1}\}} \quad \text{où} \quad \xi'_k := \pi^{X_{k-1}}(\xi_k), \quad k \geq 1. \quad (5.4)$$

Remarquons que, conditionnellement à X_{k-1} , la v.a. ξ'_k hérite de ξ_k la distribution uniforme sur E^{macro} , du fait de la symétrie par rotation de cette loi. De plus, pour tous $x \in E^{\text{micro}}$ et $y \in \{1, \dots, N\}$:

$$\begin{aligned} \mathbb{P}[X_{k-1} = x, \xi'_k = y] &= \mathbb{P}[X_{k-1} = x, \pi^x(\xi_k) = y] \\ &= \mathbb{P}[X_{k-1} = x] \mathbb{P}[\pi^x(\xi_k) = y] \\ &= \mathbb{P}[X_{k-1} = x] \mathbb{P}[\xi_k = y] \\ &= \mathbb{P}[X_{k-1} = x] \mathbb{P}[\xi'_k = y], \end{aligned}$$

prouvant que les v.a. X_{k-1} et ξ'_k sont indépendantes. Il s'en suit que la famille $\{\xi'_k, k \geq 1\}$ forme une suite de variables iid selon la loi uniforme sur E^{macro} . D'après la représentation (5.4), on déduit que $\{S_k, k \geq 0\}$ est une chaîne de Markov de matrice de transition :

$$P^{\text{macro}}(s, s+1) = \frac{N-s}{N} \quad \text{et} \quad P^{\text{macro}}(s, s-1) = \frac{s}{N} \quad \text{pour tout } 0 \leq s \leq N.$$

Exemple 5.8. (*Un modèle de dynamique des stocks*) Un industriel fabrique à chaque date $n \geq 0$ une quantité q d'un certain bien de consommation et fait face à une demande aléatoire ξ_n . On suppose que la suite de v.a. $(\xi_n)_{n \geq 1}$ est iid selon une loi à valeurs entières, et on note $p_i := \mathbb{P}[\xi_1 = i]$ pour tout $i \in \mathbb{N}$.

L'industriel répond à la demande autant que possible en utilisant sa production et ses stocks.

Si on note par X_n le niveau des stocks à la date n , on obtient la dynamique suivante :

$$X_n = (X_{n-1} + q - D_n)^+ \quad \text{pour tout } n \geq 1.$$

la suite de v.a. $(X_n)_{n \geq 0}$ forme ainsi une chaîne de Markov homogène sur $E = \mathbb{N}$, de matrice de transition

$$P(x, 0) = \mathbb{P}[D_n \geq x + q] = \sum_{i \geq x+q} p_i \quad \text{pour tout } x \in \mathbb{N},$$

$$P(x, y) = \mathbb{P}[x + q - D_n = y] = p_{q+x-y} \quad \text{pour tous } x, y \in \mathbb{N}, y \geq 1.$$

Exemple 5.9. (*Ruine du joueur*) Deux joueurs A et B , dotés initialement de montants $a, b \in \mathbb{N}$ respectivement, jouent à Pile ou Face. A chaque étape du jeu, chaque joueur effectue une mise unitaire. Le joueur B perd sa mise (en faveur du joueur A) avec probabilité p , et gagne avec probabilité $1 - p$. Les différentes étapes du jeu sont indépendantes, et le jeu continue jusqu'à ce que l'un des joueurs soit ruiné, i.e. l'un des joueurs se retrouve avec le gain $a + b$, et l'autre avec 0.

Ce jeu se modélise en introduisant une suite de variables iid $(\xi_n)_{n \geq 1}$ telles que $(1 + \xi_n)/2$ est distribuée selon une loi de Bernoulli. Pour chaque $n \geq 1$, $\xi_n = 1$ indique que le joueur A gagne l'étape n , et $\xi_n = -1$ indique que A perd l'étape n . La fortune du joueur A à chaque étape est donc donnée par :

$$X_0 = a \quad \text{et} \quad X_n = X_{n-1} + \xi_n \mathbf{1}_{\{0 < X_{n-1} < a+b\}}, \quad n \geq 1.$$

Ainsi, $(X_n)_{n \geq 1}$ est une chaîne de Markov homogène sur l'espace d'état $E = \{0, 1, \dots, a + b\}$ de matrice de transition

$$P(0, 0) = P(a + b, a + b) = 1,$$

et

$$P(n, n + 1) = 1 - P(n, n - 1) = p \quad \text{pour } 0 < n < a + b.$$

Si $p = 1/2$ on dit que le jeu est *équilibré*, et sinon qu'il est *biaisé*. On peut aussi considérer le cas limite $b = \infty$ où le joueur A ne s'arrête que quand sa fortune initiale est épuisée.

Rearquons enfin, que cet exemple rentre partiellement dans l'exemple 5.6 puisqu'on peut voir $(X_n)_{n \geq 1}$ come une marche aléatoire arrêtée.

Exemple 5.10. (*Processus de branchement*) Ces modèles ont été introduits pour modéliser les générations successives d'une population. Les individus de la population peuvent représenter les membres d'une certaine espèce (de poissons, par exemple), des virus dans un organisme, des personnes infectées par une épidémie, ou des neutrons lors d'une réaction nucléaire. Une introduction plus

détaillée sera donnée dans le chapitre 10, on se borne ici à décrire la modélisation en termes de chaîne de Markov.

On désigne par X_n la taille de la population à la date n . On note par ξ_i^n la variable aléatoire à valeurs entières représentant le nombre de descendants de chaque individu $i \leq n$ à la date n . La dynamique de la taille de la population est ainsi décrite par :

$$X_{n+1} = \sum_{i=1}^{X_n} \xi_i^n \quad \text{pour } n \geq 0,$$

où la somme est nulle sur $\{X_n = 0\}$. On suppose de plus que toutes les variables aléatoires $\{\xi_i^n, 1 \leq i \leq n, n \geq 0\}$ sont indépendantes entre elles et de même loi de probabilité. Sous cette hypothèse, $(X_n)_{n \geq 0}$ est une chaîne de Markov homogène sur l'espace d'état $E = \mathbb{N}$, de matrice de transition

$$P(0,0) = 1 \text{ et } P(x,y) = \mathbb{P} \left[\sum_{i=1}^x \xi_i^0 = y \right] \quad \text{pour tout } x \geq 1.$$

On voit ainsi que la matrice de transition est peu commode à utiliser dans le cadre de cet exemple. En fait, l'analyse de cet exemple passe par l'utilisation de la fonction génératrice :

$$G_n(x) := \mathbb{E} [x^{X_n}] = \mathbb{E} \left[\prod_{i=1}^{X_{n-1}} x^{\xi_i^n} \right] = \mathbb{E} [G_\xi(x)^{X_{n-1}}]$$

où $G_\xi(x) := \mathbb{E} [x^{\xi_1^0}]$ est la fonction génératrice commune aux ξ_i^n . On déduit la relation de récurrence :

$$G_n(x) = G_{n-1} \circ G_\xi(x) \quad \text{pour tout } n \geq 1,$$

qui caractérise ces fonctions génératrices pour toute donnée initiale G_0 .

Par exemple, si $Z_0 = 1$, alors $G_0(x) = x$, et la récurrence précédente implique que $G_n(x) = G_\xi \circ \dots \circ G_\xi$ n fois. On peut aussi remarquer que si $\mathbb{E}[\xi] < \infty$, on obtient par récurrence que $\mathbb{E}[X_n] = G'_n(1)$.

Exemple 5.11. (*Recherche par automate de comptage*) Etant donnée une suite de caractères issus d'un certain *alphabet*, il s'agit d'analyser les apparitions successives d'une certaine chaîne de caractères (ou *mot*). Par exemple, l'alphabet peut être

- $\{0, 1\}$ dans un système informatique,
- $\{A, C, G, T\}$ dans un brin d'ADN; les nucléotides Adénine (A), Cytosine (C), Guanine (G), et Thymine (T) apparaissent sur les parties codantes de l'ADN.

Supposons que l'on cherche à détecter les occurrences du mot GAG sans chevauchement (par exemple, le suite GAGAG contient une seule occurrence sans chevauchement). On définit alors l'espace d'état $E = \{\emptyset, G, GA, GAG\}$, et on parcourt la chaîne de caractère. L'état de l'automate X_n à chaque étape est donné par :

- $X_0 = \emptyset$.
- Si $X_{n-1} = \emptyset$, alors $X_n = G$ si le $n^{\text{ième}}$ caractère est G , et $X_n = \emptyset$ sinon.
- Si $X_{n-1} = G$, alors $X_n = GA$ si le $n^{\text{ième}}$ caractère est A , $X_n = G$ si le $n^{\text{ième}}$ caractère est G , et $X_n = \emptyset$ sinon.
- Si $X_{n-1} = GA$, alors $X_n = GAG$ si le $n^{\text{ième}}$ caractère est G , et $X_n = \emptyset$ sinon.
- Si $X_{n-1} = GAG$, alors $X_n = G$ si le $n^{\text{ième}}$ caractère est G , et $X_n = \emptyset$ sinon.

En supposant que les caractères A et G sont tirés de manière indépendante selon les probabilités communes p_A, p_G , on voit que $(X_n)_{n \geq 0}$ forme une chaîne de Markov homogène sur l'espace d'état E , de matrice de transition

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

Exemple 5.12. (*Processus de renouvellement*) Soit $(D_i)_{i \geq 1}$ une suite de v.a. iid à valeurs dans $\mathbb{N} \setminus \{0\}$. Pour tout $i \geq 1$, D_i représente la durée de vie du composant i que l'on installe dès que le composant $i - 1$ tombe en panne. On s'intéresse à l'âge X_n du composant en place à la date n :

$$X_n = n - \theta_{n-1} \text{ où } \theta_{n-1} := \sup \{D_1 + \dots + D_k : k \geq 1 \text{ et } D_1 + \dots + D_k \leq n\}$$

est la date d'installation du composant en place. On pose $\theta_0 = 0$. Comme les durées de vie des différents composants sont iid, la suite $(X_n)_{n \geq 1}$ forme une chaîne de Markov homogène. On calcule immédiatement les probabilités de transition correspondantes :

$$\begin{aligned} P(x, x+1) &= \mathbb{P}[D_1 > x+1 | D_1 > x] = \frac{\mathbb{P}[D_1 > x+1]}{\mathbb{P}[D_1 > x]}, \\ P(x, 0) &= 1 - P(x, x+1), \text{ pour } x \geq 1, \end{aligned}$$

et $P(0, 0) = 1 - P(0, 1) = 0$ puisqu'on a supposé que $\mathbb{P}[D_1 = 0] = 0$.

Exemple 5.13. (*File d'attente à un serveur*) Pour modéliser une file d'attente, on introduit le nombre d'arrivants ξ_n (à valeurs entières) à chaque date n , et on suppose qu'une tâche est servie à chaque date. Si on note par X_n le nombre de tâches dans la file d'attente, on a la dynamique :

$$X_n = (X_{n-1} - 1)^+ + \xi_n, \text{ pour tout } n \geq 1.$$

On suppose que les v.a. $(\xi_n)_{n \geq 1}$ sont iid. Alors $(X_n)_{n \geq 0}$ est une chaîne de Markov homogène sur l'espace d'état $E = \mathbb{N}$, de matrice de transition

$$P(x, y) = \mathbb{P}[\xi_1 = y - (x - 1)^+] \text{ pour tous } x, y \in E.$$

5.3 Problème de Dirichlet et chaînes de Markov homogènes

5.3.1 Problème de Dirichlet

Etant donnée une matrice de transition P sur E et un sous-ensemble non vide $A \subset E$, le problème de Dirichlet consiste à trouver une fonction $u : E \rightarrow \mathbb{R}$ vérifiant

$$u = f \text{ sur } A \text{ et } (I - P)u = 0 \text{ sur } E \setminus A, \quad (5.5)$$

où la fonction u est ici représentée par le vecteur colonne $(u(x))_{x \in E}$ de taille $\text{card}(E)$. On désignera par (P, f) les caractéristiques du problème de Dirichlet (5.5).

Exemple 5.14. *Considérons la marche aléatoire X sur \mathbb{Z} définie à partir d'une suite de variables iid $(\xi_n)_{n \geq 1}$ par $X_n = X_{n-1} + \xi_n$, avec $\mathbb{P}[\xi_n = -1] = \mathbb{P}[\xi_n = 1] = \frac{1}{2}$. Ceci définit clairement une chaîne de Markov de matrice de transition*

$$P(i, j) = \frac{1}{2} \mathbf{1}_{\{|i-j|=1\}} \text{ pour tous } i, j \in \mathbb{Z}.$$

Alors l'équation $(I - P)u = 0$ se traduit par :

$$-\frac{1}{2} [u(x+1) - 2u(x) + u(x-1)] = 0 \text{ pour tout } x \in \mathbb{Z}.$$

Il s'agit de la discrétisation (par différences finies) de l'équation de la chaleur stationnaire. La condition au bord $u = f$ sur A s'interprète dans ce cadre comme une source de chaleur maintenue sur le bord de A .

Notons que le problème de Dirichlet est purement déterministe. En effet la formulation (5.5) n'évoque aucun objet aléatoire, il s'agit de trouver une fonction $u(x)$ déterministe qui en est la solution. Le cadre probabiliste qui sera introduit servira pour obtenir une représentation de la solution grâce au caractère stochastique de la matrice P .

5.3.2 Temps d'atteinte d'un ensemble par une chaîne de Markov

Le but de ce paragraphe est d'exhiber la solution (unique dans une certaine classe) du problème de Dirichlet (5.5). Pour cela, on introduit une chaîne de Markov homogène $(X_n)_{n \geq 0}$ dont la matrice de transition P est celle qui apparaît dans la formulation du problème (5.5). L'existence d'une telle chaîne de Markov, pour toute donnée initiale X_0 , est assurée par la proposition 5.2.

Le premier temps d'atteinte par la chaîne du sous-ensemble A qui apparaît dans la formulation du problème de Dirichlet (5.5) est défini par :

$$T_A := \inf \{n \geq 0 : X_n \in A\}. \quad (5.6)$$

Le temps d'atteinte T_A est un temps d'arrêt. Si $A = \{x\}$, on notera simplement T_x .

Remarquons que, étant donnée une condition initiale $X_0 = x$, il n'y a *a priori* aucune raison pour que le temps d'atteinte T_A soit fini \mathbb{P}_x -p.s. ou même avec probabilité positive. En effet, la structure très générale de la matrice de transition P peut être telle que les états de l'ensemble A soient inaccessibles pour la chaîne. Ceci est par exemple le cas si $P(x, y) = 0$ pour tout $x \in E$ et $y \in A$. Nous reviendrons à cette question dans le paragraphe suivant.

5.3.3 Représentation stochastique de la solution du problème de Dirichlet

Nous commençons par un résultat de minoration des solutions potentielles du problème de Dirichlet par la fonction

$$U^A f(x) := \mathbb{E}_x [f(X_{T_A}) \mathbf{1}_{T_A < \infty}], \quad x \in E, \quad (5.7)$$

qui est bien définie pourvu que f vérifie la condition d'intégrabilité requise. Si f est bornée inférieurement, $U^A f$ est bien définie comme une quantité dans $\mathbb{R} \cup \{\infty\}$.

Lemme 5.15. *Supposons que la fonction $f : E \rightarrow \mathbb{R}$ est bornée inférieurement, et soit $u : E \rightarrow \mathbb{R}$ une sursolution positive du problème de Dirichlet (5.5), i.e. $u \geq f$ sur A , et $u \geq Pu$ sur $E \setminus A$. Alors $u \geq U^A f$.*

Démonstration. Par définition, on a $u \geq f$ sur A et

$$\begin{aligned} u(x) \geq Pu(x) &= \sum_{x_1 \in A} P(x, x_1)u(x_1) + \sum_{x_1 \in E \setminus A} P(x, x_1)u(x_1) \\ &\geq \sum_{x_1 \in A} P(x, x_1)f(x_1) + \sum_{x_1 \in E \setminus A} P(x, x_1)u(x_1). \end{aligned}$$

En intégrant cet argument et en remarquant que :

$$\sum_{x_1, \dots, x_{k-1} \in E \setminus A, x_k \in A} P(x, x_1) \dots P(x_{k-1}, x_k) f(x_k) = \mathbb{E}_x [f(X_{T_A}) \mathbf{1}_{\{T_A = k\}}],$$

on obtient pour tout $k \geq 1$:

$$\begin{aligned} u(x) &\geq \sum_{n=0}^k \mathbb{E}_x [f(X_{T_A}) \mathbf{1}_{\{T_A = n\}}] + \sum_{x_1, \dots, x_k \in E \setminus A} P(x, x_1) \dots P(x_{k-1}, x_k) u(x_k) \\ &\geq \sum_{n=0}^k \mathbb{E}_x [f(X_{T_A}) \mathbf{1}_{\{T_A = n\}}], \end{aligned}$$

grâce à la positivité de u . Le théorème de la convergence monotone permet alors de conclure que

$$u(x) \geq \mathbb{E}_x \left[\sum_{n \geq 0} f(X_{T_A}) \mathbf{1}_{\{T_A = n\}} \right] = U^A f(x).$$

◇

Théorème 5.16. *Pour toute fonction $f : E \rightarrow \mathbb{R}$ bornée inférieurement, la fonction $U^A f$ est solution du problème de Dirichlet (5.5).*

Si de plus $f : E \rightarrow \mathbb{R}$ est bornée et que $\inf_{x \in E} \mathbb{P}_x[T_A < \infty] > 0$, alors la fonction $U^A f$ est l'unique solution bornée du problème de Dirichlet (5.5) et $\mathbb{P}_x[T_A < \infty] = 1$ pour tout $x \in A$.

Démonstration. Nous commencerons par montrer que $U^A f$ est solution du problème de Dirichlet, puis nous établirons le résultat d'unicité.

1. Pour $X_0 = x \in A$, on a $T_A = 0$ et $U^A f(x) = f(x)$ par définition de la fonction $U^A f$. Pour $x \notin A$, on a $T_A \geq 1$ et il découle de la propriété des projections itérées de l'espérance conditionnelle que :

$$\begin{aligned} U^A f(x) &= \mathbb{E}_x \{ \mathbb{E} [f(X_{T_A}) \mathbf{1}_{\{T_A < \infty\}} | X_1] \} \\ &= \sum_{y \in E} P(x, y) \mathbb{E}_y [f(X_{T_A}) \mathbf{1}_{\{T_A < \infty\}}] \\ &= \sum_{y \in E} P(x, y) U^A f(y) = P U^A f(x). \end{aligned}$$

2. Soit u une deuxième solution du problème de Dirichlet (5.5). Quitte à rajouter une constante à u (et à f), on peut supposer sans perte de généralité que $u \geq 0$. Alors, la fonction

$$w_\varepsilon := U^A \mathbf{1} - \varepsilon(u - U^A f)$$

est solution du problème de Dirichlet $(P, 1)$. Du lemme 5.15 appliqué au problème de Dirichlet (P, f) , on déduit que $u - U^A f \geq 0$. De plus, comme u et U^A sont bornées, et $\inf_{x \in E} U^A \mathbf{1}(x) = \inf_{x \in E} \mathbb{P}_x[T_A < \infty] > 0$, on peut choisir $\varepsilon > 0$ suffisamment petit pour que

$$0 \leq \frac{1}{2} U^A \mathbf{1} \leq w_\varepsilon \leq U^A \mathbf{1}.$$

Ainsi w_ε est une sursolution positive bornée du problème de Dirichlet $(P, 1)$. D'après le lemme 5.15, on peut alors conclure que $w_\varepsilon \geq U^A \mathbf{1}$, et par suite $w_\varepsilon = U^A \mathbf{1}$ et $u = U^A f$.

3. En remarquant que $\mathbb{P}_x[T_A < \infty] = U^A \mathbf{1}(x)$ et que la fonction constante 1 est solution du problème de Dirichlet $(P, 1)$, on déduit du résultat d'unicité que $U^A \mathbf{1} = 1$. ◇

Remarque 5.17. Dans la démonstration précédente, la bornitude n'est utilisée que pour prouver l'unicité. Ainsi, en combinant l'étape 1 de la preuve du théorème 5.16 avec le lemme 5.15, on obtient que pour $f \geq 0$, la fonction $U^A f$ est la plus petite sursolution positive du problème de Dirichlet (5.5).

Remarque 5.18. Dans le cas d'un espace d'état E fini, le théorème 5.16 montre en particulier que si $\mathbb{P}_x[T_A < \infty] > 0$ pour tout $x \in E$, alors $\mathbb{P}_x[T_A < \infty] = 1$ pour tout $x \in E$. Cette propriété dite de *réurrence* sera étudiée plus systématiquement dans le chapitre suivant.

Cette observation s'applique à l'exemple 5.11 d'automate de comptage, où l'objet d'intérêt est le temps T_{GAG} nécessaire pour obtenir pour la première fois le mot GAG. Dans cet exemple, on vérifie immédiatement que $\mathbb{P}_x[T_{GAG}] > 0$ pour tout $x \in E$, et on voit alors que $T_{GAG} < \infty$ \mathbb{P}_x -p.s. pour tout $x \in E$.

L'exercice suivant montre comment obtenir beaucoup plus d'informations sur le temps d'arrêt T_{GAG} .

Exercice 5.19. 1. Pour une v.a. Z à valeurs entières, montrer que sa fonction génératrice vérifie :

$$\mathbb{E}[t^Z] = 1 + (t-1) \sum_{n \geq 0} \mathbb{P}[Z > n] t^n \quad \text{pour tout } t \in [0, 1].$$

2. Dans le cadre de l'exemple 5.11, on note $T := T_{GAG}$.

(a) Montrer que

$$\mathbb{P}[T > n, \xi_{n+1} = G, \xi_{n+2} = A, \xi_{n+3} = G] = \mathbb{P}[T > n] p_{AP_G}^2.$$

(b) Montrer que

$$\begin{aligned} \mathbb{P}[T > n, \xi_{n+1} = G, \xi_{n+2} = A, \xi_{n+3} = G] &= \mathbb{P}[T = n+1, \xi_{n+2} = A, \xi_{n+3} = G] \\ &\quad + \mathbb{P}[T = n+3], \end{aligned}$$

et en déduire que

$$\mathbb{P}[T > n] = \frac{\mathbb{P}[T = n+1]}{p_G} + \frac{\mathbb{P}[T = n+3]}{p_{AP_G}^2}.$$

(c) En utilisant 1, déterminer explicitement la fonction génératrice $\mathbb{E}[t^T]$ de la v.a. T .

(d) Calculer $\mathbb{E}[T]$ et $\mathbb{V}[T]$.

5.3.4 Application au problème de ruine du joueur

Ce problème, introduit dans l'exemple 5.9, décrit un jeu de Pile ou Face entre deux joueurs A et B , dotés initialement des montants $a, b \in \mathbb{N}$ respectivement. On note $N := a + b$.

A chaque étape du jeu, chaque joueur effectue une mise unitaire. Le joueur B perd sa mise (en faveur du joueur A) avec probabilité p , et gagne avec probabilité $1-p$. Les différentes étapes du jeu sont indépendantes, et le jeu continue jusqu'à ce que l'un des joueurs soit ruiné, i.e. l'un des joueurs se retrouve avec le gain $a+b$, et l'autre avec 0. La fortune du joueur A à chaque étape est décrite par une chaîne de Markov homogène $(X_n)_{n \geq 1}$ sur l'espace d'état $E = \{0, 1, \dots, N\}$ de dynamique :

$$X_0 = a \quad \text{et} \quad X_n = X_{n-1} + \xi_n \mathbf{1}_{\{0 < X_{n-1} < N\}}, \quad n \geq 1,$$

où $(\xi_n)_{n \geq 1}$ est une suite de variables iid avec $\mathbb{P}[\xi_n = 1] = 1 - \mathbb{P}[\xi_n = -1] = p$. Le jeu s'arrête dès que l'un des deux joueurs est ruiné, c'est à dire à la date aléatoire décrite par le temps d'arrêt :

$$T_{\{0, N\}} = \inf \{n \geq 0 : X_n = 0 \text{ ou } X_n = N\} = T_0 \wedge T_N,$$

Afin de savoir lequel des deux joueurs est ruiné à la fin du jeu, il faut pouvoir classer T_0 et T_N . On introduit alors la probabilité de ruine de chacun des joueurs :

$$V_N^A(x) := \mathbb{P}_x [T_0 < T_N] \quad \text{et} \quad V_N^B(x) := \mathbb{P}_x [T_N < T_0].$$

La probabilité que le jeu soit de durée finie se déduit de ces deux fonctions :

$$\mathbb{P}_x [T_0 \wedge T_N < \infty] = V_N^A(x) + V_N^B(x). \quad (5.8)$$

Avec la notation (5.7), remarquons que $V_N^A(x) = U^{\{0, N\}} \delta_0(x)$ et $V_N^B(x) = U^{\{0, N\}} \delta_N(x)$, où $\delta_i(x) = \mathbf{1}_{\{i\}}(x)$ est la masse de Dirac au point i . D'après le théorème 5.16, V_N^A et V_N^B sont solutions de l'équation :

$$-(1-p)u(x-1) + u(x) - pu(x+1) = 0, \quad 1 \leq x \leq N-1, \quad (5.9)$$

avec les conditions au bord

$$V_N^A(0) = 1, \quad V_N^A(N) = 0 \quad \text{et} \quad V_N^B(0) = N, \quad V_N^B(N) = 1. \quad (5.10)$$

L'équation (5.9) est une récurrence linéaire du second ordre dont la solution générale s'écrit en fonction des racines du polynôme caractéristique correspondant. Un calcul direct montre que ces racines sont 1 et $(1-p)/p$.

— *Jeu biaisé* Si $p \neq \frac{1}{2}$, les racines sont distinctes, et la solution générale de l'équation (5.9) est une combinaison linéaire de la fonction constante et de la fonction $x \mapsto ((1-p)/p)^x$. Les conditions au bord (5.10) donnent alors comme solution unique :

$$V_N^A(x) = \frac{\left(\frac{1-p}{p}\right)^N - \left(\frac{1-p}{p}\right)^x}{\left(\frac{1-p}{p}\right)^N - 1} \quad \text{et} \quad V_N^B(x) = \frac{\left(\frac{1-p}{p}\right)^x - 1}{\left(\frac{1-p}{p}\right)^N - 1}.$$

- *Jeu équilibré* Dans le cas $p = \frac{1}{2}$, le polynôme caractéristique admet 1 comme racine double, et l'équation (5.9) a comme solution générale l'ensemble des fonctions affines. Les conditions au bord (5.10) donnent alors :

$$V_N^A(x) = \frac{N-x}{N} \quad \text{et} \quad V_N^B(x) = \frac{x}{N},$$

ce qui correspond aux limites des expressions obtenues dans le cas biaisé. Ayant déterminé les fonctions V_N^A et V_N^B , on peut maintenant déduire de (5.8) la probabilité que le jeu soit de durée finie :

$$\mathbb{P}_x [T_0 \wedge T_N < \infty] = 1.$$

Enfin, on peut examiner le cas d'un joueur compulsif qui ne s'arrête de jouer que quand il est ruiné. Ceci correspond au cas limite $N \rightarrow \infty$. En utilisant le théorème de convergence monotone, on obtient :

$$V^A(x) = \mathbb{P}_x [T_0 < \infty] = \lim_{N \rightarrow \infty} V_N^A(x) = \begin{cases} \left(\frac{1-p}{p}\right)^x & \text{si } p > \frac{1}{2}, \\ 1 & \text{si } p \leq \frac{1}{2}. \end{cases} \quad (5.11)$$

Ainsi, le temps d'atteinte unilatéral de zéro d'une marche aléatoire sur \mathbb{N} au plus proche voisin, partant d'un point strictement positif, est fini p.s. si et seulement si $p \leq \frac{1}{2}$.

5.3.5 Monte Carlo pour le problème de Dirichlet

Revenons au problème de Dirichlet du paragraphe 5.3.1, et supposons maintenant que le temps moyen d'atteinte de l'ensemble A est fini.

$$\mathbb{E}_x [T_A] < \infty \quad \text{pour tout } x \in E. \quad (5.12)$$

En particulier, ceci implique que $T_A < \infty$ \mathbb{P}_x -p.s. pour tout $x \in E$. Le théorème de représentation 5.16 montre que pour toute fonction bornée f , décrivant la fonction u sur A , et pour toute matrice stochastique P , le problème de Dirichlet (5.5) admet une unique solution bornée donnée par

$$u(x) := U^A f(x) = \mathbb{E}_x [f(X_{T_A})]. \quad (5.13)$$

Nous nous intéressons maintenant au calcul numérique de cette solution. Une première façon de procéder est de résoudre le système (5.5) de manière algébrique, dans un cadre purement déterministe. Cette méthode se heurte rapidement à la malédiction de la dimension (*curse of dimensionality*), si E est *trop grand*. Par exemple si $E = \mathbb{Z}^d$, les méthodes algébriques sont très vite limitées même pour des dimensions assez petites.

La représentation (5.13) ouvre la porte à l'utilisation des méthodes de Monte Carlo pour l'approximation de la solution. En effet, si on disposait d'une méthode

pour simuler la chaîne de Markov, i.e. pour produire de un *échantillon* de S copies indépendantes $\{X_n^s, n \geq 0\}_{1 \leq s \leq S}$ à partir de l'état initial $X_0^s = x$, alors on pourra déduire les temps d'atteinte correspondants de l'ensemble A :

$$T_A^s := \inf \{n \geq 0 : X_n^s \in A\},$$

et on proposerait de manière naturelle l'approximation suivante de $u(x)$:

$$\hat{u}_S(x) := \frac{1}{S} \sum_{s=1}^S f(X_{T_A^s}^s). \quad (5.14)$$

Par construction, les v.a. $(f(X_{T_A^s}^s))_{1 \leq s \leq S}$ sont iid. On obtient alors par la loi forte des grand nombres la convergence

$$\hat{u}_S(x) \longrightarrow u(x) \quad \mathbb{P}_x - \text{p.s. quand } S \rightarrow \infty. \quad (5.15)$$

On peut même donner un résultat asymptotique sur l'estimation de l'erreur grâce au théorème central limite :

$$\sqrt{S}(\hat{u}_S(x) - u(x)) \longrightarrow \mathcal{N}\left(0, \mathbb{V}\left[f(X_{T_A^s}^s)\right]\right) \quad \text{en loi, quand } S \rightarrow \infty.$$

Cette estimation d'erreur n'est valable qu'asymptotiquement. Malgré cela, il est communément accepté de l'utiliser comme un bon indicateur de l'erreur d'approximation. Notons que l'on peut obtenir des bornes à distance finie (i.e. $S < \infty$) sur l'erreur en utilisant l'inégalité de Chebyshev, ou les inégalités de Markov exponentielles.

Enfin, sous la condition (5.12), on déduit de la loi des grands nombres que la durée totale des simulations pour pouvoir calculer l'estimateur (5.15) vérifie :

$$\sum_{s=1}^S T_A^s = S\mathbb{E}_x[T_A] + o(S), \quad \mathbb{P}_x - \text{p.s. quand } S \rightarrow \infty.$$

Remarque 5.20. Dans plusieurs application, on est amenés à résoudre des problèmes d'équations aux dérivées partielles dans \mathbb{R}^n du type

$$b(x) \cdot \frac{\partial u}{\partial x}(x) + \frac{1}{2} \text{Tr} \left[\sigma(x) \sigma(x)^T \frac{\partial^2 u}{\partial x \partial x^T}(x) \right] = 0, \quad x \in \mathcal{O}^c$$

$$u = f \text{ sur } \mathcal{O},$$

où \mathcal{O} est un ouvert de \mathbb{R}^n . Il s'agit d'une version générale de l'équation de la chaleur stationnaire, qu'on appelle aussi problème de Dirichlet. On peut montrer dans ce cadre un théorème de représentation similaire au théorème 5.16 en introduisant la notion de mouvement brownien et d'équations différentielles stochastiques.

En vue de l'approximation numérique de la solution de ce problème (si l'existence et l'unicité sont assurées), il est possible de mettre en place des

discrétisations par les différences finies ou les éléments finis qui conduisent à un problème de Dirichlet de la classe (5.5). On peut alors prouver la convergence de ces méthodes numériques en prouvant la convergence de la chaîne de Markov associée vers un processus stochastique continu. Ce point de vue est largement développé dans le livre de Kushner et Dupuis [10].

5.4 Temps de retours et excursions

Pour tout état $x \in E$, nous introduisons les temps aléatoires :

$$R_1^x := \inf \{k \geq 1 : X_k = x\} \quad \text{et} \quad R_n^x := \inf \{k > R_{n-1}^x : X_k = x\} \quad \text{pour } n \geq 1.$$

R_n^x représente le temps de $n^{\text{ème}}$ retour à l'état x . On pose par convention $R_0^x = 0$. On vérifie immédiatement que R_n^x est un temps d'arrêt pour chaque $n \geq 0$, et que :

$$T_x = R_0^x \mathbf{1}_{\{X_0=x\}} + R_1^x \mathbf{1}_{\{X_0 \neq x\}} \quad \text{pour tout } x \in E.$$

La trajectoire de la chaîne entre deux temps de retours successifs

$$\left\{ X_{R_{n-1}^x}, X_{R_{n-1}^x+1}, \dots, X_{R_n^x} \right\}$$

est appelée $n^{\text{ème}}$ excursion de la chaîne à l'état x . Le résultat suivant est très intuitif, et sera à l'origine des comportements asymptotiques des moyennes temporelles qui seront développés ultérieurement. Il s'agira des résultats ergodiques très importants pour les méthodes de Monte Carlo.

Pour l'énoncé de ce résultat, nous supposons que les temps de retour $R_n^x < \infty$ sont finis p.s. pour tout $n \geq 0$. Nous verrons dans le chapitre suivant qu'il existe des caractérisations simples de cette hypothèse en terme de *réurrence* de la chaîne.

Proposition 5.21. *Soit X une chaîne de Markov avec des temps de retours finis p.s. Pour tout $x \in E$. Alors, les variables aléatoires*

$$Y_n := \left(R_n^x - R_{n-1}^x, X_{R_{n-1}^x+1}, \dots, X_{R_n^x} \right), \quad n \geq 1,$$

sont indépendantes. De plus $\{Y_n, n \geq 2\}$ sont identiquement distribués suivant la loi de R_1^x sous \mathbb{P}_x .

Démonstration. Pour les deux premiers temps de retours, on écrit que pour des entiers $k, \ell \geq 1$ et $(x_1, \dots, x_{k+\ell}) \in E^{k+\ell}$ avec $x_k = x_{k+\ell} = x$ et $x_i \neq x$ pour tous les $i \notin \{k, \ell\}$:

$$\begin{aligned} & \mathbb{P}[Y_1 = (k, x_1, \dots, x_k), Y_2 = (\ell, x_{k+1}, \dots, x_{k+\ell})] \\ &= \mathbb{P}[X_i = x_i, 1 \leq i \leq k] \mathbb{P}[X_j = x_i, 1 \leq j - k \leq \ell | X_i = x_i, 1 \leq i \leq k] \\ &= \mathbb{P}[X_i = x_i, 1 \leq i \leq k] \mathbb{P}[X_j = x_i, 1 \leq j - k \leq \ell | X_k = x] \\ &= \mathbb{P}[Y_1 = (k, x_1, \dots, x_k)] \mathbb{P}_x[Y_2 = (\ell, x_{k+1}, \dots, k + \ell)]. \end{aligned}$$

L'extension à un nombre de temps de retours quelconque est immédiate, et l'identité des lois est une conséquence du caractère homogène de la chaîne de Markov. \diamond

Chapitre 6

Lois invariantes et classification des états

Dans ce paragraphe, nous considérons une chaîne de Markov homogène X de matrice de transition P . En analysant la structure de la matrice P , nous allons obtenir une classification des états qui aura une importance cruciale sur le comportement asymptotique de la chaîne. Nous nous intéressons en particulier à des situations où le comportement asymptotique de la chaîne est régi par une certaine stabilité que l'on définira précisément.

6.1 Loi invariante

Définition 6.1. Une probabilité ν sur E est représentée par un vecteur ligne $(\nu(x))_{x \in E}$. On dit que ν est une probabilité invariante pour la chaîne de Markov X si $\nu P = \nu$.

Ainsi, si ν est une probabilité invariante pour la chaîne X , et si à une certaine date n la loi marginale $\pi_n = \nu$, alors $\pi_k = \nu$ pour tout $k \geq n$, puisque $\pi_{k+1} = \pi_k P$. On dit alors que la chaîne est à l'équilibre ou encore, que la chaîne a atteint son état stationnaire. En termes d'algèbre linéaire, une probabilité invariante est un vecteur propre à gauche de la matrice P (ou un vecteur propre à droite de la matrice P^T) de composantes positives, et convenablement normalisé. La normalisation d'un vecteur propre positif, de façon à en faire une mesure de probabilité, est toujours possible en espace d'état fini. Si l'espace d'état n'est pas fini, ce problème soulève la question d'intégrabilité du vecteur propre.

L'exemple suivant montre qu'une chaîne de Markov n'admet pas toujours des probabilités invariantes.

Exemple 6.2. Considérons la marche aléatoire symétrique au plus proche voisin X sur \mathbb{Z} définie à partir d'une suite de variables iid $(\xi_n)_{n \geq 1}$:

$$X_n = X_{n-1} + \xi_n, \quad \text{avec} \quad \mathbb{P}[\xi_n = -1] = \mathbb{P}[\xi_n = 1] = \frac{1}{2}.$$

Ceci définit clairement une chaîne de Markov de matrice de transition P définie par

$$P(i, j) = \frac{1}{2} \mathbf{1}_{\{|i-j|=1\}} \quad \text{pour tous } i, j \in \mathbb{Z}.$$

Si X admettait une probabilité invariante ν , on devrait avoir $\nu(n) = [\nu(n-1) + \nu(n+1)]/2$ pour tout $n \in \mathbb{Z}$. Ceci implique que $\nu(n+1) - \nu(n) = \nu(n) - \nu(n-1) = \nu(1) - \nu(0)$, ce qui est incompatible avec le fait que $\sum_{n \in \mathbb{Z}} \nu(n) = 1$.

6.2 Loi invariante en espace d'état fini

Nous allons à présent montrer que si l'espace d'état E est fini, l'ensemble des probabilités invariantes est non vide. Ceci est une conséquence du théorème de Perron-Frobenius en algèbre linéaire, dont la démonstration suivante reproduit quelques éléments.

Théorème 6.3. *Soit E un espace d'état fini. Alors il existe au moins une probabilité invariante.*

Démonstration. Comme P est une matrice de taille $\text{card}(P)$ finie, toute valeur propre de P est aussi valeur propre de P^T . Comme le vecteur constant est un vecteur propre associé à la valeur propre 1, on déduit l'existence d'un vecteur ligne μ non nul tel que $\mu P = \mu$. Quitte à renormaliser μ , on peut supposer sans perte de généralité que $\sum_{x \in E} |\mu(x)| = 1$. Soit alors ν la mesure de probabilité sur E définie par

$$\nu(y) := |\mu(y)| \quad \text{pour tout } y \in E.$$

Alors, pour tout $y \in E$:

$$\sum_{x \in E} \nu(x) P(x, y) = \sum_{x \in E} |\mu(x)| P(x, y) \geq \left| \sum_{x \in E} \mu(x) P(x, y) \right| = |\mu P(y)| = |\mu(y)| \quad (6.1)$$

et

$$\sum_{y \in E} \left(\sum_{x \in E} |\mu(x)| P(x, y) - |\mu(y)| \right) = - \sum_{y \in E} |\mu(y)| + \sum_{x \in E} |\mu(x)| \sum_{y \in E} P(x, y) = 0.$$

On déduit alors qu'il y a égalité dans (6.1) pour tout $y \in E$, et par suite $\nu P = \nu$. \diamond

Remarque 6.4. Dans la démonstration précédente, l'argument qui permettait de passer d'un vecteur propre, associé à la valeur propre 1, à un vecteur propre à composantes positives, associé à la même valeur propre, n'est pas spécifique au cas d'un espace d'état fini. Nous utiliserons ainsi plus tard le résultat suivant : *si E est dénombrable et $\nu P = \nu$, alors $\mu(x) := |\nu(x)|$, $x \in E$, définit une mesure invariante.*

Exemple 6.5. (*Chaîne à deux états*) Une chaîne de Markov homogène à deux états a une matrice de transition

$$P = \begin{pmatrix} 1 - p_1 & p_1 \\ p_2 & 1 - p_2 \end{pmatrix},$$

où $0 \leq p_1, p_2 \leq 1$ sont deux constantes données. On cherche une solution $\nu \in [0, 1]^2$, $\nu_1 + \nu_2 = 1$, au problème de valeur propre $\nu = \nu P$, i.e.

$$p_1 \nu_1 - p_2 \nu_2 = 0, \quad \nu_1, \nu_2 \geq 0, \quad \nu_1 + \nu_2 = 1.$$

Si $p_1 > 0$ et $p_2 > 0$, ce système admet comme unique solution

$$\nu_1 = \frac{p_2}{p_1 + p_2} \quad \text{et} \quad \nu_2 = \frac{p_1}{p_1 + p_2}.$$

Si $p_1 = p_2 = 0$, alors $P = I_2$, les deux états sont absorbants, et toute loi est invariante.

Remarquons que la loi marginale de X_n vérifie la récurrence $\pi_n = \pi_{n-1}P$, ce qui donne

$$\pi_n(1) = (\pi_0(1) - \nu_1)(1 - p_1 - p_2)^n + \nu_1 \quad \text{où} \quad \nu_1 = \frac{p_2}{p_1 + p_2},$$

et $\pi_n(2)$ vérifie une formule symétrique. Le cas $p_1 = p_2 = 0$ est exclu dans cette formule, mais se traite facilement puisque $\pi_n = \pi_0$. Le cas $p_1 = p_2 = 1$ est aussi très particulier, puisqu'alors $\pi_{2n}(1) = \pi_0(1)$ et $\pi_{2n+1}(1) = 1 - \pi_0(1)$, et la chaîne est oscillante. Enfin, si $0 < p_1 + p_2 < 2$, alors π_n converge vers la loi invariante ν à vitesse géométrique de raison $1 - p_1 - p_2$.

Exemple 6.6. (*Chaîne à trois états*) Le cas des chaînes de Markov homogènes à trois états peut aussi être traité explicitement en exploitant le caractère stochastique de la matrice impliquant que 1 est une valeur propre. Dans le cas général, la matrice de transition est de la forme

$$P = \begin{pmatrix} 1 - p_1 - q_1 & p_1 & q_1 \\ p_2 & 1 - p_2 - q_2 & q_2 \\ p_3 & q_3 & 1 - p_3 - q_3 \end{pmatrix},$$

où $p_i, q_i \geq 0$ et $p_i + q_i \leq 1$ pour $i = 1, 2, 3$. Les valeurs propres de la matrice sont les racines du polynôme caractéristique qui est dans le cas présent un polynôme du troisième degré. Comme 1 est une valeur propre, on réduit la caractérisation du reste du spectre à la recherche des racines d'un polynôme du second degré, ce qui peut se faire de manière explicite.

Concernant la loi invariante, on peut vérifier que :

$$\begin{aligned} \nu(1) &= \frac{p_2 p_3 + p_2 q_3 + q_2 p_3}{p_1 q_2 + p_1 p_3 + p_1 q_3 + q_1 p_2 + q_1 q_2 + q_1 q_3 + p_2 p_3 + p_2 q_3 + q_2 p_3} \\ \nu(2) &= \frac{p_1 p_3 + p_1 q_3 + q_1 q_3}{p_1 q_2 + p_1 p_3 + p_1 q_3 + q_1 p_2 + q_1 q_2 + q_1 q_3 + p_2 p_3 + p_2 q_3 + q_2 p_3} \\ \nu(3) &= \frac{p_1 q_2 + q_1 p_2 + q_1 q_2}{p_1 q_2 + p_1 p_3 + p_1 q_3 + q_1 p_2 + q_1 q_2 + q_1 q_3 + p_2 p_3 + p_2 q_3 + q_2 p_3} \end{aligned}$$

est une loi invariante. Le calcul montre que, à l'exception de quelques cas particuliers, cette loi invariante est unique. Nous ne reportons pas ces calculs car nous aurons un moyen plus efficace dans la suite pour caractériser l'unicité des lois invariantes.

Exemple 6.7. (*Automate de comptage*) Reprenons l'exemple 5.11. Ici la chaîne est à quatre états. On peut quand même vérifier que

$$\pi = \left(\frac{1 - p_G - p_{APG}^2}{1 + p_{APG}}, \frac{p_G}{1 + p_{APG}}, \frac{p_{APG}}{1 + p_{APG}}, \frac{p_{APG}^2}{1 + p_{APG}} \right)$$

est l'unique loi invariante. Comme dans l'exemple précédent l'unicité sera justifiée par un résultat général.

6.3 Loi invariante pour les chaînes de Markov réversibles

Supposons que $\pi_n(x) > 0$ pour tout $x \in E$ et $n \geq 0$. Alors,

$$\begin{aligned} Q_n(x, y) &:= \mathbb{P}[X_n = y | X_{n+1} = x] \\ &= \frac{\mathbb{P}[(X_n, X_{n+1}) = (y, x)]}{\mathbb{P}[X_{n+1} = x]} = \frac{P(y, x)\pi_n(y)}{\pi_{n+1}(x)}. \end{aligned}$$

Il est immédiat de vérifier que pour tout entier $k \geq 2$, et pour tous $x_1 = x, x_2, \dots, x_k \in E$,

$$\mathbb{P}[X_n = y | X_{n+i} = x_i, 1 \leq i \leq k] = Q_n(x, y).$$

La matrice Q_n est la matrice de transition de la chaîne de Markov X après retournement du temps.

Définition 6.8. On dit que la chaîne de de Markov homogène X est réversible par rapport à une mesure de probabilité ν si :

$$\nu(x)P(x, y) = \nu(y)P(y, x) \quad \text{pour tous } x, y \in E,$$

i.e. si les lois marginales π_n sont données par ν , $Q_n = P$ pour tout n .

Proposition 6.9. Soit ν une mesure de probabilité par rapport à laquelle la chaîne de Markov est réversible. Alors ν est une probabilité invariante.

Démonstration. Par définition, on a $\nu(x)P(x, y) = \nu(y)P(y, x)$ pour tous $x, y \in E$. En sommant sur x , ceci implique que

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

◇

Le réversibilité est une condition très forte qui est parfois supposée vérifiée dans certaines applications en physique. Nous terminons ce paragraphe par un tel exemple.

Exemple 6.10. (*Urne d'Ehrenfest*) Revenons à la description macroscopique de ce problème décrit au paragraphe 5.7, qui étudie le répartition entre deux compartiments communiquants de N molécules d'un gaz où, à chaque date de changement, une molécule choisie au hasard (uniformément parmi toutes les molécules) change de compartiment.

Dans la description macroscopique, S_n désigne le nombre de particules dans le compartiment 1 à la date n . Etant donnée une suite de variables aléatoire iid $(\varepsilon_n)_{n \geq 1}$ suivant la loi uniforme sur $\{1, \dots, N\}$ ($\varepsilon_n = \xi'_n$ dans l'exemple 5.7), la dynamique aléatoire du processus $X = (X_n)_{n \geq 0}$ est donnée par :

$$S_n = S_{n-1} - \mathbf{1}_{[0, \varepsilon_n]}(S_{n-1}) + \mathbf{1}_{(\varepsilon_n, N]}(S_{n-1}) \quad \text{pour tout } n \geq 1,$$

et montre que X est une chaîne de Markov homogène de matrice de transition

$$P(x, x+1) = \frac{N-x}{N}, \quad P(x, x-1) = \frac{x}{N} \quad \text{pour } 0 \leq x \leq N,$$

et $P(x, y) = 0$ si $|x - y| > 1$.

Comme l'espace d'état est fini, nous savons que la chaîne admet une loi invariante. Pour l'identifier, nous considérons la loi binomiale ν de paramètres $(N, 1/2)$:

$$\nu(x) = 2^{-N} C_N^x, \quad 0 \leq x \leq N.$$

Un calcul direct montre que la chaîne est réversible par rapport à cette loi ν , prouvant ainsi que ν est une loi invariante.

6.4 Existence en espace d'état dénombrable

Dans ce paragraphe, nous construisons à la main des solutions positives du problème de valeur propre $\nu P = \nu$, et nous explicitons des conditions nécessaires et suffisantes pour que ces solutions soient des mesures invariantes.

Notre construction fait appel au temps d'arrêt défini par le premier retour en l'état $x \in E$:

$$R_x := R_1^x = \inf \{n \geq 1 : X_n = x\}, \quad x \in E,$$

et les mesures à valeurs dans $[0, \infty]$ définies par :

$$\mu_x(y) = \mathbb{E}_x \left[\sum_{n=0}^{R_x-1} \mathbf{1}_{\{X_n=y\}} \right] = \sum_{n \geq 0} \mathbb{P}_x [R_x > n, X_n = y], \quad x \in E. \quad (6.2)$$

La dernière égalité est une conséquence du théorème de convergence monotone. L'importance des mesures μ_x , $x \in E$, est justifiée par le résultat suivant.

Lemme 6.11. *Si ν est une probabilité invariante, alors*

$$\nu(y) \geq \nu(x)\mu_x(y) \quad \text{pour tous } x, y \in E.$$

Démonstration. Pour $y \neq x$, on déduit de l'invariance de ν que

$$\begin{aligned} \nu(y) &= \nu(x)P(x, y) + \sum_{y_1 \neq x} \nu(y_1)P(y_1, y) \\ &= \nu(x)P(x, y) + \sum_{y_1 \neq x} \left(\nu(x)P(x, y_1) + \sum_{y_2 \neq x} \nu(y_2)P(y_2, y_1) \right) P(y_1, y) \\ &= \dots \geq \nu(x) \sum_{k=0}^n \sum_{x_1, \dots, x_{k-1} \neq x} P(x, x_{k-1}) \dots P(x_1, y) \\ &= \nu(x) \sum_{k=0}^n \mathbb{P}_x [R_x > k, X_k = y] \end{aligned}$$

pour tout $n \geq 0$. En envoyant n à l'infini, on déduit de la définition de μ_x que :

$$\nu(y) \geq \nu(x) \sum_{k \geq 0} \mathbb{P}_x [R_x > k, X_k = y] = \nu(x)\mu_x(y).$$

◇

Définition 6.12 (Etats récurrents). *Un état $x \in E$ est dit récurrent si*

$$\mathbb{P}_x [R_x < \infty] = 1.$$

Sinon, on dit que x est transitoire. Pour un état récurrent x , on dit que

- *x est récurrent positif si $\mathbb{E}_x [R_x] < \infty$,*
- *x est récurrent nul si $\mathbb{E}_x [R_x] = \infty$.*

La classification des états récurrents entre états récurrents positifs et états récurrents nuls n'a de sens que si l'espace d'état de la chaîne n'est pas fini. En effet, si E est fini, tout état récurrent est récurrent positif.

Remarque 6.13. Introduisons le *nombre de visites de l'état x par la chaîne* :

$$N^x := \sum_{n \geq 0} \mathbf{1}_{\{X_n = x\}}.$$

1. D'après la propriété de Markov forte (ou plus précisément la proposition 5.21), on a pour tout entier $k \geq 1$:

$$\mathbb{P}_x [N^x > k] = \mathbb{P}_x [R_x < \infty]^k,$$

i.e. sous \mathbb{P}_x , la v.a. N^x est distribué selon une loi géométrique sur \mathbb{N}^* . On en déduit que :

$$\mathbb{P}_x [N^x = \infty] = \lim_{k \rightarrow \infty} \downarrow \mathbb{P}_x [R_x < \infty]^k \in \{0, 1\}.$$

Ainsi, l'événement $\{N^x = \infty\}$ est trivial dans le sens où il est soit négligeable soit de probabilité pleine sous \mathbb{P}_x :

- un état x est récurrent $\text{ssi } \mathbb{P}_x[N^x = \infty] = 1$,
- un état x est transitoire $\text{ssi } \mathbb{P}_x[N^x = \infty] = 0$.

2. Si la chaîne est démarrée à partir d'une distribution initiale arbitraire π_0 , on déduit de la propriété de Markov forte que $\mathbb{P}[N^x = \infty] = \mathbb{P}[R_x < \infty] \mathbb{P}_x[N^x = \infty]$, et par suite on a :

$$\mathbb{P}[N^x = \infty] = \mathbb{P}[R_x < \infty] \quad \text{pour tout état récurrent.}$$

3. Enfin, on a aussi respectivement sous \mathbb{P}_x et sous \mathbb{P} :

$$\begin{aligned} \mathbb{E}_x[N^x] &= \sum_{k \geq 0} \mathbb{P}_x[N^x > k] = \sum_{k \geq 0} \mathbb{P}_x[R_x < \infty]^k, \\ \mathbb{E}[N^x] &= \mathbb{P}[R_x < \infty] \mathbb{E}_x[N^x], \end{aligned}$$

ce qui montre que

- un état x est récurrent $\text{ssi } \mathbb{E}_x[N^x] = \infty$,
- $\mathbb{E}[N^x] < \infty$ pour tout état transitoire x .

Nous allons maintenant montrer que, pour tout $x \in E$, la fonction μ_x introduite dans (6.2) est un candidat naturel pour le problème de recherche de mesure invariante.

Proposition 6.14. *Soit $x \in E$. Alors :*

- (i) $\mu_x P = \mu_x$ si et seulement si x est un état récurrent,
- (ii) μ_x est une mesure finie si et seulement si x est récurrent positif. Dans ce cas, $\nu_x := \mu_x / \mathbb{E}_x[R_x]$ est une probabilité invariante pour la chaîne.

Démonstration. (i) Pour tout $y \in E$, on calcule que :

$$\begin{aligned} \mu_x P(y) &= \sum_{z \in E} \mu_x(z) P(z, y) \\ &= \sum_{z \in E} \sum_{n \geq 0} \mathbb{P}_x[X_n = z, R_x > n] P(z, y) \\ &= \sum_{n \geq 0} \mathbb{P}_x[R_x > n] \sum_{z \in E} \mathbb{P}_x[X_n = z | R_x > n] P(z, y), \end{aligned}$$

où l'inversion des sommations est légitime du fait de la positivité de tous les termes (ou le théorème de convergence monotone). Remarquons maintenant que l'événement $\{R_x > n\} \in \mathcal{F}_n^X$. Alors :

$$\begin{aligned} &\sum_{z \in E} \mathbb{P}_x[X_n = z | R_x > n] P(z, y) \\ &= \sum_{z \in E} \mathbb{P}_x[X_n = z | R_x > n] \mathbb{P}[X_{n+1} = y | X_n = z] \\ &= \sum_{z \in E} \mathbb{P}_x[X_n = z | R_x > n] \mathbb{P}[X_{n+1} = y | X_n = z, R_x > n] \\ &= \mathbb{P}_x[X_{n+1} = y | R_x > n]. \end{aligned}$$

On en déduit que :

$$\mu_x P(y) = \sum_{n \geq 0} \mathbb{P}_x [X_{n+1} = y, R_x > n] = \mathbb{E}_x \left[\sum_{n=1}^{R_x} \mathbf{1}_{\{X_n=y\}} \right],$$

qui serait égal à $\mu_x(y)$ si on pouvait décaler l'indice n d'un rang vers le bas. Ceci implique immédiatement que $\mu_x P(y) = \mu_x(y)$ pour $y \neq x$. Pour $y = x$, le décalage d'indice est possible si et seulement si $\mathbb{P}_x[R_x < \infty] = 1$, puisque dans ce cas $\mathbf{1}_{\{X_0=x\}} = \mathbf{1}_{\{X_{R_x}=x\}} = 1$, \mathbb{P}_x -p.s.

(ii) Pour passer de la mesure positive μ_x à une probabilité invariante ν_x , il faut vérifier que la mesure est finie et définir ν_x en normalisant μ_x par sa masse totale $\mu_x(E)$. Celle-ci se calcule directement :

$$\begin{aligned} \mu_x(E) &= \sum_{y \in E} \mu_x(y) = \sum_{y \in E} \mathbb{E}_x \left[\sum_{n=0}^{R_x-1} \mathbf{1}_{\{X_n=y\}} \right] = \mathbb{E}_x \left[\sum_{n=0}^{R_x-1} \sum_{y \in E} \mathbf{1}_{\{X_n=y\}} \right] \\ &= \mathbb{E}_x [R_x], \end{aligned}$$

ce qui fournit le résultat annoncé. \diamond

6.5 Classes fermées irréductibles

Afin d'aborder la question d'unicité d'une probabilité invariante, si elle existe, il est nécessaire de réduire la chaîne de Markov suivant les liens existant entre ses états, tels qu'ils sont lus sur la matrice de transition P .

En effet, supposons qu'il existe un état $x_0 \in E$ tel que $P(x_0, x_0) = 1$. Un tel état est dit absorbant, et il est clair que la masse de Dirac en ce point δ_{x_0} est une probabilité invariante. De même, si x_1 est un deuxième état absorbant, δ_{x_1} est une probabilité invariante. Enfin, toute combinaison convexe $\lambda \delta_{x_0} + (1-\lambda) \delta_{x_1}$, $\lambda \in [0, 1]$, définit une probabilité invariante.

Définition 6.15. Soient $x, y \in E$ deux états donnés. On dit que

- x communique avec y , et on note $x \rightarrow y$, s'il existe un entier $n \geq 0$ et des états $x_0 = x, x_1, \dots, x_n = y \in E$ tels que $P(x_0, x_1) \dots P(x_{n-1}, x_n) > 0$.
- x et y communiquent, et on note $x \leftrightarrow y$, si x communique avec y et y communique avec x .

Définition 6.16. (i) Une classe $E_0 \subset E$ est dite irréductible si $x \leftrightarrow y$ pour tous $x, y \in E_0$. La chaîne de Markov X est dite irréductible si l'espace d'état E est irréductible.

(ii) Une classe $E_0 \subset E$ est dite fermée si pour tous $x, y \in E$:

$$x \in E_0 \text{ et } x \rightarrow y \implies y \in E_0.$$

La restriction de la chaîne de Markov à une classe fermée E_0 est ainsi une chaîne de Markov d'espace d'état E_0 . Enfin si $E_0 = \{x_0\}$ est fermée, on dit que l'état x_0 est absorbant.

Proposition 6.17. *Soient $x, y \in E$.*

- (i) *Si $x \rightarrow y$ et x est récurrent, alors y est récurrent et $N^y = \infty$, \mathbb{P}_x -p.s.*
- (ii) *Si $x \leftrightarrow y$, alors x et y sont simultanément soit transitoires soit récurrents.*

Démonstration. l'énoncé (ii) est une conséquence immédiate de (i). Pour montrer (i), on utilise les temps de retour en x :

$$R_0^x := 0 \quad \text{et} \quad R_n^x := \inf \{k > R_{n-1}^x : X_k = x\} \quad \text{pour } n \geq 1.$$

La variable R_n^x indique le temps du $n^{\text{ième}}$ retour à l'état x . Comme x est récurrent, $R_n^x < \infty$, \mathbb{P}_x -p.s. et

$$N^y = \sum_{n \geq 1} Y_n, \quad \text{où} \quad Y_n := \sum_{k=R_{n-1}^x}^{R_n^x} \mathbf{1}_{\{X_k=y\}}.$$

Les variables aléatoires Y_n , $n \geq 1$ sont iid sous \mathbb{P}_x d'après la proposition 5.21, conséquence de la propriété de Markov forte. On calcule alors que :

$$\begin{aligned} \mathbb{P}_x [N^y < \infty] &= \mathbb{P}_x [\cup_{n \geq 0} \{Y_k = 0, k \geq n\}] \\ &= \mathbb{P}_x [Y_k = 0, k \geq 0] = \lim_{N \rightarrow \infty} \mathbb{P}_x [Y_1 = 0]^N. \end{aligned}$$

Ainsi, si $x \rightarrow y$, on a $\mathbb{P}_x [Y_1 = 0] < 1$ et par suite $\mathbb{P}_x [N^y < \infty] = 0$. De plus, la propriété de Markov forte implique que

$$1 = \mathbb{P}_x [N^y = \infty] = \mathbb{P}_x [R_y < \infty] \mathbb{P}_y [N^y = \infty],$$

en particulier, $\mathbb{P}_y [N^y = \infty] = 1$, i.e. y est récurrent. \diamond

D'après la proposition 6.17, toute classe $E_0 \subset E$ irréductible fermée est transitoire ou récurrente, i.e tous les états de E_0 sont récurrents, ou tous les états de E_0 sont transitoires. Ce résultat suggère la classification suivante des états de la chaîne :

- les états $x \in E$ qui ne communiquent pas avec eux-même sont des états transitoires,
- les états $\{x : x \leftrightarrow x\}$ qui communiquent avec eux-même peuvent être répartis en classes d'équivalence irréductibles pour la relation \leftrightarrow ,
- on identifie les classes irréductibles fermées, les autres sont nécessairement transitoires,
- on considère la restriction de la chaîne à chaque classe irréductible fermée, ramenant ainsi l'étude à une chaîne de Markov irréductible, et on cherche à déterminer si elle est récurrente ou transitoire.

Si la chaîne démarre à partir d'un état récurrent, alors elle peut être considérée comme une chaîne irréductible sur la restriction de l'espace d'état à la classe d'équivalence contenant cet état. D'après la remarque 6.13, chaque état de cette classe d'équivalence est visité une infinité de fois. De même si la chaîne démarre à partir d'un état transitoire, alors, soit elle finit par atteindre une des classes récurrentes, soit elle part à l'infini après être passée un nombre fini de fois par les états de la classe transitoire (remarque 6.13).

6.6 Récurrence positive et loi invariante

Dans ce paragraphe, nous spécialisons la discussion à une chaîne de Markov irréductible, tout en gardant à l'esprit qu'il est toujours possible de se ramener à ce contexte par une décomposition préalable des états de la chaîne. Dans ce cadre, nous allons renforcer l'énoncé de la proposition 6.14 en montrant que la récurrence positive est une condition nécessaire et suffisante pour l'existence d'une loi invariante, et nous obtiendrons l'unicité de la loi invariante comme conséquence de notre analyse.

Théorème 6.18. *Pour une chaîne de Markov irréductible $X = (X_n)_{n \geq 0}$, les deux assertions suivantes sont équivalentes :*

- (i) *X est récurrente positive,*
- (ii) *X admet une loi invariante ν .*

De plus, ν est unique, strictement positive, et est donnée par

$$\nu(x) = \frac{1}{\mathbb{E}_x[R_x]} \quad \text{pour tout } x \in E.$$

Démonstration. L'implication (i) \implies (ii) a été établie dans la proposition 6.14. Nous allons montrer l'implication inverse, ainsi que le reste de l'énoncé, en trois étapes.

Etape 1 Montrons d'abord que $\nu > 0$ sur E . Comme ν est une mesure de probabilité sur l'ensemble dénombrable E , il existe $x_0 \in E$ tel que $\nu(x_0) > 0$. Pour tout $y \in E$, on déduit de l'hypothèse d'irréductibilité de la chaîne l'existence d'un entier i tel que $P^i(x_0, y) > 0$. On utilise à présent le fait que ν est invariante $\nu = \nu P = \dots = \nu P^i$ qui implique que $\nu(y) = \sum_{z \in E} \nu(z) P^i(z, y) \geq \nu(x_0) P^i(x_0, y) > 0$.

Etape 2 Montrons maintenant que la chaîne est récurrente. Pour cela, nous allons prouver que tout état transitoire $y \in E$ vérifie :

$$y \in E \text{ est transitoire} \implies P^n(x, y) \longrightarrow 0 \text{ pour tout } x \in E. \quad (6.3)$$

Or, le fait que ν est invariante implique que $\nu(y) = \sum_{x \in E} \nu(x) P^n(x, y) \longrightarrow 0$ d'après le théorème de convergence dominée. Ceci contredit la stricte positivité de ν établie à la première étape.

Pour montrer (6.3), nous rappelons que par définition d'un état transitoire :

$$\mathbb{E}_y[N^y] = \sum_{n \geq 0} \mathbb{P}_y[X_n = y] = \sum_{n \geq 0} P^n(y, y) < \infty. \quad (6.4)$$

En particulier, $P^n(y, y) \longrightarrow 0$. D'après la propriété de Markov forte, on a aussi

pour tout $x \neq y$:

$$\begin{aligned}
P^n(x, y) &= \sum_{k=0}^n \mathbb{P}_x[R_y = k] P^{n-k}(y, y) \\
&= \sum_{k < \frac{n}{2}} \mathbb{P}_x[R_y = k] P^{n-k}(y, y) + \sum_{k \geq \frac{n}{2}} \mathbb{P}_x[R_y = k] P^{n-k}(y, y) \\
&\leq \sum_{k \geq \frac{n}{2}} P^k(y, y) + \sum_{\frac{n}{2} \leq k \leq n} \mathbb{P}_x[R_y = k] \\
&= \sum_{k \geq \frac{n}{2}} P^k(y, y) + \mathbb{E}_x \left[\mathbf{1}_{\{\frac{n}{2} \leq R_y \leq n\}} \right].
\end{aligned}$$

Le premier terme sur la droite converge vers zéro d'après la convergence de la série (6.4). Quand au second terme, on voit que sur $\{R_y < \infty\}$, on a $\mathbf{1}_{\{\frac{n}{2} \leq R_y \leq n\}} \rightarrow 0$, \mathbb{P}_x -p.s. et sur $\{R_y = \infty\}$, on a $\mathbf{1}_{\{\frac{n}{2} \leq R_y \leq n\}} = 0$ pour tout n . Ainsi, $\mathbf{1}_{\{\frac{n}{2} \leq R_y \leq n\}} \rightarrow 0$, \mathbb{P}_x -p.s. et dans \mathbb{L}^1 par le théorème de convergence dominée.

Etape 3 Soit $\tilde{\beta} := \nu - \nu(x)\mu_x$, où μ_x est la mesure définie dans (6.2), et $\beta(y) := |\tilde{\beta}(y)|$, $y \in E$. Comme $\mu_x(x) = 1$, on a $\beta(x) = 0$. Comme ν est invariante, on déduit de la proposition 6.14 et de la remarque 6.4 que β est aussi invariante. Ainsi :

$$\beta P = \beta, \quad \beta(x) = 0 \quad \text{et} \quad \beta \geq 0. \quad (6.5)$$

Comme la chaîne est irréductible, on déduit l'existence, pour tout $y \in E$, d'un entier $i \geq 1$ tel que $P^i(y, x) > 0$, et par suite $0 = \beta(x) = \sum_{z \in E} \beta(z) P^i(z, x) \geq \beta(y) P^i(y, x) \geq 0$, prouvant que $\beta(y) = 0$. Ainsi, nous avons identifié la loi invariante ν à μ_x , à une constante multiplicative près, que l'on peut déterminer en examinant la mesure de l'espace d'état :

$$1 = \nu(E) = \nu(x)\mu_x(E) = \nu(x)\mathbb{E}_x[R_x].$$

Ceci termine la preuve de l'implication (ii) \implies (i), ainsi que la caractérisation unique de la mesure invariante. \diamond

Remarque 6.19. L'étape 3 de la preuve ci-dessus utilise uniquement le fait que $\nu \geq \nu P$, i.e. la propriété de *surinvariance* de ν . En fait, en examinant la preuve de l'étape 3, on voit qu'on peut isoler le résultat d'unicité suivant :

Proposition 6.20. *Soit X une chaîne de Markov sur un espace d'état dénombrable E , et $x \in E$ un état récurrent. Alors, pour toute mesure ν sur E :*

$$\nu \geq \nu P \implies \nu = \nu(x)\mu_x.$$

Remarque 6.21. Une autre conséquence du théorème 6.18 est que toute chaîne de Markov irréductible sur un espace d'état fini est récurrente, puisque l'existence d'une loi invariante a été établie pour les chaînes à espace d'état fini. Cette observation a déjà été faite dans la remarque 5.18. Ainsi, les exemples 5.7-5.11 de l'urne d'Ehrenfest et de l'automate de comptage donnent lieu à des chaînes récurrentes.

6.7 Application aux marches aléatoires

6.7.1 Marche aléatoire au plus proche voisin sur \mathbb{Z}

Il s'agit de la chaîne de Markov sur l'espace d'état \mathbb{Z} définie par

$$X_n = X_{n-1} + \xi_n, \quad n \geq 1,$$

où $(\xi_n)_{n \geq 1}$ est une suite de v.a. iid de loi :

$$\mathbb{P}[\xi_1 = 1] = 1 - \mathbb{P}[\xi_1 = -1] = p,$$

le paramètre $p \in]0, 1[$ étant donné. La chaîne est irréductible, et par suite les états sont soit tous récurrents, soit tous transients.

En conditionnant par X_1 , on voit que

$$\mathbb{P}_0[R_0 < \infty] = \mathbb{E}_0[\mathbb{P}_{X_1}[R_0 < \infty]] = p\mathbb{P}_1[R_0 < \infty] + (1-p)\mathbb{P}_{-1}[R_0 < \infty].$$

La première probabilité $\mathbb{P}_1[R_0 < \infty]$ a été calculée dans (5.11). La deuxième probabilité $\mathbb{P}_{-1}[R_0 < \infty]$ s'en déduit par symétrie. On obtient alors

$$\mathbb{P}_0[R_0 < \infty] = 2(p \wedge (1-p)),$$

et on déduit que

$$\text{la chaîne est récurrente si et seulement si } p = \frac{1}{2}.$$

Remarque 6.22. Une autre façon de voir que la chaîne est transiente pour $p \neq \frac{1}{2}$ est de résoudre directement le problème de valeur propre $\mu P = \mu$ pour une mesure μ sur E , qui s'écrit $\mu_i = p\mu_{i+1} + (1-p)\mu_{i-1}$, et dont la solution générale est $\mu_i = A + B \left(\frac{1-p}{p}\right)^i$, A et B étant des constantes arbitraires. En imposant la positivité de la mesure μ , on obtient la restriction $A, B \geq 0$ sur les constantes. Si 0 était un état récurrent, alors on déduirait du résultat d'unicité de la proposition 6.20 que

$$A + B \left(\frac{1-p}{p}\right)^i = (A+B)\mu_0(i) \quad \text{pour tous } i \in \mathbb{Z} \text{ et } A, B \geq 0,$$

ce qui contredit le fait que $p \neq \frac{1}{2}$.

Dans le cas symétrique $p = \frac{1}{2}$, la solution générale du problème de valeur propre $\mu P = \mu$ est donnée par $\mu_i = A + Bi$, $i \in \mathbb{Z}$, où A et B sont des constantes arbitraires. En imposant la positivité de la mesure μ , on obtient la restriction $A \geq 0$ et $B = 0$. Ainsi, la seule mesure invariante est la mesure uniforme, qui ne peut être normalisée pour en extraire une probabilité. La chaîne n'a donc pas de loi invariante, et le théorème 6.18 montre que

$$\text{si } p = \frac{1}{2}, \text{ la chaîne est récurrente nulle.}$$

6.7.2 Marche aléatoire unidimensionnelle générale

On considère la marche aléatoire sur \mathbb{Z} :

$$X_n = X_{n-1} + \xi_n, \quad n \geq 1,$$

où $(\xi_n)_{n \geq 1}$ est une suite de v.a. à valeurs dans \mathbb{Z} , iid et intégrables.

Dans ce contexte plus général, on doit procéder à une décomposition de l'espace d'état en remarquant que la classe d'irréductibilité d'un état $x \in \mathbb{Z}$ est l'ensemble $x + G$, où G est le sous-groupe additif engendré par le support de la v.a. ξ_1 . Si ξ_1 est différente de la v.a. nulle, on voit immédiatement que G est infini.

Proposition 6.23. *Pour la marche aléatoire unidimensionnelle décrite ci-dessus, on a :*

- (i) si $\mathbb{E}[\xi_1] \neq 0$, tous les états sont transitoires,
- (ii) si $\mathbb{E}[\xi_1] = 0$, tous les états sont récurrents nuls.

Démonstration. La partie (i) est une conséquence de la loi forte des grands nombres, qui montre que la chaîne converge p.s. vers $+\infty$ ou $-\infty$ dès que $\mathbb{E}[\xi_1] \neq 0$.

Supposons maintenant que $\mathbb{E}[\xi_1] = 0$. Pour $\varepsilon > 0$, $n \geq 1$, et un état $x_0 \in \mathbb{Z}$ arbitraire, on calcule que :

$$\begin{aligned} \sum_{k=0}^n \mathbb{P}_{x_0} \left[\frac{|X_k|}{k} < \varepsilon \right] &= \sum_{k=0}^n \mathbb{E}_{x_0} \left[\sum_{|x| < k\varepsilon} \mathbf{1}_{\{X_k=x\}} \right] \\ &\leq \sum_{|x| < n\varepsilon} \mathbb{E}_{x_0} \left[\sum_{k=0}^n \mathbf{1}_{\{X_k=x\}} \right] \\ &\leq \sum_{|x| < n\varepsilon} \mathbb{E}_{x_0} [N^x]. \end{aligned}$$

Supposons que l'état x_0 soit transitoire, i.e. $\mathbb{E}_{x_0}[N_{x_0}] < \infty$. Remarquons que $\mathbb{E}_x[N^x] = \mathbb{E}_{x_0}[N_{x_0}]$ du fait de l'invariance par translation spatiale de la chaîne, et que $\mathbb{E}_{x_0}[N^x] = \mathbb{P}_{x_0}[T_x < \infty] \mathbb{E}_x[N^x] \leq \mathbb{E}_{x_0}[N_{x_0}]$ d'après la propriété de Markov fort. On déduit alors des inégalités précédentes que :

$$\sum_{k=0}^n \mathbb{P}_{x_0} \left[\frac{|X_k|}{k} < \varepsilon \right] \leq 2n\varepsilon \mathbb{E}_{x_0}[N_{x_0}].$$

Pour le terme de droite de cette inégalité, on utilise la loi (faible) des grands nombres pour déduire l'existence d'un entier n_0 tel que $\mathbb{P}_0 \left[\frac{|X_k|}{k} < \varepsilon \right] \geq 4\varepsilon \mathbb{E}_0[N_0]$. Alors, pour $n \geq 4n_0$, on obtient la contradiction recherchée :

$$4\varepsilon \mathbb{E}_0[N_0] \frac{3n}{4} \leq 2n\varepsilon \mathbb{E}_0[N_0].$$

Ceci montre que x_0 ne peut être transitoire.

Enfin, notons que pour toute classe irréductible E_0 , la mesure uniforme, $\mu(x) = 1$ pour tout $x \in E_0$, est invariante pour la restriction de la marche aléatoire à E_0 . La proposition 6.20 implique qu'en tout état récurrent $x_0 \in E_0$, on doit avoir $\mu = \nu(x_0)\mu_{x_0}$, c'est à dire que la loi invariante (si elle existe) doit être obtenue par normalisation de la loi uniforme. Comme E_0 est infinie, une telle normalisation est impossible et on déduit que x_0 est récurrent nul. \diamond

6.7.3 Marche aléatoire symétrique sur \mathbb{Z}^d

Rappelons que cet exemple correspond à la chaîne de Markov sur l'espace d'état \mathbb{Z}^d définie par

$$X_n = X_{n-1} + \xi_n, \quad n \geq 1,$$

où $(\xi_n)_{n \geq 1}$ est une suite de v.a. iid de loi uniforme sur $\{\pm e_i, i = 1, \dots, d\}$:

$$\mathbb{P}[\xi_1 = e_i] = \mathbb{P}[\xi_1 = -e_i] = \frac{1}{2d} \quad \text{avec} \quad e_i^j = \mathbf{1}_{\{j=i\}}, \quad i, j \leq d.$$

Il s'agit d'une chaîne de Markov irréductible. En utilisant le même argument qu'au début de la preuve de la proposition 6.23, on voit que la mesure uniforme est invariante et que le lemme 6.20 implique qu'il n'existe pas de loi invariante. Ainsi, la marche aléatoire X n'est pas récurrente positive.

Comme tous les états sont soit récurrents soit transients, il suffit d'étudier l'origine 0. D'après la partie 3 de la remarque 6.13, le caractère récurrent ou transcient de 0 est caractérisé par la finitude du temps de retour moyen à l'origine : 0 est récurrent si et seulement si $\mathbb{E}_0[N_0] = \infty$. Par ailleurs, il est clair que partant de l'origine, on ne peut y retourner (avec probabilité non nulle) qu'après un nombre pair d'étapes. Alors :

$$\mathbb{E}_0[N_0] = \sum_{n \geq 0} P^n(0, 0) = \sum_{n \geq 0} P^{2n}(0, 0). \quad (6.6)$$

Avant d'énoncer le résultat pour toute dimension, examinons le cas unidimensionnel $d = 1$. Alors $P^{2n}(0, 0) = \mathbb{P}[\sum_{i \leq 2n} \mathbf{1}_{\{\xi_i=1\}} = n] = 2^{-2n} C_{2n}^n$. En utilisant la formule de Stirling¹, on voit que $P^{2n}(0, 0) \sim (\pi n)^{-1/2}$, et on conclut que $\mathbb{E}_0[N_0] = \infty$. D'après la partie 3 de la remarque 6.13, ceci montre que 0 est un point récurrent, et par suite la marche aléatoire symétrique sur \mathbb{Z} est récurrente.

Le résultat suivant, dû à Polya (1921), montre que ce résultat de récurrence dans le cadre unidimensionnel ne s'étend pas à toute dimension.

Proposition 6.24. *La marche aléatoire au plus proche voisin symétrique sur \mathbb{Z}^d est*

- récurrente nulle si et seulement si $d \leq 2$,
- transiente si $d \geq 3$.

1. On rappelle que la formule de Stirling donne, pour tout entier naturel non nul n , un équivalent de sa factorielle : $n! = \sqrt{2\pi n} \left(\frac{n}{e}\right)^n \left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right)$.

Pour montrer cette proposition, nous utilisons le résultat suivant qui est valable pour des marches aléatoires au plus proche voisin sur \mathbb{Z}^d non nécessairement symétriques. Pour une v.a. Z à valeurs dans \mathbb{R}^d , on rappelle que $\Phi_Z(t) := \mathbb{E}[e^{it \cdot Z}]$ désigne sa fonction caractéristique.

Lemme 6.25. *Soient X une marche aléatoire au plus proche voisin sur \mathbb{Z}^d . Alors :*

$$\mathbb{E}_0[N_0] = \lim_{\lambda \nearrow 1} (2\pi)^{-d} \int_{[-\pi, \pi]^d} \frac{1}{1 - \lambda^2 \Phi_{\xi_1}(t)^2} dt. \quad (6.7)$$

Démonstration. On commence par remarquer que la probabilité $P^n(0, 0)$ peut s'exprimer en termes de la fonction caractéristique $\Phi_{X_n} := \mathbb{E}_0[e^{it \cdot X_n}]$ par :

$$P^n(0, 0) = (2\pi)^{-d} \int_{[-\pi, \pi]^d} \Phi_{X_n}(t) dt. \quad (6.8)$$

En effet, on calcule grâce au théorème de Fubini que

$$\int_{[-\pi, \pi]^d} \Phi_{X_n}(t) dt = \mathbb{E}_0 \left[\int_{[-\pi, \pi]^d} e^{it \cdot X_n} dt \right] = \mathbb{E}_0 \left[\prod_{1 \leq j \leq d} \int_{-\pi}^{\pi} e^{it_j X_n^j} dt_j \right],$$

où X_n^j est la $j^{\text{ième}}$ composante de X_n , et on déduit la formule (6.8) en remarquant que $\int_{-\pi}^{\pi} e^{it_j x} dt_j = 0$ pour tout $x \in \mathbb{Z} \setminus \{0\}$.

On continue alors le calcul (6.6) en réutilisant le théorème de Fubini ainsi que le fait que les $(\xi_n)_{n \geq 1}$ sont iid :

$$\mathbb{E}_0[N_0] = (2\pi)^{-d} \sum_{n \geq 0} \int_{[-\pi, \pi]^d} \phi_{X_{2n}}(t) dt$$

Du fait que $\phi_{X_{2n}}(0) = 1$, l'inversion de la somme et l'intégrale n'est pas immédiate. On écrit alors que :

$$\begin{aligned} \mathbb{E}_0[N_0] &= (2\pi)^{-d} \lim_{\lambda \nearrow 1} \sum_{n \geq 0} \int_{[-\pi, \pi]^d} \lambda^{2n} \phi_{X_{2n}}(t) dt \\ &= \lim_{\lambda \nearrow 1} (2\pi)^{-d} \int_{[-\pi, \pi]^d} \sum_{n \geq 0} \lambda^{2n} \prod_{1 \leq i \leq 2n} \Phi_{\xi_i}(t) dt \\ &= \lim_{\lambda \nearrow 1} (2\pi)^{-d} \int_{[-\pi, \pi]^d} \sum_{n \geq 0} \lambda^{2n} \Phi_{\xi_1}(t)^{2n} dt \\ &= \lim_{\lambda \nearrow 1} (2\pi)^{-d} \int_{[-\pi, \pi]^d} \frac{1}{1 - \lambda^2 \Phi_{\xi_1}(t)^2} dt, \end{aligned}$$

◇

Preuve de la proposition 6.24 Pour démontrer le résultat, il suffit de considérer l'origine, puisque l'irréductibilité de la chaîne implique que tous les

états sont de même nature. On utilise l'expression établie dans le lemme 6.25, en remarquant que dans le contexte de la marche aléatoire au plus proche voisin symétrique dans \mathbb{Z}^d , la fonction caractéristique Φ_{ξ_1} est donnée par :

$$\Phi_{\xi_1}(t) = \frac{1}{d} \sum_{j=1}^d \cos t_j \quad \text{pour tout } t = (t_1, \dots, t_d) \in \mathbb{R}^d.$$

Il reste à étudier la convergence de l'intégrale impropre (6.7) aux points critiques $0_{\mathbb{Z}^d}$ et $\pm\pi\mathbf{1}$, où $\mathbf{1} = \sum_{i=1}^d e_i$. Tous ces points sont étudiés de manière identique, nous présentons alors le calcul uniquement pour l'origine. Au voisinage de l'origine, la fonction $\lambda \mapsto 1 - \lambda\Phi_{\xi_1}(t)$ est positive croissante. Il suffit donc d'étudier l'intégrabilité de $(1 - \Phi_{\xi_1}(t)^2)^{-1}$. On utilise alors l'équivalent au voisinage de zéro $(1 - \Phi_{\xi_1}(t)^2)^{-1} \sim d|t|^{-2}$, qui montre que la nature (finie ou non) de l'intégrale (6.7) en ce point est la même que celle de

$$\int_{|t| \leq \varepsilon} |t|^{-2} dt = \int_0^\varepsilon r^{-2} A_d r^{d-1} dr = A_d \int_0^\varepsilon r^{d-3} dr,$$

où on est passé aux coordonnées sphériques, et A_d est l'aire de la sphère unité en dimension d . Ainsi, l'intégrale (6.7) est convergente au point critique 0 si et seulement si $d \geq 3$. \diamond

Remarque 6.26. On pourrait étudier des marches aléatoires au plus proche voisin symétriques plus générales grâce au Lemme 6.25. en effet, on peut utiliser le fait que la fonction caractéristique d'une v.a. Z dans \mathbb{R}^d , telle que $\mathbb{E}[|Z|^2] < \infty$, admet le développement au second ordre au voisinage de zéro suivant :

$$\Phi_Z(t) = 1 + it \cdot \mathbb{E}[Z] + \frac{1}{2} \text{Tr} \{ t t^T \mathbb{E}[Z Z^T] \} + o(|t|^2),$$

voir exercice 2.9. Ceci permet d'étudier le comportement de l'intégrale (6.7) en zéro. Comme les composantes de ξ_1 sont à valeurs dans $\{-1, 1\}$, les seuls autres points critiques sont $\pm\pi$, dont l'étude se ramène au point zéro en remarquant que $\Phi_{\xi_1}(\pi - h) = -\Phi_{\xi_1}(-h)$ et $\Phi_{\xi_1}(-\pi + h) = -\Phi_{\xi_1}(h)$. Dans le cas non biaisé $\mathbb{E}[\xi_1] = 0$, le développement ci-dessus montre que le résultat de la proposition 6.24 est vrai pour des cas plus généraux.

6.8 Autres exemples

Ruine du joueur Dans l'exemple 5.9, la richesse du joueur A forme une chaîne de Markov sur l'espace d'état $E = \{0, 1, \dots, N\}$ où N est la richesse totale des deux joueurs. Les états 0 et N sont absorbants puisqu'ils correspondent à la ruine d'un des deux joueurs. Donc $\{0\}$ et $\{N\}$ sont des classes fermées irréductible. La classe $E_0 = \{1, \dots, N\}$ est irréductible mais n'est pas fermée, puisqu'elle communique avec 0 et N qui n'y sont pas. Ainsi E_0 est transitoire, et $\mathbb{P}[T_{\{0,N\}} < \infty] = 1$.

Dans le cas d'un joueur compulsif $N = \infty$, l'espace d'état est $E = \mathbb{N}$, l'état absorbant 0 forme une classe fermée irréductible, et la classe $\mathbb{N} \setminus \{0\}$ des entiers non nuls est une classe irréductible non fermée, donc transitoire. La chaîne finit soit à l'infini soit en zéro, après être passée un nombre fini de fois par tout entier non nul.

Processus de branchement rappelons le cadre de l'exemple 5.10. La taille X_n de la population à la date n est régie par la dynamique

$$X_{n+1} = \sum_{i=1}^{X_n} \xi_i^n \quad \text{pour } n \geq 0,$$

où la somme est nulle sur $\{X_n = 0\}$. Les v.a. $\{\xi_i^n, 1 \leq i \leq n, n \geq 0\}$ sont iid à valeurs dans \mathbb{N} représentent le nombre de descendants de chaque individu $i \leq n$ à la date n . Ceci définit une chaîne de Markov homogène $(X_n)_{n \geq 0}$ sur l'espace d'état $E = \mathbb{N}$.

L'état 0 est absorbant et forme une classe fermée récurrente. Si $\mathbb{P}[\xi_1 = 0] > 0$, alors tout état $x \geq 1$ est transitoire puisqu'il communique avec l'état absorbant 0. Si de plus $\mathbb{P}[\xi_1 = 0] + \mathbb{P}[\xi_1 = 1] < 1$, alors la classe $\mathbb{N} \setminus \{0\}$ est irréductible, mais non fermée, et la population soit finit par s'éteindre, soit explose, après être passée un nombre fini de fois par chacun des états de cette classe.

Processus de renouvellement Reprenons le cadre de l'exemple 5.12 où l'âge X_n du composant en place est donné par

$$X_n = n - \theta_{n-1} \quad \text{où } \theta_{n-1} := \sup \{D_1 + \dots + D_k : k \geq 1 \text{ et } D_1 + \dots + D_k \leq n\}$$

est la date d'installation du composant en place, avec $T_0 = 0$, et $(D_i)_{i \geq 1}$ est une suite de v.a. iid à valeurs dans $\mathbb{N} \setminus \{0\}$. Pour chaque $i \geq 1$, D_i représente la durée de vie du composant i que l'on installe dès que le composant $i - 1$ tombe en panne. Nous avons vu que cette chaîne de Markov admet comme matrice de transition

$$P(x, x+1) = \frac{\mathbb{P}[D_1 > x+1]}{\mathbb{P}[D_1 > x]}, \quad P(x, 0) = 1 - P(x, x+1) \quad \text{pour } x \geq 1,$$

et $P(0, 0) = 1 - P(0, 1) = 0$.

On suppose que $\mathbb{P}[D_1 = x] > 0$ pour tout $x \geq 1$. Alors la chaîne est irréductible, et il suffit d'identifier la nature de l'état zéro pour caractériser la nature de la chaîne. Le temps de retour en 0 ne peut prendre que des valeurs supérieures à 2 si on démarre la chaîne à 0, et on calcule pour tout $k \geq 2$:

$$\mathbb{P}_0[R_0 = k] = \mathbb{P}_0[X_1 = 1, \dots, X_{k-1} = k-1, X_k = 0] = \frac{\mathbb{P}[D_1 = k]}{\mathbb{P}[D_1 > 1]},$$

et par suite

$$\mathbb{P}_0[R_0 = \infty] = \lim_{k \rightarrow \infty} \mathbb{P}_0[R_0 > k] = \mathbb{P}[D_1 > 1]^{-1} \lim_{k \rightarrow \infty} \mathbb{P}_0[D_1 > k] = 0.$$

L'état 0 est donc récurrent. On calcule aussi :

$$\begin{aligned}\mathbb{E}_0[R_0] &= \sum_{k \geq 2} k \mathbb{P}_0[R_0 = k] \\ &= \frac{1}{\mathbb{P}[D_1 > 1]} \sum_{k \geq 2} k \mathbb{P}[D_1 = k] = \frac{\mathbb{E}[D_1] - \mathbb{P}[D_1 = 1]}{\mathbb{P}[D_1 > 1]},\end{aligned}$$

Ainsi 0 est récurrent positif si et seulement si $\mathbb{E}[D_1] < \infty$.

Chaîne de naissance et de mort sur \mathbb{N} Il s'agit d'une marche aléatoire au plus proche voisin sur $E = \mathbb{N}$ définie par les probabilités de transitions :

$$\begin{aligned}P(0, 0) + P(0, 1) &= 1, \\ P(n, n-1) + P(n, n) + P(n, n+1) &= 1, \quad n \geq 1.\end{aligned}\tag{6.9}$$

On suppose que

$$P(0, 1) > 0, \quad P(n, n-1) > 0 \text{ et } P(n, n+1) > 0 \quad \text{pour tout } n \geq 1.\tag{6.10}$$

Alors la chaîne est irréductible.

Proposition 6.27. *Sous la condition (6.10), la chaîne de naissance et de mort (6.9) est irréductible sur \mathbb{N} et vérifie :*

(i) *il existe une unique (à une constante multiplicative près) mesure $\mu \geq 0$ invariante donnée par*

$$\mu_n = \mu_0 \prod_{i=1}^n \frac{P(i-1, i)}{P(i, i-1)} \quad \text{pour tout } n \geq 0,$$

(ii) *la chaîne est récurrente positive si et seulement si $\sum_{n \geq 0} \mu_n < \infty$, et sous cette condition l'unique loi invariante est donnée par la mesure de probabilité déduite de μ par normalisation,*

(iii) *la chaîne est transiente si et seulement si $\sum_{n \geq 0} \left(\prod_{i=1}^n \frac{P(i, i-1)}{P(i, i+1)} \right) < \infty$.*

Démonstration. (i) Le problème de valeur propre s'écrit

$$\begin{aligned}\mu_0 &= \mu_0 P(0, 0) + \mu_1 P(1, 0) \\ \mu_n &= \mu_{n-1} P(n-1, n) + \mu_n P(n, n) + \mu_{n+1} P(n, n+1), \quad n \geq 1.\end{aligned}$$

Etant donné μ_0 , la première équation détermine μ_1 . Puis, étant donnés μ_{n-1} et μ_n la $n^{\text{ième}}$ équation détermine μ_{n+1} de manière unique. On vérifie par une récurrence simple que la formule de la proposition est bien l'unique solution de ces équations.

(ii) est immédiat d'après l'unicité établie en (i) et le théorème 6.18.

(iii) La chaîne étant irréductible, il suffit d'étudier $\mathbb{P}_0[R_0 < \infty]$. On utilise la propriété des projections itérées de l'espérance conditionnelle pour obtenir

$$\mathbb{P}_0[R_0 < \infty] = P(0, 0) + P(0, 1)w(1) \text{ où } w(n) := \mathbb{P}_x[T_0 < \infty], \quad n \geq 1. \tag{6.11}$$

D'après la remarque 5.17, la fonction w est la plus petite solution positive du problème de Dirichlet

$$w(0) = 1 \quad \text{et} \quad (I - P)w = 0 \quad \text{sur} \quad \mathbb{N} \setminus \{0\}.$$

En détaillant le produit matriciel, on voit immédiatement que

$$w_{n+1} - w_n = \frac{P(n, n-1)}{P(n, n+1)}(w_n - w_{n-1}),$$

L'ensemble des solutions du problème de Dirichlet s'écrivent alors en termes de la fonction :

$$\phi(n) := \sum_{i=0}^{n-1} \prod_{j=1}^i \frac{P(j, j-1)}{P(j, j+1)}, \quad n \geq 2 \quad \phi(0) = 0, \quad \phi(1) = 1,$$

sous la forme

$$w(n) = 1 + (w(1) - 1)\phi(n)$$

où $w(1)$ doit être déterminé par la minimalité de la solution parmi les fonctions positives. Deux cas apparaissent alors :

- Si $\phi(\infty) = \infty$, alors la positivité de la solution impose $w(1) = 1$, et on déduit de (6.11) que $\mathbb{P}_0[R_0 < \infty] = 1$, 0 est un état récurrent, et la chaîne est récurrente.
- Si $\phi(\infty) < \infty$, on déduit de la minimalité de la solution parmi les fonctions positives que $w(1) = 1 - \phi(\infty)^{-1}$. On obtient par (6.11) que $\mathbb{P}_0[R_0 < \infty] = 1 - P(0, 1)\phi(\infty)^{-1}$, et la chaîne est transitoire. \diamond

Chapitre 7

Théorèmes ergodiques pour les chaînes de Markov

Dans ce chapitre, nous montrons que la loi invariante permet de caractériser le comportement asymptotique de moyennes de la forme :

$$\frac{1}{n} \sum_{i=1}^n g(X_{i-1}, X_i) \longrightarrow \mathbb{E}_\nu [g(X_0, X_1)], \mathbb{P} - \text{p.s.} \quad (7.1)$$

où g est une fonction vérifiant des conditions d'intégrabilité convenables. Nous caractérisons aussi la vitesse de convergence par le biais d'un théorème central limite. Le but de ce paragraphe est de présenter une version plus générale de ce résultat. L'argument crucial pour obtenir ces résultats est basé sur l'utilisation des temps de retour de la chaîne en un état arbitraire $x \in E$:

$$R_0^x := 0 \quad \text{et} \quad R_n^x := \inf \{k > R_{n-1}^x : X_k = x\} \quad \text{pour } n \geq 1, \quad (7.2)$$

et l'observation de la proposition 5.21 que les excursions de la chaîne entre deux temps de retours successifs R_n^x et R_{n+1}^x , $n \geq 1$, sont indépendantes et identiquement distribuées. Pour une chaîne récurrente, ces excursions se répètent une infinité de fois, ouvrant ainsi la porte à l'utilisation de la loi forte des grands nombres.

En appliquant le résultat de convergence (7.1) à la fonction $g(x, y) = \mathbf{1}_{\{y\}}$, on obtient par une application simple du théorème de convergence dominée la convergence des lois marginales en moyenne de Césaro vers la loi invariante. Nous nous intéressons alors à la convergence des lois marginales elles-mêmes vers la loi invariante. Ceci conduit à l'introduction de la notion d'*apériodicité* comme condition nécessaire et suffisante pour obtenir cette convergence.

7.1 Théorèmes ergodiques

Commençons par une version simple du théorème ergodique.

Théorème 7.1. *Soit $(X_n)_{n \geq 0}$ une chaîne de Markov irréductible. Alors, pour tous $x, y \in E$:*

$$\frac{1}{n} N_n^y := \frac{1}{n} \sum_{i=0}^n \mathbf{1}_{\{X_i=y\}} \longrightarrow \frac{1}{\mathbb{E}_y[R^y]}, \quad \mathbb{P}_x - p.s.$$

Démonstration. Si la chaîne est transiente, on a d'après la remarque 6.13 que $\mathbb{E}_x[N_y] < \infty$ et l'état y n'est visité qu'un nombre fini de fois. Le résultat du théorème est trivial dans ce cas.

Supposons maintenant que la chaîne est récurrente. Alors, les temps de retour R_n^x définis dans (7.2) sont tous finis \mathbb{P}_x -p.s.

1. D'après la proposition 5.21, les variables $R_{k+1}^x - R_k^x$ sont iid, et on déduit de la loi des grands nombres pour les v.a. positives énoncée dans le corollaire 2.39 :

$$\frac{R_n^x}{n} \longrightarrow \mathbb{E}_x[R_1^x] \quad \mathbb{P}_x - p.s. \quad (7.3)$$

où la limite est possiblement infinie.

2. Pour tout entier n , on note $k := N_n^x$ si bien que $n \in [R_k^x, R_{k+1}^x)$. Alors,

$$\frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i=x\}} = \frac{k}{n} \in \left(\frac{k}{R_{k+1}^x}, \frac{k}{R_k^x} \right],$$

et on déduit de (7.3) que

$$\frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i=x\}} \longrightarrow \frac{1}{\mathbb{E}_x[R_x]} \quad \mathbb{P}_x - p.s.$$

3. Enfin, pour $y \neq x$, on sait que la chaîne $(X_n)_{n \geq 0}$ atteint y en un temps fini R_0^y , et on applique le raisonnement précédent à la chaîne translaturée $(X_{R_0^y+n})_{n \geq 0}$ pour obtenir

$$\frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i=y\}} \longrightarrow \frac{1}{\mathbb{E}_y[R_1^y]} \quad \mathbb{P}_x - p.s.$$

◇

Remarque 7.2. Le résultat du dernier théorème 7.1 permet de retrouver l'expression de la loi invariante ν pour une chaîne irréductible récurrente positive. En effet, si X_0 est de loi $\pi_0 = \nu$, alors X_n est de loi $\pi_n = \nu$ et

$$\begin{aligned} \nu(y) &= \mathbb{E}_\nu \left[\frac{1}{n} \sum_{k=1}^n \mathbf{1}_{\{X_k=y\}} \right] \\ &= \sum_{x \in E} \nu(x) \mathbb{E}_x \left[\frac{1}{n} \sum_{k=1}^n \mathbf{1}_{\{X_k=y\}} \right] \longrightarrow \sum_{x \in E} \nu(x) \frac{1}{\mathbb{E}_y[R^y]} = \frac{1}{\mathbb{E}_y[R^y]} \end{aligned}$$

d'après le théorème 7.1 et le théorème de convergence dominée. Ceci prouve que $\nu(y) = 1/\mathbb{E}_y[R^y]$. En se rappelant de la stricte positivité de la loi invariante (cf. étape 1 de la preuve du théorème 6.18), ceci permet aussi de retrouver le résultat que l'existence d'une loi invariante implique la récurrence positive.

Nous allons maintenant présenter une version plus générale de la loi des grands nombres du théorème 7.1, dont la démonstration passe par le même type d'argument.

Théorème 7.3. *Soit X une chaîne de Markov irréductible et récurrente positive sur un espace dénombrable E , de matrice de transition P , et d'unique loi invariante ν . Alors, pour toute fonction $g : E \times E \rightarrow \mathbb{R}$ positive ou telle que $\mathbb{E}_\nu[|g(X_0, X_1)|] < \infty$, on a :*

$$\frac{1}{n} \sum_{i=1}^n g(X_{i-1}, X_i) \longrightarrow \mathbb{E}_\nu[g(X_0, X_1)] = \sum_{x \in E} \nu(x) \sum_{y \in E} P(x, y) g(x, y), \quad \text{p.s.}$$

quelle que soit la loi initiale π_0 de X_0 .

Démonstration. Nous commençons par montrer ce résultat pour une fonction $g(x, y) = f(y)$ ne dépendant que d'une seule variable, puis nous généralisons.

1. Supposons $g(x, y) = f(y)$, où $f : E \rightarrow \mathbb{R}_+$ est une fonction positive. Le résultat pour f de signe non constant se déduit facilement par décomposition $f = f^+ - f^-$. Avec les notations de la proposition 5.21, les variables aléatoires positives

$$F(Y_n) := \sum_{i=R_{n-1}^x+1}^{R_n^x} f(X_i), \quad n \geq 1,$$

sont indépendantes, et ont la même distribution pour $n \geq 2$. On déduit alors de la loi des grands nombres que

$$\frac{1}{n} \sum_{k=1}^n F(Y_k) \longrightarrow \mathbb{E}_x[F(Y_1)] = \mathbb{E}_x \left[\sum_{i=1}^{R_1^x} f(X_i) \right], \quad \mathbb{P}_x - \text{p.s.}$$

Nous avons montré dans la preuve du théorème 7.1 que $(R_n^x/n) \rightarrow 1/\nu(x)$ p.s. quelle que soit la loi initiale π_0 de X_0 . Alors :

$$\frac{1}{R_n^x} \sum_{i=1}^{R_n^x} f(X_i) = \frac{n}{R_n^x} \frac{1}{n} \sum_{k=1}^n F(Y_k) \longrightarrow \nu(x) \mathbb{E}_x[F(Y_1)].$$

Enfin, en notant $N_n := \sum_{k=1}^n \mathbf{1}_{\{X_k=x\}}$, on a $N_n \rightarrow \infty$, $R_{N_n}^x \leq n < R_{N_n+1}^x$ p.s. et on déduit de l'encadrement :

$$\frac{R_{N_n}^x}{R_{N_n+1}^x} \frac{1}{R_{N_n}^x} \sum_{i=1}^{R_{N_n}^x} f(X_i) \leq \frac{1}{n} \sum_{i=1}^n f(X_i) \leq \frac{R_{N_n+1}^x}{R_{N_n}^x} \frac{1}{R_{N_n+1}^x} \sum_{i=1}^{R_{N_n+1}^x} f(X_i),$$

ainsi que du fait que $R_n^x \rightarrow \infty$, $R_{n+1}^x/R_n^x \rightarrow 1$ p.s. que

$$\frac{1}{n} \sum_{i=1}^n f(X_i) \rightarrow \nu(x) \mathbb{E}_x [F(Y_0)].$$

2. Pour identifier cette limite, nous remarquons que pour $f(z) := \mathbf{1}_{\{z=y\}}$, on a d'après le théorème 7.1 que $\nu(y) = \nu(x) \mathbb{E}_x \left[\sum_{i=1}^{R_1^x} \mathbf{1}_{\{X_i=y\}} \right]$, et par suite

$$\nu(x) \mathbb{E}_x [F(Y_0)] = \sum_{y \in E} f(y) \nu(x) \mathbb{E}_x \left[\sum_{i=1}^{R_1^x} \mathbf{1}_{\{X_i=y\}} \right] = \sum_{y \in E} f(y) \nu(y).$$

3. Passons maintenant à des fonctions $g(x, y)$ plus générales. On décompose $g(x, y) := h(x, y) + f(x)$, où

$$f(x) := \mathbb{E}_x [g(x, X_1)] \quad \text{et} \quad h(x, y) := g(x, y) - f(x).$$

On applique le résultat établi à la première étape de cette preuve à f , et on procède de manière analogue pour la fonction h en montrant que

$$\mathbb{E} \left[\sum_{i=R_{n-1}^x+1}^{R_n^x} h(X_{i-1}, X_i) \right] = 0$$

par construction. ◇

Nous terminons ce chapitre par un théorème central limite qui précise la vitesse de convergence dans le théorème ergodique. Le schéma de démonstration est le même : en s'appuyant sur la propriété d'indépendance et d'identité de distribution des excursions, énoncée dans la proposition 5.21, nous appliquons maintenant le théorème central limite...

Théorème 7.4. *Soit X une chaîne de Markov irréductible et récurrente positive sur un espace dénombrable E , de matrice de transition P , et d'unique loi invariante ν . Soit $g : E \times E \rightarrow \mathbb{R}$ une fonction positive ou telle que $\mathbb{E}_\nu [|g|(X_0, X_1)] < \infty$. Supposons qu'il existe $x \in E$ tel que*

$$s(x)^2 := \mathbb{E}_x \left[\left(\sum_{i=1}^{R_1^x} g(X_{i-1}, X_i) - \mathbb{E}_\nu [g(X_0, X_1)] \right)^2 \right] < \infty.$$

Alors $\sigma^2 := \nu(x) s(x)^2$ est une constante (indépendante de x) et :

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n g(X_{i-1}, X_i) - \mathbb{E}_\nu [g(X_0, X_1)] \right) \rightarrow \mathcal{N}(0, \sigma^2) \text{ en loi.}$$

Démonstration. Les variables aléatoires

$$G(Y_n) := \sum_{i=R_{n-1}^x+1}^{R_n^x} g(X_{i-1}, X_i)$$

sont indépendantes pour $n \geq 1$, et de même distribution pour $n \geq 2$. La condition $s(x)^2 < \infty$ permet d'appliquer le théorème central limite pour obtenir

$$\sqrt{n} \left(\frac{1}{n} \sum_{k=1}^n G(Y_k) - \mathbb{E}_\nu [g(X_0, X_1)] \right) \longrightarrow \mathcal{N}(0, s(x)^2) \text{ en loi,}$$

où le fait que $G(Y_1)$ a une loi différente ne perturbe pas l'utilisation du théorème central limite puisque cette quantité est indépendante de n .

On note $h(x, y) := g(x, y) - \mathbb{E}_\nu [g(X_0, X_1)]$. Comme dans la preuve précédente, on note $N_n := \sum_{k=1}^n \mathbf{1}_{\{X_k=x\}}$, et on écrit que :

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n h(X_{i-1}, X_i) = \frac{1}{\sqrt{n}} \sum_{k=1}^{N_n} G(Y_k) + \frac{1}{\sqrt{n}} \sum_{i=R_{N_n}^x+1}^n h(X_{i-1}, X_i).$$

Pour le premier terme, on déduit du fait que $N_n/n \rightarrow \nu(x)$ p.s. que

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{N_n} G(Y_i) = \sqrt{\frac{N_n}{n}} \frac{1}{\sqrt{N_n}} \sum_{i=1}^{N_n} G(Y_i) \longrightarrow \mathcal{N}(0, \nu(x)s(x)^2) \text{ en loi.}$$

Quand au second terme, nous allons montrer qu'il converge vers 0 en probabilité, ce qui terminera la démonstration. En effet, notant $\xi_i := h(X_{i-1}, X_i)$, on calcule que pour $\varepsilon > 0$:

$$\begin{aligned} \mathbb{P} \left[\left| \sum_{i=R_{N_n}^x+1}^n \xi_i \right| \geq \varepsilon \sqrt{n} \right] &\leq \mathbb{P} \left[\sum_{i=R_{N_n}^x+1}^n |\xi_i| \geq \varepsilon \sqrt{n} \right] \\ &= \sum_{r=1}^{n-1} \mathbb{P} \left[\sum_{i=R_{N_n}^x+1}^n |\xi_i| \geq \varepsilon \sqrt{n}, R_{N_n}^x = n-r \right] \\ &= \sum_{r=1}^{n-1} \mathbb{P} \left[\sum_{i=n-r+1}^n |\xi_i| \geq \varepsilon \sqrt{n}, R_{N_n}^x = n-r \right] \\ &= \sum_{r=1}^{n-1} \mathbb{P}[X_{n-r} = x] \mathbb{P}_0 \left[\sum_{i=1}^r |\xi_i| \geq \varepsilon \sqrt{n}, R_1^x > r \right] \end{aligned}$$

d'après la propriété de Markov. Alors :

$$\begin{aligned} \mathbb{P} \left[\left| \sum_{i=R_{N_n}^x+1}^n \xi_i \right| \geq \varepsilon\sqrt{n} \right] &\leq \sum_{r=1}^{n-1} \mathbb{P}_0 \left[\sum_{i=1}^r |\xi_i| \geq \varepsilon\sqrt{n}, R_1^x > r \right] \\ &\leq \mathbb{E}_0 \left[\mathbf{1}_{\{\sum_{i=1}^r |\xi_i| \geq \varepsilon\sqrt{n}\}} \sum_{r \geq 1} \mathbf{1}_{\{R_1^x > r\}} \right] \\ &\leq \mathbb{E}_0 \left[R_1^x \mathbf{1}_{\{\sum_{i=1}^r |\xi_i| \geq \varepsilon\sqrt{n}\}} \right]. \end{aligned}$$

Comme $\mathbb{E}[R_1^x] < \infty$ et $R_1^x \mathbf{1}_{\{\sum_{i=1}^r |\xi_i| \geq \varepsilon\sqrt{n}\}} \rightarrow 0$, p.s. on déduit du théorème de convergence dominée que la dernière espérance converge vers zéro quand $n \rightarrow \infty$. \diamond

7.2 Convergences des lois marginales et apériodicité

Une conséquence immédiate du théorème (ergodique) 7.1 est que pour une chaîne irréductible récurrente positive, de loi invariante ν (donnée par le théorème 6.18, on a la convergence :

$$\frac{1}{n} \sum_{i=1}^n \mathbb{P}[X_i = x] = \frac{1}{n} \sum_{i=1}^n \pi_i(x) \rightarrow \nu(x), \quad \mathbb{P}_y - \text{p.s.}$$

pour tous $x, y \in E$. Ceci découle du théorème de convergence dominée.

Dans ce paragraphe, nous nous intéressons à la convergence de la suite $(\pi_n)_{n \geq 0}$ des lois marginales vers la loi invariante ν .

7.2.1 Apériodicité

L'exemple suivant montre qu'on ne peut espérer obtenir la convergence des lois marginales sans hypothèses supplémentaires.

Exemple 7.5. Les lois marginales π_n de la chaîne de Markov à deux états définie par la matrice de transition

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

oscillent entre la masse de Dirac en l'état 1 et la masse de Dirac en l'état 2. Ainsi $(\pi_n)_n$ ne converge pas.

L'exemple précédent est fondé sur une périodicité de la chaîne qui empêche clairement la convergence des lois marginales. Nous allons montrer que ce phénomène est le seul obstacle à la convergence des lois marginales.

Pour tout état $x \in E$, on définit

$$\mathbf{p}(x) := \text{PGCD}[I(x)] \quad \text{où} \quad I(x) := \{n \in \mathbb{N}^* : P^n(x, x) > 0\},$$

et PGCD désigne le *plus grand commun diviseur*.

Proposition 7.6. *Soit X une chaîne de Markov irréductible. Alors la fonction $\mathbf{p}(x) = \mathbf{p}_X$ est une constante.*

Démonstration. Soient $x, y \in E$ deux états tels que $x \longleftrightarrow y$, i.e. $P^i(x, y) > 0$ et $P^j(y, x) > 0$. Alors, une application directe de l'égalité de Chapman-Kolmogorov montre que $P^{i+j}(x, x) > 0$ et $P^{i+j+r}(x, x) > 0$ pour tout $r \in I(y)$. Alors $\mathbf{p}(x)$ divise $i + j$ et $i + j + r$, ainsi que la différence r de ces deux entiers. Comme r est arbitraire dans $I(y)$, on déduit que $\mathbf{p}(x)$ divise $\mathbf{p}(y)$. En inversant les rôles de x et y , on montre l'égalité $\mathbf{p}(x) = \mathbf{p}(y)$. \diamond

Définition 7.7. *Soit X une chaîne de Markov irréductible. On dit que X est apériodique si $\mathbf{p}_X = 1$.*

Avant d'énoncer le résultat principal de ce paragraphe, nous montrons une propriété intéressante des chaînes apériodiques.

Lemme 7.8. *Pour $x \in E$, les deux assertions suivantes sont équivalentes :*

- (i) $\mathbf{p}(x) = 1$,
- (ii) *il existe $\mathbf{n}(x) \in \mathbb{N}$ tel que $P^n(x, x) > 0$ pour tout $n \geq \mathbf{n}(x)$.*

Démonstration. L'implication (ii) \implies (i) est triviale. Pour l'implication inverse, on considère des entiers $n_1, \dots, n_k \in I(x)$ avec $\text{PGCD}[n_1, \dots, n_k] = 1$. Le théorème de Bezout assure l'existence de $q_1, \dots, q_k \in \mathbb{Z}$ tels que

$$\sum_{i=1}^k q_i n_i = 1 = a(x) - b(x) \quad \text{où} \quad a(x) := \sum_{i=1}^k q_i^+ n_i \quad \text{et} \quad b(x) := \sum_{i=1}^k q_i^- n_i.$$

Posons :

$$n(x) := b(x)^2 - 1 = (b(x) - 1)b(x) + b(x) - 1.$$

Alors pour tout $n \geq n(x)$, la division euclidienne de n par $b(x)$ s'écrit :

$$\begin{aligned} n &= db(x) + r \quad \text{avec} \quad d \geq r \quad \text{et} \quad 0 \leq r \leq b(x) - 1 \\ &= (d - r)b(x) + ra(x) = (d - r) \sum_{i=1}^k q_i^+ n_i + r \sum_{i=1}^k q_i^- n_i. \end{aligned}$$

L'égalité de Chapman-Kolmogorov montre que toute combinaison linéaire des $(n_i)_{1 \leq i \leq k}$, à coefficients dans \mathbb{N} , est dans $I(x)$. En particulier, la décomposition précédente montre que $n \in I(x)$. \diamond

Le Lemme 7.8 exprime clairement que l'apériodicité est une condition nécessaire pour la convergence des lois marginales. En effet, si $\mathbf{p}(x) \geq 2$ pour un certain point $x \in E$, alors pour tout $n \geq 1$, il existe un entier k_n tel que $P^{k_n}(x, x) = 0$, i.e. $\pi_{k_n}(x) = 0$, \mathbb{P}_x -p.s. et par conséquent la suite $(\pi_{k_n}(x))_n$ ne peut converger vers $\nu(x) > 0$.

7.2.2 Convergence et vitesse de convergence des lois marginales

Avec l'hypothèse supplémentaire d'apériodicité, nous allons maintenant énoncer et montrer le résultat de convergence des lois marginales vers la loi invariante.

Théorème 7.9. *Soit X une chaîne de Markov irréductible et récurrente positive d'unique loi invariante ν . Si X est de plus apériodique, alors :*

$$\pi_n(x) = \mathbb{P}[X_n = x] \longrightarrow \nu \quad \text{pour tout } x \in E.$$

Pour montrer ce théorème, nous isolons le résultat suivant qui a son propre intérêt.

Proposition 7.10. *Soient X^1 et X^2 deux chaînes de Markov indépendantes de même matrice de transition P irréductible apériodique. Alors la chaîne produit $Y := (X^1, X^2)$ est irréductible apériodique. Si de plus P est récurrente positive, il en est de même pour Y .*

Démonstration. Comme P est irréductible et apériodique, le lemme 7.8 s'applique à tout état $x \in E$. La matrice de transition de la chaîne produit Y est donnée par

$$(P \otimes P)(x_1, x_2, y_1, y_2) = P(x_1, y_1)P(x_2, y_2),$$

et d'après le caractère irréductible de X , il existe des entiers m_1, m_2 tels que $P^{m_i}(x_i, y_i) > 0$, $i = 1, 2$. On en déduit alors pour $n \geq \mathbf{n}(y_1) + \mathbf{n}(y_2) + m_1 + m_2$ que $P^n(x_i, y_i) \geq P_i^{m_i}(x_i, y_i)P^{n-m_i}(y_i, y_i) > 0$, et

$$(P \otimes P)^n(x_1, x_2, y_1, y_2) \geq P^n(x_1, y_1)P^n(x_2, y_2) > 0,$$

prouvant ainsi l'irréductibilité de la chaîne produit Y . L'apériodicité est immédiate puisqu'on a de même que $(P \otimes P)^n(x_1, x_2, x_1, x_2) > 0$ et $(P \otimes P)^{n+1}(x_1, x_2, x_1, x_2) > 0$, et par suite $\text{PGDCD}[I(x_1, x_2)] = 1$.

Enfin, si P est récurrente positive, elle admet une unique loi invariante ν , et il est immédiat de vérifier que $\nu \otimes \nu$ est une loi invariante pour la chaîne produit Y . Ceci prouve que Y est récurrente positive d'après le théorème 6.18. \diamond

Nous avons maintenant tous les ingrédients pour démontrer le résultat de convergence des lois marginales vers la loi invariante.

Preuve du théorème 7.9 Soient X^1 et X^2 deux chaînes indépendantes de même matrice de transition que la chaîne X , et introduisons la chaîne produit $Y := (X^1, X^2)$. Soit $T := \inf \{n \geq 0 : X_n^1 = X_n^2\}$. Notons que, bien que les deux chaînes ne soient pas issues de la même loi initiale, après le temps T , elles ont la même distribution. On calcule alors que

$$\begin{aligned} \mathbb{P}[X_n^1 = x] &= \mathbb{P}[X_n^1 = x, T > n] + \mathbb{P}[X_n^1 = x, T \leq n] \\ &= \mathbb{P}[X_n^1 = x, T > n] + \mathbb{P}[X_n^2 = x, T \leq n] \\ &\leq \mathbb{P}[T > n] + \mathbb{P}[X_n^2 = x]. \end{aligned}$$

Ceci montre que $\mathbb{P}[X_n^1 = x] - \mathbb{P}[X_n^2 = x] \leq \mathbb{P}[T > n]$. En inversant les rôles de X^1 et X^2 , on obtient alors $|\mathbb{P}[X_n^1 = x] - \mathbb{P}[X_n^2 = x]| \leq \mathbb{P}[T > n]$, et en supposant que X_0^2 est distribué selon la loi invariante ν , on a :

$$|\mathbb{P}[X_n^1 = x] - \nu(x)| \leq \mathbb{P}[T > n].$$

Pour conclure, on commence par remarquer que $\mathbb{P}[T > n] \rightarrow \mathbb{P}[T = \infty]$ quand $n \rightarrow \infty$, d'après le théorème de convergence monotone. D'après la proposition 7.10, la chaîne Y hérite la récurrence positive de la chaîne X . Par conséquent, $T = \inf_{x \in E} R_{x,x}(Y) < \infty$ p.s. où $R_{x,x}(Y)$ désigne le premier temps de retour de la chaîne Y au point (x, x) . \diamond

Nous terminons ce paragraphe par un résultat sur la vitesse de convergence dans le dernier théorème. Afin de motiver l'hypothèse du théorème 7.12 ci-dessous, nous commençons par considérer le cas d'un espace d'état fini, exploitant de manière cruciale le lemme 7.8.

Proposition 7.11. *Soit X une chaîne de Markov irréductible apériodique sur un espace d'état fini E . Alors sa matrice de transition P vérifie la Condition de Doeblin : il existe $k \in \mathbb{N}$, $\varepsilon > 0$ et une loi δ sur E tels que*

$$P^k(x, y) \geq \varepsilon \delta(y) \quad \text{pour tous } x, y \in E. \quad (7.4)$$

Démonstration. D'après le lemme 7.8 et le fait que E soit de cardinal fini, on déduit l'existence d'un entier $k \geq 1$ tel que $P^k(x, y) > 0$ pour tous $x, y \in E$. On pose alors :

$$\varepsilon := \sum_{y \in E} \min_{x \in E} P^k(x, y) \quad \text{et} \quad \delta(y) := \frac{1}{\varepsilon} \min_{x \in E} P^k(x, y),$$

et on vérifie immédiatement que la condition de Doeblin est vérifiée avec ces données. \diamond

Le résultat suivant est vrai pour un espace d'état dénombrable, non nécessairement fini.

Théorème 7.12. *Soit P une matrice de transition vérifiant la condition de Doeblin (7.4). Alors il existe une unique loi invariante $\nu \geq \varepsilon \delta$ vérifiant*

$$\sup_{x \in E} \sum_{y \in E} |P^n(x, y) - \nu(y)| \leq 2(1 - \varepsilon)^{\lfloor n/k \rfloor}.$$

Démonstration. Pour une mesure μ sur E , on note par $\|\mu\|_{\text{VT}} := \|\mu\|_{\ell^1} = \sum_{x \in E} |\mu(x)|$. Il s'agit de la variation totale de la mesure μ qui coïncide avec la norme ℓ^1 sur l'ensemble dénombrable E . Comme P est une matrice stochastique, toutes ses valeurs propres sont dans le disque unité, et on calcule pour une

mesure μ telle que $\mu(E) = 0$ que

$$\begin{aligned} \|\mu P^n\|_{\text{VT}} &= \|\mu(P^k)^{\lfloor n/k \rfloor} P^{n-k\lfloor n/k \rfloor}\|_{\text{VT}} \\ &\leq \|\mu(P^k)^{\lfloor n/k \rfloor}\|_{\text{VT}} \\ &= \sum_{y \in E} \left| \sum_{x \in E} (\mu(P^k)^{\lfloor n/k \rfloor - 1})(x) (P^k(x, y) - \varepsilon \delta(y)) \right| \end{aligned}$$

où nous avons utilisé le fait que $\mu(E) = 0$ implique $\mu Q(E) = 0$ pour toute matrice stochastique Q . EN utilisant la condition de Doeblin (7.4), on obtient alors :

$$\begin{aligned} \|\mu P^n\|_{\text{VT}} &\leq \sum_{x \in E} |\mu(P^k)^{\lfloor n/k \rfloor - 1}(x)| \sum_{y \in E} (P^k(x, y) - \varepsilon \delta(y)) \\ &= (1 - \varepsilon) \|\mu(P^k)^{\lfloor n/k \rfloor - 1}\|_{\text{VT}} \\ &\leq \dots \leq (1 - \varepsilon)^{\lfloor n/k \rfloor} \|\mu\|_{\text{VT}}. \end{aligned}$$

En particulier, pour toutes lois initiales π_0 et π'_0 , on bien sûr $(\pi_0 - \pi'_0)(E) = 0$ et :

$$\|\pi_0 P^n - \pi'_0 P^n\|_{\text{VT}} \leq (1 - \varepsilon)^{\lfloor n/k \rfloor} \|\pi_0 - \pi'_0\|_{\text{VT}} \leq 2(1 - \varepsilon)^{\lfloor n/k \rfloor}.$$

Ceci implique l'unicité de la loi invariante. Par ailleurs, en choisissant $\pi'_0 = \pi_0 P$, on voit que $(\pi_0 P^n)_{n \geq 0}$ est une suite de Cauchy dans $\ell^1(E)$, qui est un espace complet. Elle y converge donc vers une limite ν qui est une loi invariante par continuité. \diamond

7.2.3 Application à l'algorithme PageRank de Google

Le principe du moteur de recherche Google est le suivant.

- Les pages du web sont indexées par un "rampeur du web" (web crawler), un logiciel qui sonde et conserve automatiquement le contenu de chaque site web, dans un catalogue facile à parcourir.
- Ceci donne lieu à un catalogue indexé de pages web de taille gigantesque de l'ordre de $25 \cdot 10^9$ d'après Google.
- Les pages se font attribuer un ordre de priorité par la procédure PageRank qui a été développée par Larry Page et Sergey Brin en 1998. En quelques mots, cet ordre de priorité est établi en fonction de nombreux critères et principalement en fonction des liens externes (popularité de liens) pointant vers la page web, ainsi que des liens que cette dernière fait vers elle-même (liens internes).
- Le logiciel Google fournit l'ensemble des pages qui répondent à une requête donnée suivant l'ordre de priorité établi par PageRank.

Il y a quelques années, les utilisateurs devaient attendre plus longtemps les retours des moteurs de recherche. Quand ces retours étaient enfin disponibles, la liste des pages indiquées contenait souvent des informations peu importantes, et des liens inutiles apparaissaient invariablement parmi les premiers retours, alors que des liens utiles étaient enterrés dans les profondeurs de la liste des retours...

L'information disponible sur le web n'est pas structurée sur le modèle des bases de données traditionnelles, elle est plutôt auto-organisée. La taille gigantesque du web dépasse de très loin les techniques traditionnelles de recherche documentaire. L'approche originale implémentée dans la procédure PageRank a complètement révolutionné le fonctionnement des moteurs de recherche, et a permis de mieux répondre aux besoins des utilisateurs de moteurs de recherche.

PageRank a été développé à l'Université de Stanford dans le cadre d'un projet de recherche démarré en 1995 par Larry Page (d'où le nom!) et, plus tard, Sergey Brin, qui sont tous les deux diplômés d'un *Master on Computer Science* à Stanford. Cette aventure extraordinaire a conduit Page et Brin à fonder Google en 1998 mais, pour la petite histoire, ils n'ont jamais achevé le programme doctoral (PhD) de Stanford. Les recherches qui ont abouti au développement de la technologie de PageRank ont été financées en partie par la National Science Foundation. Le premier brevet (intitulé Method for Node Ranking in a Linked Database), datant de janvier 1997, est la propriété de l'Université Stanford, qui a octroyé une licence exclusive de cette technologie à Google en 1998, et ce jusqu'en 2011. L'université a reçu 1,8 Millions d'actions Google, en échange de l'utilisation du brevet; qu'elle a vendues en 2005 pour la somme de 336 Millions de Dollars US.

Nous allons décrire le principe de fonctionnement de PageRank en gardant à l'esprit que plusieurs améliorations ont été introduites depuis afin de répondre aux divers détournements des utilisateurs qui ont tenté de gonfler artificiellement leur indice de popularité. Google a ainsi réagi en instaurant des filtres et en intégrant des critères qualitatifs à l'analyse des liens (sémantique, confiance, comportement des utilisateurs).

Soit $E = (1, \dots, N)$ l'ensemble des N pages du web. Chaque page i contient des liens vers d'autres pages $I_i := \{i_k, 1 \leq k \leq d_i\} \subset E$. Pour les pages ne contenant aucun lien vers d'autres pages, on pose par convention $I_i = E$. On attribue alors un poids uniforme pour le déplacement d'une page à une autre :

$$L(i, j) = \frac{1}{d_i} \mathbf{1}_{\{j \in I_i\}} \quad \text{pour tout } i, j \in E.$$

Ceci modélise le comportement d'un navigateur ivre sur le web qui se déplace d'une page à une autre en choisissant au hasard un lien existant sur la page courante. Si la page courante ne contient aucun lien, il choisit une page au hasard. Si on note par X_n le numéro de la page visitée par notre internaute à la date n , on obtient une chaîne de Markov homogène de matrice L . En pratique, la matrice L est remise à jour régulièrement afin de prendre en compte l'évolution de la structure du web.

Le principe de PageRank est d'attribuer un rang r_i à chaque page $i \in E$ en fonction de l'importance relative des liens pointant vers i , et du rang de la page contenant un tel lien :

$$r_i = \sum_{j : i \in I_j} \frac{r_j}{d_j} \quad \text{pour tout } i \in E.$$

Pour étudier l'existence et l'unicité d'un tel vecteur de rang r , on remarque d'après la définition des $L_{i,j}$ que l'écriture matricielle de l'équation définissant le vecteur de rangs est $r = rL$, où r est un vecteur ligne et L est la matrice stochastique définie par les composantes $L_{i,j}$. Nous sommes ainsi ramenés à un problème de recherche de loi invariante pour une chaîne de Markov. Si on revient à notre internaute ivre, la question du rang se traduit par l'existence d'un état d'équilibre qu'il finirait par atteindre après un nombre infini de déplacements d'une page à une autre.

Comme l'espace d'état est fini, nous sommes assurés qu'il existe une loi invariante. Mais rien n'en garantit l'unicité, puisque la chaîne n'est en général pas irréductible. Par ailleurs, même en supposant que la chaîne soit irréductible, le calcul pratique du vecteur de rang est typiquement fait par le biais des itérations $r_n = Lr_{n-1}$ à partir d'un point initial arbitraire r_0 . En d'autres termes, ceci pose la question de la convergence des lois marginales. Enfin, en vue de l'application industrielle, la vitesse de convergence vers la loi marginale est un critère de première importance. Pour toutes ces raisons, Brin et Page ont introduit la matrice de Google G de composantes :

$$G_{i,j} := \alpha L_{i,j} + (1 - \alpha) \frac{1}{N} \quad \text{pour tous } 1 \leq i, j \leq N,$$

où α est un paramètre dans $]0, 1[$ appelé facteur de *damping*. Cette modification de la matrice peut s'interpréter par le fait que l'internaute ivre choisit avec une probabilité α de se diriger vers un de ses liens favoris (bookmarks) que l'on suppose distribués uniformément sur l'ensemble des pages.

Alors la matrice G est stochastique, irréductible, et vérifie même la condition de Doeblin puisque $G_{i,j} > 0$ pour tous $i, j \in E$. En utilisant la matrice de Google G au lieu de la matrice des liens L , nous sommes ainsi assurés de l'existence et de l'unicité d'une loi invariante, que l'on définit comme le vecteur de rang r , et de la convergence de la procédure itérative $r_n = Gr_{n-1}$ vers r à une vitesse géométrique donnée par le théorème 7.12.

Enfin, il reste à discuter le choix de la valeur du paramètre α . Il est clair que le souci de rapidité de la convergence de l'algorithme nous pousse à choisir α proche de 0, mais ceci conduirait à une matrice G qui ne reflète pas la vraie structure du web. La valeur exacte du paramètre α est un secret gardé de Google, mais il semble qu'elle se situe autour de 0,85.

A titre d'exemple, pour $N = 6$ et

$$L = \begin{pmatrix} 0 & 1/3 & 1/3 & 1/3 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 1/4 & 1/4 & 0 & 1/4 & 1/4 & 0 \\ 1/3 & 0 & 0 & 0 & 1/3 & 1/3 \\ 0 & 1/3 & 0 & 1/3 & 0 & 1/3 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{pmatrix},$$

on trouve un vecteur de rang donné par :

$$r = (0,2066, 0,1770, 0,1773, 0,1770, 0,1314, 0,1309).$$

Sous ces conditions, supposons que l'utilisateur de Google rentre une requête avec deux mots clés *mot 1* ou *mot 2*, que *mot 1* soit présent dans les pages 3, 2 et 6, et que *mot 2* soit présent dans les pages 1, 3. Alors, Google retourne comme première adresse la page 1.

7.3 Algorithme de Hastings-Metropolis

Soit ν une mesure de probabilité sur un espace d'état au plus dénombrable E . L'algorithme de Hastings-Metropolis propose une méthode pour simuler des variables aléatoires de loi ν basée sur l'interprétation de ν comme loi invariante d'une chaîne de Markov à déterminer. Notons tout de suite qu'un avantage très important, et très utile pour les applications, de l'algorithme de Metropolis est qu'il ne requiert la connaissance de la loi ν qu'à une constante près :

$$\nu = \frac{\mu}{\|\mu\|_{\ell^1}},$$

où μ est une mesure positive finie.

7.3.1 Description de l'algorithme

Le premier ingrédient est la donnée d'une *matrice de proposition*

$$\begin{aligned} Q \text{ stochastique, irréductible, et vérifiant :} \\ Q(x, y) > 0 \iff Q(y, x) > 0, \quad x, y \in E. \end{aligned} \quad (7.5)$$

qu'on considèrera comme la matrice de transition d'une chaîne de Markov $(Y_n)_{n \geq 0}$.

Le deuxième ingrédient est la donnée d'une fonction

$$h : \mathbb{R}_+ \setminus \{0\} \longrightarrow]0, 1] \quad \text{vérifiant} \quad h(y) = yh(1/y) \quad \text{pour tout } y > 0, \quad (7.6)$$

à partir de laquelle on définit la *probabilité de rejet* de la valeur courante :

$$R(x, y) := h\left(\frac{\mu(y)Q(y, x)}{\mu(x)Q(x, y)}\right) \quad \text{pour } x, y \in E \text{ tels que } x \neq y \text{ et } Q(x, y) \neq 0.$$

L'exemple typique, qui a été initialement proposé par Metropolis et ses co-auteurs en 1953, est :

$$h(y) := 1 \wedge y \quad \text{pour } y \geq 0. \quad (7.7)$$

Observons que les données ci-dessus ne font intervenir que la mesure positive μ , et non sa normalisation en mesure de probabilité ν . L'algorithme de Metropolis est défini par la simulation de la chaîne de Markov $(X_n)_{n \geq 0}$ suivante :

1. choisir une variable initiale X_0 arbitraire,
2. étant donnée la v.a. X_{n-1} ,
 - on simule la v.a. Y_n selon la loi $Q(X_{n-1}, \cdot)$,

- on accepte Y_n (et on rejette X_{n-1}) avec probabilité $R(X_{n-1}, Y_n)$, i.e. on simule une variable aléatoire uniforme U_n indépendante de (X_{n-1}, Y_n) , et on pose :

$$X_n := Y_n \mathbf{1}_{\{U_n \leq R(X_{n-1}, Y_n)\}} + X_{n-1} \mathbf{1}_{\{U_n > R(X_{n-1}, Y_n)\}}$$

La matrice de transition associée à la chaîne de Markov X est donnée par

$$\begin{aligned} P(x, y) &= \mathbb{P}[U_n \leq R(x, y)]Q(x, y) = R(x, y)Q(x, y) \text{ pour } x \neq y, \\ P(x, x) &= 1 - \sum_{x \neq y \in E} P(x, y), \end{aligned} \quad (7.8)$$

où on pose par convention $P(x, y) = 0$ pour $x \neq y$ vérifiant $Q(x, y) = 0$. La matrice P a les propriétés suivantes :

- par construction, P est une matrice stochastique, et elle hérite la propriété d'irréductibilité de Q ,
- la propriété $h(y) = yh(1/y)$ implique que P est réversible par rapport à ν :

$$\begin{aligned} \nu(y)P(y, x) &= \nu(y)Q(y, x)h\left(\frac{\mu(x)Q(x, y)}{\mu(y)Q(y, x)}\right) \\ &= \nu(x)Q(x, y)\left(\frac{\mu(x)Q(x, y)}{\mu(y)Q(y, x)}\right)^{-1}h\left(\frac{\mu(x)Q(x, y)}{\mu(y)Q(y, x)}\right) \\ &= \nu(x)Q(x, y)h\left(\frac{\mu(y)Q(y, x)}{\mu(x)Q(x, y)}\right) = \nu(x)P(x, y). \end{aligned}$$

Par suite, ν est l'unique loi invariante de $(X_n)_{n \geq 0}$.

- Enfin, si Q est apériodique ou $h < 1$, alors P est apériodique. Supposons que P n'est pas apériodique, et cherchons une contradiction. Comme P est irréductible, tous les états ont la même période, et on déduit que pour tout état $x \in E$:

$$0 = 1 - \sum_{x \neq y \in E} P(x, y) = Q(x, x) + \sum_{x \neq y \in E} Q(x, y)(1 - R(x, y)).$$

Comme tous les termes de droite sont positifs, on en déduit que

$$Q(x, x) = 0 \text{ et } Q(x, y)(1 - R(x, y)) = 0 \text{ pour tous } x, y \in E, x \neq y.$$

- Si $h < 1$ alors $R < 1$ et $Q(x, x) = 0$, $Q(x, y)(1 - R(x, y)) = 0$ pour tous $x, y \in E$, $x \neq y$, ce qui est contradictoire avec le fait que Q soit une matrice stochastique.

- Supposons maintenant que Q est apériodique. Alors pour $y \neq x$, $P(x, y) = 0$ dès que $Q(x, y) = 0$ par construction. Si $Q(x, y) > 0$, alors $R(x, y) = 1$. Par ailleurs, d'après la condition (7.5) sur la matrice de proposition Q , on a $Q(y, x) > 0$ et par suite $R(y, x) = 1$ aussi. Alors Q est réversible par rapport à ν , et comme $P(x, y) = 0$ dès que $Q(x, y) = 0$, on voit que $P = Q$, et P hérite l'apériodicité de Q .

Ainsi, sous les conditions ci-dessus, l'algorithme d'Hastings-Metropolis permet de simuler la loi ν comme loi invariante de la chaîne irréductible apériodique $(X_n)_{n \geq 0}$. Le théorème 7.12 nous garantit que la convergence des lois marginales vers ν a lieu à une vitesse géométrique.

Remarque 7.13. Soit Q une matrice de proposition stochastique et irréductible. Supposons de plus que Q est symétrique, i.e. $Q(x, y) = Q(y, x)$. Alors elle vérifie les conditions (7.5). Dans ce cas, on peut montrer que *si la loi ν n'est pas constante, alors P est apériodique*. Pour obtenir ce résultat, il suffit de suivre les arguments de la preuve précédente jusqu'à arriver, par symétrie de Q à la conclusion que $\nu(x) = \nu(y)$ pour tous $x \neq y$ avec $Q(x, y) > 0$. On déduit ensuite simplement de l'irréductibilité que ν est constante, contredisant notre hypothèse.

Remarque 7.14. Dans l'algorithme initial proposé par Metropolis et ses co-auteurs, la fonction h est donnée par (7.7) et la matrice de proposition Q est symétrique, ce qui simplifie l'étape de rejet puisqu'alors

- on rejette systématiquement X_{n-1} dès que $\nu(Y_n) \geq \nu(X_{n-1})$,
- on rejette X_{n-1} avec probabilité $\nu(Y_n)/\nu(X_{n-1})$ si $\nu(Y_n) < \nu(X_{n-1})$.

Ainsi, si la probabilité de rejet $\nu(Y_n)/\nu(X_{n-1})$ devient trop petite, les chances d'acceptation de la nouvelle observation sont trop petites, et l'algorithme est peu performant. Plus généralement, si on veut que la probabilité de rejet soit proche de 1, on doit choisir Q "pas trop loin" de la réversibilité par rapport à ν , et donc "pas trop loin" de P .

Remarque 7.15. On dit que l'algorithme de Hastings-Metropolis est indépendant si $Q(x, y) = Q(y, x)$ est indépendant de x . Dans ce cas, la variable Y_n est indépendante de X_n , mais son rejet ou acceptation en dépend.

Remarque 7.16. Si l'ensemble des états E est muni d'une structure de graphe, on choisit la matrice de proposition Q comme la matrice de transition d'une marche aléatoire symétrique sur ce graphe.

7.3.2 Simulation de la loi de Gibbs

Soit $T > 0$ une constante et $V : E \rightarrow \mathbb{R}$ une fonction donnée. La mesure de Gibbs associée à T et V est définie par

$$\Gamma_T^V(x) := \frac{1}{Z_T} \exp\left(\frac{-V(x)}{T}\right), \quad x \in E, \quad \text{où} \quad Z_T := \sum_{y \in E} \exp\left(\frac{-V(y)}{T}\right).$$

En physique statistique, V est appelée *fonction d'énergie* et T la *température*, et la loi de Gibbs apparaît naturellement comme loi d'équilibre thermodynamique du système qui a les caractéristiques (V, T) . En effet, il est facile de vérifier que Γ_T^V est solution du problème de maximisation de l'entropie

$$-\sum_{x \in E} \nu(x) \ln \nu(x) \quad \longrightarrow \quad \max !$$

parmi les loi ν qui sont d'énergie moyenne donnée $\nu V = E$, où $E = E(T)$ est une constante dépendant de T .

Dans les applications, E est de cardinal très grand, et le facteur de normalisation Z_T est difficile à calculer. L'algorithme de Hastings-Metropolis prend alors toute son importance ici, puisqu'il permet de simuler la loi de Gibbs Γ_T^V sans avoir à calculer la constante de normalisation Z_T . En effet, étant donnée une matrice de proposition Q et une fonction h vérifiant (7.5)-(7.6), la probabilité de rejet est simplement donnée par

$$R(x, y) = h \left(\exp \left(\frac{-(V(y) - V(x))}{T} \right) \frac{Q(y, x)}{Q(x, y)} \right) \quad \text{pour } x, y \in E.$$

Exemple 7.17. (*Modèle d'Ising*) L'observation expérimentale a mis en évidence que les matériaux ferromagnétiques subissaient des transitions étonnantes et discontinues dans leur comportement. Par exemple, un matériau aimanté suffisamment chauffé perd subitement ses propriétés magnétiques. De plus, ce phénomène apparaît de façon abrupte à une certaine température appelée température de Curie. On appelle souvent de tels changements de structures des transitions de phases, suivant la même terminologie que pour le passage du gazeux au liquide ou du liquide au solide, correspondants eux aussi à des températures spécifiques à la matière étudiée.

Le physicien Ising s'est intéressé à l'étude de ce problème. Devant la difficulté du problème, il a commencé par étudier le cas unidimensionnel qui ne possède malheureusement aucune propriété de transition de phase. Ce n'est que quelques années plus tard, grâce à l'algorithme de simulation d'Hastings-Metropolis, qu'on s'est rendu compte que le modèle permettait de produire des phénomènes de transition de phase en dimension supérieure à 1.

Le modèle d'Ising est une version simplifiée d'un système magnétique. On considère N particules positionnées sur une grille régulière, comme les atomes d'un cristal. On affecte à chaque particule un spin, correspondant à son orientation. Les orientations possibles sont 1 ou -1 . L'espace d'état est donc $\{-1, 1\}^N$, et on définit l'énergie du système en tout point $x \in E$ par :

$$V(x) = -\frac{1}{2} \sum_{x, y \in E} J_{x, y} x y - \sum_{x \in E} B_x x,$$

où le vecteur $B = (B_x)_{x \in E}$ modélise l'effet d'un champs extérieur d'aimantation, et la matrice $J = (J_{x, y})_{x, y \in E}$ est typiquement choisie symétrique à diagonale nulle et modélise l'interaction entre les particules : $J_{x, y} > 0$ signifie que l'interaction est attractive du type *ferromagnétique*, puisque la maximisation de l'entropie conduit à favoriser les configurations où les spins en x et y ont des orientations identiques. De même, $J_{x, y} < 0$ signifie que l'interaction est répulsive du type *antiferromagnétique*.

Dans cet exemple, la constante de normalisation, appelée *fonction de partition*, est très difficile à calculer car l'espace d'état $\{-1, 1\}^N$ est gigantesque. Comme exemple de matrice de proposition Q , on peut considérer celle qui à partir d'une position x choisit uniformément l'état suivant en changeant l'orientation d'exactly un spin.

7.4 Recuit simulé

7.4.1 Description de la méthode

Soit E un espace d'état fini, et $V : E \rightarrow \mathbb{R}$ donnée. On s'intéresse à la recherche des minima globaux de V

$$\mathcal{V}_{\min} = \{x : V(x) \leq V(y) \text{ pour tout } y \in E\},$$

dans des cas où des minima locaux peuvent conduire les méthodes habituelles à être piégées. Dans cette littérature, les minima locaux sont appelés *puits de potentiel*.

L'idée essentielle est l'observation que la mesure de Gibbs associée à l'énergie V et la température T vérifie :

$$\lim_{T \rightarrow 0} \Gamma_T^V(x) = \frac{\mathbf{1}_{\mathcal{V}_{\min}}(x)}{\text{card}(\mathcal{V}_{\min})}. \quad (7.9)$$

Ceci suggère de se donner un *schéma de température* $(T_n)_{n \geq 1}$, qui converge vers 0, et d'utiliser l'algorithme de Hastings-Metropolis avec la température T_n à chaque étape n . Ceci définit une chaîne de Markov non homogène $(X_n)_{n \geq 0}$ de matrices de transition $(P_n)_{n \geq 0}$ définies par (7.8) en remplaçant T par T_n . Pour toute loi initiale π_0 , la loi marginale de X_n est notée π_n , et la loi de Gibbs $\Gamma_{T_n}^V$ est la loi invariante d'une chaîne de Markov homogène (fictive) de matrice de transition P_n . On a alors

$$\pi_n = \pi_0 P_1 \dots P_n \text{ et } \Gamma_{T_n}^V = \Gamma_{T_n}^V P_n \text{ pour tout } n \geq 1. \quad (7.10)$$

Il faut alors faire attention à la vitesse de convergence vers 0 du schéma de températures. En effet, dans le contexte présent, nous avons les deux asymptotiques $n \rightarrow \infty$ et $T_n \rightarrow 0$, et on doit choisir convenablement la vitesse de convergence de T_n vers 0 afin de garantir la convergence de la chaîne vers l'ensemble \mathcal{V}_{\min} .

La terminologie provient de l'analogie avec la métallurgie. Le recuit est une technique qui consiste à faire fondre plusieurs fois un métal, puis à le laisser refroidir lentement afin d'en améliorer les qualités mécaniques.

7.4.2 Un résultat de convergence

Nous allons montrer un résultat de convergence simple, et nous renvoyons vers [7] pour des résultats plus élaborés. Pour la procédure de rejet, on prendra la fonction h de (7.7). Alors, étant donnée une matrice de proposition Q symétrique, la matrice de transition de la chaîne est donnée par

$$P_n(x, y) = Q(x, y) e^{-(V(y) - V(x))^+ / T_n} \text{ pour } x \neq y.$$

Théorème 7.18. *Soit Q une matrice de proposition symétrique et vérifiant la condition de Doeblin sur l'espace d'état fini E : il existe $k \in \mathbb{N}$, $\varepsilon > 0$ et une loi δ sur E tels que :*

$$Q^k(x, y) \geq \varepsilon \delta(y) \quad \text{pour tous } x, y \in E. \quad (7.11)$$

Soit $T_n = (b \ln n)^{-1}$. Alors, il existe $b_0 > 0$ tel que pour tout $b \leq b_0$, $\eta > 0$ et toute loi initiale π_0 , on a :

$$\lim_{n \rightarrow \infty} \|\pi_n - \Gamma_{T_n}^V\|_{\ell^1} = 0 \quad \text{et} \quad \lim_{n \rightarrow \infty} \mathbb{P} \left[V(X_n) \geq \eta + \min_E V \right] = 0.$$

Démonstration. Remarquons d'abord que les mesures de Gibbs sont invariantes par addition d'une constante, i.e. $\Gamma_T^V = \Gamma_T^{V+c}$ pour toute constante $c \in \mathbb{R}$. On peut alors supposer sans perte de généralité que l'énergie $V \geq 0$. Pour alléger les notations, on écrit simplement $\Gamma_n := \Gamma_{T_n}^V$ et $\|\cdot\| := \|\cdot\|_{\ell^1}$.

1 En utilisant (7.10), on obtient pour des entiers n et r :

$$\begin{aligned} \|\pi_{n+r} - \Gamma_{n+r}\| &= \left\| (\pi_n - \Gamma_n) P_{n+1} \dots P_{n+r} + \sum_{j=n}^{n+r-1} (\Gamma_j - \Gamma_{j+1}) P_{j+1} \dots P_{n+r} \right\| \\ &\leq \|(\pi_n - \Gamma_n) P_{n+1} \dots P_{n+r}\| + \sum_{j=n}^{n+r-1} \|(\Gamma_j - \Gamma_{j+1}) P_{j+1} \dots P_{n+r}\| \\ &\leq \|(\pi_n - \Gamma_n) P_{n+1} \dots P_{n+r}\| + \sum_{j=n}^{n+r-1} \|(\Gamma_j - \Gamma_{j+1})\|. \end{aligned} \quad (7.12)$$

Dans les deux étapes suivantes, nous traitons les deux termes de droite séparément.

2 Pour tout n , on calcule par définition de la mesure de Gibbs :

$$\begin{aligned} \|\Gamma_{j+1} - \Gamma_j\| &\leq \sum_{x \in E} \left| \frac{1}{Z_{T_{j+1}}} e^{-V(x)/T_{j+1}} - \frac{1}{Z_{T_j}} e^{-V(x)/T_j} \right| \\ &\leq \frac{1}{Z_{T_{j+1}}} \sum_{x \in E} \left| e^{-V(x)/T_{j+1}} - e^{-V(x)/T_j} \right| \\ &\quad + \left| \frac{1}{Z_{T_{j+1}}} - \frac{1}{Z_{T_j}} \right| \sum_{x \in E} e^{-V(x)/T_j} \\ &= \frac{1}{Z_{T_{j+1}}} \sum_{x \in E} \left| e^{-V(x)/T_{j+1}} - e^{-V(x)/T_j} \right| + \frac{1}{Z_{T_{j+1}}} |Z_{T_j} - Z_{T_{j+1}}| \\ &\leq \frac{2}{Z_{T_{j+1}}} \sum_{x \in E} \left| e^{-V(x)/T_{j+1}} - e^{-V(x)/T_j} \right| \\ &= \frac{2}{Z_{T_{j+1}}} \sum_{x \in E} e^{-V(x)/T_{j+1}} \left| 1 - e^{-V(x)(T_j^{-1} - T_{j+1}^{-1})} \right|. \end{aligned}$$

Comme $T_{j+1} \leq T_j$, ceci donne :

$$\begin{aligned} \|\Gamma_{j+1} - \Gamma_j\| &\leq \frac{2}{Z_{T_{j+1}}} \sum_{x \in E} e^{-V(x)/T_{j+1}} V(x) \left(\frac{1}{T_j} - \frac{1}{T_{j+1}} \right) \\ &\leq 2\|V\|_\infty \left(\frac{1}{T_j} - \frac{1}{T_{j+1}} \right). \end{aligned} \quad (7.13)$$

3 Pour le premier terme de droite dans l'inégalité (7.12), nous montrons maintenant que pour $r = k$, l'entier de la condition de Doeblin, on a l'estimation :

$$\|(\pi_n - \Gamma_n)P_{n+1} \dots P_{n+k}\| \leq (1 - (n+k)^{-bv}) \|\pi_n - \Gamma_n\|, \quad (7.14)$$

où b est la constante qui définit le schéma de température $(T_n)_{n \geq 1}$ et

$$v := \max_{x, y \in E} (V(x) - V(y)) \mathbf{1}_{\{Q(x, y) > 0\}}.$$

En effet, pour $x \neq y$ dans E et $j \in \mathbb{N}$, on a :

$Q(x, y) \geq P_j(x, y) = e^{-(V(y) - V(x))^+ / T_j} Q(x, y) \geq e^{-v/T_j} Q(x, y) = j^{-bv} Q(x, y)$,
 et $P_j(x, x) = 1 - \sum_{y \neq x} P_j(x, y) \geq 1 - \sum_{y \neq x} Q(x, y) = Q(x, x) \geq j^{-bv} Q(x, y)$.
 Ainsi $P_j(x, y) \geq j^{-bv} Q(x, y)$ pour tous $x, y \in E$, et on déduit de la condition de Doeblin vérifiée par Q que

$$P_{n+1} \dots P_{n+k}(x, y) \geq Q^k(x, y) \prod_{j=n+1}^{n+k} j^{-bv} \geq (n+k)^{-bv} \varepsilon \delta(y),$$

prouvant que la matrice de transition $\bar{P} := P_{n+1} \dots P_{n+k}$ vérifie la condition de Doeblin avec la même loi δ et une constante $\bar{\varepsilon} := \varepsilon(n+k)^{-bv}$. En répétant l'argument de la démonstration du théorème 7.12 pour la matrice \bar{P} , on voit alors que $\|(\pi_n - \Gamma_n)P_{n+1} \dots P_{n+k}\| \leq (1 - \bar{\varepsilon}) \|\pi_n - \Gamma_n\|$.

4 Dans cette étape, nous allons montrer que

$$\|\pi_{ki} - \Gamma_{ki}\| \longrightarrow 0 \quad \text{quand } i \rightarrow \infty, \quad (7.15)$$

où k est l'entier appraisant dans la condition de Doeblin. Pour celà, on considère $r = k$ et $n = ki$, $i \in \mathbb{N}$, et on commence par déduire de (7.12)-(7.13)-(7.14) que

$$\|\pi_{(i+1)k} - \Gamma_{(i+1)k}\| \leq (1 - \alpha_i) \|\pi_{ik} - \Gamma_{ik}\| + \beta_i,$$

où les α_i, β_i sont donnés par :

$$\alpha_i := \varepsilon((i+1)k)^{-bv} \quad \text{et} \quad \beta_i := 2b\|V\|_\infty \ln \left(\frac{i+1}{i} \right).$$

Alors :

$$\begin{aligned} \|\pi_{ik} - \Gamma_{ik}\| \prod_{j=0}^i (1 - \alpha_j)^{-1} &\leq \|\pi_{(i-1)k} - \Gamma_{(i-1)k}\| \prod_{j=0}^{i-1} (1 - \alpha_j)^{-1} + \beta_i \prod_{j=0}^i (1 - \alpha_j)^{-1} \\ &\leq \dots \leq \|\pi_0 - \Gamma_0\| + \sum_{l=1}^i \beta_l \prod_{j=0}^l (1 - \alpha_j)^{-1}, \end{aligned}$$

et on obtient, en notant $A_l := \prod_{j=0}^l (1 - \alpha_j)$,

$$\|\pi_{ik} - \Gamma_{ik}\| \prod_{j=0}^i (1 - \alpha_j)^{-1} \leq \frac{\|\pi_0 - \Gamma_0\|}{A_i} + \frac{1}{A_i} \sum_{l=1}^i \frac{\beta_l}{\alpha_l} (A_l - A_{l-1}).$$

D'après la définition des α_i , on voit que pour

$$b \leq b_1 := (vk)^{-1}, \quad (7.16)$$

la serie $\sum \alpha_i$ diverge, et par suite $A_i \rightarrow \infty$ quand $i \rightarrow \infty$. On calcule aussi directement que $\alpha_i/\beta_i \sim i^{bvk-1} \rightarrow 0$ pour $b < b_1$. Pour tout $\gamma > 0$, la suite positive $(\alpha_i/\beta_i)_i$ est bornée (par M) et $\alpha_i/\beta_i \leq \gamma$ à partir d'un certain rang i_0 , et par suite :

$$\frac{1}{A_i} \sum_{l=1}^i \frac{\beta_l}{\alpha_l} (A_l - A_{l-1}) \leq \frac{1}{A_i} (MA_{i_0} + \varepsilon(A_i - A_{i_0})) \rightarrow 0$$

quand $i \rightarrow \infty$, terminant la preuve de (7.15).

5 Pour $n = ki + r$, $i \in \mathbb{N}$ et $r = 1, \dots, k-1$, on déduit de l'inégalité (7.12), de l'estimation (7.13), et du fait que le spectre de la matrice de transition est borné par 1, que

$$\|\pi_{ki+r} - \Gamma_{ki+r}\| \leq \|\pi_{ki} - \Gamma_{ki}\| + 2\|V\|_\infty \ln \left(\frac{ki+r}{ki} \right) \rightarrow 0,$$

terminant ainsi la preuve de la convergence de $\pi_n - \Gamma_n$ vers zero. Enfin, pour $\eta > 0$, on obtient par l'inégalité de Chebyshev que

$$\begin{aligned} \mathbb{P} \left[V(X_n) - \min_E V \geq \eta \right] &\leq \eta^{-1} \mathbb{E} \left[V(X_n) - \min_E V \right] \\ &= \eta^{-1} \left(\pi_n V - \min_E V \right) \\ &= \eta^{-1} \left(\Gamma_n V - \min_E V \right) + \eta^{-1} (\pi_n V - \Gamma_n V). \end{aligned}$$

Le premier terme à droite de la dernière égalité tend vers 0 d'après (7.9), et le second terme tend vers 0 d'après la convergence de $\pi_n - \Gamma_n$ vers zero. \diamond

7.4.3 Application au voyageur de commerce

Un voyageur de commerce doit visiter N clients, dans N villes différentes, et revenir à son point de départ. Son objectif est de minimiser le trajet à parcourir. Une méthode de recherche systématique du minimum consisterait à explorer tous les chemins possibles et à comparer leur longueur. Ceci conduirait à une complexité numérique gigantesque puisque l'ensemble des trajets possibles est donné par l'ensemble des permutations des N villes, soit $N!$, qui croît trop vite avec N . Par exemple, pour 30 villes, le nombre de trajets est de l'ordre de 10^{30} .

D'autres méthodes du type *descente de gradient* risquent d'être piégées dans des puits de potentiel. Nous allons montrer comment mettre en place la méthode du recuit simulé dans ce contexte.

Ici, l'espace d'état E est l'ensemble des permutations de $\{1, \dots, N\}$. La fonction à minimiser est

$$V(x) := \sum_{i=0}^N \text{dist}(x_i, x_{i+1}) \quad \text{pour tout } x \in E,$$

où $x_0 = x_{N+1} = 0$ indique son lieu de résidence. La matrice de proposition est construite en supposant que, partant d'une permutation $x \in E$, les seules transitions sont les suivantes :

- on échange deux villes x_i et x_j sélectionnées au hasard,
- on garde le même trajet.

On dira que deux chemins x et y sont voisins, et on note $x \sim y$, si $x \neq y$ et s'ils peuvent être obtenus l'un de l'autre par un seul échange de deux villes. On considère la matrice de proposition :

$$Q(x, y) = \frac{2}{N^2} \mathbf{1}_{\{x \sim y\}} + \frac{1}{N} \mathbf{1}_{\{x=y\}} \quad \text{pour } x, y \in E.$$

Un autre choix possible, qui donne de meilleurs résultats numériques, consiste à garder la même forme pour la matrice de transition, et à définir les trajets voisins par $x \approx y$ s'il existe $1 \leq i < j \leq N$ tels que

$$y = (x_1, \dots, x_{i-1}, x_j, x_{j-1}, \dots, x_{i+1}, x_{j+1}, \dots, x_N).$$

Il est clair que Q est irréductible. Notre choix de matrice de proposition vérifiant $Q(x, x) > 0$, qui peut sembler curieux à première vue, assure que Q est apériodique, et vérifie donc la condition de Doeblin d'après la proposition 7.11.

Chapitre 8

Martingales en temps discret

Dans tout ce chapitre $(\Omega, \mathcal{A}, \mathbb{F}, \mathbb{P})$ désigne un espace probabilisé filtré, et on considèrera des processus aléatoires réels.

8.1 Martingales et temps d'arrêt

Dans le domaine des jeux de hasard, on trouve la définition suivante : *une martingale est une technique permettant d'augmenter les chances de gain. Toutefois, en moyenne, un joueur utilisant une martingale ne gagnera pas plus qu'un autre ; la martingale permet de perdre moins souvent, mais le montant des pertes potentielles est plus important.* Ce concept a un côté mystérieux lié à des histoires colportées sur l'utilisation de certains joueurs particulièrement doués d'une martingale dont ils détiennent jalousement le secret. Bien sûr, tout ceci relève plutôt du mythe que de la réalité. En effet, l'analyse précise des stratégies dites de martingale révèle des caractéristiques de moyenne de gain et de risque qui ne justifient pas les bienfaits prétendus.

La notion de filtration permet de donner une définition mathématique précise pour cette notion.

Définition 8.1. (*Martingale*) Soit $X = (X_n)_{n \geq 0}$ un processus aléatoire adapté sur l'espace probabilisé filtré $(\Omega, \mathcal{A}, \mathbb{F}, \mathbb{P})$. On dit que X est une surmartingale (resp. sous-martingale) si X_n est \mathbb{P} -intégrable pour tout n et

$$E[X_n | \mathcal{F}_{n-1}] \leq (\text{resp. } \geq) X_{n-1} \quad \text{pour tout } n \geq 1.$$

On dit qu'un processus X est une martingale s'il est à la fois surmartingale et sous-martingale.

Exemple 8.2. Soit Y une v.a. réelle intégrable. Alors $X_n := \mathbb{E}[Y | \mathcal{F}_n], n \in \mathbb{N}$, est une martingale. Ceci est une conséquence immédiate de la propriété des projections itérées de l'espérance conditionnelle.

Exemple 8.3. Soit $(X_n)_n$ une chaîne de Markov homogène sur un espace d'état E , de matrice de transition P , et soit $f : E \rightarrow \mathbb{R}$ vérifiant $P^n|f| < \infty$ pour tout n .

- Si $f = Pf$, on dit que f est harmonique, et $(f(X_n))_n$ est une martingale.
- Si $f \geq Pf$, on dit que f est surharmonique, et $(f(X_n))_n$ est une surmartingale.
- Si $f \leq Pf$, on dit que f est sous-harmonique, et $(f(X_n))_n$ est une sous-martingale.

Exemple 8.4. Soit $(\xi_n)_n$ une suite de v.a. indépendantes intégrables avec $\mathbb{E}[\xi_n] = 0$. Alors le processus

$$S_n := \sum_{i \leq n} \xi_i, \quad n \in \mathbb{N},$$

est une martingale. les v.a. $(\xi_n)_n$ peuvent représenter le gain d'un joueur pratiquant un jeu au hasard à répétition de manière indépendante. Alors ξ_n est le gain (ou perte) du joueur à la $n^{\text{ème}}$ partie.

Exemple 8.5. Soit M une martingale et ϕ un processus \mathbb{F} -adapté borné. Alors, le processus X défini par :

$$X_0 = 0 \quad \text{et} \quad X_n = \sum_{k=1}^n \phi_{k-1}(M_k - M_{k-1}), \quad n \geq 1,$$

est une martingale. Cet exemple a aussi une interprétation naturelle en terme de jeu de hasard en considérant $\xi_k := M_k - M_{k-1}$ comme le gain de la $k^{\text{ème}}$ partie, et ϕ_{k-1} le capital mis en jeu par le joueur pour cette partie.

L'exemple précédent permet de construire une grande classe de martingales à partir de la donnée d'une martingale M . Cependant, il faut bien noter que le processus ϕ n'y est pas tout à fait quelconque. En l'absence de condition d'intégrabilité sur ϕ , on est ramené à la notion de martingale locale; voir définition 8.19 ci-dessous.

Exercice 8.6. Vérifier que les processus $(S_n)_n$ et $(X_n)_n$ des exemples 8.4 et 8.5 sont bien des martingales.

Exercice 8.7. Montrer que

1. toute martingale a une espérance constante,
2. toute surmartingale (resp. sous-martingale) d'espérance constante est une martingale.

Exercice 8.8. Soient $(X_n)_n$ une martingale, et $g : \mathbb{R} \rightarrow \mathbb{R}$ une application telle que $\mathbb{E}[|g(X_n)|] < \infty$. Montrer que si g est convexe, alors le processus aléatoire $(g(X_n))_n$ est une sous-martingale.

Pour un processus aléatoire $X = (X_n)_{n \geq 0}$, on définit le processus arrêté au temps d'arrêt ν par :

$$X_n^\nu := X_{n \wedge \nu} \quad \text{pour tout} \quad n \geq 0.$$

Lemme 8.9. *Soient X une surmartingale (resp. sous-martingale, martingale) et ν un temps d'arrêt sur $(\Omega, \mathcal{A}, \mathbb{F}, \mathbb{P})$. Alors le processus arrêté X^ν est une surmartingale (resp. sous-martingale, martingale).*

Démonstration. Il suffit de montrer le résultat pour le cas surmartingale. Remarquons d'abord que X^ν est \mathbb{F} -adapté puisque pour tous $n \geq 0$ et $B \in \mathcal{E}$,

$$(X_n^\nu)^{-1}(B) = \left[\bigcup_{k \leq n-1} \{\nu = k\} \cap (X_k)^{-1}(B) \right] \cup [\{\nu \leq n-1\}^c \cap (X_n)^{-1}(B)].$$

Pour $n \geq 1$, $|X_n^\nu| \leq \sum_{k \leq n} |X_k| \in \mathcal{L}^1(\Omega, \mathcal{F}_n, \mathbb{P})$. Remarquons que du fait que ν est un temps d'arrêt, on a :

$$X_\nu \mathbf{1}_{\{\nu \leq n-1\}} = \sum_{k \leq n-1} X_k \mathbf{1}_{\{\nu = k\}} \in \mathcal{L}^0(\mathcal{F}_{n-1}) \text{ et } \{\nu > n-1\} = \{\nu \leq n-1\}^c \in \mathcal{F}_{n-1}.$$

On calcule alors directement que

$$\begin{aligned} \mathbb{E}[X_n^\nu | \mathcal{F}_{n-1}] &= \mathbb{E}[X_\nu \mathbf{1}_{\{\nu \leq n-1\}} + X_n \mathbf{1}_{\{\nu > n-1\}} | \mathcal{F}_{n-1}] \\ &= X_\nu \mathbf{1}_{\{\nu \leq n-1\}} + \mathbb{E}[X_n \mathbf{1}_{\{\nu > n-1\}} | \mathcal{F}_{n-1}] \\ &= X_\nu \mathbf{1}_{\{\nu \leq n-1\}} + \mathbf{1}_{\{\nu > n-1\}} \mathbb{E}[X_n | \mathcal{F}_{n-1}] \\ &\leq X_\nu \mathbf{1}_{\{\nu \leq n-1\}} + \mathbf{1}_{\{\nu > n-1\}} X_{n-1} = X_{n-1}^\nu. \end{aligned}$$

◇

Théorème 8.10. (Arrêt, Doob) *Soit X une martingale (resp. surmartingale) et $\underline{\nu}, \bar{\nu}$ deux temps d'arrêt bornés dans \mathcal{T} vérifiant $\underline{\nu} \leq \bar{\nu}$ p.s. Alors :*

$$\mathbb{E}[X_{\bar{\nu}} | \mathcal{F}_{\underline{\nu}}] = X_{\underline{\nu}} \quad (\text{resp. } \leq X_{\underline{\nu}}).$$

Démonstration. Nous montrons le résultat pour les martingales; le cas des surmartingales se traite de manière identique. Par hypothèse, il existe $N \in \mathbb{N}$ tel que $\bar{\nu} \leq N$.

(i) Commençons par montrer que $E[X_N | \mathcal{F}_\nu] = X_\nu$ pour tout $\nu \in \mathcal{T}$. Soit A un événement arbitraire de \mathcal{F}_ν . Alors l'événement $A \cap \{\nu = n\} \in \mathcal{F}_n$ et par suite :

$$\mathbb{E}[(X_N - X_\nu) \mathbf{1}_{A \cap \{\nu = n\}}] = \mathbb{E}[(X_N - X_n) \mathbf{1}_{A \cap \{\nu = n\}}] = 0$$

puisque X est une martingale. En sommant sur les n , on obtient alors :

$$0 = \sum_{n=0}^N \mathbb{E}[(X_N - X_\nu) \mathbf{1}_{A \cap \{\nu = n\}}] = \mathbb{E}[(X_N - X_\nu) \mathbf{1}_A].$$

Comme A est arbitraire dans \mathcal{F}_ν , ceci prouve que $E[X_N - X_\nu | \mathcal{F}_\nu] = 0$.

(ii) D'après le lemme 8.9 le processus arrêté $X^{\bar{\nu}}$ est une martingale. En lui appliquant le résultat du (i), on voit que :

$$\mathbb{E}[X_{\bar{\nu}} | \mathcal{F}_{\underline{\nu}}] = \mathbb{E}[X_N^{\bar{\nu}} | \mathcal{F}_{\underline{\nu}}] = X_{\underline{\nu}}^{\bar{\nu}} = X_{\underline{\nu}}$$

puisque $\nu \leq \bar{\nu}$. ◇

Par définition une martingale a une espérance constante puisque $\mathbb{E}[X_n] = \mathbb{E}[X_0]$ pour tout $n \in \mathbb{N}$. Le résultat suivant donne une sorte de réciproque grâce à la notion de temps d'arrêt.

Proposition 8.11. *Soit $X = (X_n)_n$ un processus aléatoire \mathbb{F} -adapté, $\mathbb{E}[|X_n|] < \infty$ pour tout $n \in \mathbb{N}$. Alors, X est une martingale si et seulement si*

$$\mathbb{E}[X_\nu] = \mathbb{E}[X_0] \quad \text{pour tout temps d'arrêt } \nu \text{ borné.}$$

Démonstration. La condition nécessaire est une application immédiate du théorème d'arrêt de Doob. Pour la condition suffisante, on fixe $n \in \mathbb{N}$, on considère un événement $A \in \mathcal{F}_n$ arbitraire, et on introduit

$$\nu := n\mathbf{1}_A + (n+1)\mathbf{1}_{A^c}.$$

Notons que $\{\nu = n\} = A \in \mathcal{F}_n$ et $\{\nu = n+1\} = A^c \in \mathcal{F}_n \subset \mathcal{F}_{n+1}$. Pour $k \in \mathbb{N} \setminus \{n, n+1\}$, on voit que $\{\nu = k\} = \emptyset \in \mathcal{F}_k$. Ceci prouve que ν est un temps d'arrêt borné, et par suite $0 = \mathbb{E}[X_0 - X_{n+1}] = \mathbb{E}[X_\nu - X_{n+1}] = \mathbb{E}[X_n\mathbf{1}_A - X_{n+1}\mathbf{1}_A] = \mathbb{E}[(X_{n+1} - X_n)\mathbf{1}_A]$. Comme $A \in \mathcal{F}_n$ est arbitraire, on déduit que $\mathbb{E}[X_{n+1} - X_n | \mathcal{F}_n] = 0$ par définition de l'espérance conditionnelle. ◇

Exercice 8.12. (*Identité de Wald*) Soient $(X_n)_{n \geq 1}$ une suite de v.a. intégrables iid, et τ un temps d'arrêt intégrable. On note $S_n := \sum_{i \leq n} X_i$ pour tout $n \geq 1$.

1. Montrer que

$$\mathbb{E}[S_\tau] = \mathbb{E}[\tau]\mathbb{E}[X_1].$$

Indication : introduire le processus $M_n := S_n - n\mathbb{E}[X_1]$, $n \geq 1$.

2. Si de plus $\mathbb{E}[X_1] = 0$, montrer que

$$\text{Var}[S_\tau] = \mathbb{E}[\tau]\text{Var}[X_1].$$

8.2 Martingales fermées

Dans ce paragraphe, nous étudions les martingales introduites dans l'exemple 8.2.

Définition 8.13. *Un processus aléatoire $X = (X_n)_{n \geq 0}$ est une martingale fermée s'il existe une v.a. réelle intégrable Y telle que $X_n = \mathbb{E}[Y | \mathcal{F}_n]$ pour tout $n \in \mathbb{N}$.*

Théorème 8.14. *Toute martingale fermée est uniformément intégrable.*

Démonstration. Soient $Y \in \mathcal{L}^1(\mathcal{A}, \mathbb{P})$ et $X_n = \mathbb{E}[Y | \mathcal{F}_n]$, $n \in \mathbb{N}$. Pour $\varepsilon > 0$, on déduit du lemme 1.26 qu'il existe $\delta > 0$ tel que

$$\mathbb{E}[|Y| \mathbf{1}_A] \leq \varepsilon \quad \text{pour tout } A \in \mathcal{A} \text{ de mesure } \mathbb{P}[A] \leq \delta. \quad (8.1)$$

d'après l'inégalité de Jensen, on a

$$|X_n| \leq \mathbb{E}[|Y| | \mathcal{F}_n] \quad \text{pour tout } n \geq 0. \quad (8.2)$$

Alors, en utilisant l'inégalité de Chebychev et la propriété des projections itérées de l'espérance conditionnelle :

$$\mathbb{P}[|X_n| \geq c] \leq \frac{\mathbb{E}[|X_n|]}{c} \leq \frac{\mathbb{E}[\mathbb{E}[|Y| | \mathcal{F}_n]]}{c} = \frac{\mathbb{E}[|Y|]}{c} \leq \delta \quad \text{pour } c \text{ assez grand.}$$

On peut donc appliquer (8.1) à l'événement $A := \{|X_n| \geq c\}$ pour obtenir :

$$\varepsilon \geq \mathbb{E}[|Y| \mathbf{1}_{\{|X_n| \geq c\}}] = \mathbb{E}[\mathbb{E}[|Y| | \mathcal{F}_n] \mathbf{1}_{\{|X_n| \geq c\}}] \geq \mathbb{E}[|X_n| \mathbf{1}_{\{|X_n| \geq c\}}],$$

où nous avons utilisé la propriété des projections itérées de l'espérance conditionnelle et (8.2). \diamond

Dans le chapitre suivant, nous obtiendrons l'équivalence dans l'énoncé du théorème 8.14.

8.3 Inégalités de martingales

Nous avons vu que les martingales fermées sont résumées par leur valeur finale, dans le sens que la donnée de la valeur terminale détermine toute la martingale. Le résultat suivant donne un contrôle du maximum de la trajectoire en terme de la valeur finale. Ce résultat est très utile, et est en partie responsable de l'importance des martingales dans la théorie des processus stochastiques.

Théorème 8.15. (*Inégalité maximale de Doob*) Soit $M = \{M_n, n \geq 0\}$ une sous-martingale, et $M_n^* := \sup_{k \leq n} M_k$ son processus de maximum courant.

(i) Pour tout $c > 0$, on a :

$$c\mathbb{P}[M_n^* \geq c] \leq \mathbb{E}[M_n \mathbf{1}_{\{M_n^* \geq c\}}] \quad \text{pour tout } n \in \mathbb{N}.$$

(ii) Soit $p > 1$, et supposons que la sous-martingale M est positive et $M_n \in \mathcal{L}^p$ pour tout $n \geq 0$. Alors, $M_n^* \in \mathcal{L}^p$ et

$$\|M_n^*\|_p \leq \frac{p}{p-1} \|M_n\|_p \quad \text{pour tout } n \in \mathbb{N}.$$

Démonstration. En introduisant le temps d'arrêt $T_c := \inf\{k > 0 : M_k \geq c\}$, on calcule que :

$$\begin{aligned} \mathbb{E}\left[M_n \mathbf{1}_{\{M_n^* \geq c\}}\right] &= \mathbb{E}\left[M_n \mathbf{1}_{\{T_c \leq n\}}\right] = \mathbb{E}\left[\mathbb{E}[M_n | \mathcal{F}_{T_c}] \mathbf{1}_{\{T_c \leq n\}}\right] \\ &\geq \mathbb{E}\left[M_{T_c} \mathbf{1}_{\{T_c \leq n\}}\right] \\ &\geq c \mathbb{E}\left[\mathbf{1}_{\{T_c \leq n\}}\right] \\ &= c \mathbb{P}\left[M_n^* \geq c\right]. \end{aligned}$$

(ii) On note $q := p/(p-1)$. On déduit de l'inégalité du (i) que

$$L := \int_0^\infty pc^{p-1} \mathbb{P}[M_n^* \geq c] dc \leq R := \int_0^\infty pc^{p-2} \mathbb{E}\left[M_n \mathbf{1}_{\{M_n^* \geq c\}}\right] dc.$$

Comme $M \geq 0$, on déduit du théorème de Fubini que

$$L = \mathbb{E}\left[\int_0^{M_n^*} pc^{p-1} dc\right] = \mathbb{E}[(M_n^*)^p]$$

et

$$\begin{aligned} R &= \mathbb{E}\left[M_n \int_0^{M_n^*} pc^{p-2} dc\right] \\ &= q \mathbb{E}\left[M_n (M_n^*)^{p-1}\right] \leq q \|M_n\|_p \|(M_n^*)^{p-1}\|_q = q \|M_n\|_p \mathbb{E}[(M_n^*)^p]^{1/q}, \end{aligned}$$

par l'inégalité de Hölder et le fait que $p^{-1} + q^{-1} = 1$. Ainsi, $\mathbb{E}[(M_n^*)^p] \leq q \|M_n\|_p \mathbb{E}[(M_n^*)^p]^{1/q}$, qui donne exactement l'inégalité voulue. \diamond

8.4 Décomposition des surmartingales

Nous commençons par un résultat qui permet d'extraire une martingale de tout processus aléatoire.

Proposition 8.16. (*Décomposition de Doob*) Soit $(X_n)_n$ un processus aléatoire intégrable. Alors il existe une martingale $(M_n)_n$ et un processus \mathbb{F} -prévisible $(V_n)_n$, tels que $M_0 = V_0 = 0$ et :

$$X_n = X_0 + M_n + V_n \quad \text{pour tout } n \geq 0.$$

De plus, cette décomposition est unique.

Démonstration. Pour l'unicité, on considère une autre décomposition avec $(M'_n)_n, (V'_n)_n$, alors $M_n - M'_n = V'_n - V_n$ est prévisible et par suite, pour tout $n \geq 1$, $M_n - M'_n = M_{n-1} - M'_{n-1} = \dots = M_0 - M'_0 = 0$ et $V_n = V'_n$. On note $\Delta X_n = X_n - X_{n-1}$, $\Delta M_n = M_n - M_{n-1}$, $\Delta V_n = V_n - V_{n-1}$. Si la

décomposition existe, alors $\Delta V_n = \Delta X_n - \Delta M_n$, et comme M est une martingale et V est prévisible, $\Delta V_n = \mathbb{E}[\Delta X_n | \mathcal{F}_{n-1}]$. Ceci suggère une définition unique pour le processus prévisible V et de M :

$$V_n := \sum_{i=1}^n \mathbb{E}[\Delta X_i | \mathcal{F}_{i-1}] \quad \text{et} \quad M_n := \sum_{i=1}^n (\Delta X_i - \mathbb{E}[\Delta X_i | \mathcal{F}_{i-1}]), \quad n \geq 1.$$

Cet unique candidat vérifie bien les conditions de la proposition. \diamond

Remarque 8.17. Avec la décomposition de Doob, on voit que

- X est une surmartingale si et seulement si V est décroissant,
- X est une sous-martingale si et seulement si V est croissant,
- X est une martingale si et seulement si $V = 0$.

L'unicité de la décomposition de Doob est liée de manière cruciale au caractère prévisible du processus A . La décomposition suivante, différente de la décomposition de Doob, relâche cette condition.

Proposition 8.18. Soient $X = (X_n)_n$ une martingale de carré intégrable, et $\Delta X_n := X_n - X_{n-1}$, $n \geq 1$. Alors :

$$X_n^2 = X_0^2 + N_n + [X]_n \quad \text{où} \quad N_n := 2 \sum_{i=1}^n X_{i-1} \Delta X_i, \quad [X]_n := \sum_{i=1}^n (\Delta X_i)^2, \quad n \geq 1,$$

et $N_0 = [X]_0 = 0$. Dans cette décomposition $(N_n)_n$ est une martingale nulle en zéro, et $([X]_n)_n$ est un processus \mathbb{F} -adapté croissant intégrable appelé variation quadratique de la martingale X .

Démonstration. Il s'agit d'un calcul immédiat. \diamond

8.5 Martingales locales

Définition 8.19. On dit qu'un processus $X = \{X_n, n \geq 0\}$ est une martingale locale s'il existe une suite de temps d'arrêt $(\tau_n)_n$ telle que $\tau_n \rightarrow \infty$ \mathbb{P} -p.s. et le processus arrêté X^{τ_n} est une martingale pour tout n .

Exemple 8.20. Soit M une martingale et ϕ un processus \mathbb{F} -adapté. Alors, le processus X défini par :

$$X_0 = 0 \quad \text{et} \quad X_n = \sum_{k=1}^n \phi_{k-1} (M_k - M_{k-1}) \quad \text{pour} \quad n \geq 1,$$

est une martingale locale. En effet, il suffit de considérer la suite de temps d'arrêt

$$\tau_n = \inf \{t : |\phi_t| \geq n\},$$

ave la convention habituelle $\inf \emptyset = \infty$. Alors, pour tout $n \geq 1$,

$$X_t^{\tau_n} = \sum_{k=1}^t \phi_{k-1} \mathbf{1}_{\{k-1 < \tau_n\}} (M_k - M_{k-1}), \quad t = 1, \dots, T.$$

Comme $\tilde{\phi}_k := \phi_k \mathbf{1}_{\{k < \tau_n\}}$ est \mathcal{F}_k -mesurable et borné, on est alors ramené à l'exemple 8.5, et le processus arrêté X^{τ_n} est une martingale.

En général, une martingale locale n'est pas une martingale. Le résultat suivant précise la différence entre ces deux notions dans le cas du temps discret fini.

Lemme 8.21. *Soit $X = \{X_n, n = 0, \dots, N\}$ une martingale locale telle que :*

$$E[X_N^-] < \infty.$$

Alors, X est une martingale.

Démonstration. Soit $(\tau_n)_n$ une suite de temps d'arrêt tendant vers l'infini p.s. telle que le processus arrêté X^{τ_n} est une martingale pour tout n . Nous décomposons la démonstration en deux étapes.

1- Montrons d'abord que X_k est intégrable pour tout k :

- sur l'événement $\{\tau_n > N - 1\}$, on a $X_{N-1}^- \leq \mathbb{E}[X_N^- | \mathcal{F}_{N-1}]$; en faisant tendre n vers l'infini, ceci montre que X_{N-1}^- hérite l'intégrabilité de X_N^- ,
- en itérant ce raisonnement de manière rétrograde, on obtient l'intégrabilité de X_k^- pour tout $k = 0, \dots, N$,
- Quant à la partie positive de X , on utilise le lemme de Fatou pour obtenir :

$$\begin{aligned} \mathbb{E}[X_k^+] &= \mathbb{E} \left[\liminf_{n \rightarrow \infty} X_{k \wedge \tau_n}^+ \right] \\ &\leq \liminf_{n \rightarrow \infty} \mathbb{E} [X_{k \wedge \tau_n}^+] = \liminf_{n \rightarrow \infty} \mathbb{E} [X_{k \wedge \tau_n} + X_{k \wedge \tau_n}^-]; \end{aligned}$$

en se rappelant que le processus arrêté X^{τ_n} est une martingale, on voit alors que :

$$\mathbb{E}[X_k^+] \leq X_0 + \liminf_{n \rightarrow \infty} \mathbb{E} [X_{k \wedge \tau_n}^-] \leq X_0 + \sum_{j=0}^N E[X_j^-] < \infty.$$

2- En utilisant le fait que le processus arrêté X^{τ_n} est une martingale, on voit que :

$$\mathbb{E}[X_k | \mathcal{F}_{k-1}] = \mathbb{E}[X_k^{\tau_n} | \mathcal{F}_{k-1}] = X_{k-1}^{\tau_n} = X_{k-1} \quad \text{sur l'événement } \{\tau_n > k - 1\}.$$

En faisant tendre n vers l'infini, on obtient $E[X_k | \mathcal{F}_{k-1}] = X_{k-1}$ \mathbb{P} -p.s. \diamond

Nous terminons ce chapitre par un exemple de martingale locale stricte.

Exemple 8.22. Soit $(A_n)_{n \geq 1}$ une partition mesurable de Ω avec $P[A_n] = 2^{-n}$. Soit $(Z_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes des A_n avec $P[Z_n = 1] = P[Z_n = -1] = 1/2$. On définit

$$\mathcal{F}_0 := \sigma(A_n, n \geq 1) \quad \text{et} \quad \mathcal{F}_1 := \sigma(A_n, Z_n, n \geq 1).$$

On considère maintenant le processus

$$X_0 := 0 \quad \text{et} \quad X_1 := Y_\infty = \lim_{n \rightarrow \infty} Y_n \quad \text{avec} \quad Y_n := \sum_{1 \leq p \leq n} 2^p Z_p \mathbf{1}_{A_p}.$$

Il est clair que $X = \{X_0, X_1\}$ n'est pas une martingale puisque $X_1 = Y_\infty$ n'est pas intégrable. Dans la suite de cet exemple, nous allons montrer que X est une martingale locale. On définit la suite de temps d'arrêt qui tend vers l'infini :

$$\tau_n := +\infty \quad \text{on} \quad \cup_{1 \leq p \leq n} A_p \quad \text{et} \quad \tau_n := 0 \quad \text{ailleurs.}$$

Pour tout n , le processus arrêté X^{τ_n} est donné par :

$$X_0^{\tau_n} = 0 \quad \text{et} \quad X_1^{\tau_n} = Y_n.$$

Comme la variable aléatoire Y_n est bornée et indépendante de \mathcal{F}_0 , il est clair que $\mathbb{E}[Y_n | \mathcal{F}_0] = 0$, et par suite X^{τ_n} est une martingale.

Chapitre 9

Convergence des martingales

9.1 Convergence des martingales de carré intégrable

Nous commençons par étudier le cas d'une martingale $M = \{M_n, n \geq 0\}$ de carré intégrable. L'espace de Hilbert \mathbb{L}^2 étant de dimension infinie, nous savons que la boule unité n'est pas compacte. En d'autres termes, la bornitude d'une suite ne garantit pas sa convergence lelong d'une sous-suite. la structure de martingale va cependant rajouter l'ingrédient qui va rendre la convergence de la suite possible. La remarque fondamentale est que pour de telles martingales les accroissements

$$\Delta M_n := M_n - M_{n-1}, \quad n \geq 1,$$

sont orthogonaux au sens du produit scalaire de \mathbb{L}^2 :

$$\mathbb{E}[(\Delta M_i)(\Delta M_j)] = \mathbb{E}[\Delta M_i \mathbb{E}\{\Delta M_j | \mathcal{F}_{j-1}\}] = 0 \quad \text{pour } 1 \leq i < j, \quad (9.1)$$

d'après la propriété des projection itérées de l'espérance conditionnelle. En particulier, ceci implique que

$$\begin{aligned} \mathbb{E}[M_n^2] &= \mathbb{E} \left[\left(M_0 + \sum_{i=1}^n \Delta M_i \right)^2 \right] \\ &= \mathbb{E}[M_0^2] + \sum_{i=1}^n \mathbb{E}[(\Delta M_i)^2] + 2 \sum_{i < j} \mathbb{E}[(\Delta M_i)(\Delta M_j)] \\ &= \mathbb{E}[M_0^2] + \sum_{i=1}^n \mathbb{E}[(\Delta M_i)^2] \quad \text{pour tout } n \geq 0. \end{aligned}$$

Ainsi la suite de réels $(\mathbb{E}[M_n^2])_{n \geq 0}$ est croissante. Si elle est bornée, alors cette suite converge vers une valeur positive finie.

Théorème 9.1. *Soit $\{M_n, n \geq 0\}$ une martingale bornée dans \mathbb{L}^2 , c'est à dire que $\sup_{n \geq 0} \mathbb{E}[M_n^2] < \infty$. Alors il existe une v.a. $M_\infty \in \mathbb{L}^2$ telle que*

- (i) $M_n \longrightarrow M_\infty$ dans \mathbb{L}^2 ,
- (ii) $M_n \longrightarrow M_\infty$ p.s.

Démonstration. (i) Comme la martingale est bornée dans \mathbb{L}^2 , la suite $(\mathbb{E}[M_n^2])_{n \geq 0}$ est convergente dans \mathbb{R} . D'après l'orthogonalité des accroissements dans \mathbb{L}^2 , on voit que pour $n, p \geq 0$

$$\begin{aligned} \mathbb{E}[(M_{n+p} - M_n)^2] &= \sum_{i=n+1}^{n+p} \mathbb{E}[(\Delta M_i)^2] \leq \sum_{i \geq n+1} \mathbb{E}[(\Delta M_i)^2] \\ &= \mathbb{E}[M_{n+p}^2] - \mathbb{E}[M_n^2] \longrightarrow 0 \end{aligned}$$

quand $n \rightarrow \infty$, du fait de la convergence de la suite $(\mathbb{E}[M_n^2])_{n \geq 0}$. Ainsi, $(M_n)_{n \geq 0}$ forme une suite de Cauchy dans l'espace de Hilbert \mathbb{L}^2 , d'où l'existence d'une v.a. limite M_∞ dans \mathbb{L}^2 .

(ii) Utilisons la caractérisation de la convergence p.s. du théorème 2.17, et fixons $\varepsilon > 0$. En appliquant successivement l'inégalité de Chebychev et le théorème de la convergence monotone, on obtient :

$$\begin{aligned} \mathbb{P}\left[\sup_{k \geq n} |M_k - M_\infty| \geq \varepsilon\right] &\leq \frac{1}{\varepsilon^2} \mathbb{E}\left[\sup_{k \geq n} |M_k - M_\infty|^2\right] \\ &\leq \frac{2}{\varepsilon^2} \left(\mathbb{E}[|M_n - M_\infty|^2] + \mathbb{E}\left[\sup_{k \geq n} |M_k - M_n|^2\right] \right) \\ &= \frac{2}{\varepsilon^2} \left(\mathbb{E}[|M_n - M_\infty|^2] + \lim_{N \rightarrow \infty} \uparrow \mathbb{E}\left[\sup_{n \leq k \leq N} |M_k - M_n|^2\right] \right). \end{aligned}$$

Remarquons que le processus $(|M_k - M_n|)_{k \geq n}$ est une sous-martingale positive. On déduit alors de l'inégalité maximale de Doob, établie dans le théorème 8.15, que :

$$\begin{aligned} \mathbb{P}\left[\sup_{k \geq n} |M_k - M_\infty| \geq \varepsilon\right] &\leq \frac{2}{\varepsilon^2} \left(\mathbb{E}[|M_n - M_\infty|^2] + 4 \lim_{N \rightarrow \infty} \uparrow \mathbb{E}[|M_N - M_n|^2] \right) \\ &= \frac{10}{\varepsilon^2} \mathbb{E}[|M_n - M_\infty|^2] \longrightarrow 0, \end{aligned}$$

quand $n \rightarrow \infty$ d'après le résultat de convergence dans \mathbb{L}^2 du (i). \diamond

9.2 Loi des Grands Nombres

Comme application du théorème 9.1, nous allons maintenant montrer la loi des grands nombres pour les suites de v.a. iid intégrables. ceci renforce le résultat vu dans le cours de première année où la loi des grands nombres a été établie pour les suites de v.a. iid de carré intégrable. Nous commençons par une loi des grands nombres pour les martingales de carré intégrables.

Théorème 9.2. *Soit $\{M_n, n \geq 0\}$ une martingale de carré intégrable telle que*

$$\sum_{n \geq 1} \frac{1}{n^2} \mathbb{E} [|\Delta M_n|^2] < \infty.$$

Alors, $\frac{1}{n}M_n \rightarrow 0$ p.s. et dans \mathbb{L}^2 .

Démonstration. 1- Le processus $X_n := \sum_{k=1}^n \frac{1}{k} \Delta M_k$, $n \geq 1$, est une martingale vérifiant les conditions du théorème 9.1. Il existe donc une v.a. $X_\infty \in \mathbb{L}^2$ telle que $X_n \rightarrow X_\infty$ dans \mathbb{L}^2 et p.s.

2- Pour conclure, il suffit de reproduire l'argument classique du lemme de Kronecker pour les suites déterministes :

$$\begin{aligned} \frac{1}{n}M_n &= \frac{1}{n} \sum_{i=1}^n i(X_i - X_{i-1}) \\ &= \frac{1}{n} \left(\sum_{i=1}^n iX_i - \sum_{i=1}^n (i-1)X_{i-1} - \sum_{i=1}^n X_{i-1} \right) \\ &= X_n - \frac{1}{n} \sum_{i=1}^n X_{i-1} \rightarrow 0, \text{ dans } \mathbb{L}^2 \text{ et p.s.} \end{aligned}$$

d'après la première étape. ◇

Pour une suite de v.a. indépendantes, la marche aléatoire correspondante (définie par les sommes partielles) est une martingale à laquelle on peut appliquer le théorème précédent si la condition de convergence est vérifiée. Le résultat suivant construit sur cet argument pour montrer la version la plus forte de la loi des grands nombres.

Théorème 9.3. *(Loi forte des Grands Nombres) Soit $(X_n)_{n \geq 0}$ une suite iid de v.a. intégrables. Alors*

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

Démonstration. Sans perte de généralité, on suppose $\mathbb{E}[X_1] = 0$. On note

$$M_n := \sum_{i=1}^n (X_i \mathbf{1}_{\{|X_i| \leq i\}} - \mathbb{E}[X_i \mathbf{1}_{\{|X_i| \leq i\}}]), \quad n \geq 1.$$

Alors, $\{M_n, n \geq 1\}$ est une martingale de carré intégrable, et

$$\begin{aligned} \sum_{k=1}^n \frac{1}{k^2} \mathbb{E} [|\Delta M_k|^2] &\leq \sum_{k=1}^n \frac{1}{k^2} \mathbb{E} [|X_1|^2 \mathbf{1}_{\{|X_1| \vee 1 \leq k\}}] \\ &\leq \mathbb{E} \left[|X_1|^2 \sum_{k=1}^n \frac{1}{k^2} \mathbf{1}_{\{k \geq |X_1| \vee 1\}} \right] \\ &\leq \mathbb{E} \left[|X_1|^2 \int_{|X_1| \vee 1}^{\infty} \frac{dt}{t^2} \right] = E \left[\frac{|X_1|^2}{|X_1| \vee 1} \right] \leq \mathbb{E}[|X_1|] < \infty. \end{aligned}$$

On déduit du théorème 9.2 que $\frac{1}{n}M_n \rightarrow 0$, p.s. Puisque $\mathbb{E}[X_1 \mathbf{1}_{\{|X_1| \leq n\}}] \rightarrow \mathbb{E}[X_1] = 0$ par le théorème de convergence dominée, on voit alors que

$$\frac{1}{n} \sum_{i=1}^n X_i \mathbf{1}_{\{|X_i| \leq i\}} \rightarrow 0, \text{ p.s.}$$

Remarquons maintenant que

$$\sum_{i \geq 1} \mathbb{P}[|X_i| \geq i] = \sum_{i \geq 1} i \mathbb{P}[i \leq |X_1| < i+1] \leq \mathbb{E}[|X|] < \infty.$$

D'après le lemme 2.36 de Borel-Cantelli, on déduit alors que $\mathbb{P}[\limsup_n \{|X_n| \geq n\}] = 0$. Comme $\limsup_n \{|X_n| \geq n\} = \bigcap_{n \geq 1} \bigcup_{k \geq n} \{|X_k| \geq k\}$, ceci veut dire que, p.s., il existe $N(\omega) \in \mathbb{N}$ tel que pour tout $i \geq N(\omega)$, $|X_i| \leq i$, et par suite

$$\lim (\text{p.s.}) \frac{1}{n} \sum_{i=1}^n X_i = \lim (\text{p.s.}) \frac{1}{n} \sum_{i=1}^n X_i \mathbf{1}_{\{|X_i| \leq i\}} = 0.$$

◇

9.3 Convergence des sous-martingales

Dans ce paragraphe, $\{X_n, n \geq 0\}$ désigne une sous-martingale. Pour $a < b$, on définit la suite de temps d'arrêt

$$\tau_0 = 0, \theta_{n+1} := \inf \{i \geq \tau_n : X_i \leq a\}, \text{ et } \tau_{n+1} := \inf \{i \geq \theta_{n+1} : X_i \geq b\}.$$

Alors pour tout $n \geq 0$, la variable aléatoire

$$U_n^{a,b} := \max \{j : \tau_j \leq n\} \tag{9.2}$$

représente le nombre de traversées du niveau b en partant en dessous du niveau a avant la date n . Nous dirons plus simplement que U_n est le nombre de traversées montantes de l'intervalle $[a, b]$ avant la date n .

Lemme 9.4. Soient $\{X_n, n \geq 0\}$ une sous-martingale, et $a < b$. Alors la moyenne du nombre de traversées montantes de l'intervalle $[a, b]$ défini en (9.2) vérifie :

$$\mathbb{E}[U_n^{a,b}] \leq \frac{1}{b-a} \mathbb{E}[(X_n - a)^+].$$

Démonstration. On note $Y_n := (X_n - a)^+, n \geq 0$. Comme $\theta_{n+1} > n$, on a

$$\begin{aligned} Y_n &= Y_{n \wedge \theta_1} + \sum_{i=1}^n (Y_{n \wedge \tau_i} - Y_{n \wedge \theta_i}) + \sum_{i=1}^n (Y_{n \wedge \theta_{i+1}} - Y_{n \wedge \tau_i}) \\ &\geq \sum_{i=1}^n (Y_{n \wedge \tau_i} - Y_{n \wedge \theta_i}) + \sum_{i=1}^n (Y_{n \wedge \theta_{i+1}} - Y_{n \wedge \tau_i}). \end{aligned}$$

Par définition des traversées montantes, on a $Y_{\theta_i} = 0, Y_{\tau_i} \geq b - a$ sur $\{\tau_i \leq n\}$, et $Y_{n \wedge \tau_i} - Y_{n \wedge \theta_i} \geq 0$, p.s. Alors :

$$(b-a)\mathbb{E}[U_n^{a,b}] \leq \mathbb{E}[Y_n] - \sum_{i=1}^n \mathbb{E}[Y_{n \wedge \theta_{i+1}} - Y_{n \wedge \tau_i}] \leq \mathbb{E}[Y_n],$$

où la dernière égalité découle du théorème d'arrêt et du fait que le processus $\{Y_n, n \geq 0\}$ est une sous-martingale, propriété qu'on vérifie immédiatement grâce à l'inégalité de Jensen. \diamond

Nous donnons à présent une généralisation du théorème 9.1.

Théorème 9.5. Soit $\{X_n, n \geq 0\}$ une sous-martingale. Si $\sup_{n \geq 0} \mathbb{E}[X_n^+] < \infty$, alors il existe une v.a. $X \in \mathbb{L}^1$ telle que

$$X_n \longrightarrow X, \text{ p.s.}$$

Démonstration. Le processus comptant le nombre de traversées montantes $\{U_n, n \geq 0\}$ est croissant, on note alors $U^{a,b} := \lim_{n \rightarrow \infty} U_n^{a,b}$. D'après le théorème de convergence monotone et le lemme 9.4 :

$$\begin{aligned} \mathbb{E}[U^{a,b}] &= \lim_{n \rightarrow \infty} \mathbb{E}[U_n^{a,b}] \leq \frac{1}{b-a} \sup_{n \geq 0} \mathbb{E}[(X_n - a)^+] \\ &\leq \frac{1}{b-a} \left(\sup_{n \geq 0} \mathbb{E}[X_n^+] + |a| \right) < \infty. \end{aligned}$$

En particulier, ceci prouve que $U^{a,b} < \infty$, p.s. L'événement

$$N^{a,b} := \left\{ \liminf_{n \rightarrow \infty} X_n \leq a < b \leq \limsup_{n \rightarrow \infty} X_n \right\}$$

est alors négligeable (i.e. de probabilité nulle). En prenant l'union sur les rationnels, on voit que

$$N := \left\{ \liminf_{n \rightarrow \infty} X_n < \limsup_{n \rightarrow \infty} X_n \right\} = \bigcup \{N^{a,b} : a, b \in \mathbb{Q}, a < b\}$$

est négligeable, comme union dénombrable de négligeables. Ceci montre bien que $X_\infty := \lim_n X_n$ existe p.s.

Il reste à montrer que $X_\infty \in \mathbb{L}^1$. D'abord, par le lemme de Fatou :

$$\mathbb{E}[X_\infty^+] \leq \liminf_{n \rightarrow \infty} \mathbb{E}[X_n^+] \leq \sup_{n \geq 0} \mathbb{E}[X_n^+] < \infty.$$

Comme X est une sous-martingale, on a $\mathbb{E}[X_n] \geq \mathbb{E}[X_0]$, et on obtient en réutilisant le lemme de Fatou que :

$$\begin{aligned} \mathbb{E}[X_\infty^-] &\leq \liminf_{n \rightarrow \infty} \mathbb{E}[X_n^-] = \liminf_{n \rightarrow \infty} \{\mathbb{E}[X_n^+] - \mathbb{E}[X_n]\} \\ &\leq -\mathbb{E}[X_0] + \sup_{n \geq 0} \mathbb{E}[X_n^+] < \infty. \end{aligned}$$

◇

Remarque 9.6. (i) Une sous-martingale (X_n) est bornée dans \mathbb{L}^1 si et seulement si $\sup_{n \geq 0} \mathbb{E}[X_n^+] < \infty$. En effet, en raisonnant comme à la fin de la preuve précédente, on voit que $\mathbb{E}[|X_n|] \leq 2\mathbb{E}[X_n^+] - \mathbb{E}[X_0]$.

(ii) Bien que la limite X_∞ dans le théorème 9.5 soit dans \mathbb{L}^1 , la convergence n'a pas lieu dans \mathbb{L}^1 en général. L'exercice suivant fournit un contre-exemple.

Exercice 9.7. Soient $(\xi_n)_{n \geq 0}$ une suite de v.a. iid de loi de Bernoulli de paramètre $p := \mathbb{E}[\xi_1]$, et

$$S_0 = 1 \text{ et } S_n := e^{-n+2 \sum_{i=1}^n \xi_i} \text{ pour tout } n \geq 1.$$

1. Montrer que S_n est une martingale si et seulement si $p = (1 + e)^{-1}$.
2. Montrer que $(S_n)^{1/n}$ converge p.s. vers une constante $s < 1$, et en déduire que $S_n \rightarrow 0$, p.s.
3. Calculer $\mathbb{E}[S_n]$ pour tout n , et en déduire que $S_n \not\rightarrow 0$ dans \mathbb{L}^1 .

Pour que la convergence dans le théorème 9.5 ait lieu dans \mathbb{L}^1 , il est nécessaire de rajouter une condition sur la martingale qui permet, en particulier, d'éviter le phénomène de perte de masse observé dans l'exercice 9.7. Le résultat suivant caractérise cette condition dans le cadre des martingales. On notera $\mathcal{F}_\infty := \sigma(\cup_{n \geq 0} \mathcal{F}_n)$.

Théorème 9.8. Soit $M = \{M_n, n \geq 0\}$ une martingale.

1. Les deux assertions suivantes sont équivalentes.
 - (i) M est fermée, c'est à dire qu'il existe $M_\infty \in \mathbb{L}^1$, telle que $M_n = \mathbb{E}[M_\infty | \mathcal{F}_n]$ pour tout $n \geq 0$, et $M_n \rightarrow M_\infty$ dans \mathbb{L}^1 et p.s.
 - (ii) M est uniformément intégrable.
2. Supposons que M soit définie à partir d'une v.a. $Y \in \mathbb{L}^1$ par $M_n = \mathbb{E}[Y | \mathcal{F}_n]$, $n \geq 0$. Alors $M_\infty = \mathbb{E}[Y | \mathcal{F}_\infty]$.

Démonstration. 1. L'implication (i) \implies (ii) a déjà été établie dans le théorème 8.14. Supposons maintenant que M est U.I. Alors $\sup_n \mathbb{E}[|M_n|] < \infty$ d'après l'exercice 2.21, et on se retrouve dans le cadre du théorème 9.5 qui assure l'existence d'une v.a. $M_\infty \in \mathbb{L}^1$ telle que $M_n \rightarrow M_\infty$ p.s. Dans le reste de la preuve, nous allons montrer que cette convergence a lieu dans \mathbb{L}^1 et que $M_n = \mathbb{E}[M_\infty | \mathcal{F}_n]$ pour tout $n \geq 0$.

• Pour $c > 0$, on définit la fonction $f_c(x) := x\mathbf{1}_{|x| \leq c} + c\mathbf{1}_{x > c} - c\mathbf{1}_{x < -c}$, $x \in \mathbb{R}$, et on décompose

$$\begin{aligned} \mathbb{E}[|M_n - M_\infty|] &\leq \mathbb{E}[|f_c(M_n) - M_n|] + \mathbb{E}[|f_c(M_\infty) - M_\infty|] \\ &\quad + \mathbb{E}[|f_c(M_n) - f_c(M_\infty)|]. \end{aligned}$$

Soit $\varepsilon > 0$. Comme $|f_c(x) - x| \leq |x|\mathbf{1}_{|x| \geq c}$, l'uniforme intégrabilité assure l'existence d'un $c_0 > 0$ tel que pour $c \geq c_0$ assez grand

$$\mathbb{E}[|f_c(M_n) - M_n|] + \mathbb{E}[|f_c(M_\infty) - M_\infty|] \leq \varepsilon \quad \text{pour tout } n \geq 0.$$

Comme la fonction f_c est continue bornée, on déduit du théorème de convergence dominée que

$$\mathbb{E}[|f_c(M_n) - f_c(M_\infty)|] \leq \varepsilon \quad \text{pour } n \text{ assez grand.}$$

On a finalement $\mathbb{E}[|M_n - M_\infty|] \leq 2\varepsilon$, prouvant la convergence dans \mathbb{L}^1 de M_n vers M_∞ .

• Soit A un événement arbitraire dans \mathcal{F}_n . Comme M est une martingale, on a pour tout $N \geq n$:

$$|\mathbb{E}[(M_n - M_\infty)\mathbf{1}_A]| = |\mathbb{E}[(M_N - M_\infty)\mathbf{1}_A]| \leq |\mathbb{E}[|M_N - M_\infty|]| \rightarrow 0, \quad N \rightarrow \infty.$$

Donc, $\mathbb{E}[(M_n - M_\infty)\mathbf{1}_A] = 0$ pour tout $A \in \mathcal{F}_n$, ce qui montre bien que $M_n = \mathbb{E}[M_\infty | \mathcal{F}_n]$.

2. Supposons maintenant que $M_n = \mathbb{E}[Y | \mathcal{F}_n]$, $n \geq 0$, pour une certaine v.a. $Y \in \mathbb{L}^1$. Alors, la v.a. \mathcal{F}_∞ -mesurable $X := M_\infty - \mathbb{E}[Y | \mathcal{F}_\infty]$ vérifie $\mathbb{E}[X\mathbf{1}_A] = 0$ pour tout événement $A \in \cup_{n \geq 0} \mathcal{F}_n$. Comme $\cup_{n \geq 0} \mathcal{F}_n$ est stable par intersection, on déduit du résultat d'unicité des mesures (Proposition 1.5) que $X = 0$. \diamond

L'exercice suivant fournit une approche martingale pour la démonstration de la loi du zéro-un du théorème 2.37.

Exercice 9.9. Soit $(X_n)_{n \geq 0}$ une suite de v.a. indépendantes, de filtration canonique $\mathcal{F}_n := \sigma(X_i, i \leq n)$, $n \geq 0$. On note $\mathcal{T} := \cap_n \sigma(X_m, m > n)$ la σ -algèbre de queue associée.

1. Montrer que $\mathbb{E}[\mathbf{1}_A | \mathcal{F}_n] = \mathbb{P}[A]$ pour tout $A \in \mathcal{T}$.
2. En déduire que $\mathbf{1}_A = \mathbb{P}[A]$ pour tout $A \in \mathcal{T}$.

9.4 Théorème central limite martingale

Dans ce paragraphe, nous montrons un théorème central limite qui généralise celui vu dans le cours de première année, et rappelé dans le théorème 2.40. Pour faire le lien rappelons-en l'énoncé. Soient $(X_i)_{i \geq 0}$ une suite de v.a. iid centrées avec $\mathbb{E}[X_1^2] = 1$, et $M_n := \sum_{i \leq n} X_i$ la marche aléatoire correspondante. Alors $n^{1/2}M_n$ converge en loi vers une gaussienne centrée réduite.

Dans l'énoncé ci-dessus du théorème central limite, $\{M_n, n \geq 0\}$ est une martingale dont les accroissements sont iid. Le résultat suivant fournit une généralisation en relâchant cette hypothèse sur les accroissements.

Théorème 9.10. Soit $\{M_n, n \geq 0\}$ une martingale telle que

$$\mathbb{E}[(\Delta M_n)^2 | \mathcal{F}_{n-1}] = 1 \quad \text{et} \quad K := \sup_{n \geq 1} \mathbb{E}[|\Delta M_n|^3] < \infty.$$

Alors

$$\frac{1}{\sqrt{n}} M_n \longrightarrow \mathcal{N}(0, 1) \text{ en loi.}$$

Démonstration. D'après l'exercice 2.9, nous avons le développement suivant de la fonction caractéristique conditionnelle

$$\begin{aligned} \varphi_{n,j} &:= \mathbb{E}\left[e^{i \frac{u}{\sqrt{n}} \Delta M_j} | \mathcal{F}_{j-1}\right] \\ &= 1 + i \frac{u}{\sqrt{n}} \mathbb{E}[\Delta M_j | \mathcal{F}_{j-1}] - \frac{u^2}{2n} \mathbb{E}[(\Delta M_j)^2 | \mathcal{F}_{j-1}] - i \frac{u^3}{6n^{3/2}} \mathbb{E}[\bar{X}_j^3 | \mathcal{F}_{j-1}] \\ &= 1 - \frac{u^2}{2n} - i \frac{u^3}{6n^{3/2}} \mathbb{E}[\bar{X}_j^3 | \mathcal{F}_{j-1}], \end{aligned}$$

d'après les conditions du théorème. Ici $\bar{X}_j(\omega)$ est une valeur aléatoire entre 0 et $\Delta M_j(\omega)$. On en déduit que la fonction caractéristique de M_n s'écrit comme :

$$\begin{aligned} \Phi_{M_k/\sqrt{n}}(u) &:= \mathbb{E}\left[e^{i \frac{u}{\sqrt{n}} M_k}\right] \\ &= \mathbb{E}\left[e^{i \frac{u}{\sqrt{n}} M_{k-1}} e^{i \frac{u}{\sqrt{n}} \Delta M_k}\right] \\ &= \mathbb{E}\left[e^{i \frac{u}{\sqrt{n}} M_{k-1}} \varphi_{n,k}(u)\right] \\ &= \mathbb{E}\left[e^{i \frac{u}{\sqrt{n}} M_{k-1}} \left(1 - \frac{u^2}{2n} - i \frac{u^3}{6n^{3/2}} \bar{X}_j^3\right)\right] \\ &= \left(1 - \frac{u^2}{2n}\right) \Phi_{M_{k-1}/\sqrt{n}}(u) - i \frac{u^3}{6n^{3/2}} \mathbb{E}\left[\bar{X}_j^3 e^{i \frac{u}{\sqrt{n}} M_{k-1}}\right]. \end{aligned}$$

En utilisant le fait que $|\bar{X}_j| \leq |\Delta M_j|$ ainsi que la condition du théorème sur le moment d'ordre 3, Ceci implique que pour tout $k \geq 1$:

$$\left| \left(1 - \frac{u^2}{2n}\right)^{n-k} \Phi_{M_k/\sqrt{n}}(u) - \left(1 - \frac{u^2}{2n}\right)^{n-(k-1)} \Phi_{M_{k-1}/\sqrt{n}}(u) \right| \leq K \frac{u^3}{6n^{3/2}},$$

et par suite :

$$\left| \Phi_{M_n/\sqrt{n}}(u) - \left(1 - \frac{u^2}{2n}\right)^n \right| \leq n K \frac{u^3}{6n^{3/2}} = K \frac{u^3}{6\sqrt{n}}.$$

En envoyant n vers l'infini, on obtient :

$$\Phi_{M_n/\sqrt{n}}(u) \longrightarrow e^{-u^2/2} = \Phi_{\mathcal{N}(0,1)}(u) \text{ pour tout } u \in \mathbb{R}.$$

et on conclut grâce au théorème 2.28 de convergence de Lévy. \diamond

9.5 Application : l'algorithme de Robbins-Monro

Dans cette section, nous développons un algorithme stochastique de recherche de racine d'une équation définie à travers d'une espérance. Afin de motiver la démarche, nous commençons par en rappeler la version déterministe.

9.5.1 Méthode itérative de recherche de racine

Soit $f : \mathbb{R}^d \longrightarrow \mathbb{R}^d$ une fonction continue bornée. Contrairement à la continuité, la condition de bornitude n'est pas essentielle et peut aisément être remplacée par une condition de croissance linéaire.

On suppose que f est nulle en un certain point θ^* que l'on cherche à déterminer à l'aide de l'algorithme :

$$\theta_n = \theta_{n-1} - \gamma_n f(\theta_{n-1}), \quad n \geq 1, \quad (9.3)$$

où $(\gamma_n)_n \subset \mathbb{R}$ est une suite réelle à choisir convenablement.

On dit que f est *séparante* pour le point θ^* si

$$(\theta - \theta^*) \cdot f(\theta) > 0 \quad \text{pour tout } \theta \in \mathbb{R}^d, \theta \neq \theta^*. \quad (9.4)$$

Proposition 9.11. *Soient $\theta^* \in \mathbb{R}^d$ et $f : \mathbb{R}^d \longrightarrow \mathbb{R}^d$ une fonction continue, bornée, nulle en θ^* et séparante pour le point θ^* . Soit $\theta_0 \in \mathbb{R}^d$ une valeur initiale arbitraire et $(\theta_n)_{n \geq 1}$ la suite définie par (9.3). On suppose que*

$$\gamma_n > 0, \quad \sum_{n \geq 1} \gamma_n = \infty \quad \text{et} \quad \sum_{n \geq 1} \gamma_n^2 < \infty.$$

Alors $\theta_n \longrightarrow \theta^*$ quand $n \rightarrow \infty$.

Démonstration. Nous commençons par montrer que la suite $(\theta_n)_{n \geq 1}$ est convergente, puis nous montrons que sa limite est donnée par la racine θ^* .

(a) Comme f est séparante pour le point θ^* et $\gamma_n > 0$, on a :

$$\begin{aligned} |\theta_n - \theta^*|^2 &= |\theta_{n-1} - \theta^*|^2 + 2(\theta_{n-1} - \theta^*) \cdot (\theta_n - \theta_{n-1}) + |\theta_n - \theta_{n-1}|^2 \\ &= |\theta_{n-1} - \theta^*|^2 - 2\gamma_n f(\theta_{n-1}) \cdot (\theta_{n-1} - \theta^*) + \gamma_n^2 |f(\theta_{n-1})|^2 \\ &\leq |\theta_{n-1} - \theta^*|^2 + \gamma_n^2 |f(\theta_{n-1})|^2. \end{aligned}$$

Alors la suite

$$x_n := |\theta_n - \theta^*|^2 - \sum_{k=1}^n \gamma_k^2 |f(\theta_{k-1})|^2, \quad n \geq 1,$$

est décroissante. Comme f est bornée, on déduit des conditions sur la suite $(\gamma_n)_{n \geq 1}$ que $x_n \geq -|f|_\infty \sum_{k \geq 1} \gamma_k^2$. Ainsi $x_n \rightarrow x_\infty$ et

$$|\theta_n - \theta^*| \rightarrow \ell := x_\infty + \sum_{k \geq 1} \gamma_k^2 |f(\theta_{k-1})|^2,$$

où la dernière série est bien convergente sous nos conditions.

(b) Comme f est continue et séparante pour θ^* , on

$$\eta := \inf_{\delta < |\theta - \theta^*| < 2\ell} (\theta - \theta^*) \cdot f(\theta) > 0.$$

Alors, si $\ell \neq 0$, on a $\delta < |\theta_n - \theta^*| < 2\ell$ à partir d'un certain rang N et

$$\sum_{n \geq 1} \gamma_k f(\theta_{k-1}) \cdot (\theta_{k-1} - \theta^*) \geq \sum_{n \geq N} \gamma_k f(\theta_{k-1}) \cdot (\theta_{k-1} - \theta^*) \geq \eta \sum_{n \geq N} \gamma_k = \infty,$$

ce qui est faux d'après le calcul suivant :

$$\begin{aligned} \sum_{n \geq 1} \gamma_k f(\theta_{k-1}) \cdot (\theta_{k-1} - \theta^*) &= - \sum_{n \geq 1} (\theta_k - \theta_{k-1}) \cdot (\theta_{k-1} - \theta^*) \\ &= -\frac{1}{2} \sum_{n \geq 1} (|\theta_k - \theta^*|^2 - |\theta_k - \theta_{k-1}|^2 - |\theta_{k-1} - \theta^*|^2) \\ &= \frac{1}{2} \left(\sum_k \gamma_k^2 |f(\theta_{k-1})|^2 - \ell + |\theta_0 - \theta^*|^2 \right) < \infty. \end{aligned}$$

On conclut alors que $\ell = 0$. ◇

9.5.2 L'algorithme de Robins-Monro

Suposons que la fonction f soit donnée par

$$f(\theta) = \mathbb{E}[F(\theta, X)],$$

où X est une variable aléatoire et F une fonction bornée. Soient $(X_i)_{i \geq 1}$ une suite iid de v.a. de même loi que X . Les X_i peuvent être vues comme des simulations indépendantes de X . On notera $\mathcal{F}_i := \mathcal{F}_i^X = \sigma(X_j, j \leq i)$, $i \geq 1$, la filtration canonique de $(X_i)_{i \geq 1}$.

Dans le cas où la fonction f ne peut être calculée explicitement, il semble naturel d'utiliser les simulations $(X_i)_{i \geq 1}$ pour construire un algorithme numérique convergent vers la racine θ^* de f .

L'idée la plus simple est d'approximer la fonction f à chaque étape de l'algorithme (9.3) par

$$\hat{f}_n(\theta_n) := \frac{1}{N_n} \sum_{k=1}^{N_n} F(\theta_n, X_{\bar{N}_{n-1}+k}),$$

où $N_n \rightarrow \infty$ et $\bar{N}_{n-1} = N_1 + \dots + N_{n-1}$. Intuitivement, un tel algorithme est raisonnable car la loi des grands nombres devrait nous assurer la convergence pour une bonne vitesse de convergence de la suite $(N_n)_n$ vers l'infini.

De façon surprenante, l'algorithme de Robins-Monro propose $N_n = 1$ à toute étape :

$$\begin{aligned} \theta_{n+1} &= \theta_n - \gamma_{n+1} Y_{n+1}, \\ Y_{n+1} &= f(\theta_n) + \xi_{n+1} \end{aligned} \quad (9.5)$$

où

$$\xi_{n+1} = F(\theta_n, X_{n+1}) - f(\theta_n) \quad \text{et par suite} \quad Y_{n+1} = F(\theta_n, X_{n+1}).$$

Avant d'énoncer notre résultat de convergence, remarquons que la bornitude de F et l'indépendance des X_i impliquent que

$$\sup_n \|\xi_n\|_\infty < \infty \quad \text{et} \quad \mathbb{E}[\xi_n | \mathcal{F}_{n-1}] = 0 \quad \text{pour tout } n \geq 1. \quad (9.6)$$

Nous allons maintenant montrer la convergence de cet algorithme en suivant de très près la démonstration de la version déterministe de la proposition 9.11. Plus précisément, la première étape de la démonstration de la proposition 9.11 introduit la suite déterministe $(x_n)_n$ décroissante et minorée, donc convergente. Cet argument sera remplacé par l'introduction d'une surmartingale minorée dont la convergence est alors assurée par notre théorème 9.5.

Proposition 9.12. *Soit $f : \mathbb{R}^d \rightarrow \mathbb{R}^d$ une fonction continue, bornée, nulle en un point $\theta^* \in \mathbb{R}^d$ et séparante pour θ^* . Soit $\theta_0 \in \mathbb{R}^d$ une valeur initiale arbitraire et $(\theta_n)_{n \geq 1}$ la suite définie par (9.5), où la suite de variables aléatoires $(\xi_n)_{n \geq 1}$ vérifie (9.6). On suppose que*

$$\gamma_n > 0, \quad \sum_{n \geq 1} \gamma_n = \infty \quad \text{et} \quad \sum_{n \geq 1} \gamma_n^2 < \infty. \quad (9.7)$$

Alors $\theta_n \rightarrow \theta^*$ p.s. quand $n \rightarrow \infty$.

Démonstration. 1- Nous allons d'abord montrer que la suite $(\theta_n)_n$ converge p.s. Sous la première condition dans (9.7), on voit que les variables aléatoires

$$S_n := \sum_{k=0}^{n-1} \gamma_{k+1}^2 \mathbb{E}[|Y_{k+1}|^2 | \mathcal{F}_k] = 2 \sum_{k=0}^{n-1} \gamma_{k+1}^2 \left(|f|^2_\infty + \sup_n \|\xi_n\|_\infty^2 \right) = : C < \infty, \quad \text{p.s.}$$

d'après (9.6). Alors, il existe une v.a. positive finie S_∞ telle que

$$S_n \rightarrow S_\infty := \sum_{k \geq 0} \gamma_{k+1}^2 \mathbb{E}[|Y_{k+1}|^2 | \mathcal{F}_k] \quad \text{p.s.}$$

Nous introduisons ensuite le processus

$$X_n := |\theta_n - \theta^*|^2 - S_n \quad \text{pour } n \geq 1,$$

et on vérifie que $\{X_n, n \geq 1\}$ est une surmartingale :

$$\begin{aligned} \mathbb{E}[X_{n+1}|\mathcal{F}_n] &= -S_{n+1} + \mathbb{E}[|\theta_{n+1} - \theta^*|^2|\mathcal{F}_n] \\ &= -S_{n+1} + |\theta_n - \theta^*|^2 + \mathbb{E}[|\theta_{n+1} - \theta_n|^2|\mathcal{F}_n] \\ &\quad - 2\gamma_{n+1}(\theta_n - \theta^*) \cdot f(\theta_n), \end{aligned}$$

et du fait que la fonction f est séparante pour le point θ^* :

$$\begin{aligned} \mathbb{E}[X_{n+1}|\mathcal{F}_n] &\leq -S_{n+1} + |\theta_n - \theta^*|^2 + \mathbb{E}[|\theta_{n+1} - \theta_n|^2|\mathcal{F}_n] \\ &= -S_{n+1} + |\theta_n - \theta^*|^2 + \gamma_{n+1}^2 \mathbb{E}[|Y_{n+1}|^2|\mathcal{F}_n] \\ &= -S_n + |\theta_n - \theta^*|^2 = X_n \quad \text{pour tout } n \geq 1. \end{aligned}$$

Comme $X_n \geq -C$, on applique le théorème 9.5 à la sous-martingale $-X$, et on obtient la convergence p.s. de X_n vers une v.a. $X_\infty \in \mathbb{L}^1$. Alors :

$$|\theta_n - \theta^*|^2 \longrightarrow L := X_\infty + S_\infty, \quad \text{p.s.}$$

2- Nous allons maintenant montrer que cette limite L est nulle. Pour cela, on commence par remarquer que, par la continuité de f et le fait qu'elle soit séparante pour θ^* ,

$$\eta := \inf_{\delta < |x - x_0| < 2L} (x - x_0) \cdot f(x) > 0 \quad \text{pour tout } \delta > 0.$$

En introduisant $\tau := \inf\{n : |\theta_n - \theta^*| \leq 2L\}$, ceci implique que

$$\begin{aligned} \sum_{k=0}^{\infty} \gamma_{k+1} f(\theta_k) \cdot (\theta_k - \theta^*) &\geq \sum_{k \geq \tau} \gamma_{k+1} f(\theta_k) \cdot (\theta_k - \theta^*) \\ &\geq \eta \sum_{k \geq \tau} \gamma_k = \infty \quad \text{sur } A_\delta := \{L > \delta\}, \end{aligned}$$

d'après la deuxième condition sur la suite $(\gamma_n)_n$ dans (9.7). Nous montrons maintenant que cette série ne peut prendre la valeur infinie. en effet :

$$0 \leq \sum_{k=0}^n 2\gamma_{k+1} \mathbb{E}[f(\theta_k) \cdot (\theta_k - \theta^*)] = \sum_{k=0}^n 2\gamma_{k+1} \mathbb{E}[Y_{k+1} \cdot (\theta_k - \theta^*)]$$

d'après (9.6). Alors :

$$\begin{aligned} 0 &\leq \sum_{k=0}^n 2\mathbb{E}[(\theta_k - \theta_{k+1}) \cdot (\theta_k - \theta^*)] \\ &= \sum_{k=0}^n \mathbb{E}[|\theta_{k+1} - \theta^*|^2 - |\theta_k - \theta^*|^2 - |\gamma_{k+1} Y_{k+1}|^2] \\ &= \mathbb{E}[|\theta_{n+1} - \theta^*|^2] - \mathbb{E}[|\theta_0 - \theta^*|^2] - \mathbb{E}[S_{n+1}] \\ &= \mathbb{E}[X_{n+1}] - |\theta_0 - \theta^*|^2 \leq X_0 - |\theta_0 - \theta^*|^2, \end{aligned}$$

d'après la propriété de surmartingale de X . On déduit alors que $\mathbb{P}[A_\delta] = 0$ pour tout $\delta > 0$, et par suite

$$\mathbb{P}[\cap_{0 < \delta \in \mathbb{Q}} A_\delta] = 0 \quad \text{et} \quad L = \lim_{n \rightarrow \infty} |\theta_n - \theta^*|^2 = 0.$$

◇

Chapitre 10

Introduction aux processus de branchement

Dans ce chapitre nous proposons une introduction aux processus de Galton-Watson qui constituent l'exemple de base des processus de branchement. Il s'agit d'un modèle très simple d'évolution d'une population décrite par une chaîne de Markov à valeurs dans \mathbb{N} . A chaque génération, tout individu de la population engendre un certain nombre de descendants, qui sont les seuls constituants de la génération suivante, et dont le nombre est indépendant de la taille de la population et des autres individus.

Les premières questions d'intérêt concernent la probabilité d'extinction ou au contraire d'explosion de la population.

Les processus de branchement ont de nombreuses applications, notamment en physique, en biologie et en télécommunications. Le problème a été initialement introduit par Francis Galton en 1873, et a été traité avec succès par Henry Watson la même année. Ces auteurs menaient une étude sur la probabilité d'extinction des noms de famille, motivés par l'observation soucieuse que de nombreuses familles de l'aristocratie britannique étaient éteintes ou risquaient de s'éteindre.

10.1 Processus de Galton-Watson

Etant donnée une suite $(\xi_i^n, 1 \leq i \leq n)_{n \geq 1}$ de v.a. iid de même loi qu'une v.a. générique ξ à valeurs dans \mathbb{N} , on définit le processus de Galton-Watson $(Z_n)_n$ par

$$Z_0 = 1 \quad \text{et} \quad Z_n := \sum_{i=1}^n \xi_i^n, \quad n \geq 1.$$

On note par $p_j := \mathbb{P}[\xi = j]$ pour tout $j \geq 0$. Alors $(Z_n)_n$ est une chaîne de Markov homogène sur l'espace d'état \mathbb{N} , de matrice de transition

$$P(i, j) := \mathbb{P}[Z_{n+1} = j | Z_n = i] = \sum_{j_1 + \dots + j_i = j} p_{j_1} \dots p_{j_i}$$

L'état 0 est absorbant. Afin d'éviter les situations triviales, on supposera toujours que

$$p_0 + p_1 < 1. \quad (10.1)$$

Cette condition implique que la chaîne est transiente. en effet

- si $p_0 > 0$, ceci est une conséquence immédiate du fait que p_0 est absorbant,
- si $p_0 = 0$, on calcule directement que

$$\begin{aligned} \mathbb{P}[Z_{n+i} \neq k, i \geq 1 | Z_n = k] &= 1 - \mathbb{P}[\cup_{i \geq 1} \{Z_{n+i} = k\} | Z_n = k] \\ &\geq 1 - \mathbb{P}[Z_{n+1} = k | Z_n = k] = 1 - p_1^n > 0 \end{aligned}$$

du fait que $p_0 = 0$ et $p_1 < 1$ d'après (10.1).

Dans le contexte actuel, il est clair que la matrice de transition n'est pas très commode à manipuler. Le problème est ainsi abordé par le biais de la *fonction génératrice* :

$$G(z) := G_\xi(z) := \mathbb{E}[z^\xi] = \sum_{n \geq 0} p_n z^n \quad \text{pour tout } z \in \mathbb{C},$$

dont nous n'aurons à utiliser que la restriction à \mathbb{R} . Le fonction génératrice de la chaîne $(Z_n)_n$ est alors donnée par

$$\begin{aligned} G_n(z) := G_{Z_n}(z) &:= \mathbb{E}[z^{Z_n}] \\ &= \mathbb{E}[\mathbb{E}\{z^{Z_n} | \mathcal{F}_{n-1}\}] \\ &= \mathbb{E}[G(z)^{Z_{n-1}}] \end{aligned}$$

du fait de l'indépendance des $\{\xi_i^{n-1}, 1 \leq i \leq n-1\}$. Comme $Z_0 = 1$, on obtient ainsi la fonction génératrice du processus de Galton-Watson :

$$G_0(z) = z \quad \text{et} \quad G_n(z) = \underbrace{G \circ \dots \circ G}_n(z), \quad z \in \mathbb{C}.$$

10.2 Propriétés de la fonction génératrice

La fonction génératrice permet d'obtenir les moments de la v.a sous-jacente quand ils existent. Dans le contexte actuel, on a :

$$\mathbb{E}[Z_1] = \mathbb{E}[\xi] = \sum_{n \geq 1} n p_n = G'(1)$$

et, plus généralement,

$$\mathbb{E}[Z_n] = G'_n(1) = \mathbb{E}[\xi]^n. \quad (10.2)$$

La dernière formule peut être vérifiée facilement par récurrence.

De même, si $\mathbb{E}[\xi^2] < \infty$, on calcule directement que

$$\text{Var}[Z_n] = \begin{cases} \frac{\text{Var}[\xi]\mathbb{E}[\xi]^{n-1}(\mathbb{E}[\xi]^n-1)}{\mathbb{E}[\xi]-1} & \text{if } \mathbb{E}[\xi] \neq 1, \\ n\text{Var}[\xi] & \text{if } \mathbb{E}[\xi] = 1. \end{cases} \quad (10.3)$$

Les moments d'ordre supérieurs, s'ils existent peuvent aussi être calculés par des dérivations d'ordre supérieur.

Les propriétés suivantes de G peuvent être vérifiées sans peine.

Lemme 10.1. *Supposons $p_0 + p_1 < 1$. Alors G est une fonction croissante strictement convexe, $G(0) = p_0$, $G(1) = 1$, et*

- (i) *si $\mathbb{E}[\xi] \leq 1$, G n'a pas de point fixe sur $[0, 1[$,*
- (ii) *si $\mathbb{E}[\xi] > 1$, G admet un unique point fixe sur $[0, 1[$,*
- (iii) *soient $\pi := \min\{t \in [0, 1] : G(t) = t\}$, et $0 \leq s < \pi < t < 1$. Alors*

$$G_n(s) \uparrow \pi, \quad G_n(t) \downarrow \pi, \quad \text{et} \quad G_n(\pi) = \pi \quad \text{pour tout } n \geq 1.$$

(iv) *les fonctions G_n sont différentiables, convergent simplement sur $[0, 1[$, et*

$$G'_n(s) \geq G'(s)^n, \quad G'_n(t) \leq G'(t)^n \quad \text{pour tous } 0 \leq s < \pi \leq t < 1.$$

10.3 Loi de reproduction géométrique

Dans ce paragraphe, nous considérons l'exemple où la v.a. de reproduction ξ est distribuée suivant une loi géométrique :

$$\mathbb{P}[\xi = n] = (1-p)p^n, \quad 0 < p < 1, \quad \text{pour tout } n \geq 0.$$

La fonction génératrice se calcule immédiatement :

$$G(z) = \frac{1-p}{1-pz} \quad \text{pour tout } pz \neq 1,$$

et la moyenne est donnée par

$$\mathbb{E}[\xi] = G'(1) = \frac{p}{1-p}.$$

L'équation $G(z) = z$ admet deux racines 1 et $\frac{1-p}{p}$, et par suite, avec les notations du lemme 10.1 :

$$\pi = 1 \wedge \frac{1-p}{p} = \mathbf{1}_{\{p \leq \frac{1}{2}\}} + \frac{1-p}{p} \mathbf{1}_{\{p > \frac{1}{2}\}}. \quad (10.4)$$

Pour calculer les fonctions génératrices G_n , on remarque que

$$\frac{G(z) - \frac{1-p}{p}}{G(z) - 1} = \frac{1-p}{p} \frac{z - \frac{1-p}{p}}{z-1},$$

ce qui donne par itération immédiate

$$\frac{G_n(z) - \frac{1-p}{p}}{G_n(z) - 1} = \left(\frac{1-p}{p}\right)^n \frac{z - \frac{1-p}{p}}{z-1}, \quad (10.5)$$

d'où l'on peut tirer une expression explicite des fonctions génératrices G_n . En particulier, on obtient :

$$G_n(0) = \begin{cases} \frac{\mathbb{E}[\xi]^n - 1}{\mathbb{E}[\xi]^{n+1} - 1} & \text{si } \mathbb{E}[\xi] \neq 1, \\ \frac{n}{n+1} & \text{si } \mathbb{E}[\xi] = 1. \end{cases}$$

Par définition, on a $G_n(0) = \mathbb{P}[Z_n = 0]$. Comme 0 est un état absorbant, la suite d'événements $(\{Z_n = 0\})_{n \geq 1}$ est croissante. On déduit alors que l'événement d'extinction $\cup_{n \geq 1} \{Z_n = 0\}$ a pour probabilité :

$$\mathbb{P}[\cup_{n \geq 1} \{Z_n = 0\}] = \lim_{n \rightarrow \infty} \uparrow \mathbb{P}[Z_n = 0] = \begin{cases} \lim_{n \rightarrow \infty} \uparrow \frac{\mathbb{E}[\xi]^n - 1}{\mathbb{E}[\xi]^{n+1} - 1} & \text{si } \mathbb{E}[\xi] \neq 1, \\ 1 & \text{si } \mathbb{E}[\xi] = 1. \end{cases}$$

Ceci conduit aux conclusions suivantes :

- si $\mathbb{E}[\xi] \leq 1$, i.e. $p \leq \frac{1}{2}$, il y a extinction avec probabilité 1,
- si $\mathbb{E}[\xi] > 1$, i.e. $p > \frac{1}{2}$, la probabilité d'extinction de la population est $\mathbb{E}[\xi]^{-1} = \frac{1-p}{p}$,

qui se résument par

$$\mathbb{P}[\text{extinction}] = \mathbb{P}[\cup_{n \geq 1} \{Z_n = 0\}] = \pi,$$

où π est le plus petit point fixe de G dans l'intervalle $[0, 1]$ calculé dans (10.4).

10.4 Probabilité d'extinction et classification

La remarque concernant l'événement d'extinction du paragraphe précédent n'est pas spécifique au cas d'une loi de reproduction géométrique. L'état 0 étant absorbant, on a l'interprétation suivante des événements

$$\{Z_n = 0\} = \{\text{extinction à la génération } n\},$$

la suite des événements $(\{Z_n = 0\})_n$ est croissante, et

$$\cup_{n \geq 1} \{Z_n = 0\} = \{\text{extinction}\}.$$

La probabilité d'extinction de la population peut ainsi être calculée par le biais de la fonction génératrice puisque le théorème de convergence monotone permet d'obtenir

$$\mathbb{P}[\text{extinction}] = \lim_{n \rightarrow \infty} \uparrow \mathbb{P}[Z_n = 0] = \lim_{n \rightarrow \infty} \uparrow G_n(0).$$

De même on introduit l'événement

$$\{\text{explosion}\} := \{Z_n \rightarrow \infty\}.$$

En vue des propriétés de la fonction génératrice G du lemme 10.1 (iii), et du fait que tous les états $k \geq 1$ sont transients, on obtient immédiatement le résultat général suivant pour les processus de Galton-Watson.

Proposition 10.2. *Avec π désignant le plus petit point fixe de G sur $[0, 1]$:*

$$\mathbb{P}[\text{extinction}] = 1 - \mathbb{P}[\text{explosion}] = \pi.$$

Ainsi, on a la classification suivante :

- *cas sous-critique* : si $\mathbb{E}[\xi] < 1$, il y a extinction p.s.
- *cas critique* : si $\mathbb{E}[\xi] = 1$, il y a extinction p.s.
- *cas sur-critique* : si $\mathbb{E}[\xi] > 1$, il y a extinction avec probabilité $\pi \in]0, 1[$, et explosion avec probabilité $1 - \pi$.

Le résultat suivant donne des informations plus précises sur la vitesse d'extinction ou d'explosion.

Proposition 10.3. (a) *Dans le cas sous-critique $\mathbb{E}[\xi] < 1$, on a $Z_n = 0$ à partir d'un certain rang, et*

$$\mathbb{P}[Z_n \neq 0] \leq \mathbb{E}[\xi]^n \quad \text{pour tout } n \geq 1.$$

(b) *Dans le cas surcritique $\mathbb{E}[\xi] > 1$, pour tout $a > 1$, on a $Z_n \notin [1, a]$ à partir d'un certain rang, et*

$$1 - \mathbb{P}[Z_n > a | \text{non extinction}] \leq \frac{\pi^{1-a}}{1-\pi} G'(\pi)^n \quad \text{pour tout } n \geq 1.$$

(c) *Dans le cas critique $\mathbb{E}[\xi] = 1$, sous l'hypothèse supplémentaire $\mathbb{E}[\xi^2] < \infty$, on a*

$$\lim_{n \rightarrow \infty} n \mathbb{P}[Z_n \neq 0] = \frac{2}{\text{Var}[\xi]}.$$

Démonstration. 1- Pour $a > 0$ et $n \geq 0$, on considère l'événement $A_n := \{1 \leq Z_n \leq a\}$, et on calcule pour un paramètre arbitraire $z \in]0, \pi[$ que :

$$\begin{aligned} \mathbb{P}[A_n] &= \sum_{k=1}^a \mathbb{P}[Z_n = k] \\ &\leq z^{-a} \sum_{k=1}^a z^k \mathbb{P}[Z_n = k] \\ &\leq z^{-a} (G_n(a) - G_n(0)) \\ &\leq z^{1-a} G'_n(z) \leq z^{1-a} G'_n(\pi), \end{aligned}$$

où les deux dernières inégalités sont des conséquences de la convexité de G_n . Dans les cas sous-critiques et sur-critiques, on rappelle que $\pi < 1$ et $G'_n(\pi) < 1$.

L'inégalité précédente implique alors que $\sum_{n \geq 1} \mathbb{P}[A_n] < \infty$, et on déduit du lemme de Borel-Cantelli que $\mathbb{P}[\limsup_n A_n] = \bar{0}$.

En revenant à la définition des A_n , ceci montre exactement que $Z_n \notin [1, a]$ à partir d'un certain rang. En particulier, dans le cas sous-critique ou l'extinction a lieu p.s., on obtient que $Z_n = 0$ à partir d'un certain rang.

2- L'estimation de la probabilité d'extinction du (a) est une conséquence de la convexité de G_n :

$$\mathbb{P}[Z_n \neq 0] = 1 - G_n(0) = G_n(1) - G_n(0) \leq G'_n(1) = \mathbb{E}[\xi]^n.$$

Quand à l'estimation de la probabilité conditionnelle du (b), on la déduit de la première étape de cette preuve :

$$\begin{aligned} 1 - \mathbb{P}[Z_n > a | \text{non extinction}] &= \frac{\mathbb{P}[\text{non extinction}] - \mathbb{P}[\text{non extinction et } Z_n > a]}{\mathbb{P}[\text{non extinction}]} \\ &= \frac{\mathbb{P}[\text{non extinction et } Z_n \leq a]}{\mathbb{P}[\text{non extinction}]} \\ &\leq \frac{\mathbb{P}[1 \leq Z_n \leq a]}{\mathbb{P}[\text{non extinction}]} \leq \frac{z^{1-a} G'_n(\pi)}{1 - \pi}. \end{aligned}$$

3- Pour $\mathbb{E}[\xi] = 1$, on a $G'(1) = 1$ et $G''(1) = \sigma^2 := \text{Var}[\xi]$, et le développement de G au voisinage de 1 s'écrit

$$G(z) = z + \frac{\sigma^2}{2}(1-z)^2 + (1-z)^2 \circ(z), \quad z \in [0, 1],$$

où \circ est une fonction tendant vers 0 pour $z \rightarrow 1$. En ré-utilisant le fait que $G'(1) = 1$, ceci conduit à :

$$\frac{1}{1-G(z)} - \frac{1}{1-z} = \frac{\sigma^2}{2} + \bar{\circ}(z),$$

où $\bar{\circ}(z)$ est encore une fonction tendant vers 0 pour $z \rightarrow 1$. Ceci implique que

$$\begin{aligned} \frac{1}{n} \left(\frac{1}{1-G_n(z)} - \frac{1}{1-z} \right) &= \frac{1}{n} \sum_{k=1}^n \left(\frac{1}{1-G_k(z)} - \frac{1}{1-G_{k-1}(z)} \right) \\ &= \frac{\sigma^2}{2} + \frac{1}{n} \sum_{k=1}^n \bar{\circ}(G_{k-1}(z)) \end{aligned}$$

Comme $G_n(z) \rightarrow 1$ pour tout $z < 1$, on déduit du critère d'Abel que

$$\lim_{n \rightarrow \infty} \frac{1}{n(1-G_n(z))} = \frac{\sigma^2}{2}.$$

◇

10.5 Comportement asymptotique

Nous allons maintenant utiliser les théorèmes de convergence des martingales pour déduire des propriétés asymptotiques du processus de Galton-Watson. On considère la filtration canonique de $(Z_n)_n$ définie par

$$\mathcal{F}_n := \sigma \{Z_i, i \leq n\}.$$

L'observation essentielle est que le processus normalisé :

$$M_n := \frac{Z_n}{\mathbb{E}[\xi]^n}, \quad n \geq 0,$$

est une martingale positive partant de $M_0 = 1$. Par application du théorème 9.5 de convergence des sous-martingales à $-M$, on déduit l'existence d'une v.a. $M_\infty \in \mathbb{L}^1$ telle que

$$M_n \longrightarrow M_\infty, \text{ p.s.}$$

Ce résultat n'est pas encore assez significatif car il ne donne pas d'information sur la nature de la limite M_∞ .

Dans les cas critiques et sous critique, on sait que l'événement d'extinction a lieu p.s., et par suite $M_\infty = 0$ p.s. (en particulier, $M_n \not\rightarrow M_\infty$ dans \mathbb{L}^1).

Avant de continuer notre analyse dans le cas général, revenons à l'exemple d'une variable de reproduction géométrique dans lequel les calculs peuvent être conduits explicitement.

Exemple 10.4. (*Retour au cas géométrique*) Reprenant le cadre du paragraphe 10.3 dans le cadre sur-critique, on note $m := \mathbb{E}[\xi] = p/(1-p) > 1$, et on calcule la transformée de Laplace de M_n en utilisant l'expression de G_n obtenue de (10.5) :

$$\begin{aligned} \mathbb{E}[e^{-\lambda M_n}] &= G_n(e^{-\lambda m^{-n}}) \\ &= 1 - \frac{m^n(m-1)(1 - e^{-\lambda m^{-n}})}{m(m^n - 1)(1 - e^{-\lambda m^{-n}}) + m - 1} \\ &\longrightarrow 1 - \frac{\lambda(m-1)}{\lambda m + m - 1}. \end{aligned}$$

Ceci caractérise la distribution de la v.a. M_∞ , et on peut vérifier que

$$\mathbb{P}[M_\infty \leq x] = \frac{1}{m} + \left(1 - \frac{1}{m}\right) \left(1 - e^{(1-\frac{1}{m})x}\right), \quad x \geq 0.$$

Afin de mieux comprendre la nature de la limite M_∞ dans le cas sur-critique, on suppose dans le résultat suivant que la variable de reproduction ξ est de carré intégrable afin d'obtenir une convergence dans un sens plus fort.

Proposition 10.5. *Supposons que $m := \mathbb{E}[\xi] > 1$ et $\sigma^2 := \text{Var}[\xi^2] < \infty$. Alors*

- (i) $M_n \rightarrow M_\infty$ dans \mathbb{L}^2 ,
- (ii) $\mathbb{E}[M_\infty] = 1$, $\text{Var}[M_\infty] = \frac{\sigma^2}{m^2 - m}$, et $\mathbb{P}[M_\infty = 0] = \pi$.

Démonstration. On calcule directement en utilisant (10.2) et (10.3) que

$$\mathbb{E}[M_n^2] = \frac{\mathbb{E}[Z_n^2]}{m^{2n}} = 1 + \frac{\sigma^2(1 - m^{-n})}{m(m-1)} \leq 1 + \frac{\sigma^2}{m(m-1)},$$

et on déduit du théorème 9.1 de convergence des martingales de carré intégrable que $M_n \rightarrow M_\infty$ dans \mathbb{L}^2 . ceci implique que les moments d'ordre 1 et 2 de M_∞ sont les limites des moments correspondants de M_n , justifiant les expressions données dans la proposition.

Enfin, comme $\mathbb{E}[M_\infty] = 1$, on voit que $r = \mathbb{P}[M_\infty = 0] < 1$, et on obtient en conditionnant par Z_1 , et en utilisant l'indépendance des individus de chaque génération,

$$r = \sum_{k \geq 0} \mathbb{P}[M_\infty = 0 | Z_1 = k] \mathbb{P}[Z_1 = k] = \sum_{k \geq 0} \mathbb{P}[M_\infty = 0]^k p_k = G(r),$$

et par suite $r = \pi$. ◇

Pour terminer ce chapitre, notons que des résultats plus fort peuvent être obtenus sous des conditions plus faibles en utilisant la transformée de Laplace de Z_n , comme dans l'exemple de la variable de reproduction géométrique, et en remarquant que :

$$\varphi_n(\lambda) := \mathbb{E}[e^{-\lambda M_n}] = G_n\left(e^{-\lambda/\mathbb{E}[\xi]^n}\right) = G \circ G_{n-1}\left(e^{-\lambda/\mathbb{E}[\xi]^n}\right) = G\left(\varphi_n\left(\frac{\lambda}{\mathbb{E}[\xi]}\right)\right).$$

Par convergence dominée, on obtient alors que la transformée de Laplace $\varphi_\infty(\lambda) := \mathbb{E}[e^{-\lambda M_\infty}]$ de la limite M_∞ satisfait l'équation d'Abel :

$$\varphi_\infty(\lambda) = G\left(\varphi_\infty\left(\frac{\lambda}{\mathbb{E}[\xi]}\right)\right), \lambda \geq 0.$$

Chapitre 11

Arrêt optimal

11.1 Arrêt optimal en horizon fini

Soit $(\Omega, \mathcal{A}, \mathbb{F}, \mathbb{P})$ un espace probabilisé filtré. Dans ce paragraphe, on considère une maturité finie $T \in \mathbb{N}$ et un processus $X = \{X_t, t = 0, \dots, T\}$ \mathbb{F} -adapté intégrable.

On note par \mathcal{T}^T l'ensemble des temps d'arrêt à valeurs dans $\mathbb{N} \cap [0, T]$, et on cherche à résoudre le problème d'arrêt optimal

$$V^T := \sup_{\tau \in \mathcal{T}^T} \mathbb{E}[X_\tau]. \quad (11.1)$$

On dira qu'un temps d'arrêt $\tau^* \in \mathcal{T}^T$ est optimal si $V^T = \mathbb{E}[X_{\tau^*}]$.

Pour résoudre ce problème, on introduit le processus Y défini par la récurrence arrière :

$$Y_T = X_T \quad \text{et} \quad Y_t = \max\{X_t, \mathbb{E}[Y_{t+1} | \mathcal{F}_t]\}, \quad t = 0, \dots, T-1.$$

Vérifions d'abord par une récurrence arrière que ce processus est bien intégrable. A la date finale, on a $Y_T = X_T \in \mathbb{L}^1$. Si on suppose que $Y_t \in \mathbb{L}^1$, alors $\mathbb{E}|Y_{t-1}| \leq \mathbb{E}|X_{t-1}| + \mathbb{E}|\mathbb{E}\{Y_t | \mathcal{F}_{t-1}\}| \leq \mathbb{E}|X_{t-1}| + \mathbb{E}|Y_t|$. Ainsi Y est intégrable.

Le processus Y est appelé *enveloppe de Snell* du processus X . Il apparaît naturellement dans ce problème, et on s'attend à ce que Y_0 soit égal à V_0 , ce que l'on démontrera ci-dessous. En effet, Si le problème d'arrêt optimal était posé au temps T , le seul choix possible de temps d'arrêt est $\tau = T$, d'où la définition $Y_T = X_T$. A la date $T-1$, nous avons le choix entre l'arrêt en $T-1$ ou en T , et on choisit la stratégie d'arrêt en comparant le gain en s'arrêtant, X_{T-1} , et le gain espéré si on continuait, $\mathbb{E}[Y_T | \mathcal{F}_{T-1}] = \mathbb{E}[X_T | \mathcal{F}_{T-1}]$. Ceci explique la définition de Y_{T-1} . En procédant de manière rétrograde date par date, on comprend la logique derrière l'enveloppe de Snell.

Proposition 11.1. *Supposons que X est intégrable. Alors, l'enveloppe de Snell Y est une surmartingale. C'est la plus petite surmartingale majorant le processus X .*

Démonstration. Nous avons déjà vérifié que Y est intégrable. La définition de Y montre immédiatement qu'il s'agit d'une surmartingale majorant X . Soit \tilde{Y} une surmartingale majorant X . Montrons par récurrence arrière que $\tilde{Y}_t \geq Y_t$ pour tout $t = 0, \dots, T$, p.s.

Par définition, $\tilde{Y}_T \geq X_T = Y_T$. Supposons maintenant que $\tilde{Y}_t \geq Y_t$. Comme \tilde{Y} est une surmartingale, on a :

$$\tilde{Y}_{t-1} \geq \mathbb{E}[\tilde{Y}_t | \mathcal{F}_{t-1}] \geq \mathbb{E}[Y_t | \mathcal{F}_{t-1}].$$

De plus \tilde{Y} majore X et par suite :

$$\tilde{Y}_{t-1} \geq \max\{X_{t-1}, \mathbb{E}[Y_t | \mathcal{F}_{t-1}]\} = Y_{t-1}.$$

◇

Le résultat suivant montre que l'enveloppe de Snell permet de résoudre le problème d'arrêt optimal V^T , et exhibe un temps d'arrêt optimal.

Proposition 11.2. *La variable aléatoire*

$$\tau^* := \inf\{t = 0, \dots, T : Y_t = X_t\}$$

est un temps d'arrêt tel que le processus arrêté $Y^{\tau^} := Y_{\cdot \wedge \tau^*}$ est une martingale et :*

$$Y_0 = \sup_{\tau \in \mathcal{T}^T} \mathbb{E}[X_\tau] = \mathbb{E}[X_{\tau^*}].$$

Démonstration. Puisque $Y_T = X_T$, la variable τ^* définit bien un temps d'arrêt dans \mathcal{T}^T , comme premier temps d'atteinte du processus $Y - X$ du niveau 0.

Montrons maintenant que le processus arrêté Y^{τ^*} est une martingale. Par définition de τ^* , on a :

$$Y_{t+1}^{\tau^*} - Y_t^{\tau^*} = (Y_{t+1} - Y_t)\mathbf{1}_{\{\tau^* \geq t+1\}}.$$

Par définition du processus Y , on a $Y_t > X_t$ et $Y_t = \mathbb{E}[Y_{t+1} | \mathcal{F}_t]$ sur l'ensemble $\{\tau^* \geq t+1\}$ (\mathcal{F}_t -mesurable comme complémentaire de $\{\tau^* \leq t\}$). D'où :

$$Y_{t+1}^{\tau^*} - Y_t^{\tau^*} = (Y_{t+1} - \mathbb{E}[Y_{t+1} | \mathcal{F}_t])\mathbf{1}_{\{\tau^* \geq t+1\}},$$

et le résultat est obtenu en prenant l'espérance conditionnelle à \mathcal{F}_t et en utilisant le fait que $\{\tau^* \geq t+1\} \in \mathcal{F}_t$.

Comme le processus arrêté Y^{τ^*} est une martingale, on a :

$$Y_0 = \mathbb{E}[Y_T^{\tau^*}] = \mathbb{E}[Y_{T \wedge \tau^*}] = \mathbb{E}[Y_{\tau^*}] = \mathbb{E}[X_{\tau^*}].$$

Par ailleurs, pour tout $\tau \in \mathcal{T}$, le processus arrêté Y^τ est une surmartingale, d'après le théorème d'arrêt de Doob, et par suite :

$$Y_0 \geq \mathbb{E}[Y_T^\tau] = \mathbb{E}[Y_\tau] \geq \mathbb{E}[X_\tau]$$

par définition de Y . On a ainsi montré la dernière partie de la proposition. ◇

11.2 Exemples en horizon fini

11.2.1 Le problème du parking

Vous conduisez le long d'une rue infinie vers votre lieu de rendez-vous se situant dans un quartier bien animé. Le stationnement dans la rue est autorisé, mais bien sûr peu de places sont libres. Alors, si vous avez la possibilité de vous garer, à quelle distance de votre lieu de rendez-vous décidez-vous de prendre la place ?

Modélisation. On considère la modélisation discrète suivante.

1. Vous démarrez à l'origine. Des emplacements de stationnement sont disponibles à tous les entiers. On considère une suite $(\xi_n)_{n \geq 0}$ de v.a. iid de loi de Bernoulli de paramètre p , où $\xi_n = 1$ si et seulement si l'emplacement au point n est déjà occupé. Le lieu de rendez-vous se trouve au point entier $R > 0$.
2. Si l'emplacement n est disponible, i.e. $\xi_n = 0$, et que vous décidez de vous y garer, vous subissez le coût $|R - n|$ correspondant à l'effort de faire la distance restante en marchant.
3. Quand vous arrivez au niveau du point n , vous ne pouvez pas savoir si des places de stationnement sont disponibles au niveau $n + 1$ ou plus loin. Si vous décidez de passer au point $n + 1$, vous ne pouvez retourner aux points précédents.
4. Enfin, si vous arrivez en R , votre point de rendez-vous sans vous être garé, vous prenez la première place de stationnement libre qui se présente. Ainsi, si l'emplacement R est occupé, le coût espéré que vous subissez est :

$$-X_R := \sum_{j \geq 1} j p^{j-1} (1-p) = \frac{1}{1-p}.$$

Avant d'arriver en R le processus de coût s'écrit

$$-X_n := (R - n) \mathbf{1}_{\{\xi_n = 0\}} + \infty \mathbf{1}_{\{\xi_n = 1\}},$$

où le coût infini pour $\xi_n = 1$ signifie que vous ne pouvez occuper la place à aucun coût fini.

Le problème d'arrêt optimal consiste à chercher le temps d'arrêt qui minimise le coût de l'effort de l'agent, ou en inversant les signes :

$$\sup_{\tau \in \mathcal{T}^R} \mathbb{E}[X_\tau].$$

L'enveloppe de Snell est donnée par $Y_R = X_R$ et

$$Y_n = \max \{X_n, \mathbb{E}[Y_{n+1} | \mathcal{F}_n]\}, \quad n < R.$$

Un simple raisonnement par récurrence arrière, utilisant l'indépendance des ξ_n , montre que Y_n est une fonction de ξ_n et :

$$\mathbb{E}[Y_{n+1}|\mathcal{F}_n] = \mathbb{E}[Y_{n+1}] = : f(R-n),$$

où f est une fonction à déterminer. Remarquons que f est croissante du fait de la décroissance de $n \mapsto \mathbb{E}[Y_n]$, conséquence de la propriété de surmartingale de Y .

Par ailleurs, supposons que $\xi_n = 0$ et qu'il est optimal d'arrêter en n . Alors $Y_n = X_n = -(R-n) \geq \mathbb{E}[Y_{n+1}|\mathcal{F}_n] = f(R-n)$ et :

$$-(R-n-1) > f(R-n) \geq f(R-n-1) = \mathbb{E}[Y_{n+2}|\mathcal{F}_{n+1}].$$

Par conséquent, si $\xi_{n+1} = 0$, il est également optimal de s'arrêter en $n+1$. Ceci montre que la stratégie d'arrêt optimale est de la forme

$$\tau^* = \inf\{n \geq R-r^* : \xi_n = 0\}$$

où r^* est une constante à déterminer. On note $\ell(r)$ la performance espérée en utilisant une stratégie de seuil avec paramètre r , i.e.

$$\ell(r) = \mathbb{E}[Y_{\tau(r)}], \quad \text{où } \tau(r) = \inf\{n \geq R-r : \xi_n = 0\}$$

Alors

$$\ell(0) = -(1-p) \times 0 + pX_R = \frac{p}{1-p},$$

et, pour $r \geq 1$, $\ell(r) = -(1-p)r + p\ell(r-1)$. On déduit alors que

$$-\ell(r) = r+1 + \frac{2p^{r+1}-1}{1-p}, \quad r \in \mathbb{N}.$$

Pour maximiser $\ell(r)$, on remarque que la fonction $r \mapsto \ell(r+1) - \ell(r) = -1 + 2p^{r+1}$ est décroissante en r . On en déduit alors que

$$r^* = \inf\{r \geq 0 : -1 + 2p^{r+1} \leq 0\}.$$

A titre d'exemple, on peut voir que pour $p \leq 0.5$, il faut chercher à se garer en arrivant à destination, et que pour $p = 0.9$, il faut chercher la première place disponible dès qu'on arrive à 6 places de la destination.

11.2.2 Problème de la secrétaire

Ce problème a généré une littérature très abondante du fait de ses nombreuses applications. Nous allons nous intéresser ici au problème de base, et nous ne rentrerons pas dans les différentes ramifications qui ont été développées.

Notons que ce problème est aussi connu sous le nom de *problème de la princesse*, où on remplace la recherche d'une secrétaire par la recherche d'une épouse. Le jeu Googol est aussi une autre manifestation de ce problème, où on

remplace la recherche de la secrétaire ou de l'épouse par la recherche de la carte la plus rénumératrice.

Description du problème.

1. Un nombre N connu de candidats sont en course pour un poste de secrétaire vacant.
2. Les candidats peuvent être tous comparés strictement (une fois rencontrés) et rangés du meilleur au moins intéressant.
3. Les candidats sont auditionnés un à un dans un ordre arbitraire, les $N!$ façons d'ordonner les candidats sont équiprobables.
4. A l'issue de l'audition de chaque candidat, et sur la base de son rang relatif par rapport aux candidats précédemment auditionnés, on doit
 - soit sélectionner pour le poste, terminant ainsi la procédure de sélection,
 - soit refuser définitivement sa candidature, sans possibilité de le rappeler ultérieurement, et passer au candidat suivant.
5. L'objectif est de sélectionner LE meilleur candidat, i.e. le gain est défini par 1 si on sélectionne le meilleur candidat, et 0 sinon.

Modélisation. La difficulté de ce problème est liée à la non observabilité du processus de gain $\{\hat{X}_n, 1 \leq n \leq N\}$ qui vaut 1 pour LE meilleur candidat et 0 pour tous les autres. Pour contourner cette difficulté, on prend comme variables les rangs relatifs $(\xi_n)_{1 \leq n \leq N}$. Précisément, ξ_n désigne le rang du $n^{\text{ème}}$ candidat auditionné parmi les n candidats auditionnés. Ainsi $\xi_1 = 1$, et $\mathbb{P}[\xi_2 = 1] = \mathbb{P}[\xi_2 = 2] = 1/2$, etc...

Pour chaque $n \geq 1$, la v.a. ξ_n est distribuée selon la loi uniforme sur $\{1, \dots, n\}$, et est indépendante des $\xi_i, i < n$.

On note $\mathcal{F}_n := \sigma(\xi_i, i \leq n)$ la filtration canonique correspondante.

Nous passons maintenant à l'écriture du gain espérée après l'audition du $n^{\text{ème}}$:

- Si $\xi_n \neq 1$, alors le gain du recruteur est $X_n := \hat{X}_n = 0$, puisqu'on est sûr que le $n^{\text{ème}}$ candidat n'est pas le meilleur candidat, n'étant par le meilleur dans le sous-groupe de ceux qui ont été auditionnés. Il n'y a donc aucun avantage à arrêter la procédure de sélection.
- Si $\xi_n = 1$ Le $n^{\text{ème}}$ candidat auditionné est admissible. Dans ce cas, si on décide de sélectionner le candidat, et donc d'arrêter la procédure de sélection, le gain du recruteur est 1 si ce candidat se révèle être LE meilleur, et 0 sinon. Par conséquent, le gain espéré après l'audition est la probabilité que le meilleur candidat se situe parmi les n premiers auditionnés, i.e.

$$X_n := \mathbb{E}[\hat{X}_n | \mathcal{F}_n] = \frac{n}{N} \mathbf{1}_{\{\xi_n=1\}}, \quad 1 \leq n \leq N.$$

Résolution. On représente X_n par le vecteur $(X_n(i), 1 \leq i \leq n)$ où i indique la réalisation de ξ_n . Ainsi

$$X_n(1) = \frac{n}{N} \quad \text{et} \quad X_n(i) = 0 \quad \text{pour tout} \quad i = 2, \dots, n.$$

L'enveloppe de Snell de X s'avère être facile à calculer pour n assez grand. Notons par

$$n^* := \inf \left\{ n : \sum_{k=n+1}^N \frac{1}{k-1} < 1 \right\}.$$

Alors, on peut montrer par récurrence que les composantes $\{Y_n, n \geq n^*\}$ sont données par

$$Y_n(1) = X_n(1) = \frac{n}{N} \quad \text{et} \quad Y_n(i) = \frac{n}{N} \sum_{k=n+1}^N \frac{1}{k-1} \quad \text{pour } i = 2, \dots, n.$$

Le même calcul montre que

$$Y_{n^*-1}(1) > X_{n^*-1}(1),$$

i.e. il n'est pas optimal de retenir le $(n^* - 1)$ ème candidat, même s'il est meilleur que tous ses prédécesseurs. Comme le gain généré par l'arrêt de la procédure de recrutement $X_n(1)$ est croissant par rapport à n , on voit que la dernière inégalité stricte implique que :

$$Y_{n^*-2} - X_{n^*-2} \geq \mathbb{E}[Y_{n^*-1} | \mathcal{F}_{n^*-2}] - X_{n^*-2} \geq \mathbb{E}[Y_{n^*-1} - X_{n^*-1} | \mathcal{F}_{n^*-2}] > 0,$$

i.e. il n'est jamais optimal d'arrêter la procédure à une date $n < n^*$. Ainsi, la stratégie d'arrêt optimale est donnée par :

$$\tau^* = N \wedge \inf \{n \geq n^* : \xi_n = 1\}.$$

Pour se donner une idée de n^* , on considère l'approximation de $\sum_{n+1}^N (k-1)^{-1}$ par l'intégrale $\int_n^N t^{-1} dt$, pour N grand, ce qui donne $n^* \sim Ne^{-1} \approx .37 N$. Ainsi, la stratégie optimale consiste à rejeter systématiquement les malheureux candidats qui se présentent parmi les 37% premiers, puis sélectionner le premier candidat qui sera rangé premier parmi ses prédécesseurs.

11.3 Arrêt optimal en horizon infini

Nous nous intéressons maintenant au problème d'arrêt optimal

$$V := \sup_{\tau \in \mathcal{T}} \mathbb{E}[X_\tau],$$

où \mathcal{T} est l'ensemble des temps d'arrêt à valeurs dans \mathbb{N} , et $\{X_t, t \in \mathbb{N}\}$ est un processus vérifiant :

$$\mathbb{E} \left[\sup_{t \geq 0} |X_t| \right] < \infty. \quad (11.2)$$

L'enveloppe de Snell ne peut plus être introduite comme dans le premier paragraphe car il n'y a pas un temps final auquel on pourrait la définir sans ambiguïté pour repartir sur la définition rétrograde.

Nous allons par contre utiliser la fait que la définition du premier paragraphe est basée sur l'idée de considérer le problème d'arrêt optimal à toute date intermédiaire, en restreignant bien sûr l'ensemble des temps d'arrêt à ceux prenant des valeurs dans les dates restantes. Ceci conduit naturellement à une définition du genre :

$$"Y_t := \sup_{\tau \in \mathcal{T}_t} \mathbb{E}[X_\tau],"$$

où \mathcal{T}_t désigne l'ensemble de tous les temps d'arrêt à valeurs dans $\mathbb{N} \cap [t, \infty[$. Mais cette définition pose un problème de mesurabilité car le suprémum d'une famille non dénombrable de v.a. n'est pas mesurable en général.

Exemple 11.3. Sur $[0, 1]$ muni de la mesure de Lebesgue, on considère un ensemble $A \subset [0, 1]$ et la famille de fonctions $\{X_t, t \in A\}$, où $X_t(s) := \mathbf{1}_{\{s=t\}}$ pour tout $s \in [0, 1]$. Pour tout t , $X_t = 0$ p.p. est mesurable. Mais $\sup_{t \in A} X_t = \mathbf{1}_A$ n'est mesurable que si l'ensemble A l'est.

Théorème 11.4. (*suprémum essentiel*) Soit \mathcal{Z} une famille de v.a. $Z : \Omega \rightarrow \mathbb{R} \cup \{\infty\}$ sur une espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. Alors il existe une unique (p.s.) v.a. $\bar{Z} : \Omega \rightarrow \mathbb{R} \cup \{\infty\}$ telle que :

- (a) $\bar{Z} \geq Z$, p.s. pour tout $Z \in \mathcal{Z}$,
- (b) Pour toute v.a. Z' vérifiant (a), on a $\bar{Z} \leq Z'$, p.s.

De plus, il existe une suite $(Z_n)_{n \in \mathbb{N}} \subset \mathcal{Z}$ telle que $\bar{Z} = \sup_{n \in \mathbb{N}} Z_n$.

La v.a. \bar{Z} est appelée *suprémum essentiel* de la famille \mathcal{Z} , et est notée $\text{ess sup } \mathcal{Z}$.

Démonstration. L'unicité de \bar{Z} est une conséquence immédiate de (b). Pour l'existence, on considère l'ensemble \mathcal{D} des familles dénombrables de \mathcal{Z} . Pour tout $D \in \mathcal{D}$, on définit la v.a. $Z_D := \sup\{Z : Z \in D\}$, et on introduit la v.a. $\zeta := \sup\{\mathbb{E}[Z_D] : D \in \mathcal{D}\}$.

1- Montrons d'abord qu'il existe $D^* \in \mathcal{D}$ tel que $\zeta = \mathbb{E}[Z_{D^*}]$. En effet, si $(D_n)_n \subset \mathcal{D}$ est une suite maximisante, i.e. $\mathbb{E}[Z_{D_n}] \rightarrow \zeta$, alors $D^* := \cup_n D_n \in \mathcal{D}$ vérifie bien $\mathbb{E}[Z_{D^*}] = \zeta$. On note $\bar{Z} := Z_{D^*}$.

2- La v.a. \bar{Z} vérifie de manière évidente la propriété (b). Montrons qu'elle vérifie (a). Pour tout $Z \in \mathcal{Z}$, la famille $D := D^* \cup \{Z\} \subset \mathcal{D}$, on a $Z_D = Z \vee \bar{Z}$, et $\zeta = \mathbb{E}[\bar{Z}] \leq \mathbb{E}[Z \vee \bar{Z}] \leq \zeta$. Par suite $Z \vee \bar{Z} = \bar{Z}$, et $Z \leq \bar{Z}$, p.s. \diamond

Définition 11.5. Soit $\{X_t, t \in \mathbb{N}\}$ un processus \mathbb{F} -adapté vérifiant (11.2). Son *enveloppe de Snell* est définie par le processus

$$Y_t := \text{ess sup}_{\tau \in \mathcal{T}_t} \mathbb{E}[X_\tau | \mathcal{F}_t], \quad t \geq 0.$$

Proposition 11.6. (i) Soit Y l'enveloppe de Snell d'un processus X vérifiant (11.2). Alors on a l'équation de la programmation dynamique :

$$Y_t = \max \{X_t, \mathbb{E}[Y_{t+1} | \mathcal{F}_t]\} \quad \text{pour tout } t \geq 0,$$

et Y est la plus petite surmartingale qui majore X .

(ii) Soit $\tau^* := \inf \{t \geq 0 : Y_t = X_t\}$. Si $\tau^* < \infty$ p.s. alors $V = Y_0 = \mathbb{E}[X_{\tau^*}]$, i.e. τ^* est un temps d'arrêt optimal.

Démonstration. (i) Le processus Y est intégrable d'après la condition d'intégrabilité (11.2). Pour $t \in \mathbb{N}$, on a évidemment $Y_t \geq X_t$, et

$$Y_t = \operatorname{ess\,sup}_{\tau \in \mathcal{T}_{t+1}} \mathbb{E}[X_\tau | \mathcal{F}_t] \quad \text{sur} \quad \{Y_t > X_t\}. \quad (11.3)$$

D'après le théorème 11.4, il existe une suite $(\tau_n)_n \subset \mathcal{T}_{t+1}$ telle que $Y_{t+1} = \sup_n \mathbb{E}[X_{\tau_n} | \mathcal{F}_{t+1}]$. On définit la suite

$$\hat{\tau}_n := \sum_{i=1}^n \tau_i \mathbf{1}_{A_i}, \quad n \geq 1, \quad \text{où} \quad A_i = \left\{ \mathbb{E}[X_{\tau_i} | \mathcal{F}_{t+1}] = \max_{j \leq n} \mathbb{E}[X_{\tau_j} | \mathcal{F}_{t+1}] \right\}.$$

Alors $\hat{\tau}_n \in \mathcal{T}_{t+1}$ et on vérifie

$$Y_{t+1} = \lim_{n \rightarrow \infty} \uparrow \mathbb{E}[X_{\hat{\tau}_n} | \mathcal{F}_{t+1}].$$

Par convergence monotone, ceci implique que

$$\mathbb{E}[Y_{t+1} | \mathcal{F}_t] = \lim_{n \rightarrow \infty} \mathbb{E}[X_{\hat{\tau}_n} | \mathcal{F}_t] \leq \sup_{\tau \in \mathcal{T}_{t+1}} \mathbb{E}[X_\tau | \mathcal{F}_t].$$

Mais, la propriété des projections itérées de l'espérance conditionnelle implique que $\mathbb{E}[Y_{t+1} | \mathcal{F}_t] \geq \mathbb{E}[X_\tau | \mathcal{F}_t]$ pour tout $\tau \in \mathcal{T}_{t+1}$, et par suite

$$\mathbb{E}[Y_{t+1} | \mathcal{F}_t] = \sup_{\tau \in \mathcal{T}_{t+1}} \mathbb{E}[X_\tau | \mathcal{F}_t].$$

En combinant avec (11.3), ceci montre que

$$Y_t = \mathbb{E}[Y_{t+1} | \mathcal{F}_t] \quad \text{sur} \quad \{Y_t > X_t\},$$

terminant le preuve de l'équation de la programmation dynamique.

Enfin, une fois l'équation de la programmation dynamique établie, la propriété de surmartingale de Y , ainsi que l'optimalité du temps d'arrêt τ^* , sont obtenues par les mêmes arguments que dans le cas de l'horizon fini. \diamond

11.4 Application aux options américaines perpétuelles

dans cette section, nous devançons les applications aux mathématiques financières du chapitre 12 afin de présenter une application de l'arrêt optimal en horizon infini.

Soit $(\xi_n)_{n \geq 1}$ une suite de v.a. iid de distribution

$$\mathbb{P}[\xi_n = 1] = 1 - \mathbb{P}[\xi_n = -1] = \frac{1}{2}, \quad n \geq 1.$$

La marche aléatoire associée est définie par

$$M_0 = 0 \quad \text{et} \quad M_n := \sum_{i=1}^n \xi_i, \quad n \geq 1.$$

On considère un actif financier dont le processus de prix est défini par

$$S_n := S_0 e^{\sigma M_n}, \quad n \geq 0. \quad (11.4)$$

On note par $\mathcal{S} := S_0 e^{\sigma \mathbb{Z}}$ l'espace d'état de cette chaîne de Markov. On choisit ensuite un paramètre $r > 0$ désignant le taux d'intérêt sur une période, de façon à ce que

$$\{e^{-rn} S_n, n \geq 0\} \quad \text{est une martingale.} \quad (11.5)$$

Un calcul direct donne la valeur unique de r :

$$2e^r = e^\sigma + e^{-\sigma}.$$

Une option de vente américaine sur l'actif $\{S_n, n \geq 0\}$ de prix d'exercice $K > 0$ et de maturité N est un contrat qui donne le droit à son détenteur de vendre à toute date $n \leq N$ l'actif sous-jacent au prix d'exercice K .

Dans cette section, on considère une option de vente américaine perpétuelle, c'est à dire que sa maturité N est infinie. De tels contrats n'existent pas sur le marché, mais ils ont un intérêt théorique car, comme on va le voir, on peut trouver une expression explicite pour leur prix.

Dans le chapitre 12, nous allons voir que le prix d'un tel contrat vérifie l'équation de la programmation dynamique

$$Y_n = \max \{X_n, e^{-r} \mathbb{E}[Y_{n+1} | \mathcal{F}_n]\} \quad \text{où} \quad X_n := (K - S_n)^+, \quad n \geq 0,$$

voir la remarque 12.9. Nous sommes ainsi ramenés à la résolution du problème d'arrêt optimal en horizon infini

$$V(S_0) := Y_0 := \sup_{\tau \in \mathcal{T}} \mathbb{E} [e^{-r\tau} (K - S_\tau)^+].$$

Lemme 11.7. *Il existe $s^* \in \mathcal{S} \cap [0, K[$ tel que $V(s) = K - s$ si et seulement si $s \leq s^*$.*

Démonstration. Supposons au contraire qu'il existe $0 \leq s_1 < s_2$ tels que

$$V(s_i) = (K - s_i)^+ \quad \text{mais} \quad V(s) > (K - s)^+ \quad \text{sur} \quad]s_1, s_2[. \quad (11.6)$$

Comme $(K - s)^+ = 0$ pour $s \geq K$, on doit avoir $s_2 \leq K$, et par suite

$$(K - s)^+ = K - s \quad \text{pour} \quad s \in [s_1, s_2]. \quad (11.7)$$

Soit

$$\hat{s} = \frac{s_1 + s_2}{2}, \quad T_{s_i} = \inf \left\{ t \geq 0; S_t^{0, \hat{s}} = s_i \right\},$$

où $S^{0,\hat{s}}$ désigne le processus S de (11.4) avec condition initiale $S_0 = \hat{s}$. Alors, d'après les résultats de la section précédente, on sait que le temps d'arrêt $\tau^* := T_{s_1} \wedge T_{s_2}$ est optimal pour le problème d'arrêt optimal $V(\hat{s})$. Par la propriété de linéarité (11.7) sur l'intervalle $[s_1, s_2]$, on obtient :

$$\begin{aligned} V(\hat{s}) &= \mathbb{E} \left[e^{-r\tau^*} \left(K - S_{\tau^*}^{0,\hat{s}} \right) \right] \\ &= K \mathbb{E} \left[e^{-r\tau^*} \right] - \mathbb{E} \left[e^{-r\tau^*} S_{\tau^*}^{0,\hat{s}} \right] \\ &= K \mathbb{E} \left[e^{-r\tau^*} \right] - \hat{s} < K - \hat{s}, \end{aligned}$$

où nous avons utilisé le théorème d'arrêt pour la martingale $\{e^{-rn} S_n, n \geq 0\}$, d'après (11.5), ainsi que le fait que $r > 0$. Ceci est une contradiction puisque l'enveloppe de Snell d'un processus en est un majorant. \diamond

Le résultat précédent montre que la stratégie d'arrêt optimale est donnée par l'atteinte de la barrière s^* . Ceci réduit considérablement la difficulté du problème d'arrêt optimal, puisqu'il suffit maintenant d'optimiser sur le choix de la barrière qui définit l'arrêt :

$$V(S_0) = \sup_{b \in [0, K[} \mathbb{E} \left[e^{-rT_b} (K - S_{T_b})^+ \right] = \mathbb{E} \left[e^{-rT_{s^*}} (K - S_{T_{s^*}})^+ \right],$$

où $T_b := \inf\{t \geq 0 : S_t \leq b\}$. En revenant à la marche aléatoire, on ré-écrit le problème sous la forme :

$$V(S_0) = \sup_{\beta < \kappa} \mathbb{E} \left[e^{-r\theta_\beta \mathbf{1}_{\{\beta \leq 0\}}} \left(K - S_0 e^{\sigma M_{\theta_\beta \mathbf{1}_{\{\beta \leq 0\}}}} \right)^+ \right],$$

où on a posé $\beta := \frac{1}{\sigma} \ln(b/S_0)$, et

$$\kappa := \frac{1}{\sigma} \ln(K/S_0), \quad \text{et} \quad \theta_\beta := \inf \{n \geq 0 : M_n = \beta\}$$

Il est clair qu'il suffit de considérer les valeurs entières de β . Alors, du fait que $M_{\theta_\beta \mathbf{1}_{\{\beta \leq 0\}}} = -\beta^-$:

$$\begin{aligned} V(S_0) &= \sup_{\beta \in \mathbb{Z} \cap]-\infty, \kappa[} \left(K - S_0 e^{-\sigma\beta^-} \right) \mathbb{E} \left[e^{-r\theta_\beta \mathbf{1}_{\{\beta \leq 0\}}} \right] \\ &= \max \left\{ K - S_0, \sup_{\beta \in \mathbb{Z}_-} \left(K - S_0 e^{\sigma\beta} \right) \mathbb{E} \left[e^{-r\theta_\beta} \right] \right\} \\ &= \max \left\{ K - S_0, \sup_{\beta \in \mathbb{N}} \left(K - S_0 e^{-\sigma\beta} \right) \mathbb{E} \left[e^{-r\theta_\beta} \right], \right\} \end{aligned}$$

par symétrie de la marche aléatoire. On est ramené au calcul de la transformée de Laplace du temps d'atteinte θ_β de la marche aléatoire de la barrière $\beta \in \mathbb{N}$.

Lemme 11.8. *Pour tout $\beta \in \mathbb{N}$ et $r > 0$, on a*

$$\mathbb{E} [e^{-r\theta_\beta}] = \left(e^r + \sqrt{e^{2r} - 1} \right)^{-\beta}.$$

Démonstration. Notons par $U(n)$ la transformée de Laplace recherchée quand la marche aléatoire est démarrée à partir de la position n . En conditionnant par rapport à M_1 , on voit que U est solution de l'équation aux différences :

$$U(n) = \frac{1}{2} e^{-r} (U(n-1) + U(n+1)), \quad n < \beta.$$

La solution générale de cette équation aux différences est

$$U(n) = A \left(e^r + \sqrt{e^{2r} - 1} \right)^n + B \left(e^r - \sqrt{e^{2r} - 1} \right)^n,$$

où A et B sont des constantes. Comme $U(-\infty) = 0$, on voit que $B = 0$. Enfin, la condition au bord $U(\beta) = 1$ détermine la constante A et on obtient :

$$U(n) = \left(e^r + \sqrt{e^{2r} - 1} \right)^{n-\beta}.$$

Dans notre contexte la marche aléatoire est démarrée à partir de l'origine. Alors la transformée de Laplace recherchée est donnée par $U(0)$. \diamond

Exercice 11.9. *Terminer la résolution explicite du problème d'évaluation de l'option de vente américaine perpétuelle.*

Chapitre 12

Introduction aux mathématiques financières

Ce chapitre est une introduction aux modèles d'évaluation et de couvertures des produits dérivés en mathématiques financières. Les produits dérivés sont des contrats définis par des paiements aléatoires intervenant à des dates précises. Ces paiements aléatoires représentent généralement une garantie pour l'acheteur contre un risque dont il veut se prémunir. Par exemple, ces contrats peuvent être conçus pour atténuer le risque de taux de change pour un exportateur ou un importateur, le risque de grande variation du prix d'une matière première représentant la base de la production d'un industriel (pétrole, gaz), ou le risque de variation des taux d'intérêt pour un emprunteur.

Les produits dérivés ont connu un développement extraordinaire depuis l'ouverture des marchés organisés dans les années quatre-vingt. En France, il s'agit du marché à terme des Instruments Financiers (MATIF, 1986) et du Marché des Options Négociables de Paris (MONEP, 1987). Mais leur échange ne se limite pas à ce type de marchés, et repose en grande partie sur le marché de gré à gré. Ainsi les institutions financières conçoivent des produits standardisés pour des clients potentiels intéressés par des garanties particulières, ou répondent à la demande de certains clients pour leur proposer des contrats de garantie adaptés à la nature des risques encourus. Il s'agit là de l'activité de structuration.

Dans la première section, nous introduisons les produits dérivés de base, ainsi que les relations fondées sur le concept de non arbitrage, sans aucune hypothèse sur la modélisation probabiliste des risques. Dans la section suivante, nous développons l'approche de l'évaluation et de la couverture dans le modèle binomiale où l'incertitude est réduite au cadre le plus simple. Par passage à la limite en temps continu, ceci conduit à la célèbre formule de Black-Scholes qui est couramment utilisée par les praticiens de ce marché. La dernière section reprend une modélisation plus générale, et développe la théorie d'évaluation et de couverture de manière plus systématique. La théorie des martingales joue un rôle clé dans tous les développements.

12.1 Contrats d'options et principe de non domination

12.1.1 Options européennes et américaines

Les exemples les plus populaires de produits dérivés sont les options d'achat (call) et les options de ventes (put) européennes et américaines. D'autres exemples seront présentés dans la section 12.1.5.

Une *option d'achat européenne* sur un actif sous-jacent S^i est un contrat où le vendeur s'engage à livrer l'actif risqué S^i à la maturité T à un *prix d'exercice*, ou *strike*, $K > 0$. A la date T , l'acheteur a la possibilité, et non l'obligation, d'exercer l'option, i.e. d'acheter l'actif risqué S^i au prix d'exercice K . Bien sûr, l'acheteur n'exercera l'option que si le prix de l'actif S^i à la date T est supérieur à K . Ainsi le gain, ou le paiement, aléatoire généré par le contrat est donné par :

$$B = (S_T^i - K)^+ = \max\{S_T^i - K, 0\},$$

i.e. si le prix de l'actif S^i à la date T est supérieur à K , alors le contrat représente pour l'acheteur le gain $S_T^i - K$ correspondant au bénéfice réalisé en achetant l'actif aux prix K plutôt qu'au prix du marché. Par contre, si le prix de marché est inférieur à K , l'acheteur n'a aucune raison de faire appel à son contrat, et celui-ci représente alors une valeur nulle.

Une *option de vente européenne* sur l'actif sous-jacent S^i est un contrat dont le vendeur s'engage à acheter l'actif sous-jacent S^i à la maturité T à un *prix d'exercice*, ou *strike*, $K > 0$. A la date T , l'acheteur du contrat a la possibilité, et non l'obligation, d'exercer l'option, i.e. de vendre l'actif S^i au vendeur du contrat au prix K . Bien sûr, le détenteur du contrat ne va exercer ce droit que si le prix de marché à la date T est inférieur à K . Ainsi, ce contrat représente pour son détenteur le paiement aléatoire

$$B = (K - S_T^i)^+ = \max\{K - S_T^i, 0\},$$

Une *option d'achat (resp. de vente) américaine* de maturité T et de prix d'exercice $K > 0$ diffère de sa contrepartie européenne par la possibilité qu'elle offre d'exercice à toute date avant la maturité, et non seulement à la maturité.

Le vendeur d'un produit dérivé exige une compensation pour le risque qu'il accepte de porter. En d'autres termes, l'acheteur se doit de payer une prime contre la signature du contrat. L'objet principal est la détermination de cette prime, et de la gestion du risque qui doit être effectuée par le vendeur pour faire face au paiement aléatoire qu'il s'est engagé à payer. Dans les paragraphes suivants, nous introduisons le *principe de non domination* qui permet d'ores et déjà d'obtenir des propriétés des prix des options, sans hypothèse de modèle.

Nous considérerons des options d'achat et de vente de prix d'exercice K , de maturité T , écrite sur un sous-jacent de prix S . A toute date $t \leq T$, les prix de l'option d'achat américaine et européenne sont respectivement notés :

$$C(t, S_t, T, K) \quad \text{et} \quad c(t, S_t, T, K).$$

De même, les prix de l'option de vente américaine et européenne sont respectivement notés :

$$P(t, S_t, T, K) \quad \text{et} \quad p(t, S_t, T, K).$$

La valeur intrinsèque de l'option d'achat et de vente sont respectivement :

$$\begin{aligned} C(t, S_t, t, K) &= c(t, S_t, t, K) = (S_t - K)^+, \\ P(t, S_t, t, K) &= p(t, S_t, t, K) = (K - S_t)^+, \end{aligned}$$

c'est la valeur reçue lors de l'exercice immédiat de l'option. L'option est dite *dans la monnaie* (resp. *en dehors de la monnaie*) si sa valeur intrinsèque est positive. Si $K = S_t$, on dit que l'option est *à la monnaie*. Ainsi, une option d'achat est dans la monnaie si $S_t > K$, alors qu'une option de vente est dans la monnaie si $S_t < K$.

Enfin,, une *obligation zéro-coupon* est un contrat défini par le paiement fixe 1 à la maturité T . On notera par $B_t(T)$ son prix à la date t . La donnée des prix des obligations zéro-coupon pour toutes maturités détermine le prix de tout contrat défini par des flux de paiement déterministes F_1, \dots, F_n intervenant aux maturités $t < T_1 < \dots < T_n$:

$$F_1 B_t(T_1) + \dots + F_n B_t(T_n).$$

12.1.2 Principe de non domination et premières propriétés

On suppose qu'il n'y a pas d'imperfections de marché telles que les coûts de transaction, les taxes, les contraintes de portefeuille.

Principe de non domination *Soit X le gain généré par une stratégie de portefeuille de coût initial x . Si $X \geq 0$ dans tous les états du monde, alors $x \geq 0$.*

1 Appliquons ce principe d'abord aux options américaines. Le choix d'exercice de l'option américaine à la date T donne le même paiement que sa contrepartie européenne. Alors, le portefeuille consistant en une *position longue* (i.e. achat) dans l'option américaine et une *position courte* (i.e. vente) de sa contrepartie européenne à au moins un paiement nul à la maturité T . On déduit alors du principe de domination que

$$C(t, S_t, T, K) \geq c(t, S_t, T, K) \quad \text{et} \quad P(t, S_t, T, K) \geq p(t, S_t, T, K)$$

2 Par un argument similaire, nous montrons à présent que *les prix des options européennes et américaines d'achat (resp. de vente) sont des fonctions décroissantes (resp. croissantes) du prix d'exercice*, i.e. pour tous $K_1 \geq K_2$:

$$\begin{aligned} C(t, S_t, T, K_1) &\leq C(t, S_t, T, K_2) \quad \text{et} \quad c(t, S_t, T, K_1) \leq c(t, S_t, T, K_2) \\ P(t, S_t, T, K_1) &\geq P(t, S_t, T, K_2) \quad \text{et} \quad p(t, S_t, T, K_1) \leq p(t, S_t, T, K_2) \end{aligned}$$

Justifions ceci dans le cas des options d'achat américaines. Si le détenteur de l'option de plus faible prix d'exercice adopte la stratégie d'exercice optimale de l'option de plus fort prix d'exercice, le paiement de l'option à faible prix d'exercice est plus grand dans tous les états du monde. Ainsi, la valeur de l'option à plus faible prix d'exercice ne peut être inférieure au prix de l'option au prix d'exercice plus élevé.

3 *Les prix des options américaines et européennes sont convexes par rapport à K .* Justifions ceci pour le cas des options d'achat américaines. A un instant arbitraire $u \in [t, T]$ et pour tout $\lambda \in [0, 1]$, on obtient de la convexité de la valeur intrinsèque que

$$\lambda(S_u - K_1)^+ + (1 - \lambda)(S_u - K_2)^+ - (S_u - \lambda K_1 + (1 - \lambda)K_2)^+ \geq 0.$$

On considère alors le portefeuille X consistant en une position longue de λ options avec prix d'exercice K_1 , une position longue de $(1 - \lambda)$ options avec prix d'exercice K_2 , et une position courte d'une option de prix d'exercice $\lambda K_1 + (1 - \lambda)K_2$. Si les deux premières options sont exercées à la date d'exercice optimale de la troisième, le paiement obtenu sera positif d'après l'inégalité de convexité. Donc, la valeur à la date t du portefeuille est positive.

4 Nous montrons maintenant le résultat suivant de sensibilité du prix de l'option d'achat européenne par rapport au prix d'exercice :

$$-B_t(T) \leq \frac{c(t, S_t, T, K_2) - c(t, S_t, T, K_1)}{K_2 - K_1} \leq 0$$

L'inégalité de droite est une conséquence de la décroissance du prix de l'option d'achat européenne c en fonction de K . Pour l'inégalité de gauche, on considère le portefeuille X consistant en une position courte de l'option d'achat européenne de prix d'exercice K_1 , et une position longue de l'option d'achat européenne de prix d'exercice K_2 , et d'une position longue de $K_2 - K_1$ obligations zéro-coupon. La valeur de ce portefeuille à la maturité T est

$$X_T = -(S_T - K_1)^+ + (S_T - K_2)^+ + (K_2 - K_1) \geq 0.$$

Par le principe de non domination, ceci implique que $-c(S_t, \tau, K_1) + c(S_t, \tau, K_2) + B_t(\tau)(K_2 - K_1) \geq 0$, qui est l'inégalité recherchée.

5 *Les options d'achat et de vente américaines sont des fonctions croissantes de la maturité, i.e. pour $T_1 \geq T_2$:*

$$C(t, S_t, T_1, K) \geq C(t, S_t, T_2, K) \text{ et } P(t, S_t, T_1, K_1) \geq P(t, S_t, T_2, K_2)$$

Ceci est une conséquence immédiate du fait que toutes les stratégies d'exercice autorisées par l'option de maturité la plus courte sont aussi autorisées par l'option de plus longue maturité. Notons que cet argument est spécifique aux options américaines.

12.1.3 Parité call-put

Quand l'actif sous-jacent ne verse pas de dividendes avant la maturité T , les prix des options d'achat et de vente européennes sont liés par la relation :

$$p(t, S_t, T, K) = c(t, S_t, T, K) - S_t + KB_t(T).$$

En effet, soit X le portefeuille consistant d'une position longue de l'option de vente européenne et d'une unité de l'actif sous-jacent, et d'une position courte de l'option d'achat européenne et de K obligations zéro-coupon. La valeur de ce portefeuille à la maturité T est

$$X_T = (K - S_T)^+ + S_T - (S_T - K)^+ - K = 0.$$

Le principe de non domination dit alors que la valeur à toute date t de ce portefeuille est positive, ce qui donne l'inégalité recherchée.

Notons que cet argument est spécifique aux options européennes. Nous allons voir que ce résultat n'est pas vrai dans le cas américain.

Enfin, si l'actif sous-jacent verse des dividendes, l'argument ci-dessus doit être adapté en incluant les versements de dividendes du fait de la détention de l'actif sous-jacent dans le portefeuille. Si les dividendes sont connus à l'avance, i.e. non aléatoires, alors l'adaptation de la parité call-put est immédiate. Par contre, si les dividendes sont aléatoires, la question devient beaucoup plus délicate...

12.1.4 Bornes sur les prix des calls et exercice anticipé

1 D'après la décroissance du prix du call américain en fonction du prix d'exercice, on a

$$c(S_t, \tau, K) \leq C(S_t, \tau, K) \leq S_t$$

• Si l'actif sous-jacent ne verse pas de dividendes, nous avons la borne inférieure :

$$C(t, S_t, T, K) \geq c(t, S_t, T, K) \geq (S_t - KB_t(T))^+.$$

En effet, considérons le portefeuille X consistant d'une position longue du call européen, d'une position longue de K obligations zéro-coupon de maturité T , et d'une position courte d'une unité de l'actif sous-jacent. La valeur de ce portefeuille à la maturité T est positive, et on obtient l'inégalité recherchée par le principe de non domination.

2 *Supposons que les taux d'intérêt sont strictement positifs. Alors, il n'est jamais optimal d'exercer avant la maturité un call américain sur un actif sous-jacent ne versant pas de dividendes avant la maturité.*

En effet, soit u instant arbitraire dans $[t, T)$,

- le call américain paie $S_u - K$ s'il est exercé à la date u ,
- mais $S_u - K < S - KB_u(T)$ d'après la stricte positivité des taux d'intérêt.
- Comme $C(S_u, u, K) \geq S_u - KB_u(T)$, d'après la borne inférieure du paragraphe précédent, l'option américaine vaut plus que sa valeur d'exercice, donc l'exercice anticipé n'est jamais optimal.

3 Supposons que le prix de l'actif sous-jacent peut être aussi proche que possible de zéro. Alors, l'exercice anticipé d'un put américain avant la maturité peut être optimal.

Supposons que le prix de l'actif sous-jacent à la date u vérifie $S_u < K - KB_u(T)$.

- Notons que la valeur maximale que le put américain peut générer est K .

- La valeur d'exercice immédiat à la date u est $K - S_u > K - [K - KB_u(T)] = KB_u(T) \equiv$ valeur actualisée du paiement maximum généré par le put à la maturité,

Ainsi, dans ce cas, il n'est pas optimal d'attendre la maturité pour exercer l'option.

12.1.5 Quelques exemples populaires de produits dérivés

Exemple 12.1. (*options sur panier*) Soit I un ensemble d'indices dans $\{1, \dots, n\}$, et $a_i \in \mathbb{R}$, $i \in I$. le paiement d'une option d'achat (resp. vente) sur le panier I est défini par

$$B = \left(\sum_{i \in I} a_i S_T^i - K \right)^+ \quad \text{resp.} \quad \left(K - \sum_{i \in I} a_i S_T^i \right)^+ .$$

◇

Exemple 12.2. (*Option sur sous-jacent non échangeable*) Soit $U_t(\omega)$ la réalisation à la date t d'une variable d'état observable. Alors le paiement du call (resp. put) sur U est défini par :

$$B = (U_T - K)^+ \quad \text{resp.} \quad (K - U_T)^+ .$$

Par exemple, un *call sur température* correspond au cas où U_t est la température à la date t observée à un endroit précisé dans le contrat.

Exemple 12.3. (*Option asiatique*) Un call asiatique sur l'actif S^i de maturité $T > 0$ et de prix d'exercice $K > 0$ est défini par le paiement à la maturité :

$$\left(\bar{S}_T^i - K \right)^+ ,$$

où \bar{S}_T^i est le processus de prix moyen sur la période $[0, T]$. Dans cette définition, \bar{S}_T^i peut être défini comme la moyenne arithmétique sur un ensemble de dates discrètes (spécifiées dans le contrat), ou la moyenne arithmétique continue...

Exemple 12.4. (*Options à barrières*) Soient $B, K > 0$ deux paramètres, et $T > 0$ une maturité donnée. Il y a quatre types d'options d'achat à barrières sur l'actif S^i avec prix d'exercice K , barrière B et maturité T :

— Si $B > S_0$:

- Un *call up and out* est défini par le paiement à la maturité T :

$$\text{UOC}_T = (S_T - K)^+ \mathbf{1}_{\{\max_{[0,T]} S_t \leq B\}}.$$

Il s'agit du paiement d'un call européen si le prix de l'actif sous-jacent n'a pas atteint la barrière sur toute la période précédant la maturité. Si la barrière est dépassée, le paiement devient nul.

- Un *call up and in* est défini par le paiement à la maturité T :

$$\text{UIC}_T = (S_T - K)^+ \mathbf{1}_{\{\max_{[0,T]} S_t > B\}}.$$

Il s'agit du paiement d'un call européen si le prix du sous-jacent a passé la barrière B avant la maturité. Sinon, le paiement est nul. Remarquons que

$$\text{UOC}_T + \text{UIC}_T = C_T$$

est le paiement de l'option d'achat européenne correspondante.

- Si $B < S_0$:

- Un *down and in Call* est défini par le paiement à la maturité T :

$$\text{DIC}_T = (S_T - K)^+ \mathbf{1}_{\{\min_{[0,T]} S_t \geq B\}}.$$

Il s'agit du paiement d'un call européen si le prix de l'actif sous-jacent n'a pas atteint la barrière B avant la maturité T . Sinon, le paiement est nul.

- Un *down and out call* est défini par le paiement à la maturité T :

$$\text{DOC}_T = (S_T - K)^+ \mathbf{1}_{\{\min_{[0,T]} S_t < B\}}.$$

Il s'agit du paiement d'un call européen si le prix de l'actif sous-jacent traverse la barrière B avant la maturité. Sinon, le paiement est nul. Comme pour les barrière up, on a encore,

$$\text{DOC}_T + \text{DIC}_T = C_T$$

est le paiement du call européen correspondant.

◇

Exemple 12.5. (*Options de vente à barrières*) Remplacer tous les calls par des puts dans les exemples précédents.

12.2 Du modèle binomial à la formule de Black-Scholes

12.2.1 Modèle binomial à une période

Nous allons maintenant étudier le modèle de marché financier le plus simple. Soit $\Omega = \{\omega_u, \omega_d\}$, \mathcal{F} la σ -algèbre des parties de Ω , et \mathbb{P} la mesure de probabilité sur (Ω, \mathcal{F}) telle que $0 < \mathbb{P}(\omega_u) < 1$.

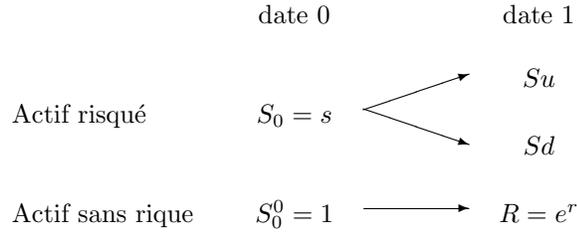
Le marché financier contient un actif sans risque de prix

$$S_0^0 = 1, \quad S_1^0(\omega_u) = S_1^0(\omega_d) = e^r,$$

et un actif risqué de prix

$$S_0 = s, \quad S_1(\omega_u) = su, \quad S_1(\omega_d) = sd,$$

où s, r, u et d sont des paramètres strictement positifs tels que $u > d$. Ce marché est communément représenté par l'arbre binomial :



Les prix actualisés expriment la valeur des actifs en termes de celle de l'actif sans risque, et sont donnés par

$$\tilde{S}_0 = S_0, \quad \tilde{S}_0^0 = 1, \quad \text{et} \quad \tilde{S}_1 = \frac{S_1}{R}, \quad \tilde{S}_1^0 = 1.$$

Une stratégie auto-financée est un couple $(x, \theta) \in \mathbb{R}^2$, où x désigne le capital initial et θ la position en actif risqué. La valeur d'un tel portefeuille à la date 1 est donnée par :

$$X_1^{x, \theta} := (x - \theta S_0)R + \theta \tilde{S}_1,$$

ou, en termes de valeur actualisée :

$$\tilde{X}_1^{x, \theta} := x + \theta(\tilde{S}_1 - \tilde{S}_0).$$

(i) Condition de non arbitrage : Une opportunité d'arbitrage est une stratégie de portefeuille $\theta \in \mathbb{R}$ telle que

$$X_1^{0, \theta}(\omega_i) \geq 0, \quad i \in \{u, d\}, \quad \text{et} \quad \mathbb{P}[X_1^{0, \theta} > 0] > 0.$$

On peut montrer (exercice!) que l'exclusion de toutes les stratégies d'arbitrage est équivalente à

$$d < R < u. \tag{12.1}$$

Alors, en introduisant la mesure de probabilité \mathbb{Q} équivalente à \mathbb{P} définie par :

$$\mathbb{Q}[S_1 = uS_0] = 1 - \mathbb{Q}[S_1 = dS_0] = q := \frac{R - d}{u - d}, \tag{12.2}$$

on voit que les prix actualisés vérifient

$$\tilde{S} \text{ est une martingale sous } \mathbb{Q}, \text{ i.e. } \mathbb{E}^{\mathbb{Q}}[e^{-r} S_1] = S_0. \tag{12.3}$$

La mesure de probabilité Q est appelée *probabilité neutre au risque*, ou *probabilité martingale équivalente*.

(i) Couverture des produits dérivés : Un actif contingent, ou un produit dérivé, est défini par ses paiements $B_u := B(\omega_u)$ et $B_d := B(\omega_d)$ à la date 1 dans chacun des états du monde.

Dans notre contexte d'arbre binomial, il se trouve qu'il existe un couple $(x_0, \theta^0) \in \mathbb{R} \times \mathcal{A}$ tel que $X_1^{x_0, \theta^0} = B$. En effet, cette égalité de v.a. définit un système linéaire de deux équations à deux inconnues qu'on résout immédiatement :

$$x^0(B) = q \frac{B_u}{R} + (1-q) \frac{B_d}{R} = \mathbb{E}^Q[\tilde{B}] \quad \text{et} \quad \theta^0(B) = \frac{B_u - B_d}{su - sd}.$$

Le portefeuille $(x_0(B), \theta^0(B))$ vérifie $X_T^{x_0, \theta^0} = B$, et est alors appelé stratégie de *réplication parfaite* de B .

(iii) Prix de non arbitrage : Supposons que l'actif contingent B est disponible sur le marché à l'achat et à la vente à la date 0, au prix de marché $p(B)$, et montrons que, sous la condition de non arbitrage, le prix $p(B)$ est nécessairement donné par

$$p(B) = x_0(B) = \mathbb{E}^Q[\tilde{B}].$$

(iii-a) En effet, si $p(B) < V(B)$, on considère la stratégie de portefeuille suivante :

- à la date 0, on paie $p(B)$ pour l'achat de l'actif contingent, générant la réception du paiement B à la date 1,
- effectuer la stratégie auto-financée $(-V(B), -\theta)$, en payant $-V(B)$ à la date 0, et générant le paiement $-B$ à la date 1.

Le capital initial nécessaire pour mettre en place cette stratégie de portefeuille est $p(B) - V(B) < 0$. À la date 1, la valeur induite par cette stratégie est nulle. C'est donc une opportunité d'arbitrage dans le marché financier élargi par l'actif contingent, contredisant la condition de non arbitrage.

(iii-b) Si $p(B) > V(B)$, on considère la stratégie de portefeuille :

- à la date 0, on reçoit $p(B)$ en vendant l'actif contingent, s'engageant ainsi au paiement B à la date 1,
- effectuer la stratégie auto-financée $(V(B), \theta)$, dont le coût initial est $V(B)$ à la date 0, et qui génère le paiement B à la date 1.

Cette stratégie de portefeuille a un coût initial de $-p(B) + V(B) < 0$, et génère un paiement nul à la date 1. C'est encore une opportunité d'arbitrage, contredisant le principe de non arbitrage sur le marché financier élargi par l'actif contingent.

12.2.2 Le modèle de Cox-Ross-Rubinstein

Ce paragraphe présente la version dynamique du modèle binomial du paragraphe précédent. Nous découpons un intervalle de temps $[0, T]$ en n périodes,

et nous chercherons à analyser la limite de nos résultats quand la longueur de la période T/n tend vers zéro. Il s'agit de la limite *temps continu*.

Soit $\Omega = \{-1, 1\}^{\mathbb{N}}$ et \mathcal{F} la σ -algèbre des parties de Ω . Soit $(Z_k)_{k \geq 0}$ une suite de v.a. iid de distribution $\mathbb{P}[Z_k = 1] = \mathbb{P}[Z_k = -1] = 1/2$. On considère la σ -algèbre triviale $\mathcal{F}_0 = \{\emptyset, \mathcal{F}\}$, $\mathcal{F}_k = \sigma(Z_0, \dots, Z_k)$ et $\mathbb{F}^n = \{\mathcal{F}_0, \dots, \mathcal{F}_n\}$.

Soit $T > 0$ une maturité fixée, et $(b_n, \sigma_n)_{n \geq 1}$ la suite définie par :

$$b_n = b \frac{T}{n} \quad \text{et} \quad \sigma_n = \sigma \left(\frac{T}{n} \right)^{1/2},$$

où b et σ sont des paramètres strictement positifs donnés.

Remarque 12.6. Tous les résultats ultérieurs sont valables avec une suite (b_n, σ_n) vérifiant :

$$nb_n \longrightarrow bT \quad \text{et} \quad \sqrt{n}\sigma_n \longrightarrow \sigma\sqrt{T} \quad \text{quand } n \rightarrow \infty.$$

Pour $n \geq 1$, le processus de prix de l'actif risqué est donné par $S^n = \{S_k^n, k = 0, \dots, n\}$ défini par :

$$S_0^n = s \quad \text{et} \quad S_k^n = s \exp \left(kb_n + \sigma_n \sum_{i=1}^k Z_i \right), \quad k = 1, \dots, n.$$

L'actif sans risque est défini par un taux d'intérêt constant r . Le rendement d'un investissement unitaire sur la période de longueur T/n est donc :

$$R_n := e^{r(T/n)}.$$

Pour tout $n \geq 1$, la construction ci-dessus définit un marché financier de n périodes de longueur T/n .

Afin de garantir que ces marchés financiers n'admettent pas d'opportunité d'arbitrage, on suppose que :

$$d_n < R_n < u_n \quad \text{où} \quad u_n = e^{b_n + \sigma_n}, \quad d_n = e^{b_n - \sigma_n}, \quad n \geq 1. \quad (12.4)$$

Sous cette condition, La probabilité neutre au risque \mathbb{Q}_n est définie par :

$$\mathbb{Q}_n[Z_i = 1] = 1 - \mathbb{Q}_n[Z_i = -1] = q_n := \frac{R_n - d_n}{u_n - d_n}.$$

et on a que

$$\left\{ \tilde{S}_k^n := S_k^n e^{-rkT/n}, 0 \leq k \leq n \right\} \quad \text{est une} \quad \mathbb{Q}_n - \text{martingale.}$$

Evaluation et couverture Pour tout $n \geq 1$, on considère l'actif contingent

$$B^n := g(S_n^n) \quad \text{où} \quad g(s) = (s - K)^+ \quad \text{et} \quad K > 0.$$

A la date $n - 1$, le problème d'évaluation et de couverture est réduit à celui du modèle à une période du paragraphe précédent, dont on tire que le prix de

non arbitrage de l'actif contingent à la date $n - 1$ et la stratégie de couverture parfaite correspondante sont donnés par

$$B_{n-1}^n := \mathbb{E}_{n-1}^{\mathbb{Q}_n}[B^n] \quad \text{et} \quad \theta_{n-1}^n = \frac{B^n(u_n) - B^n(d_n)}{u_n S_{n-1}^n - d_n S_{n-1}^n}.$$

où $\mathbb{E}_{n-1}^{\mathbb{Q}_n}$ désigne l'espérance sous \mathbb{Q}_n conditionnellement à \mathcal{F}_{n-1} . Par un raisonnement analogue, paracourant l'arbre binomial de manière rétrograde, période par période, on définit l'actif contingent B_k^n à chaque date k comme le prix de non arbitrage de l'actif B_{k+1}^n , ainsi que la stratégie de couverture parfaite :

$$B_k^n := \mathbb{E}_k^{\mathbb{Q}_n}[B_{k+1}^n] \quad \text{et} \quad \theta_k^n = \frac{B_{k+1}^n(u_n) - B_{k+1}^n(d_n)}{u_n S_k^n - d_n S_k^n}, \quad k \leq n-1. \quad (12.5)$$

Remarque 12.7. La stratégie de couverture est donnée par l'approximation des différences finies (sur l'arbre binomial) de la dérivée partielle du prix de l'actif contingent par rapport au prix de l'actif risqué sous-jacent.

Par la propriété des projections itérées de l'espérance conditionnelle, on déduit que le prix de non arbitrage de l'option d'achat européenne est :

$$p^n(B^n) = e^{-rT} \mathbb{E}^{\mathbb{Q}_n} [(S_n^n - K)^+].$$

Sous la probabilité \mathbb{Q}_n , les v.a. $(1 + Z_i)/2$ sont iid selon la loi de Bernoulli de paramètre q_n . Alors :

$$\mathbb{Q}_n \left[\sum_{i=1}^n \frac{1 + Z_i}{2} = j \right] = C_n^j q_n^j (1 - q_n)^{n-j} \quad \text{pour } j = 0, \dots, n.$$

Le prix de non arbitrage de l'options d'achat s'écrit alors :

$$p^n(B^n) = e^{-rT} \sum_{j=0}^n g(su_n^j d_n^{n-j}) C_n^j q_n^j (1 - q_n)^{n-j}.$$

Remarque 12.8. La mesure de probabilité initiale \mathbb{P} n'intervient plus ni dans la formule du prix, ni dans l'expression du portefeuille de couverture. Ceci est dû au fait que le prix de non arbitrage coïncide avec le coût de réplcation parfaite qui, par définition ne fait intervenir la mesure de référence qu'à travers les états qu'elle charge.

Remarque 12.9. La démarche précédente peut être appliquée aux options américaines. Soit $\{B_k, 0 \leq k \leq n\}$ le processus de paiements associés à l'option, i.e. on gagne B_k si on exerce à la date k . On peut voir alors que le prix de l'option américaine et la stratégie de couverture correspondante sont donnés par la modification suivante de (12.5) :

$$\begin{aligned} \hat{B}_n &= B_n \quad \text{et} \quad \hat{B}_k^n := \max \left\{ B_k, e^{-rT/n} \mathbb{E}_k^{\mathbb{Q}_n} [\hat{B}_{k+1}^n] \right\}, \\ \hat{\theta}_k^n &= \frac{\hat{B}_{k+1}^n(u_n) - \hat{B}_{k+1}^n(d_n)}{u_n S_k^n - d_n S_k^n}, \quad k = 0, \dots, n-1. \end{aligned}$$

12.2.3 Limite en temps continu

Dans ce paragraphe, nous analysons la limite du modèle du paragraphe précédente quand $n \rightarrow \infty$, et nous obtiendrons la célèbre formule de Black et Scholes [2] comme limite du prix de non arbitrage du paragraphe précédent.

Bien que les calculs suivants seront faits dans le cas des options d'achat européennes, tous les résultats de convergence sont valables pour une classe plus large d'actifs contingents.

Introduisons la suite :

$$\eta_n := \inf\{j = 0, \dots, n : su_n^j d_n^{n-j} \geq K\},$$

et soit

$$B(n, p, \eta) := \text{Prob}[\text{Bin}(n, p) \geq \eta],$$

où $\text{Bin}(n, p)$ est une loi binomiale de paramètres (n, p) , i.e. la somme de n v.a. indépendantes de loi de Bernoulli de paramètre p .

Nous commençons par le résultat simple suivant.

Lemme 12.10. *Pour $n \geq 1$, on a :*

$$p^n(B^n) = sB\left(n, \frac{q_n u_n}{R_n}, \eta_n\right) - Ke^{-rT} B(n, q_n, \eta_n).$$

Démonstration. En utilisant l'expression de $p^n(B^n)$ obtenue dans le paragraphe précédent, on voit que

$$\begin{aligned} p^n(B^n) &= R_n^{-n} \sum_{j=\eta_n}^n (su_n^j d_n^{n-j} - K) C_n^j q_n^j (1 - q_n)^{n-j} \\ &= s \sum_{j=\eta_n}^n C_n^j \left(\frac{q_n u_n}{R_n}\right)^j \left(\frac{(1 - q_n) d_n}{R_n}\right)^{n-j} - \frac{K}{R_n^n} \sum_{j=\eta_n}^n C_n^j q_n^j (1 - q_n)^{n-j}, \end{aligned}$$

et on obtient le résultat voulu en remarquant que $q_n u_n + (1 - q_n) d_n = R_n$. \diamond

Ainsi, pour analyser la limite de la suite $(p^n(B^n))_n$, nous devons analyser les termes $B(n, q_n u_n / R_n, \eta_n)$ et $B(n, q_n, \eta_n)$. Nous détaillons les calculs uniquement pour le second terme, le premier se traitant de la même façon.

L'outil technique central pour obtenir les résultats asymptotiques est le suivant.

Lemme 12.11. *Soit $(X_{k,n})_{1 \leq k \leq n}$ une suite triangulaire de v.a. iid selon la loi de Bernoulli de paramètre π_n :*

$$\mathbb{P}[X_{k,n} = 1] = 1 - \mathbb{P}[X_{k,n} = 0] = \pi_n.$$

Alors :

$$\frac{\sum_{k=1}^n X_{k,n} - n\pi_n}{\sqrt{n\pi_n(1 - \pi_n)}} \longrightarrow \mathcal{N}(0, 1) \quad \text{en loi.}$$

Nous reportons la preuve de ce lemme à la fin de ce paragraphe.

Exercice 12.12. Utiliser le lemme 12.11 pour prouver que

$$\ln\left(\frac{S_n^n}{s}\right) \longrightarrow \mathcal{N}(bT, \sigma^2 T) \quad \text{en loi sous } \mathbb{P}.$$

Ceci montrer que le modèle de Cox-Ross-Rubinstein peut être vu comme une approximation en temps discret d'un modèle en temps continu où l'actif risqué est distribué selon la loi log-normale avec les caractéristiques ci-dessus.

Théorème 12.13. *Dans le contexte du modèle de Cox-Ross-Rubinstein, le prix de non arbitrage $p^n(B^n)$ d'un call européen converge, quand $n \rightarrow \infty$, vers le prix de Black-Scholes :*

$$p(B) = s \mathbf{N}\left(\mathbf{d}_+(s, \tilde{K}, \sigma^2 T)\right) - \tilde{K} \mathbf{N}\left(\mathbf{d}_-(s, \tilde{K}, \sigma^2 T)\right)$$

où

$$\tilde{K} := Ke^{-rT}, \quad \mathbf{d}_\pm(s, k, v) := \frac{\ln(s/k) \pm \frac{\sqrt{v}}{2}}{\frac{\sqrt{v}}{2}},$$

et $\mathbf{N}(x) = \int_{-\infty}^x e^{-v^2/2} dv / \sqrt{2\pi}$ est la fonction de répartition de la loi normale centrée réduite $\mathcal{N}(0, 1)$.

Démonstration. Sous la probabilité \mathbb{Q}_n , les v.a. $B_i := (Z_i + 1)/2$, $i \geq 0$, définissent une suite iid selon la loi de Bernoulli de paramètre q_n . Alors

$$B(n, q_n, \eta_n) = \mathbb{Q}_n \left[\sum_{j=1}^n B_j \geq \eta_n \right].$$

Nous allons détailler les calculs seulement pour ce terme.

1. Par définition de η_n , on a

$$su_n^{\eta_n - 1} d_n^{n - \eta_n + 1} \leq K \leq su_n^{\eta_n} d_n^{n - \eta_n}.$$

Alors,

$$2\eta_n \sigma \sqrt{\frac{T}{n}} + n \left(b \frac{T}{n} - \sigma \sqrt{\frac{T}{n}} \right) = \ln\left(\frac{K}{s}\right) + O\left(n^{-1/2}\right),$$

et par un calcul direct :

$$\eta_n = \frac{n}{2} + \sqrt{n} \frac{\ln(K/s) - bT}{2\sigma\sqrt{T}} + o(\sqrt{n}). \quad (12.6)$$

On calcule aussi que :

$$nq_n = \frac{1}{2} + \frac{\left(r - b - \frac{\sigma^2}{2}\right)T}{2\sigma\sqrt{T}} \sqrt{n} + o(\sqrt{n}). \quad (12.7)$$

d'après (12.6) et (12.7), on obtient :

$$\lim_{n \rightarrow \infty} \frac{\eta_n - nq_n}{\sqrt{nq_n(1 - q_n)}} = -\mathbf{d}_-(s, \tilde{K}, \sigma^2 T).$$

2. Par le lemme 12.11, appliqué à la suite (Z_1, \dots, Z_n) , on voit que :

$$\mathcal{L}^{\mathbb{Q}_n} \left(\frac{\sum_{k=1}^n Z_k - nq_n}{\sqrt{nq_n(1 - q_n)}} \right) \rightarrow \mathcal{N}(0, 1),$$

où $\mathcal{L}^{\mathbb{Q}_n}(Z)$ désigne la distribution sous \mathbb{Q}_n de la v.a. Z . Alors :

$$\begin{aligned} \lim_{n \rightarrow \infty} B(n, q_n, \eta_n) &= \lim_{n \rightarrow \infty} \mathbb{Q}_n \left[\frac{\sum_{k=1}^n Z_k - nq_n}{\sqrt{nq_n(1 - q_n)}} \geq \frac{\eta_n - nq_n}{\sqrt{nq_n(1 - q_n)}} \right] \\ &= 1 - \mathbf{N} \left(-\mathbf{d}_-(s, \tilde{K}, \sigma^2 T) \right) = \mathbf{N} \left(\mathbf{d}_-(s, \tilde{K}, \sigma^2 T) \right). \end{aligned}$$

◇

Preuve du lemme 12.11. Soit

$$Y_j := \frac{X_{j,n} - \pi_n}{\sqrt{n\pi_n(1 - \pi_n)}} \quad \text{et} \quad \Sigma Y_n := \sum_{k=1}^n Y_j.$$

Les v.a. Y_j étant iid, on a la factorisation suivante de la fonction caractéristique :

$$\Phi_{\Sigma Y_n}(t) = (\Phi_{Y_1}(t))^n.$$

Par ailleurs, on calcule directement que $\mathbb{E}[Y_j] = 0$ and $\mathbb{E}[Y_j^2] = 1/n$. Alors, on déduit du résultat de développement des fonctions caractéristiques 2.9 que :

$$\Phi_{Y_1}(t) = 1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right).$$

envoyant n vers l'infini, ceci donne :

$$\lim_{n \rightarrow \infty} \Phi_{\Sigma Y_n}(t) = e^{-t^2/2} = \Phi_{\mathcal{N}(0,1)}(t),$$

montrant bien la convergence en loi de ΣY_n vers la loi normale centrée réduite.

◇

12.3 Evaluation et couverture dans un modèle général en temps discret

12.3.1 Formulation du modèle

On considère maintenant un modèle en temps discret général sur un espace de probabilité filtré $(\Omega, \mathcal{F}, \mathbb{F} = \{\mathcal{F}_n\}_{0 \leq n \leq N}, \mathbb{P})$.

Le marché financier est constitué d'un actif sans risque de processus de prix $\{S_n^0, 0 \leq n \leq N\}$ strictement positif, et d actifs risqués de processus de prix $S = (S^1, \dots, S^d)$. Mathématiquement, le caractère non risqué du processus S^0 et risqué du processus S sont exprimés en imposant que

$$S^0 \text{ est } \mathbb{F} - \text{prévisibles et } S \text{ est } \mathbb{F} - \text{adapté.}$$

Comme dans la modélisation dans l'arbre binomial, on obtient une formulation plus simple du modèle en considérant les prix actualisés

$$\tilde{S}_n^0 := 1 \text{ et } \tilde{S}_n := \frac{S_n}{S_n^0} \text{ pour } n = 0, \dots, N.$$

Une stratégie de portefeuille est un processus $\theta = \{\theta_0, \dots, \theta_{N-1}\}$ \mathbb{F} -adapté à valeurs dans \mathbb{R}^d . On note par \mathcal{A} l'ensemble des portefeuilles.

Pour chaque n , la $i^{\text{ème}}$ composante θ_n^i désigne le nombre d'unités d'actif S^i , $1 \leq i \leq d$, détenues en portefeuille à la date n . Le montant investi dans le marché risqué est ainsi égal à $\sum_{i=1}^d \theta_n^i S_n^i = \theta_n \cdot S_n$. L'investisseur, dont la richesse à la date n est X_n , équilibre donc son budget grâce à l'actif sans risque, i.e. le montant investi en actif sans risque est $\theta_n^0 := X_n - \theta_n \cdot S_n$.

Ici, $\theta_n^0 \geq 0$ signifie que l'investisseur place son argent sur le compte sans risque, et $\theta_n^0 < 0$ signifie que l'investisseur effectue un emprunt à la banque au taux d'intérêt sans risque.

Etant donné un portefeuille $\theta \in \mathcal{A}$, la dynamique de la richesse de l'investisseur est donnée par

$$\Delta X_n^\theta := X_n^\theta - X_{n-1}^\theta = \frac{\theta_{n-1}^0}{S_{n-1}^0} \Delta S_n^0 + \theta_{n-1} \cdot \Delta S_n, \quad 1 \leq n \leq N,$$

où $\Delta S_n^i := S_n^i - S_{n-1}^i$, $0 \leq i \leq d$. Comme le portefeuille ne subit ni des entrées ni des sorties d'argent, on dit qu'il s'agit d'un *portefeuille auto-financé*.

De manière équivalente, le processus de richesse actualisée de l'investisseur $\tilde{X}_n := X_n/S_n^0$, $0 \leq n \leq N$, vérifie la dynamique

$$\Delta \tilde{X}_n^\theta = \theta_{n-1} \cdot \Delta \tilde{S}_n, \quad 1 \leq n \leq N.$$

12.3.2 Probabilités neutres au risque

Enfin, nous introduisons l'ensemble des mesures de probabilité équivalentes à \mathbb{P} sous lesquelles le processus de prix actualisés \tilde{S} est une martingale :

$$\mathcal{M}(S) := \left\{ \mathbb{Q} \sim \mathbb{P} : \tilde{S} \text{ est une } \mathbb{Q} - \text{martingale} \right\}.$$

Comme dans le cas de l'arbre binomial, l'absence d'arbitrage peut être caractérisée par la condition $\mathcal{M}(S) \neq \emptyset$. Dans le cadre de ces notes, nous montrons uniquement le sens suivant de cette équivalence.

Lemme 12.14. *Supposons que $\mathcal{M}(S) \neq \emptyset$, et soit $\theta \in \mathcal{A}$ un portefeuille tel que $X_0 = 0$ et $X_N^\theta \geq 0$, p.s. Alors $X^\theta = 0$, p.s.*

Démonstration. On ré-écrit les conditions $X_0 = 0$ et $X_N^\theta \geq 0$ de manière équivalente en terme du processus de richesse actualisée

$$\tilde{X}_0 = 0 \quad \text{et} \quad \tilde{X}_N^\theta \geq 0, \text{ p.s.}$$

Pour $\mathbb{Q} \in \mathcal{M}(S)$, le processus X^θ est une \mathbb{Q} -martingale locale. Comme $\tilde{X}_N^\theta \geq 0$, on déduit du lemme 8.21 que \tilde{X} est une \mathbb{Q} -martingale. Alors $X_0 = 0 = \mathbb{E}^\mathbb{Q}[\tilde{X}_N]$ et par suite $\tilde{X}_N = 0$, \mathbb{Q} -p.s. et donc \mathbb{P} -p.s. du fait que \mathbb{P} et \mathbb{Q} sont des probabilités équivalentes. \diamond

12.3.3 Evaluation et couverture

Un actif contingent est une v.a. \mathcal{F}_T -mesurable. Pour tout actif contingent B , on introduit le problème de sur-couverture

$$V(B) := \inf \{X_0 : \exists \theta \in \mathcal{A}, X_N^\theta \geq B \text{ p.s.}\},$$

et on note par $p(B)$ le prix de marché (à déterminer) de l'actif contingent B . Par un raisonnement d'absence d'arbitrage analogue à celui fait dans le cadre de l'arbre binomial de la section 12.2 (iii), on voit immédiatement que

$$-V(-B) \leq p(B) \leq V(B). \quad (12.8)$$

Ces inégalités sont appelées *bornes de non arbitrage* sur le prix de marché.

Le résultat principal qui généralise la situation observée dans le modèle binomial est le suivant.

Théorème 12.15. *Supposons que $\mathcal{M}(S) \neq \emptyset$, et soit B un actif contingent dont la valeur actualisée $\tilde{B} := B/S_N^0$ vérifie $\mathbb{E}^\mathbb{Q}[|\tilde{B}^-|] < \infty$ pour tout $\mathbb{Q} \in \mathcal{M}(S)$. Alors :*

$$V(B) = \sup_{\mathbb{Q} \in \mathcal{M}(S)} \mathbb{E}^\mathbb{Q}[\tilde{B}].$$

De plus, si $V(B) < \infty$, il existe $\theta^ \in \mathcal{A}$ tel que $X_0^{\theta^*} = V(B)$ et $X_N^{\theta^*} \geq B$, p.s.*

Avant de montrer ce résultat remarquons que dans le cas où $\mathcal{M}(S)$ est réduit à un singleton, on obtient immédiatement du résultat précédent que

$$\text{si } \mathcal{M}(S) = \{\mathbb{Q}\} \quad \text{alors pour } B \in \mathbb{L}^1(\mathbb{Q}), \quad p(B) = V(B) = \mathbb{E}^\mathbb{Q}[\tilde{B}].$$

C'est justement ce qui se passe pour le modèle binomial qui a été étudié dans les sections précédentes.

La démonstration du théorème 12.15 fait appel au résultat suivant que nous admettons.

Lemme 12.16. *Supposons que $\mathcal{M}(S) \neq \emptyset$. Alors, l'ensemble*

$$K := \left\{ U \in \mathbb{L}^0 : \exists \theta \in \mathcal{A}, X_0 = 0 \text{ et } \tilde{X}_N^\theta \geq U \right\} \quad (12.9)$$

est fermé en probabilité, i.e. pour toute suite $(U_n)_n$ d'éléments de K qui converge en probabilité vers une v.a. U , on a $U \in K$.

Preuve du théorème 12.15 Sans perte de généralité, on suppose

$$\tilde{B} \in \mathbb{L}^1(\mathbb{P}) \quad \text{et} \quad \tilde{S}_n \in L^1(\mathbb{P}) \quad \text{pour tout } n \leq N. \quad (12.10)$$

En effet, cette condition est satisfaite sous la probabilité $\tilde{\mathbb{P}} \sim \mathbb{P}$ définie par la densité

$$\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}} := \frac{e^{-|\tilde{B}| - \sum_{n=1}^N |\tilde{S}_n|}}{\mathbb{E} \left[e^{-|\tilde{B}| - \sum_{n=1}^N |\tilde{S}_n|} \right]},$$

et le problème est inchangé en remplaçant \mathbb{P} par $\tilde{\mathbb{P}}$.

(i) Pour montrer que $V(B) \geq \sup_{\mathbb{Q} \in \mathcal{M}(S)} \mathbb{E}^{\mathbb{Q}} [\tilde{B}]$, nous allons montrer que $X_0 \geq \mathbb{E}^{\mathbb{Q}} [\tilde{B}]$ pour tout $\mathbb{Q} \in \mathcal{M}(S)$, et $X_0 \in \mathbb{R}$, $\theta \in \mathcal{A}$ tels que

$$X_N^\theta \geq B, \quad \mathbb{P} - \text{p.s.} \quad (12.11)$$

Fixons donc $\mathbb{Q} \in \mathcal{M}(S)$ arbitraire et (X_0, θ) vérifiant (12.11). Comme \tilde{S} est une \mathbb{Q} -martingale, le processus de richesse actualisé $\tilde{X}^{x, \theta}$ est une \mathbb{Q} -martingale locale. En utilisant (12.11), on voit que $\mathbb{E}^{\mathbb{Q}} \left[\left(\tilde{X}_N^\theta \right)^- \right] \leq \mathbb{E}^{\mathbb{Q}} [\tilde{B}^-] \leq \|\tilde{B}^-\|_\infty < \infty$ d'après la condition du théorème. On déduit alors du lemme 8.21 que le processus \tilde{X}^θ est une \mathbb{Q} -martingale. Prenant les espérances dans (12.11), on obtient :

$$X_0 = \mathbb{E}^{\mathbb{Q}} [\tilde{X}_N^\theta] \geq \mathbb{E}^{\mathbb{Q}} [\tilde{B}].$$

(ii) Si $\sup_{\mathbb{Q} \in \mathcal{M}(S)} \mathbb{E}^{\mathbb{Q}} [\tilde{B}] = \infty$, alors l'étape précédente montre que $V(B) = \infty$ et la preuve est terminée. Il reste donc à montrer l'inégalité inverse $V(B) \leq \sup_{\mathbb{Q} \in \mathcal{M}(S)} \mathbb{E}^{\mathbb{Q}} [\tilde{B}]$ quand

$$\sup_{\mathbb{Q} \in \mathcal{M}(S)} \mathbb{E}^{\mathbb{Q}} [\tilde{B}] < \infty. \quad (12.12)$$

Fixons un paramètre $b < V(B)$. Alors, avec les notations du lemme 12.16,

$$\tilde{B} - b \notin K_1 := K \cap \mathbb{L}^1.$$

En effet, si au contraire $\tilde{B} - b \in K_1$, alors il existe $\theta \in \mathcal{A}$ tel que $\tilde{X}_N^\theta \geq \tilde{B} - b$, et $X_0 = 0$, i.e. et $X_0 = b$ et $X_N^\theta \geq B$, et par suite $b \geq V(B)$ par définition de $V(B)$.

Comme K est fermé en probabilité d'après le lemme 12.16, on voit immédiatement que l'ensemble K_1 est fermé dans \mathbb{L}^1 . On vérifie aussi que K est convexe. On peut alors Appliquer le théorème de séparation de Hahn-Banach aux convexes non vides $\{\tilde{B} - b\}$ et K_1 , pour déduire l'existence d'un couple $(Z, \alpha) \in \mathbb{L}^\infty \times \mathbb{R}$, avec $\mathbb{P}[Z = 0] < 1$, tel que

$$\mathbb{E}[ZU] \leq \alpha < \mathbb{E}[Z(\tilde{B} - b)] \quad \text{pour tout } U \in K \cap \mathbb{L}^1. \quad (12.13)$$

Remarquons que K_1 est un cône contenant l'origine. Alors, on peut choisir $\alpha = 0$. On remarque aussi que la v.a. $Y := -\mathbf{1}_{Z < 0} \in K_1$. Alors $\mathbb{E}[ZY] = -\mathbb{E}[Z\mathbf{1}_{Z < 0}] \leq 0$ d'après (12.13), et par suite $Z \geq 0$, p.s. Comme $\mathbb{P}[Z = 0] < 1$, ceci implique aussi que $\mathbb{E}[Z] > 0$, et on peut supposer que $\mathbb{E}[Z] = 1$ par une normalisation évidente. Enfin, soit ξ une v.a. bornée \mathcal{F}_{t-1} -mesurable à valeurs dans \mathbb{R}^d . Alors, on déduit de (12.10) que la v.a. $\xi \cdot \Delta \tilde{S}_t \in K$. L'inégalité de séparation (12.13) implique alors que $\mathbb{E}[Z\xi \cdot \Delta \tilde{S}_t] \leq 0$, et comme ξ est une v.a. \mathcal{F}_{t-1} -mesurable arbitraire, ceci montre que $\mathbb{E}\left[Z\Delta \tilde{S}_t | \mathcal{F}_{t-1}\right] = 0$. Ainsi, la v.a. Z vérifie :

$$Z \geq 0, \mathbb{E}\left[Z\Delta \tilde{S}_t | \mathcal{F}_{t-1}\right] = 0 \quad \text{et} \quad \mathbb{E}[Z\tilde{B}] \geq b. \quad (12.14)$$

Utilisons maintenant l'hypothèse $\mathcal{M}(S) \neq \emptyset$, et soit $\mathbb{Q}^1 \in \mathcal{M}(S)$ de densité $Z^1 := d\mathbb{Q}^1/d\mathbb{P}$ par rapport à \mathbb{P} . Pour tout $\varepsilon \in (0, 1]$, on introduit la probabilité $\mathbb{Q}^\varepsilon \sim \mathbb{P}$ de densité par rapport à \mathbb{P} :

$$Z^\varepsilon = d\mathbb{Q}^\varepsilon/d\mathbb{P} := \varepsilon Z^1 + (1 - \varepsilon)Z.$$

On vérifie immédiatement que $\mathbb{Q}^\varepsilon \in \mathcal{M}(S)$. Compte-tenu de l'inégalité $b \leq \mathbb{E}[Z\tilde{B}]$, ceci implique que :

$$\begin{aligned} b &\leq \mathbb{E}^{\mathbb{Q}^\varepsilon}[\tilde{B}] + \mathbb{E}[(Z - Z^\varepsilon)\tilde{B}] \\ &= \mathbb{E}^{\mathbb{Q}^\varepsilon}[\tilde{B}] + \varepsilon\mathbb{E}[(Z - Z^1)\tilde{B}] \\ &\leq \sup_{\mathbb{Q} \in \mathcal{M}(S)} \mathbb{E}^{\mathbb{Q}}[\tilde{B}] + \varepsilon\mathbb{E}[(Z - Z^1)\tilde{B}], \end{aligned}$$

et en envoyant ε vers zéro, on obtient :

$$b \leq \sup_{\mathbb{Q} \in \mathcal{P}} \mathbb{E}^{\mathbb{Q}}[\tilde{B}],$$

et l'inégalité recherchée est obtenue en envoyant $b \nearrow V(B)$.

(iii) Il reste à montrer le résultat d'existence. Par définition du problème de sur-couverture $V(B)$, on peut trouver $\theta^n \in \mathcal{A}$, pour tout $n \geq 0$, telle que

$$X_0^{\theta^n} V(B) + \frac{1}{n}, X_T^{\theta^n} \geq \tilde{B}, \quad \mathbb{P} - \text{p.s.}$$

Alors la v.a. $C^n := B - V(B) - n^{-1} \in K$. Il est clair que $C^n \rightarrow B - V(B)$ en probabilité. Alors, on déduit du lemme 12.16 que $B - V(B) \in K$, c'est à dire qu'il existe $\theta^* \in \mathcal{A}$ telle que $X_0 = 0$ et $X_N^{\theta^*} \geq B - V(B)$ ou, de manière équivalente, $X_0 = V(B)$ et $X_N^{\theta^*} \geq B$, p.s. \diamond

Bibliographie

- [1] K.B. Athreya et P.E. Ney (1972), *Branching Processes*, Springer Verlag.
- [2] F. Black et M. Scholes (1973), The pricing of options and corporate liabilities, *Journal of Political Economy*, 81, 637-654.
- [3] J.C. Cox, S.A. Ross, et M. Rubinstein (1979), Option pricing : a simplified approach, *Journal of Financial Economics*, 7, 229-263.
- [4] M. Benaim et N. El Karoui (2006), *Promenade Aléatoire : Chaînes de Markov et simulations ; martingales et stratégies*, Les Editions de l'Ecole Polytechnique.
- [5] K. L. Chung (1967), *Markov Chains with Stationary Transition Probabilities*, 2nd edition. Springer-Verlag.
- [6] J.-F. Delmas et B. Jourdain (2006), *Modèles aléatoires : applications aux sciences de l'ingénieur et du vivant*, Springer-Verlag. Springer-Verlag.
- [7] M. Dufflo (1996), *Algorithmes stochastiques*. Springer-Verlag.
- [8] C. Graham (2008), *Chaînes de Markov : cours et exercices corrigés*. Dunod.
- [9] J. Jacod et P. Protter (2002), *Probability Essentials*, 2nd Edition. Universitext, Springer-Verlag.
- [10] H.J. Kushner et P. Dupuis (1992), *Numerical Methods for Stochastic Control Problems in Continuous Time*, Springer-Verlag.
- [11] S. Méléard (2006), *Aléatoire*. Polycopié de cours de première année, Ecole Polytechnique.
- [12] R.C. Merton (1973), Theory of Rational Option Pricing, *Bell Journal of Economics*, 4, 141-183.
- [13] J. Neveu (1972), *Martingales à temps discret*. Masson et CIE.
- [14] A. Shiryaev (1989), *Probability*, 2nd Edition. Graduate Texts in Mathematics 95, Springer-Verlag.
- [15] D. Williams (1991), *Probability with martingales*, Cambridge Mathematical Textbooks.

Index

- π -système, 12
- d -système, 27
- à la monnaie, 177
- Algèbre, 11
- Algèbre
 - σ -algèbre, 11
 - σ -algèbre borelienne, 12
- Apériodicité, 114
- Black et Scholes, formule, 187
- Borel-Cantelli, lemme, 51
- Borel-Cantelli, lemme
 - premier, 15
- Bornes de non arbitrage
 - bornes, 190
- Chaîne de Markov, 71
- Chaîne de Markov
 - homogène, 75
- Classes monotones, théorème, 16, 31
- Convergence
 - dans \mathbb{L}^p , 40
 - en loi, 45
 - en probabilité, 41
- Convergence des martingales, théorème, 145
- Convergence dominée, théorème, 19
- Convergence monotone, théorème, 18
- dans la monnaie, 177
- Dirichlet, problème
 - définition, 81
 - Monte Carlo, 86
 - représentation stochastique, 82
- Doebelin, condition, 117
- Doob, inégalités maximales, 135
- Doob, théorème d'arrêt, 133
- en dehors de la monnaie, 177
- Espérance conditionnelle
 - convergence dominée, 59
 - convergence monotone, 59
 - Fatou, 59
 - inégalité de Jensen, 59
 - projections itérées, 60
 - dans \mathbb{L}^2 , 58
- Espaces produits, 24
- Excursions, 88
- exercice anticipé, 179
- Fatou, lemme
 - inverse pour les ensembles, 15
 - pour les ensembles, 14
 - pour les fonctions, 18
- Fermée, classe, 96
- Filtration, 67
- Filtre de Kalman-Bucy, 61
- Fonction mesurable, 15
- Fubini, théorème, 26
- Galton-Watson
 - Extinction, probabilité, 158
- Galton-Watson, processus
 - définition, 155
- Hastings-Metropolis, algorithme, 121
- Inégalité de Chebyshev, 23
- Inégalité de Hölder, 24
- Inégalité de Jensen, 36
- Inégalité de Markov, 23
- Inégalité de Minkowski, 24
- Inégalité de Schwarz, 24
- Indépendantes
 - σ -algèbres, 49
 - variables aléatoires, 50

- Intégrale
 - changement de variable, 26
- Intégration
 - fonctions réelles, 19
- intégration
 - fonctions positives, 16
- Irréductible, classe, 96
- La forte des grands nombres, 143
- Loi du zéro-un, 51
- Loi forte des grands nombres, 52
- Loi invariante
 - définition, 89
 - et récurrence positive, 98
 - existence, espace dénombrable, 93
 - existence, espace fini, 90
- Lois marginales, convergence, 116
- Marche aléatoire
 - définition, 75
 - sur \mathbb{Z} , récurrence, 100
 - sur \mathbb{Z}^d , récurrence, 102
 - unidimensionnelle, récurrence, 101
- Martingale
 - définition, 131
 - fermée, 134, 146
 - variation quadratique, 137
- Martingale locale, 137
- Matrice de transition, 71
- Mesure, 12
- Mesure
 - à densité, 22
 - de Lebesgue, 13
 - extérieure, 28
 - image, 22
- mesure
 - σ -additive, 12
 - additive, 12
- Mesures
 - égalité, 12
 - extension de Carathéodory, 12
- modèle binomial, 181
 - condition de non arbitrage, 182
 - couverture, 183
 - probabilité neutre au risque, 183
- Modèle binomial
 - multi-périodique, 183
 - uni-périodique, 181
- modèle de Cox-Ross-Rubinstein, 183
 - limite en temps continu, 186
- non domination, 177
- obligation zéro-coupon, 177
- option américaine, 176
- option asiatique, 180
- option d'achat européenne, 176
- options à barrières, 180
- Options américaines perpétuelles, 170
- Options européennes et américaines, 176
- options sur panier, 180
- parité call-put, 179
- Principe de non domination, 177
- principe de non domination, 176
- prix d'exercice, 176
- probabilité martingale équivalente, 183
- Probabilités neutres au risque, 189
- Processus
 - adapté, 68
 - prévisible, 68
- Propriété de Markov forte, 74
- Récurrent, état
 - nul, 94
 - positif, 94
- Recuit simulé, 125
- Robbins-Monro, algorithme, 149
- Scheffé, lemme, 20
- Snell, enveloppe de
 - horizon fini, 163
 - horizon infini, 169
- strike, 176
- Suprémum essentiel, 169
- Surmartingales, décomposition, 136
- Temps d'arrêt, 68
- Temps de retour, 88
- théorème central limite, 53
- Théorème central limite
 - pour les chaînes de Markov, 112
 - pour les martingales, 148

Théorème ergodique, 111

Uniformément intégrable, famille, 42

valeur intrinsèque, 177

Variable aléatoire, 33

Variable aléatoire

 densité, 35

 distribution, 34

 espérance, 35

 fonction caractéristique, 37

 fonction de répartition, 34

Vecteur gaussien, loi conditionnelle, 62