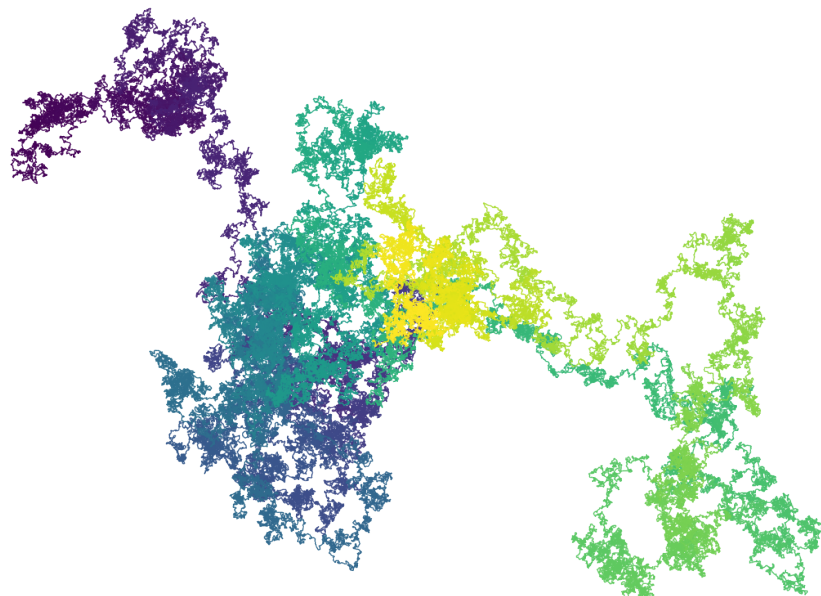


Probabilités et chaînes de Markov



Ronan Herry

Master de mathématiques
Université de Rennes
Premier semestre du M1

Contact : ronan.herry@univ-rennes.fr

Mis à disposition sous licence
Creative Commons



Ce document est le support du cours *Probabilités et chaînes de Markov* de la première année du master de mathématiques fondamentales de l'Université de Rennes. Il forme un tout avec la feuille d'exercices qui l'accompagne et il est conçu pour être lu en préparation ou en regard des séances.

Probabilités et chaînes de Markov. Les probabilités occupent une place centrale en mathématiques. Les solides fondations de la théorie de la mesure, posées au début du XX^e siècle par Lebesgue, Borel, Carathéodory, Radon et bien d'autres, ont permis l'émergence, sous l'impulsion de Kolmogorov dans les années 30, d'une théorie axiomatique des probabilités. Ce domaine autonome a conduit à l'émergence de nouveaux champs de recherche fondamentaux : processus stochastiques, théorèmes limites, matrices aléatoires, percolation, géométrie aléatoire, équations différentielles stochastiques. Les probabilités servent également de point de rencontre entre l'analyse, la physique mathématique, les systèmes dynamiques, la combinatoire, la théorie des groupes et la géométrie.

Les chaînes de Markov sont des objets simples et fondamentaux de la théorie. Ce sont des systèmes aléatoires à temps discret *sans mémoire* : le futur ne dépend du passé qu'à travers le présent. Elles donnent lieu à une théorie riche, capables de décrire des phénomènes aussi éloignés que le mélange d'un jeu de cartes, la dynamique d'une population, la géométrie d'un groupe, le classement des pages du web ou la simulation numérique de lois complexes. Cette ubiquité fait des chaînes de Markov un langage commun aux mathématiques théoriques et pratiques. Les étudier, c'est aussi apprendre à faire dialoguer deux points de vue dont l'unité est l'un des fils conducteurs de ce cours : un point de vue *probabiliste*, celui des trajectoires et des temps d'arrêt, et un point de vue *analytique*, celui des opérateurs, des fonctions harmoniques et des mesures invariantes.

Prérequis. Ce cours s'adresse à des étudiantes et étudiants ayant validé une licence de mathématiques fondamentales. En particulier, il suppose acquis :

- la théorie de la mesure et de l'intégration de niveau L3 (théorème de Radon–Nikodym, théorèmes de convergence, théorème de Fubini, espaces L^p), rappelée sans démonstration dans l'annexe **B** ;
- les bases d'un cours de topologie générale (voir annexe **A** pour un rappel succinct) ; en revanche, les espaces polonais, qui forment le cadre naturel de ce cours, y sont définis et n'ont pas à être connus au préalable ;
- le socle probabiliste de L3 (indépendance, fonctions caractéristiques, théorème central limite et théorème de Lévy) rappelé, et parfois redémontré, dans l'annexe **C** et au début du chapitre **1** ;
- un minimum d'analyse hilbertienne, au niveau de la projection orthogonale dans L^2 .

Ce cours ne suppose pas connue la théorie des martingales : c'est un choix assumé, sur lequel je reviens ci-dessous.

Public cible. Ce cours s'adresse à des étudiantes et des étudiants de première année de master, en particulier de l'Université de Rennes, dont les projets peuvent être indécis et variés. J'ai, tant bien que mal, cherché à faire en sorte que chacun y trouve son compte.

1. *Aux candidates et candidats à l'agrégation*, le cours donne des bases solides pour le tronc commun d'analyse et de probabilités et, pour celles et ceux qui la choisissent, pour l'option probabilités–statistiques. Je recommande de viser à minima le niveau *Approfondissement* décrit dans les compétences, de s'entraîner à traiter les problèmes plus longs, de développer sa culture générale et de contextualiser les résultats du cours par rapport aux autres cours du cursus. Si l'on souhaite prendre l'option probabilités–statistiques, il est aussi fortement recommandé de s'investir dans les TPs.

2. *À celles et ceux qui se destinent à la recherche*, ce cours fournit des fondations robustes en théorie de la mesure générale, théorie ergodique, théorie du potentiel. Ces fondations sont prolongées par les encadrés *Pour aller plus loin* et les exercices du dernier niveau. Même si l'on ne se destine pas à la recherche en probabilités, ces concepts interviennent dans de nombreux champs de recherche.

3. *À celles et ceux qui s'orienteront vers des domaines appliqués ou industriels*, le cours donne les fondations conceptuelles d'outils devenus quotidiens en modélisation, en science des données et en apprentissage : chaînes de Markov, méthodes de Monte-Carlo, temps de mélange. On retrouvera ces exemples concrets en TD et en TP.

Place dans le cursus. Ce cours ne se lit pas isolément : il s'inscrit dans le master de mathématiques de Rennes et dialogue avec plusieurs autres enseignements. Au premier semestre, il s'appuie sur les enseignements suivants, auxquels il fait écho :

- *Analyse, distributions, Fourier* : lien entre mesures et distributions, aspects topologiques, et correspondance entre transformée de Fourier et fonction caractéristique ;
- *Analyse fonctionnelle* : les espaces de Hilbert et les opérateurs, présents dans la construction L^2 de l'espérance conditionnelle, dans les opérateurs de Markov et dans le point de vue spectral ;
- *Statistiques mathématiques* : de nombreux résultats de convergence probabiliste y sont utilisés ;
- plus ponctuellement, *Théorie des groupes et géométrie* : à travers les marches aléatoires sur les groupes.

Au second semestre, le cours obligatoire *Probabilités et martingales* s'appuie fortement sur celui-ci : c'est là qu'est développée la théorie des martingales, délibérément non supposée ici, de sorte que ce cours y prépare directement. Enfin, plusieurs UE optionnelles le prolongent : *Apprentissage statistique* et *Modèles aléatoires*, où les chaînes de Markov et les méthodes de Monte-Carlo sont centrales, et, plus indirectement, *Équations aux dérivées partielles*, dont le problème de Dirichlet et les fonctions harmoniques de la section 3.3 sont l'analogue discret.

Partis pris. Des choix de fond distinguent ce texte d'une présentation plus classique.

1. Les chaînes de Markov y sont étudiées sur un *espace d'états quelconque* dès le départ, et non selon la progression habituelle : fini $\rightarrow$ dénombrable $\rightarrow$ général. Le contenu conceptuel sur la stationnarité, l'harmonicité et l'ergodicité est le même dans tous les cas. Et je trouve plus clair de l'aborder dans un cadre général plutôt que de le répéter dans une généralité croissante. Le cours ne se spécialise au cas fini ou dénombrable qu'au chapitre 4, pour les résultats fins de temps long.

-
2. L'espérance conditionnelle et les lois conditionnelles font l'objet d'une construction rigoureuse (chapitre 2) : c'est le cœur technique du cours.
 3. La théorie ergodique est énoncée pour des systèmes dynamiques généraux puis spécialisée aux chaînes, si bien que la loi des grands nombres et la structure des mesures stationnaires en découlent d'un même geste.

Organisation du document. Le texte est divisé en deux parties *Probabilités* et *Chaînes de Markov*, soit quatre chapitres, complétés par une introduction et deux annexes de rappels. Les encadrés *Pour aller plus loin* sont des ouvertures : ils pointent vers des développements avancés ou des questions de recherche, ne sont pas exigibles à l'examen, mais montrent où conduisent les idées du cours. Le poly et sa feuille d'exercices sont indissociables : de nombreux résultats sont mis en pratique, parfois démontrés, dans le TD et les deux documents se renvoient l'un à l'autre. Chaque chapitre se termine avec une liste d'exercices qui permette d'évaluer si vous maîtriser le chapitre, vaguement ranger par thème. **Cherchez les exercices par vous-même** : c'est là que le cours s'apprend. Ils sont gradués en quatre niveaux de difficulté, expliqués en tête de la feuille (voir *Comment utiliser cette feuille*), que je ne reproduis pas ici. Les références croisées fonctionnent : si vous consultez le polycopié en version électronique, un clic sur le numéro de la feuille renvoie directement à l'exercice dans le livret.

Références. La bibliographie en fin de volume référence les ouvrages utiles. Un commentaire bibliographique contextualise ces ouvrages. J'ai également ajouté une liste (subjective) de théorèmes importants, un index de notation ainsi qu'une liste des figures.

Avertissement. Comme tout document mathématique, celui-ci contient sans doute des erreurs, coquilles ou inconsistances. Je remercie les lecteur·ices qui m'en ont déjà signalées ainsi que celles et ceux qui le feront dans le futur. Me contacter à ronan.herry@univ-rennes.fr.

Chaque compétence est déclinée en trois niveaux de maîtrise : *Socle* (savoir-faire attendu de tous, dans un cas de référence), *Approfondissement* (mise en œuvre dans le cadre général du cours) et *Ouverture* (maîtrise des démonstrations, de leurs limites et de leurs généralisations). Il n'est pas envisageable de valider ce cours si l'on ne maîtrise pas complètement le niveau Socle.

Compétences techniques

CONVERGENCE ÉTROITE ET COMPACTITÉ DES LOIS

Socle. Maîtriser la convergence en loi et le théorème de Prokhorov dans le cas de $\mathbb{R}^d$: tension d'une loi, extraction de sous-suites convergentes.

Approfondissement. Transposer la convergence étroite, l'équitension et le théorème de Prokhorov au cadre général des espaces polonais.

Ouverture. Connaître les démonstrations (régularité des mesures, Prokhorov), leurs hypothèses, et discuter leurs limites et généralisations.

FONCTIONS CARACTÉRISTIQUES ET THÉORÈME CENTRAL LIMITE

Socle. Calculer les fonctions caractéristiques usuelles, et appliquer le théorème de Lévy et le théorème central limite à des sommes de variables i.i.d. sur $\mathbb{R}$.

Approfondissement. Utiliser le théorème de continuité de Lévy en dimension quelconque et le critère d'équitension par la fonction caractéristique.

Ouverture. Connaître la preuve du théorème de Lévy (équitension, Prokhorov, injectivité de Fourier), ses limites et ses variantes.

ESPÉRANCE ET LOIS CONDITIONNELLES

Socle. Calculer une espérance conditionnelle pour les variables aléatoires discrètes, à densité, indépendantes et gaussiennes.

Approfondissement. Manipuler l'espérance conditionnelle par rapport à une tribu et les lois conditionnelles (noyaux, désintégration) sur des espaces généraux.

Ouverture. Maîtriser la construction (Radon–Nikodym, propriété universelle), les hypothèses d'existence des lois conditionnelles et leurs limitations.

MODÉLISATION PAR UNE CHAÎNE DE MARKOV

Socle. Décrire et manipuler une chaîne sur un espace fini ou dénombrable : noyau de transition, graphe.

Approfondissement. Modéliser sur un espace d'états général au moyen d'un noyau markovien, et en construire la version canonique.

Ouverture. Justifier l'existence de la chaîne et la propriété de Markov forte, et discuter le choix de l'espace d'états.

FONCTIONS HARMONIQUES ET PROBLÈME DE DIRICHLET

Socle. Résoudre un problème de temps d'atteinte concret (ruine du joueur, marche sur un graphe fini) par un système linéaire.

Approfondissement. Formuler et résoudre le problème de Dirichlet, le potentiel et l'équation de Poisson sur des espaces généraux.

Ouverture. Connaître la caractérisation des probabilités d'atteinte comme plus petite solution positive, sa preuve, et le parallèle avec le laplacien et les EDP.

RÉCURRENCE, TRANSIENCE ET MESURES STATIONNAIRES

Socle. Classifier les états d'une chaîne finie ou dénombrable et déterminer une mesure stationnaire.

Approfondissement. Établir l'existence et l'unicité d'une mesure stationnaire sur un espace général (Krylov–Bogolioubov, fonctions de Lyapunov, réversibilité).

Ouverture. Connaître les preuves (dichotomie récurrence/transience, critères de Lyapunov) et discuter la non-unicité et les phénomènes de type Liouville.

MARCHES ALÉATOIRES

Socle Connaître le résultat de Pólya et pouvoir faire des calculs pour la marche aléatoire simple sur $\mathbb{Z}^d$.

Approfondissement Connaître la preuve du théorème de Pólya et savoir faire des calculs sur des groupes dénombrables généraux.

Ouverture Maîtriser les notions de bords d'un groupe et d'un arbre, connaître le lien entre croissance d'un groupe et transience.

COMPORTEMENT EN TEMPS LONG

Socle. Appliquer la convergence à l'équilibre d'une chaîne finie irréductible et apériodique (Perron–Frobenius) et en calculer la limite.

Approfondissement. Établir la convergence sur un espace dénombrable ou général (couplage) et les moyennes ergodiques (théorème de Birkhoff), y compris dans le cas périodique.

Ouverture. Connaître les preuves (couplage, lemme ergodique maximal) et situer la théorie quantitative : coefficient de Dobrushin, trou spectral, cutoff.

Compétences transversales

RIGUEUR ET DOUBLE POINT DE VUE

Socle. Mener une démonstration probabiliste rigoureuse appuyée sur la théorie de la mesure dans des situations standard.

Approfondissement. Choisir, selon le problème, entre le point de vue trajectorien et le point de vue opératoire, et passer de l'un à l'autre.

Ouverture. Analyser de façon critique une démonstration (hypothèses nécessaires, contre-exemples) et situer un résultat dans le paysage mathématique (théorie ergodique, potentiel, EDP, inégalités fonctionnelles).

MODÉLISATION ALÉATOIRE

Socle. Reconnaître, dans une situation concrète, une structure sans mémoire relevant d'une chaîne de Markov, et en identifier les états et les transitions élémentaires.

Approfondissement. Traduire un problème appliqué (algorithme, dynamique de population, marche sur un réseau) en un modèle markovien pertinent (choix de l'espace d'états et du noyau), puis réinterpréter les résultats mathématiques (mesure stationnaire, temps d'atteinte, vitesse de mélange) dans les termes du problème d'origine.

Ouverture. Discuter la pertinence et les limites d'un modèle, comparer des modélisations concurrentes d'un même phénomène, et apprécier l'écart entre le modèle et la réalité qu'il représente.

Avant-propos	i
Compétences visées	v
Compétences techniques	v
Convergence étroite et compacité des lois	v
Fonctions caractéristiques et théorème central limite	v
Espérance et lois conditionnelles	v
Modélisation par une chaîne de Markov	v
Fonctions harmoniques et problème de Dirichlet	vi
Récurrence, transience et mesures stationnaires	vi
Marches aléatoires	vi
Comportement en temps long	vi
Compétences transversales	vi
Rigueur et double point de vue	vi
Modélisation aléatoire	vii
Table des matières	viii
0 Introduction	1
0.1 Présentation générale	1
0.2 Résumé et plan du cours	2
0.2.1 L'espace des mesures de probabilités	2
0.2.2 Espérance et loi conditionnelle	2
0.2.3 Chaînes de Markov	3
0.2.4 Étude des mesures stationnaires et convergences en temps long	4
0.3 Exemples	4
0.3.1 Un modèle d'urne avec une force de rappel	4
0.3.2 Le cavalier errant	5
0.3.3 Marche aléatoire sur un groupe abélien	6
0.3.4 Modèles de population	6
Première partie Probabilités	7
1 Les convergences en loi et étroite	9
1.1 Rappels	9
1.1.1 Cas général	9
1.1.2 Cas des variables numériques	11
1.2 Cadre topologique	11
1.2.1 Espaces complets et séparables	11

	Hypothèse et notations	11
	Exemples et contre-exemples	15
1.2.2	Fonctions semi-continues	17
	Fonction semi-continue inférieurement : définition et exemples	18
	Caractérisation de la semi-continuité inférieure	18
	Fonction semi-continue supérieurement	19
	Approximation d'intégrales	20
1.3	Propriétés de la convergence étroite et de la convergence en loi	20
1.3.1	Le lemme du portemanteau	20
1.3.2	Tension et régularité des mesures de probabilités	22
	Tension	22
	Régularité	23
1.3.3	Distance de Fortet–Mourier	24
	Distance de Fortet–Mourier et épaissis	25
	Compatibilité de la distance et de la convergence étroite	25
	Définition topologie de la convergence étroite	26
1.3.4	Compacité pour la convergence étroite	28
	Cas général	28
	Cas des variables aléatoires numériques	33
	Exercices fondamentaux	35
2	Conditionnement	37
2.1	Espérance conditionnelle par rapport à une tribu	37
2.1.1	Définition	37
	Existence d'un objet universel	37
	Définition à partir de la propriété universelle	38
2.1.2	Propriétés	38
	Propriétés qui généralisent celles de l'espérance	38
	Propriétés spécifiques	40
2.1.3	Conditionnement par rapport à une variable aléatoire	41
	Lemme de Doob	41
	L'espérance conditionnelle comme une fonction	42
	Conditionnement discret	42
2.2	Lois conditionnelles et noyaux de Markov	43
2.2.1	Noyaux de Markov	43
	Opérateurs associés	44
2.2.2	Loi conditionnelles	44
	Définition de la loi conditionnelle	44
	La loi conditionnelle est portée par les fibres	47
2.2.3	Lien entre loi conditionnelle et espérance conditionnelle	48
	Le théorème abstrait	48
	Application à des couples de variables aléatoires	49
2.2.4	Calculs explicites d'espérances conditionnelles	50
	Cas des variables aléatoires indépendantes	50
	Cas des variables aléatoires à densité	51
	Cas des vecteurs gaussiens	51
2.3	Applications : les théorèmes ergodiques de Birkhoff et de Choquet	52
2.3.1	Mesures ergodiques	53
2.3.2	Démonstration du théorème de Birkhoff	55
	Exercices fondamentaux	57

Deuxième partie Chaînes de Markov	59
3 Chaînes de Markov	61
3.1 Définitions et premières propriétés	61
3.1.1 Définition	61
Cas des espaces finis	62
3.1.2 Unicité, existence et caractérisation	63
Loi d'une chaîne de Markov	63
Les chaînes de Markov comme système récurrent	64
Exemples importants	64
Existence	65
3.2 Analyse stochastique	66
3.2.1 Processus canonique	66
L'espace canonique	66
Applications coordonnées	66
Chaîne de Markov canonique	67
Autres conditions initiales	67
3.2.2 Reformulation de la propriété de Markov	67
Opérateurs de translation	67
Espérance aléatoire	67
Propriété de Markov faible	68
3.2.3 Renforcement de la propriété de Markov	68
Temps d'arrêt	69
Temps d'atteinte	69
Propriété de Markov forte	70
3.3 Analyse harmonique	70
3.3.1 Fonctions harmoniques	70
Fonctions harmoniques sur tout l'espace	70
Harmonicité sur un domaine	71
3.3.2 Temps d'atteinte et problème de Dirichlet	71
3.3.3 Potentiel d'une chaîne de Markov	74
Noyau potentiel	74
Équation de Poisson	74
3.4 Mesures stationnaires	75
3.4.1 Définitions et premiers résultats	75
Stationnarité	75
Réversibilité	75
3.4.2 Mesures stationnaires et théorie ergodique	76
Lien avec les systèmes dynamiques	76
Mesures ergodiques	77
3.4.3 Propriété de Liouville et non-unicité	78
Exercices fondamentaux	79
4 Mesures stationnaires et temps longs	81
4.1 Cas des espaces finis	81
4.1.1 Matrices irréductibles	81
Irréductibilité	81
Périodicité	82
Théorème de Perron–Frobenius	83
4.2 Cas des espaces dénombrables	85

4.2.1	Classification des états	85
	Classes de communication	85
	États transitoires et récurrents	86
	Classe de récurrence	87
	Récurrence et transience de la marche aléatoire simple et symétrique sur $\mathbb{Z}^d$	89
4.2.2	Comportement en temps long	90
	Mesures stationnaires	90
	Loi des grands nombres	92
	Théorème de Birkhoff	94
	Convergence en loi	95
4.3	Cas général	96
4.3.1	Existence de mesures stationnaires	96
	Critère par compacité	97
	Critère par fonctions de Lyapunov	98
4.3.2	Unicité de la mesure stationnaire	99
	Phénomène de contraction déterministe	99
	Phénomènes probabilistes	100
4.3.3	Ouverture : vitesse de convergence	101
	Distance en variation totale.	102
	Contraction et coefficient de Dobrushin	102
	Trou spectral et approche L^2	103
	Exercices fondamentaux	103
Troisième partie Rappels		105
A Topologie générale		107
B Théorie de la mesure		109
B.1	Rappels d'intégration	109
B.2	Rappels et compléments de théorie de la mesure	110
C Vecteurs gaussiens et théorème central de la limite		119
C.1	Mesures gaussiennes sur la droite réelle	119
C.1.1	Loi gaussienne standard	119
C.1.2	Loi gaussienne avec moyenne et variance	120
C.2	Vecteurs gaussiens	120
C.2.1	Vecteur gaussien standard	120
C.2.2	Vecteurs gaussiens avec moyenne et covariance	121
C.3	Le théorème central limite	122
Commentaires bibliographiques		123
Liste des théorèmes importants		127
Table des figures		128
Index des notations		131
Bibliographie		133

0.1 Présentation générale

Vu de loin, ce cours cherche à répondre à une question en apparence simple mais d'une grande richesse : que peut-on dire d'une suite de variables aléatoires (X_n) ? À partir de l'intuition développée pour les suites déterministes, différents comportements peuvent être étudiés.

- (a) Le comportement *local*. Si on connaît X_n que peut-on dire de X_{n+1} , X_{n+2} , etc.?
- (b) Le comportement *global*. En moyenne, combien de fois va-t-on visiter tel état?
- (c) Le comportement *asymptotique*. La suite converge-t-elle?

Dans ce cours, nous tenterons de répondre à certaines de ces questions pour les *chaînes de Markov*. Il s'agit des suites (X_n) pour lesquelles on impose une structure locale par une relation de récurrence à un pas :

$$X_{n+1} = f(X_n, U_{n+1}), \quad (1)$$

avec (U_n) une suite de variables aléatoires indépendantes et identiquement distribuées. On résume souvent cette propriété en disant que « le futur ne dépend du passé qu'à travers le présent ». Comme dans le cas des suites récurrentes ou des équations différentielles ordinaires, nous verrons que cette prescription locale caractérise complètement la suite.

	Déterministe	Aléatoire
Temps discret	Suite récurrente $x_{n+1} = f(x_n)$	Chaîne de Markov $X_{n+1} = f(X_n, U_{n+1})$
Temps continu	EDO $\dot{x}_t = f(x_t)$	Processus de Markov $dX_t = f(X_t)dU_t$

TABLE 0.1 : Les chaînes de Markov en regard d'autres systèmes dynamiques

L'étude de (1) se révèle particulièrement compliquée si on s'impose une étude au niveau des trajectoires¹, c'est-à-dire ω par ω . En suivant l'idée clef de la physique statistique, dans ce cours on passera très vite au niveau de l'étude des lois. En écrivant $\mu_n := \text{loi}[X_n]$, on verra (théorème 3.9) que la suite (μ_n) satisfait elle-même l'équation de récurrence *linéaire*

$$\mu_{n+1}(A) = \int q(x, A) \mu_n(dx), \quad (2)$$

où $(q(x, A))_{x \in \mathbb{X}, A \in \mathfrak{B}(\mathbb{X})}$ est un *noyau de probabilité* qui ne dépend que de f et de la loi de U_0 . En passant de X_n à μ_n , on passe de $\mathbb{X}$ à $\mathcal{P}(\mathbb{X})$ qui est un espace plus compliqué, mais on a récupéré de la linéarité. Par chance, on sait traiter les équations linéaires à un grand niveau de généralité.

1. Une telle étude sera partiellement entreprise dans le cadre du cours *Probabilités et martingales* au second semestre

0.2 Résumé et plan du cours

L'objectif principal de ce cours est de définir et d'étudier les chaînes de Markov et leurs mesures *stationnaires*, c'est-à-dire les solutions de (2).

0.2.1 L'ESPACE DES MESURES DE PROBABILITÉS

Pour cela on commencera au chapitre 1 par définir une topologie et des outils analytiques sur $\mathcal{P}(\mathbb{X})$ quand $\mathbb{X}$ est un espace topologique assez général (un espace *polonais*). Plus précisément, nous définissons la topologie *étroite*, sur $\mathcal{P}(\mathbb{X})$ comme celle induite par la distance

$$d_{BL}(\mu, \nu) := \sup \left\{ \left| \int \varphi d\mu - \int \varphi d\nu \right| : \|\varphi\|_\infty \leq 1, \text{Lip}(\varphi) \leq 1 \right\},$$

où Lip est la constante de Lipschitz de φ . Cette topologie a plusieurs propriétés sympathiques.

- (théorème 1.31) Elle correspond à la convergence en loi

$$X_n \xrightarrow[n \rightarrow \infty]{\text{loi}} X_\infty \Leftrightarrow d_{BL}(\text{loi}[X_n], \text{loi}[X_\infty]) \xrightarrow[n \rightarrow \infty]{} 0.$$

- (théorème 1.38) Les compacts de $\mathcal{P}(\mathbb{X})$ s'expriment en terme des compacts de $\mathbb{X}$: une famille (μ_i) est relativement compact, si et seulement si elle uniformément presque a presque supporté sur un compact : pour tout ε on peut trouver un compact K tel que pour tout i

$$\mu_i(K) \geq 1 - \varepsilon.$$

- (théorème 1.45) Dans le cas où $\mathbb{X} := \mathbb{R}^d$, elle permet de préciser la convergence en loi en termes de convergence de la fonction caractéristique vers une fonction continue en 0.

0.2.2 ESPÉRANCE ET LOI CONDITIONNELLE

Dans le chapitre 2, on continue à s'intéresser aux propriétés des variables aléatoires ou des lois de probabilités, non plus du point de vue analytique-topologique, mais réellement du point de vue probabiliste. Plus précisément on s'intéresse à la question suivante : étant un couple de variables aléatoires (X, Y) de loi jointe π , que peut-on dire de la loi de Y sachant que l'on a observé X ? Pour répondre à cette question, nous introduisons deux objets importants.

Nous définissons d'abord, sur un espace probabilisé quelconque, l'*espérance conditionnelle* $\mathbf{E}[F | X]$ d'une variable aléatoire numérique F sachant que l'on a observé une variable aléatoire X à valeurs quelconques. C'est (théorème 2.1) la *variable aléatoire* qui donne la meilleure approximation $\sigma(X)$ -mesurable de F . En particulier, si F est indépendant de X , alors $\mathbf{E}[F | X] = \mathbf{E}[F]$ et $\mathbf{E}[\varphi(X) | X] = \varphi(X)$.

Au-delà de ces cas simples, nous verrons qu'il est utile de considérer, la notion plus fine de *loi conditionnelle*. Intuitivement, si (X, Y) est un couple de variables aléatoires de loi π , la loi conditionnelle de Y sachant X est une *mesure de probabilité* aléatoire mesurable par rapport à X , donc de la forme $q(X, \cdot)$, et qui satisfait

$$\mathbf{P}[X \in A, Y \in B] = \mathbf{E}[q(X, B)1_A(X)].$$

En d'autres termes, la loi jointe π se *désintègre*

$$\pi(dx dy) = q(x, dy)\mu(dx),$$

où μ est la loi de X . L'existence, l'unicité et les bonnes propriétés des lois conditionnelles dépendent, contrairement à l'espérance conditionnelle, des hypothèses topologiques sur l'espace

de probabilité Ω sur lequel sont définies les variables aléatoires (théorème 2.16). Les hypothèses du chapitre 1 (espace polonais) sont suffisantes.

L'espérance conditionnelle et la loi conditionnelle sont, comme on peut s'y attendre, liées. En particulier, on peut calculer l'espérance conditionnelle à partir de la loi conditionnelle (théorème 2.19), ce qui permet de conclure dans de nombreux cas concrets.

La construction complète et rigoureuse de l'espérance conditionnelle et des lois conditionnelles nous permet à la section 2.3 d'énoncer et de démontrer facilement le théorème ergodique de Birkhoff (théorème 2.36) pour les *systèmes dynamiques*, c'est-à-dire les itérées successives d'une transformation mesurable $\theta : \Omega \rightarrow \Omega$ qui préserve une mesure de probabilité $\mathbf{P}$. Étant donné une variable aléatoire $F : \Omega \rightarrow \mathbb{R}$, on s'intéresse aux valeurs prises par F le long de l'orbite de θ : on étudie la suite $(F \circ \theta^n)_{n \in \mathbb{N}}$. Le théorème ergodique nous dit alors que, sous des conditions sur $\mathbf{P}$, les moyennes temporelles convergent vers les moyennes spatiales

$$\frac{1}{n} \sum_{k=0}^{n-1} F \circ \theta^k \xrightarrow[n \rightarrow \infty]{p.s.} \mathbf{E}[F].$$

La condition sur $\mathbf{P}$ est dite d'*ergodicité*, et elle demande que toutes les fonctions invariantes par θ , c'est-à-dire $F \circ \theta = F$, soient $\mathbf{P}$ -presque sûrement constantes.

0.2.3 CHAÎNES DE MARKOV

Dans le chapitre 3, nous commençons l'étude générale des chaînes de Markov. Plutôt que d'utiliser la relation de récurrence (1), on va définir les chaînes de Markov à partir de la relation suivante sur les lois conditionnelles

$$\mathbf{loi}[X_{n+1} | X_n, \dots, X_0] = \mathbf{loi}[X_{n+1} | X_n] = q(X_n, \cdot), \quad (3)$$

pour un certain noyau q . La famille $(q(x, A))_{x \in \mathbb{X}, A \in \mathfrak{B}(\mathbb{Y})}$, utilisé pour présenter la loi conditionnelle, est un noyau de probabilité déjà rencontré informellement en (2). Ceci explique l'importance des noyaux dans ce cours et les liens entre chaînes de Markov et l'espérance conditionnelle. La relation (3) traduit l'intuition que nous avons formée plus haut : « le futur ne dépend du passé qu'à travers le présent ». Fort heureusement pour ce cours, les chaînes de Markov existent (théorème 3.13) et ont uniquement déterminées (en loi) par le noyau q et la loi initiale de X_0 (théorème 3.9). Bien que définis de manière plus abstraite, les systèmes récurrents $X_{n+1} = f(X_n, U_{n+1})$ sont effectivement des chaînes de Markov (théorème 3.11) et constituent la classe d'exemple à laquelle on se ramènera toujours en pratique.

Comme expliqué plus haut, l'étude de la suite (X_n) se révèle compliquée si on la regarde comme une suite aléatoire sur $\mathbb{X}$. Nous rencontrerons le premier basculement conceptuel à la section 3.2 : au lieu de considérer une seule suite, on va regarder des suites de suites. Plus précisément, on va considérer $\Omega := \mathbb{X}^{\mathbb{N}}$ l'espace des suites à valeurs dans $\mathbb{X}$ et l'opérateur de décalage

$$\theta : \Omega \ni (x_n) \rightarrow (x_{n+1}) \in \Omega.$$

Cet opérateur de décalage permet de formaliser la propriété d'absence de mémoire, appelée aussi propriété de Markov (théorème 3.22). L'opérateur (déterministe) θ permet aussi de voir les chaînes de Markov comme un système dynamique à condition de trouver une mesure $\mathbf{P}$ sur Ω . Comme la loi de la chaîne de Markov est complètement déterminée par q (qui est supposée d'aunée) et la loi initiale, cela revient à trouver une loi initiale qui est un point fixe, on parlera de *mesure stationnaire*, de (2). On abrègera souvent cette équation en

$$\mu_{n+1} = \mu_n \mathbf{Q},$$

où Q est un opérateur agissant à droite sur les mesures par

$$\mu Q(A) := \int q(x, A) \mu(dx).$$

En reformulant de manière plus probabiliste, si X_0 est de loi μ alors X_1 est de loi μQ . On regardera également son action « duale » à gauche sur les fonctions

$$Q\varphi(x) := \int \varphi(y) q(x, dy).$$

Probabilistiquement, si la chaîne est commencée en $X_0 = x$, alors $Q\varphi(x) = \mathbf{E}[\varphi(X_1)]$. La définition et l'étude des premières propriétés de cet opérateur sont entreprises à la section 3.3. Nous y donnons une exposition superficielle de l'analyse harmonique et de la théorie du potentiel. Cela permet malgré tout d'obtenir des résultats précis sur les temps d'atteinte (théorème 3.29 et proposition 3.32).

Finalement, à la section 3.4, nous étudions les mesures stationnaires. Nous explicitons le lien avec les systèmes dynamiques. Plus précisément, si μ est une probabilité sur $\mathbb{X}$, alors en considérant la chaîne de loi initiale μ , on peut relever μ en une probabilité $\mathbf{P}^\mu$ sur Ω . L'opérateur de décalage θ préserve $\mathbf{P}^\mu$ si et seulement si μ est stationnaire. Ainsi, le théorème ergodique s'applique aux chaînes de Markov (théorème 3.52) et généralise la loi des grands nombres. Une question importante est alors : parmi toutes les mesures de probabilité stationnaires, quelles sont celles qui sont ergodiques, c'est-à-dire telles que $\mathbf{P}^\mu$ est ergodique ?

0.2.4 ÉTUDE DES MESURES STATIONNAIRES ET CONVERGENCES EN TEMPS LONG

Motivé par cette dernière question, nous entreprenons dans chapitre 4 une étude des mesures stationnaires. Nous travaillons avec un niveau croissant de généralité. Dans le cas où $\mathbb{X}$ est fini, résoudre $\mu Q = \mu$ revient à trouver les espaces propres associés à la valeur propre 1 d'une matrice avec une forme particulière (coefficients positifs et lignes qui somment à 1). Le théorème de Perron–Frobenius (théorème 4.9) donne une réponse précise à cette question : il caractérise toutes les valeurs propres sur le cercle unité à partir d'une condition de périodicité ce qui permet de conclure précisément sur le comportement de Q^n quand $n \rightarrow \infty$ et donc sur la convergence en loi de la chaîne de Markov (X_n) (théorème 4.10). Dans le cas où $\mathbb{X}$ est infini l'équation $\mu Q = \mu$ n'est plus matricielle et il serait nécessaire d'utiliser des techniques d'analyse fonctionnelle si l'on souhaitait la résoudre analytiquement. Cette approche dépasse le cadre du cours, forte heureusement nous disposons de méthodes probabilistes. Dans le cas où $\mathbb{X}$ est dénombrable cela passe par la classification des états (théorème 4.18) en termes du nombre de visites ou du noyau de Green : les états récurrents sont visités une infinité de fois, les états transferts seulement un nombre fini. Pour les états récurrents, nous donnons une construction probabiliste de l'unique mesure ergodique (corollaire 4.23) d'où découle la loi des grands nombres (théorème 4.29) et les résultats de convergence en loi (théorème 4.30). Finalement, dans le cas général, nous nous contentons d'énoncer quelques critères suffisants pour l'existence : critères de Krylov–Bogolioubov (théorème 4.32) et Lyapunov (théorème 4.38); et pour l'unicité : critères de contraction déterministe (théorème 4.40) et du point accessible (théorème 4.49).

0.3 Exemples

0.3.1 UN MODÈLE D'URNE AVEC UNE FORCE DE RAPPEL

On considère une urne avec des boules jaunes et des boules rouges. À chaque étape on tire une boule au hasard et on la remplace par une boule de l'autre couleur. La suite (X_n) du nombre

de boules rouges au temps n est une chaîne de Markov. Il y a une force de rappel : plus il y a de boules d'une couleur plus la probabilité que cette couleur diminue à l'étape suivante est grande. Dans le cas où il y a sept boules au total, on peut schématiser la situation de la manière suivante.

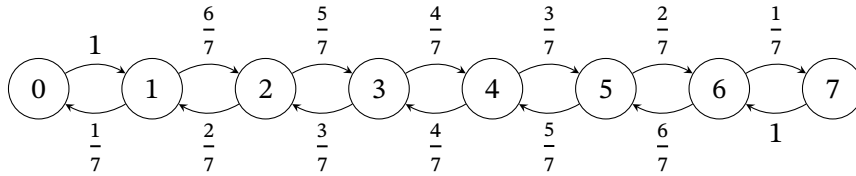


FIGURE 0.1 : Transition d'une urne avec force de rappel

Ici, les flèches représentent la probabilité de passer d'un état à l'autre.

Voici quelques questions auxquelles vous pourrez répondre grâce à ce cours. En partant de $x \in \{0, \dots, 7\}$ boules rouges dans l'urne au temps initial,

- Quelle est la probabilité d'avoir $y \in \{0, \dots, 7\}$ boules rouges après n étapes?
- Cette probabilité dépend-elle de x quand $n \rightarrow \infty$?
- Quelle est la durée moyenne avant qu'il n'y ait plus aucune boule rouge dans l'urne?

0.3.2 LE CAVALIER ERRANT

On considère sur un échiquier standard (8 cases par 8 cases) un cavalier qui peut bouger en L de deux cases dans une direction puis d'une case dans une autre, mais il n'a pas le droit de sortir de l'échiquier.

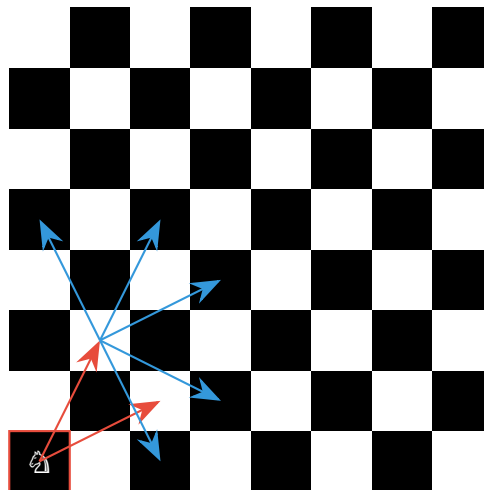


FIGURE 0.2 : Les premiers (en rouge) et deuxièmes (en bleu) coups possibles pour un cavalier

Le cavalier commence sur une case donnée x puis choisi uniformément au hasard un déplacement parmi tous ces déplacements autorisés. Le cavalier va-t-il finir par avoir visité tout l'échiquier? Si oui, en combien de temps? Sinon, quelle case ne peut-il pas atteindre? Après un grand nombre de déplacements quel est la probabilité que le cavalier soit sur une des quatre cases centrales?

0.3.3 MARCHE ALÉATOIRE SUR UN GROUPE ABÉLIEN

Soit $\mathbb{G}$ un groupe abélien, qu'on suppose ici dénombrable pour fixer les idées. On considère une loi de probabilité μ sur $\mathbb{G}$ et on définit $X_0 := x$ pour un certain $x \in \mathbb{G}$ puis $X_{n+1} = X_n + U_{n+1}$ où (U_n) est une suite de variables aléatoires indépendantes de même loi μ . On appelle la suite (X_n) une marche aléatoire sur $\mathbb{G}$ de loi μ . D'autres questions auxquelles nous essayerons de répondre sont par exemple.

- La marche revient-elle en x ? Si oui, au bout de combien de temps?
- Tous les éléments du groupe sont-ils visités par la marche?
- Quelle est la loi de X_n quand n est grand?

0.3.4 MODÈLES DE POPULATION

Soit μ une loi de probabilité sur $\mathbb{N}$. On considère une suite $(U_{n,j})$ de variables aléatoires indépendantes de même loi μ . On définit une chaîne de Markov à valeurs dans $\mathbb{N}$ par $X_0 := x \in \mathbb{N}$ puis

$$X_{n+1} := \sum_{j=1}^{X_n} U_{n,j}.$$

On interprète X_n quand le nombre d'individus d'une certaine génération dans une population et $U_{n,j}$ est le nombre d'enfants de l'individu $j \in \{1, \dots, X_n\}$. Si $X_n = 0$ pour un certain n alors $X_{n+k} = 0$ pour tout $k \in \mathbb{N}$, on dit que la population est *éteinte*. L'étude des chaînes de Markov permet de répondre aux questions suivantes.

- La population s'éteint-elle?
- Si oui, en combien de temps?
- Sinon, quel est le nombre moyen d'individus après de nombreuses générations?

Première partie

Probabilités

Convergence en loi et convergence étroite

L'objectif de cette partie est de munir, sous des conditions topologiques raisonnables sur $\mathbb{X}$, l'espace des mesures de probabilités $\mathcal{P}(\mathbb{X})$ de la distance *bornée Lipschitz* d_{BL} , aussi parfois appelée distance de Fortet–Mourier. Nous verrons que la topologie induite par d_{BL} , appelée *topologie étroite*, caractérise la convergence en loi :

$$\begin{aligned} d_{BL}(\mu_n, \mu) \rightarrow 0 &\Leftrightarrow \forall \varphi \in \mathcal{C}_b(\mathbb{X}), \mathbf{E}[\varphi(X_n)] \xrightarrow{n \rightarrow \infty} \mathbf{E}[\varphi(X)] \\ &\Leftrightarrow \forall \varphi \in \mathcal{C}_b(\mathbb{X}), \int \varphi d\mu_n \xrightarrow{n \rightarrow \infty} \int \varphi d\mu, \end{aligned}$$

avec $\mathbf{loi}[X_n] = \mu_n$ et $\mathbf{loi}[X] = \mu$. Nous énoncerons au théorème 1.38 un critère important de compacité pour la topologie étroite dû à Prokhorov. C'est un critère crucial pour étudier des limites de suites de variables aléatoires. Au passage, nous revoyons et détaillons certaines notions liées à la convergence en loi.

1.1 Rappels

1.1.1 CAS GÉNÉRAL

Lois Pour une transformation mesurable $f : (\mathbb{X}, \mathfrak{X}) \rightarrow (\mathbb{Y}, \mathfrak{Y})$ et une mesure μ sur $\mathbb{X}$, on définit la *mesure image* de μ par f

$$(f_{\#}\mu)(A) := \mu(f^{-1}(A)), \quad \forall A \in \mathfrak{Y}.$$

La loi d'une variable aléatoire X à valeurs dans $\mathbb{X}$ est $X_{\#} \mathbf{P}$. Autrement dit,

$$\mu(B) := \mathbf{P}[X \in B], \quad \forall B \in \mathfrak{X}.$$

Soit (X_n) une suite de variables aléatoires à valeurs dans $\mathbb{X}$. On note (μ_n) les lois respectives.

Inadéquation de la convergence simple La convergence simple des mesures, c'est-à-dire

$$\mu_n(B) \xrightarrow{n \rightarrow \infty} \mu_{\infty}(B), \quad \forall B \in \mathfrak{X},$$

est en général beaucoup trop forte. Par exemple, si les X_n sont à valeurs dans des ensembles discrets, il ne pourra jamais y avoir convergence au sens précédent de (μ_n) vers μ_{∞} absolue continue par rapport à la mesure de Lebesgue. En particulier, on ne pourrait pas obtenir le théorème central de la limite pour des variables de Rademacher !

Convergence en loi Pour disposer d'une notion de convergence plus satisfaisante, nous supposons que $\mathbb{X}$ est un espace topologique et que $\mathfrak{X}$ est la tribu borélienne $\mathfrak{B}(\mathbb{X})$.

Définition 1.1 (Convergence en loi des variables aléatoires). La suite de variables aléatoires (X_n) converge en loi vers variable aléatoire X_∞ si

$$\lim_{n \rightarrow \infty} \mathbf{E}[\varphi(X_n)] = \mathbf{E}[\varphi(X_\infty)], \quad \forall \varphi \in \mathcal{C}_b(\mathbb{X}).$$

Dans ce cas, on note

$$X_n \xrightarrow[n \rightarrow \infty]{\text{loi}} X_\infty.$$

Avec un léger abus de notation, on note parfois

$$X_n \xrightarrow[n \rightarrow \infty]{\text{loi}} \mu_\infty,$$

avec $\mu_\infty := \mathbf{loi}[X_\infty]$.

La convergence en loi ne dépend que des lois (μ_n) . Cela conduit à une notion de convergence pour les mesures.

Définition 1.2 (Convergence étroite pour les mesures de probabilités). La suite de mesures de probabilités $(\mu_n) \in \mathcal{P}(\mathbb{X})$ converge *étroitement* vers $\mu_\infty \in \mathcal{P}(\mathbb{X})$ si

$$\int \varphi d\mu_n \xrightarrow[n \rightarrow \infty]{} \int \varphi d\mu_\infty, \quad \forall \varphi \in \mathcal{C}_b(\mathbb{X}).$$

Dans ce cas on note

$$\mu_n \xrightarrow[n \rightarrow \infty]{\mathcal{P}(\mathbb{X})} \mu_\infty.$$

On a que $(X_n)_{n \in \mathbb{N}}$ converge en loi vers μ_∞ si et seulement si $(\mathbf{loi}[X_n])_{n \in \mathbb{N}}$ converge étroitement vers μ_∞ . Contrairement à la notion de convergence en loi, qui ne permet pas de définir une topologie satisfaisante sur les variables aléatoires, le point de vue de la convergence étroite permet de définir une topologie « sympathique » sur $\mathcal{P}(\mathbb{X})$ ce que nous ferons dans la suite.

Ensembles caractéristiques Tester la convergence contre *toutes* les fonctions continues bornées est parfois trop compliqué, il est donc utile de disposer d'ensembles de fonctions plus petits qui caractérisent la loi.

Définition 1.3 (Ensembles caractéristiques). Soit $\mathcal{F} \subset \mathcal{C}_b(\mathbb{X})$. On dit que $\mathcal{F}$ caractérise la convergence étroite (ou la convergence en loi) si pour tout $\mu_\infty, \mu_0, \mu_1, \dots \in \mathcal{P}(\mathbb{X})$

$$\int \varphi d\mu_n \xrightarrow[n \rightarrow \infty]{} \int \varphi d\mu_\infty, \quad \forall \varphi \in \mathcal{F},$$

implique que (μ_n) converge étroitement vers μ_∞ .

Lemme 1.4. Soit $\mathcal{F} \subset \mathcal{C}_b(\mathbb{X})$ un ensemble qui caractérise la convergence en loi et $\mathcal{F}' \subset \mathcal{F}$ dense dans $\mathcal{F}$ (pour la norme uniforme), alors $\mathcal{F}'$ caractérise aussi la convergence en loi.

Démonstration. Soient $\mu_\infty, \mu_0, \mu_1, \dots \in \mathcal{P}(\mathbb{X})$ et $\varphi \in \mathcal{F}$. Par hypothèse, il existe $(\varphi_k) \subset \mathcal{F}'$ tel que $\|\varphi - \varphi_k\|_\infty \rightarrow 0$. Il suffit d'écrire

$$\begin{aligned} \left| \int \varphi (d\mu_n - d\mu_\infty) \right| &\leq \left| \int (\varphi - \varphi_k) d\mu_n \right| + \left| \int \varphi_k (d\mu_n - d\mu_\infty) \right| + \left| \int (\varphi - \varphi_k) d\mu_\infty \right| \\ &\leq \left| \int \varphi_k (d\mu_n - d\mu_\infty) \right| + 2\|\varphi - \varphi_k\|_\infty. \end{aligned}$$

On prend ensuite $\limsup_{n \rightarrow \infty}$ puis $k \rightarrow \infty$. □

Il ne faut pas se faire d'illusion sur l'utilité pratique de ce résultat dans un cadre abstrait : sauf quand $\mathbb{X}$ est déjà *petit*, c'est-à-dire compact, il n'existe jamais d'ensemble dénombrable dense dans $\mathcal{C}_b(\mathbb{X})$. Sur des exemples concrets il est toutefois très utile.

1.1.2 CAS DES VARIABLES NUMÉRIQUES

À support compact Par le théorème de Stone–Weierstraß, les polynômes sont denses dans $\mathcal{C}(K)$ pour tout compact K de $\mathbb{R}$. Au vu du lemme 1.4, on obtient le résultat suivant.

Corollaire 1.5. *Soit (X_n) une suite de variables aléatoires à valeurs réelles. On suppose qu’il existe un compact $K \subset \mathbb{R}$ tel que pour tout $n \in \mathbb{N}$, $X_n \in K$. Soit μ_∞ une mesure de probabilités sur K . Alors (X_n) converge en loi vers μ_∞ si et seulement si*

$$\mathbf{E}[X_n^p] \xrightarrow{n \rightarrow \infty} \int x^p \mu_\infty(dx).$$

Fonctions caractéristiques et fonctions tests Dans le cas non compact on ne peut plus utiliser les polynômes. Voici toutefois deux ensembles caractérisant plus petit que $\mathcal{C}_b(\mathbb{R}^d)$.

Théorème 1.6. *Les ensembles suivants caractérisent la convergence en loi sur $\mathbb{R}^d$:*

- (i) $\mathcal{C}_c^\infty(\mathbb{R}^d)$, l’ensemble des fonctions lisses à support compact ;
- (ii) l’ensemble $\{x \mapsto e^{ix \cdot t} : t \in \mathbb{R}^d\}$, autrement dit (X_n) converge en loi vers μ_∞ si et seulement si

$$\hat{\mu}_n(t) := \mathbf{E}[e^{it \cdot X_n}] \xrightarrow{n \rightarrow \infty} \int e^{it \cdot x} \mu_\infty(dx) =: \hat{\mu}_\infty(t), \quad \forall t \in \mathbb{R}^d.$$

Démonstration. Le premier point vient de la densité de $\mathcal{C}_c^\infty(\mathbb{R}^d)$ dans $\mathcal{C}_b(\mathbb{R}^d)$. Pour le deuxième point, voir l’exercice 11. □

Remarque 1.7 (Théorème de Lévy). La convergence ponctuelle de la fonction caractéristique n’est pas suffisante pour garantir la convergence. Autrement dit, si (μ_n) est une suite de mesure de probabilité et f une fonction telle que $(\hat{\mu}_n)$ converge ponctuellement vers f , alors on n’a pas forcément la convergence en loi. En effet, il faut s’assurer que f est bien la fonction caractéristique d’une certaine loi. Nous démontrons à la fin de ce chapitre, un théorème de Paul Lévy, qui assure que c’est le cas si et seulement si f est continue en 0.

1.2 Cadre topologique

1.2.1 ESPACES COMPLETS ET SÉPARABLES

Hypothèse et notations

La convergence en loi est définie de manière ad hoc au niveau des variables aléatoires mais n’a aucune chance d’être associée à une topologie sympathique. En effet, deux variables différentes mais de même loi seraient indistinguables du point de vue de la loi et en particulier cette topologie ne serait même pas séparée. Une idée naturelle est de passer de l’espace des variables aléatoires à l’espace des lois, c’est-à-dire l’ensemble des mesures de probabilité sur $\mathbb{X}$. Puisque que la convergence en loi s’obtient en testant contre des fonctions continues (et bornées), cette convergence est sensible à la topologie de $\mathbb{X}$. On pourra retenir comme principe que la topologie sur $\mathcal{P}(\mathbb{X})$ hérite des propriétés de celle sur $\mathbb{X}$. On va supposer que $\mathbb{X}$ lui-même est muni d’une topologie sympathique.

Hypothèse 1. $\mathbb{X}$ est un espace topologique séparable et sa topologie est métrisable par une distance d telle que $(\mathbb{X}, d)$ est complet. Dans ce cas on dit que $\mathbb{X}$ est un espace polonais.

Dans la suite $\mathbb{X}$ est toujours équipé de sa tribu Borélienne $\mathfrak{B}(\mathbb{X})$. On note $\mathfrak{K}(\mathbb{X})$ l'ensemble des compacts de $\mathbb{X}$ et $\mathfrak{O}(\mathbb{X})$ l'ensemble des ouverts. On désigne par $\mathcal{C}_b(\mathbb{X})$ l'ensemble des fonctions continues et bornées de $\mathbb{X}$ dans $\mathbb{R}$, $\mathcal{P}(\mathbb{X})$ est l'ensemble des mesures de probabilités boréliennes sur $\mathbb{X}$.

Remarque 1.8 (On peut toujours prendre une distance bornée). Soit d une distance qui métrise complètement la topologie de $\mathbb{X}$, alors $d' := d \wedge 1$, est aussi une distance qui métrise complètement la topologie. On peut donc toujours supposer que la distance est majorée par 1.

Remarque 1.9 (La question de la distance). Avec la définition donnée, on prendra garde de penser qu'il existe une distance canonique associée à $\mathbb{X}$. Par exemple, l'espace $\mathbb{X} :=]-\frac{\pi}{2}, \frac{\pi}{2}[$ muni de la distance usuel n'est pas complet. Par contre, il est complet pour la distance

$$d(x, y) := |\tan(x) - \tan(y)|, \quad \forall x, y \in \mathbb{X},$$

et la distance d induit bien la topologie usuelle sur $\mathbb{X}$, qui est donc polonais.

Les espaces polonais sont stables par de nombreuses opérations.

Théorème 1.10. *Les espaces suivants sont polonais.*

- (i) *Les espaces homéomorphes à un espace polonais.*
- (ii) *Les fermés et les ouverts d'un espace polonais.*
- (iii) *Les intersections, les unions et les produits dénombrables d'espaces polonais.*

Démonstration. On se contentera de donner une expression de la distance dans chacun des cas. On laissera au lecteur le soin de vérifier la séparabilité et que les expressions donnent bien des distances qui métrisent complètement la topologie. Dans toute la preuve $(\mathbb{X}, d)$, $(\mathbb{X}_n, d_n)$ sont des espaces polonais munis d'une distance complète qui métrise leur topologie respective.

(i) Soient $\mathbb{Y}$ un espace topologique et $f : \mathbb{X} \rightarrow \mathbb{Y}$ un homéomorphisme. Alors on peut prendre

$$\rho(y_1, y_2) := d(f^{-1}(y_1), f^{-1}(y_2)), \quad \forall y_1, y_2 \in \mathbb{Y}.$$

(ii) Pour $F \subset \mathbb{X}$ fermé, alors d convient. Pour $U \subset \mathbb{X}$ ouvert, alors on définit

$$\rho(x, y) := d(x, y) + \frac{|\delta(x) - \delta(y)|}{|\delta(x) - \delta(y)| + 1},$$

avec

$$\delta(x) := \inf\{d(x, y) : y \in \mathbb{X} \setminus U\}, \quad \forall x \in \mathbb{X}.$$

(iii) Pour le produit on peut considérer

$$\rho(x, \tilde{x}) := \sum_{n \in \mathbb{N}} 2^{-n} d_n(x_n, \tilde{x}_n), \quad \forall x, \tilde{x} \in \mathbb{X}^{\mathbb{N}}.$$

Pour l'union, en supposant sans perte de généralité que $d_n \leq 1$ pour tout $n \in \mathbb{N}$, il suffit de poser

$$\rho(x, \tilde{x}) := \begin{cases} d_n(x, \tilde{x}), & \text{si } \exists n \in \mathbb{N} : x, \tilde{x} \in X_n, \\ 1, & \text{sinon,} \end{cases} \quad \forall x, \tilde{x} \in \bigcup_{n \in \mathbb{N}} X_n.$$

Pour l'intersection, on considère

$$\rho(x, \tilde{x}) := \sum_{n \in \mathbb{N}} 2^{-n} d_n(x, \tilde{x}), \quad \forall x, \tilde{x} \in \bigcap_{n \in \mathbb{N}} X_n.$$

□

En fait on a une caractérisation précise des espaces polonais.

Théorème 1.11. *Un espace $\mathbb{X}$ est polonais si et seulement s'il est homéomorphe à une intersection dénombrable d'ensembles ouverts de $[0, 1]^{\mathbb{N}}$.*

Démonstration. C'est suffisant par le théorème 1.10. Montrons que c'est nécessaire. Comme d'habitude, fixons une distance $d \leq 1$ qui métrise complètement la topologie de $\mathbb{X}$. Fixons une suite (x_n) dense dans $\mathbb{X}$. Alors,

$$f : \mathbb{X} \ni x \mapsto (d(x, x_n))_n \in [0, 1]^{\mathbb{N}},$$

est un homéomorphisme (exercice!). Par le (i) du théorème 1.10, supposons sans perte de généralité que $\mathbb{X} \subset [0, 1]^{\mathbb{N}}$. Posons $\tilde{\mathbb{X}}$ l'adhérence de $\mathbb{X}$ dans $[0, 1]^{\mathbb{N}}$. Définissons, pour $n \in \mathbb{N}^*$, l'ensemble U_n des points $y \in \tilde{\mathbb{X}}$ qui possèdent un voisinage ouvert V , dans $\tilde{\mathbb{X}}$, tel que

$$\sup\{d(z, z') : z, z' \in V \cap \mathbb{X}\} \leq \frac{1}{n}. \quad (1.1)$$

Alors, pour tout $n \in \mathbb{N}^*$, U_n est un ouvert de $\tilde{\mathbb{X}}$ qui contient $\mathbb{X}$. Soit maintenant $y \in \cap U_n$. Par construction, y est adhérent à $\mathbb{X}$ dans $[0, 1]^{\mathbb{N}}$. On peut donc trouver une suite (y_n) d'éléments de $\mathbb{X}$ tel que $y_n \rightarrow y$ dans $[0, 1]^{\mathbb{N}}$. Pour tout $n \in \mathbb{N}^*$, il existe un voisinage V_n de y dans $[0, 1]^{\mathbb{N}}$ tel que V_n satisfait (1.1). Quitte à considérer des intersections finies, supposons que les (V_n) sont décroissants; quitte à extraire, supposons que $y_m \in V_n$ pour tout $m \geq n$. La suite (y_n) est donc de Cauchy pour d . Puisque $(\mathbb{X}, d)$ est complet, on trouve que $y \in \mathbb{X}$. Nous avons montré que $\mathbb{X} = \cap U_n$. Notons que les (U_n) sont ouverts dans $\tilde{\mathbb{X}}$, mais a priori pas dans $[0, 1]^{\mathbb{N}}$. Toutefois, d'une part, il existe des ouverts $(\tilde{U}_n)$ dans $[0, 1]^{\mathbb{N}}$ tel que

$$U_n = \tilde{U}_n \cap \tilde{\mathbb{X}}, \quad \forall n \in \mathbb{N}^*.$$

D'autre, dans un espace métrique, tout fermé est intersection dénombrable d'ouverts. D'où des ouverts (W_m) de $[0, 1]^{\mathbb{N}}$ tels que $\tilde{\mathbb{X}} = \cap W_m$. Finalement $\mathbb{X} = \cap (\tilde{U}_n \cap W_m)$. $\square$

Pour aller plus loin 1 (Espaces de Cantor et hypothèse du continu). Le théorème 1.11 montre que le *cube de Hilbert* $[0, 1]^{\mathbb{N}}$ est l'objet final de la catégorie des espaces polonais (où les flèches sont les homéomorphismes). On a un résultat suivant qui montre que l'*espace de Cantor* $\{0, 1\}^{\mathbb{N}}$ est l'objet initial des espaces polonais non dénombrables.

Théorème 1.12. *Tout espace polonais non dénombrable contient, à homéomorphisme près, une copie de $\{0, 1\}^{\mathbb{N}}$.*

Démonstration. Comme d'habitude on va supposer que $\mathbb{X}$ est équipé d'une métrique complète de diamètre ≤ 1 . Puisque $\mathbb{X}$ est métrique et séparable, il existe une base dénombrable $(V_n)_{n \in \mathbb{N}}$ de la topologie de $\mathbb{X}$. Posons

$$N := \{n \in \mathbb{N} : V_n \text{ est dénombrable}\}.$$

$$\mathbb{X}^* := \mathbb{X} \setminus \bigcup_{n \in N} V_n.$$

Maintenant, je prétends (exercice!) que

$$\mathbb{X}^* = \{x \in \mathbb{X} : \text{tout voisinage de } x \text{ est non dénombrable}\}.$$

De plus, $\mathbb{X}^*$ est fermé (complémentaire d'une union d'ouverts), donc (théorème 1.10 (ii)) polonais, et non dénombrable, car $\mathbb{X}$ est non dénombrable et $\bigcup_{n \in N} V_n$ est dénombrable. On va

montrer que $\mathbb{X}^*$ contient une copie homéomorphique de $\{0, 1\}^{\mathbb{N}}$. C'est le schéma usuel de Cantor. Pour $s \in \bigcup_{n \in \mathbb{N}} \{0, 1\}^n$, on définit récursivement les points $x_s \in \mathbb{X}$ et les nombres réels $r_s \in]0, 2^{-|s|}]$. On notera $F_s := \bar{B}(x_s, r_s)$. On choisit arbitrairement $x_\emptyset \in \mathbb{X}$ et $r_\emptyset \in]0, 1]$. Supposons avoir construit les x_s et les r_s pour tout $|s| \leq n$. Pour un certain $|s| = n$, $B(x_s, r_s) \cap \mathbb{X}^*$ est un ouvert non-vide de $\mathbb{X}^*$ donc contient au moins deux (en fait une infinité non dénombrable) de points, disons $x_{s_0} \neq x_{s_1}$. On peut trouver r_{s_0} et $r_{s_1} \in]0, 2^{-|s|-1}]$ tels que

$$\begin{aligned} F_{s_0} \cap F_{s_1} &= \emptyset, \quad \text{et} \\ F_{s_0} \cup F_{s_1} &\subset B(x_s, r_s). \end{aligned}$$

Pour $s \in \{0, 1\}^{\mathbb{N}}$ et $n \in \mathbb{N}$, on écrit $s_{\upharpoonright n} \in \{0, 1\}^n$ pour la restriction de s aux n premiers bits. Alors $(F_{s_{\upharpoonright n}})_n$ est une suite décroissante de fermés non-vides dont le diamètre tend vers 0. Ainsi $\bigcap_{n \in \mathbb{N}} F_{s_{\upharpoonright n}}$ est réduit à un point, disons $f(s)$. La fonction f est injective (exercice!) et continue : si $s_{\upharpoonright n} = t_{\upharpoonright n}$, alors $d(f(s), f(t)) \leq 2^{-n+1}$. $\square$

Puisque $\{0, 1\}^{\mathbb{N}}$ et $[0, 1]^{\mathbb{N}}$ ont tous les deux la puissance du continu, on voit que les espaces polonais non dénombrables ont tous la puissance du continu. En particulier, l'hypothèse du continu est vraie pour les espaces polonais.

Pour aller plus loin 2 (Espace de Borel standard). L'encadré précédent montre que pour un espace polonais $\mathbb{X}$,

$$\text{card}(\mathbb{X}) \in \mathbb{N} \cup \{\aleph_0\} \cup \{\aleph_1\},$$

avec possiblement de nombreux espaces non homéomorphes pour chaque cardinal. Si l'on remplace l'équivalence des espaces topologiques (homéomorphisme) par l'équivalence des espaces mesurables (bijection mesurable d'inverse mesurable), alors on obtient une classification surprenante : il n'existe qu'un seul représentant par cardinal.

Théorème 1.13 (Kuratowski). *Soient $\mathbb{X}$ et $\mathbb{Y}$ des espaces polonais. Alors il existe une bijection mesurable d'inverse mesurable si et seulement si $\text{card } \mathbb{X} = \text{card } \mathbb{Y}$.*

Démonstration. Si les espaces n'ont pas le même cardinal, alors il ne peut pas exister de bijection entre eux. S'ils ont le même cardinal, on discute suivant deux cas. Dans le cas d'un cardinal dénombrable (fini ou infini), la tribu borélienne d'un espace polonais dénombrable coïncide avec grossière (exercice!). Ainsi, n'importe quelle énumération de $\mathbb{X}$ convient. Dans le cas d'un cardinal non dénombrable, on sait que $\text{card } \mathbb{X} = \aleph_0$ et que

$$\{0, 1\}^{\mathbb{N}} \hookrightarrow \mathbb{X} \hookrightarrow [0, 1]^{\mathbb{N}},$$

où les injections sont continues, donc mesurables. Pour conclure il suffit de montrer qu'il existe une injection mesurable $[0, 1]^{\mathbb{N}} \hookrightarrow \{0, 1\}^{\mathbb{N}}$. Pour cela, il suffit de construire une injection mesurable $\varphi : [0, 1] \hookrightarrow \{0, 1\}^{\mathbb{N}}$. Le produit tensoriel $\varphi^{\otimes \mathbb{N}}$ donne alors une injection

$$[0, 1]^{\mathbb{N}} \hookrightarrow (\{0, 1\}^{\mathbb{N}})^{\mathbb{N}} \simeq \{0, 1\}^{\mathbb{N}}.$$

La dernière identification venant de l'équivalence mesurable $\mathbb{N}^2 \simeq \mathbb{N}$. Maintenant pour φ il suffit de prendre la décomposition dyadique usuelle. La mesurabilité est facile et laissée en exercice. $\square$

Théorème 1.14 (Théorème de Baire pour les espaces polonais). *Dans un espace polonais, toute intersection dénombrable d'ouverts denses est encore dense.*

Remarque 1.15. Par passage au complémentaire, dans un espace polonais, toute union dénombrable de fermés d'intérieur vide, est encore d'intérieur vide.

Démonstration. En fait nous allons montrer la propriété pour un espace complètement métrisable sans utiliser la séparabilité. Comme d'habitude, on va fixer d une distance ≤ 1 qui métrise complètement la topologie. Soit (U_n) une suite d'ouverts denses et soit V un ouvert quelconque. En posant $U := \cap_n U_n$ on veut montrer que $U \cap V$ est non vide. On construit par récurrence une suite de boules fermées (F_n) de rayon $\leq (n+1)^{-1}$ telle que

$$\cap_n F_n \subset U \cap V.$$

Or par le théorème des fermés emboîtés (l'espace est complet) l'intersection est non vide. Par densité, l'ouvert $U_0 \cap V$ est non vide et contient donc une boule fermée F_0 de rayon ≤ 1 . En supposant F_n construit, on utilise que $U_{n+1} \cap \text{int}(F_n)$ est également un ouvert non vide et donc contient une boule fermée. $\square$

Pour aller plus loin 3 (Espaces de Baire). Un espace topologique est dit *de Baire*, s'il a la propriété du théorème, c'est-à-dire si toute union dénombrable d'ouverts denses est dense. Nous avons montré que les espaces polonais, et plus généralement les espaces complètement métrisables, sont de Baire. La même preuve, en remplaçant les fermés par des compacts, montre que les espaces localement compacts sont également de Baire.

Exemples et contre-exemples

De nombreux espaces rencontrés en analyse sont polonais.

Espaces discrets Commençons par le cas de la topologie discrète, c'est-à-dire celle induite par la distance $d(x, y) = 1_{x \neq y}$.

Proposition 1.16. *Un espace muni de la topologie discrète est polonais si et seulement s'il est dénombrable.*

Démonstration. Un espace dénombrable muni de la topologie discrète est homéomorphe à $\mathbb{N}$. D'où (théorème 1.10), pour traiter le cas dénombrable, il suffit de montrer que $\mathbb{N}$ muni de sa topologie est polonais. Or, $\mathbb{N}$ est un fermé de $\mathbb{R}$ qui est polonais. Inversement, un espace non dénombrable muni de la topologie discrète n'est jamais séparable, donc jamais polonais. $\square$

Espaces dénombrables Un espace dénombrable muni de la topologie discrète est toujours polonais. Quand on considère des espaces dénombrables munis d'une topologie induites, ils peuvent être ou ne pas être polonais. Par exemple, l'espace

$$\{0\} \cup \bigcup_{n \in \mathbb{N}^*} \{n^{-1}\},$$

muni de la topologie induite de $\mathbb{R}$ est un fermé de $\mathbb{R}$ donc polonais (théorème 1.10). Par contre, $\mathbb{Q}$ muni de la topologie induite de $\mathbb{R}$ n'est pas polonais. En effet, on peut écrire

$$\mathbb{Q} = \bigcup_{q \in \mathbb{Q}} \{q\}.$$

Donc c'est un ensemble maigre, c'est-à-dire réunion dénombrable de fermés d'intérieur vide. Or, par le théorème de Baire, un ensemble polonais ne peut pas être maigre dans lui-même. Notons que $\mathbb{R} \setminus \mathbb{Q}$ est polonais, bien qu'il ne soit pas complet pour la distance usuelle, car il est homéomorphe à $\mathbb{N}^{\mathbb{N}}$ par identification avec une fraction continue infinie.

Espaces propres Commençons avec le cas des espaces à boules fermées compactes.

Lemme 1.17. *Soit (X, d) un espace métrique avec la propriété de Heine–Borel : les boules fermées sont compactes, Alors (X, d) est complet et séparable. En particulier, l'espace topologique X est polonais.*

Démonstration. Pour la complétude, considérons une suite de Cauchy. La suite est en particulier bornée et, par hypothèse, elle est relativement compacte. Elle admet donc au moins un point d'adhérence, mais comme les suites de Cauchy ont au plus un point d'adhérence, on conclut que la suite converge.

Pour la séparabilité, montrons d'abord qu'une boule fermée B est séparable. Comme B est compacte et que nous travaillons dans un espace métrique, elle est aussi précompacte. En particulier pour tout $n \in \mathbb{N}^*$, il existe une famille finie $F_n \subset X$ telle que

$$B \subset \bigcup_{x \in F_n} B(x, 1/n).$$

Maintenant $\bigcup_{n \in \mathbb{N}^*} F_n$ est une famille dénombrable dense. Pour conclure, on écrit $X = \bigcup_n B(x_0, n)$ pour un $x_0 \in X$. $\square$

Comme application de ce critère, donnons quelques exemples classiques d'espaces qui satisfont nos hypothèses. Dans tous les exemples suivants, les espaces ont des boules fermées compactes. Donc, par le lemme 1.17, ils satisfont l'hypothèse 1.

- Exemple 1** (Cas de la dimension finie). (a) Les espaces euclidiens et les espaces projectives munis de leur distance usuelle.
- (b) Les sphères et les espaces hyperboliques munis de leur distance géodésique.
- (c) Les variétés riemanniennes compactes munies de la distance riemannienne.
- (d) Les groupes de Lie connexes et de dimension finie équipés de leur distance invariante à gauche canonique.

Exemple 2 (Espaces de Montel). On se gardera bien de penser que la compacité des boules fermées est un phénomène fini-dimensionnel. Ce que l'on sait, c'est qu'un espace de Banach de dimension fini n'a jamais de boule fermée compacte, mais il existe d'autres espaces vectoriels topologiques qui ont des boules fermées compactes. On appelle de tels espaces des espaces de *Montel*. Bien sûr leur topologie ne peut jamais être induite par une norme. Dans tous les exemples suivants ce sont des espaces de Fréchet : la topologie est induite par une famille dénombrable de semi-normes, à partir de laquelle on peut construire une distance.

- (d) Soit $\mathcal{H}$ l'espace des fonctions holomorphes sur $\mathbb{C}$ considérons la distance

$$d_{\mathcal{H}}(f, g) := \sum_{n \in \mathbb{N}} 2^{-n} \sup_{|z| \leq n} |f(z) - g(z)|,$$

Cette distance induit la topologie de la convergence uniforme sur les compacts. Par le théorème de Montel, les fermés bornés pour cette distance sont compacts. Ce résultat se généralise aux fonctions holomorphes sur un domaine ouvert $U \subset \mathbb{C}$ à condition de modifier correctement.

- (e) De manière similaire, munissons l'espace $\mathcal{D}$ des fonctions infiniment dérivables à support compact sur $\mathbb{R}$ de la distance

$$d_{\mathcal{D}}(f, g) := \sum_{k, n \in \mathbb{N}} 2^{-n-k} \sup_{|x| \leq n} |f^{(k)}(x) - g^{(k)}(x)|,$$

qui induit la topologie la convergence uniforme sur les compacts pour toutes les dérivées. Le théorème d'Ascoli permet de montrer que les boules fermées sont compactes.

- (f) L'espace des distributions, c'est-à-dire le dual $\mathcal{D}'$ de l'espace des fonctions $\mathcal{D}$ à support compact muni de la topologie précédente. Plus précisément, posons

$$d_{\mathcal{D}'}(T, S) := \sum_{n \in \mathbb{N}} 2^{-n} \sup\{|T(f) - S(f)| : f \in B_{\mathcal{D}}(0, n)\},$$

avec $B_{\mathcal{D}}(0, n)$ la boule de centre 0 et de rayon n pour la distance $d_{\mathcal{D}}$ précédemment définie. Alors, cette distance engendre bien la topologie duale et on peut montrer qu'elle est complète et que les bornés sont compacts. Voir, par exemple, [Sch65, III § 1 Thms. VIII & XII].

Pour aller plus loin 4 (Espaces métriques localement compacts). Le cas des boules fermées compactes est plus restrictif que la compacité locale. Même s'il permet de couvrir la plupart des exemples classiques, nous présentons tout de même ce cas-ci. Un espace métrique $\mathbb{X}$ est localement compact si pour tout $x \in \mathbb{X}$, il existe $r_x > 0$ tel que la boule fermée $\bar{B}(x, r_x)$ soit compacte. Dans ce cas, le même argument que dans le théorème donne la complétude. Par contre, l'argument de séparabilité ne marche plus. Par exemple tout espace non dénombrable muni de la distance discrète est localement compact, mais n'est pas séparable. Plus généralement, dans la démonstration nous avons utilisé que *toutes* les boules $\bar{B}(x_0, n)$ sont séparables, alors que la compacité nous donne une seule boule. Si on suppose de plus que $\mathbb{X}$ est σ -compact, c'est-à-dire s'écrit comme une union dénombrable de compacts, alors on peut faire de nouveau marcher la preuve de la séparabilité. Un critère classique (exercice!) est le suivant : si $\mathbb{X}$ est métrisable, localement compact et connexe, alors $\mathbb{X}$ est σ -compact et donc polonais.

Espaces de Banach Rappelons que dans un espace normé de dimension infinie, les boules fermées ne sont jamais compactes pour la topologie induite par la norme. Il existe toutefois de nombreux exemples d'espaces de Banach séparables. Puisque, par définition, un espace de Banach est complet, les espaces de Banach séparables sont des espaces polonais.

Exemple 3 (Espaces de Banach séparables). Les espaces suivants sont des espaces de Banach séparables.

- (a) Quand $1 \leq p < \infty$, les espaces L^p par rapport à mesure sigma finie sur une tribu engendré par une famille dénombrable. En particulier, les espaces de fonctions $L^p(\mathbb{R}^d, dx)$ et les espaces de suites ℓ^p .
- (b) Tous les espaces de Hilbert qui admettent une base dénombrable.
- (c) L'espace des fonctions continues de $[0, 1]$ dans $\mathbb{R}$ muni de la norme uniforme.

Contre-exemples Finissons avec les contre-exemples classiques.

Exemple 4 (Espaces de Banach non séparables). Les espaces suivants sont des espaces de Banach *non* séparables.

- (a) L'espace de Banach des fonctions continues bornées sur $\mathbb{R}$ muni de la norme uniforme n'est pas séparable.
- (b) Les espaces de Banach ℓ^∞ et $L^\infty(\mathbb{R}, dx)$ ne sont pas séparables.

1.2.2 FONCTIONS SEMI-CONTINUES

Les fonctions continues jouent un rôle important dans la définition de la convergence étroite. Nous relaxons ici la notion de continuité et travaillons avec une classe plus générale : les fonctions semi-continues. Celles-ci se sont stables par passage au supremum, opération naturelle en théorie de la mesure.

Fonction semi-continue inférieurement : définition et exemples

Définition 1.18 (Semi-continuité inférieure). Une fonction $\varphi : \mathbb{X} \rightarrow [-\infty, +\infty]$ est *semi-continue inférieurement* si son épigraphe est fermé, où

$$\text{epi}(\varphi) := \{(x, t) \in \mathbb{X} \times \mathbb{R} : \varphi(x) \leq t\}.$$

Exemple 5. Étudions la semi-continuité inférieure de certaines fonctions.

- (i) La fonction $\varphi = -\infty$ est semi-continue inférieurement car $\text{epi}(\varphi) = \mathbb{X} \times \mathbb{R}$.
- (ii) La fonction $\varphi = +\infty$ est semi-continue inférieurement car $\text{epi}(\varphi) = \emptyset$.
- (iii) Si φ est à valeurs réelles et continue, alors φ est semi-continue inférieurement.
- (iv) Pour $B \in \mathfrak{B}(\mathbb{X})$, on a $\text{epi}(1_B) = B \times \{1 \leq t < \infty\} \cup B^C \times \mathbb{R}_+$ qui est fermé si et seulement si $B \in \mathfrak{D}(\mathbb{X})$.
- (v) Pour $B \in \mathfrak{B}(\mathbb{X})$, on pose $\iota_B = +\infty 1_{B^C}$, alors $\text{epi}(\iota_B) = B \times \mathbb{R}_+$ qui est fermé si et seulement si B est fermé.
- (vi) Le supremum d'une famille quelconque de fonctions semi-continues inférieurement est encore semi-continue inférieurement car

$$\text{epi}(\sup_{i \in I} \varphi_i) = \bigcap_{i \in I} \text{epi}(\varphi_i),$$

qui est fermé comme intersection quelconque de fermés.

- (vii) Le minimum d'une famille finie de fonctions semi-continues inférieurement est encore semi-continue inférieurement car

$$\text{epi}(\min_{i=1}^l \varphi_i) = \bigcup_{i=1}^l \text{epi}(\varphi_i),$$

qui est fermé comme union finie de fermés.

Caractérisation de la semi-continuité inférieure

Comme pour la continuité, on peut caractériser la semi-continuité inférieure de manière séquentielle ou en termes de graphe. Pour une fonction φ on définit

$$\begin{aligned} \liminf_{y \rightarrow x} \varphi(y) &:= \lim_{\varepsilon \rightarrow 0} \inf\{\varphi(y) : y \in B(x, \varepsilon)\} = \inf\left\{\liminf_{n \rightarrow \infty} \varphi(x_n) : x_n \rightarrow x\right\}; \\ \limsup_{y \rightarrow x} \varphi(y) &:= \lim_{\varepsilon \rightarrow 0} \sup\{\varphi(y) : y \in B(x, \varepsilon)\} = \sup\left\{\limsup_{n \rightarrow \infty} \varphi(x_n) : x_n \rightarrow x\right\}. \end{aligned}$$

Par définition, on a toujours

$$\liminf_{y \rightarrow x} \varphi(y) \leq \varphi(x) \leq \limsup_{y \rightarrow x} \varphi(y).$$

On a égalité dans les inégalités précédentes si et seulement si φ est continue si et seulement si

$$\limsup_{y \rightarrow x} \varphi(y) \leq \varphi(x) \leq \liminf_{y \rightarrow x} \varphi(y).$$

Nous montrons maintenant que φ est semi-continue inférieurement si et seulement si la dernière inégalité est vraie, c'est-à-dire que la moitié de l'égalité de continuité est vraie (d'où le nom).

Lemme 1.19. Soit $\varphi : \mathbb{X} \rightarrow [-\infty, +\infty]$. Il y a équivalence entre :

- (i) φ est semi-continue inférieurement.
- (ii) Pour tout $t \in \mathbb{R}$, le sous-ensemble de niveau suivant est fermé :

$$\{x \in \mathbb{X} : \varphi(x) \leq t\},$$

- (iii) Pour tout $x \in \mathbb{X}$ tel que $\varphi(x) < \infty$,

$$\liminf_{y \rightarrow x} \varphi(y) \geq \varphi(x).$$

Si de plus φ est minorée ces conditions sont aussi équivalentes à

- (iv) φ est la limite monotone d'une suite croissante de fonctions Lipschitz et bornées.
- (v) φ est un supremum de fonctions continues et bornées. Plus précisément,

$$\varphi = \sup\{\psi : \psi \in \mathcal{C}_b(\mathbb{X}), \psi \leq \varphi\}.$$

Démonstration. (i) $\Rightarrow$ (ii). Immédiat.

(ii) $\Rightarrow$ (iii). Nous montrons la caractérisation séquentielle de la limite inférieure. Soit (x_n) une suite qui converge vers $x \in \mathbb{X}$. Quitte à extraire, supposons que $\varphi(x_n) \rightarrow \liminf_{n \rightarrow \infty} \varphi(x_n) =: \ell$. Pour tout $t > \ell$, pour n assez grand $\varphi(x_n) \leq t$, donc comme l'ensemble de sous-niveau est fermé $\varphi(x) \leq t$, et finalement $\varphi(x) \leq \ell$.

(iii) $\Rightarrow$ (i). Nous montrons que l'épigraphe est séquentiellement fermé, ce qui suffit puisque $\mathbb{X}$ est polonais. Soit (x_n, t_n) une suite d'éléments de l'épigraphe qui converge vers (x, t) . Par l'hypothèse sur la limite inférieure, on trouve

$$t = \liminf_{n \rightarrow \infty} t_n \geq \liminf_{n \rightarrow \infty} \varphi(x_n) \geq \varphi(x).$$

(iii) $\Rightarrow$ (iv). Sans perte de généralité, supposons $\varphi \geq 0$ et qu'il existe $x_0 \in \mathbb{X}$ tel que $\varphi(x_0) < \infty$. Posons

$$\varphi_k(x) := \inf\{\varphi(y) + kd(x, y) : y \in \mathbb{X}\}, \quad \forall k \in \mathbb{N}, x \in \mathbb{X}.$$

Par construction (φ_k) est une suite croissante. De plus $\varphi_0 = \inf \varphi \geq 0$. Pour tout x, y et $z \in \mathbb{X}$,

$$\varphi_k(x) - \varphi_k(y) \leq \varphi(z) + kd(x, y) - \varphi_k(y) \leq \varphi(z) + kd(y, z) + kd(x, y) - \varphi_k(y).$$

En prenant l'infimum sur $z \in \mathbb{X}$ on trouve $\varphi_k(x) - \varphi_k(y) \leq kd(x, y)$, puis par un argument symétrique que φ_k est k -Lipschitz donc en particulier continue. On a alors

$$\varphi = \sup_{k \in \mathbb{N}} \varphi_k \wedge k.$$

(iv) $\Rightarrow$ (v). Immédiat.

(v) $\Rightarrow$ (i). C'est vrai même si φ n'est pas minorée : un supremum de fonctions semi-continues inférieurement est toujours semi-continue inférieurement et les fonctions continues sont semi-continues inférieurement. $\square$

Fonction semi-continue supérieurement

Définition 1.20 (Semi-continuité supérieure). Une fonction $\varphi : \mathbb{X} \rightarrow [-\infty, +\infty]$ est semi-continue supérieurement si $-\varphi$ est semi-continue inférieurement.

Nous laissons en exercice d'adapter les résultats précédents au cas semi-continue supérieurement.

Approximation d'intégrales

Pour illustrer le l'intérêt des fonctions semi-continues inférieurement, finissons par ce résultat d'approximation.

Lemme 1.21. Soient $\mu \in \mathcal{P}(\mathbb{X})$ et φ une fonction semi-continue inférieurement positive. Alors,

$$\int \varphi d\mu = \sup \left\{ \int h d\mu : h \in \mathcal{C}_b(\mathbb{X}), 0 \leq h \leq \varphi \right\}.$$

Démonstration. C'est le théorème de convergence monotone avec le lemme 1.19 (iv). □

1.3 Propriétés de la convergence étroite et de la convergence en loi**1.3.1 LE LEMME DU PORTEMANTEAU**

Passons maintenant à plusieurs reformulations de la convergence étroite.

Théorème 1.22 (Portemanteau). Soient (μ_n) une suite dans $\mathcal{P}(\mathbb{X})$ et $\mu_\infty \in \mathcal{P}(\mathbb{X})$. Il y a équivalence entre :

- (i) (μ_n) converge étroitement vers μ .
- (ii) Pour toute fonction semi-continue inférieurement et minorée φ ,

$$\liminf_{n \rightarrow \infty} \int \varphi d\mu_n \geq \int \varphi d\mu.$$

- (iii) Pour toute fonction semi-continue supérieurement et majorée φ ,

$$\limsup_{n \rightarrow \infty} \int \varphi d\mu_n \leq \int \varphi d\mu.$$

- (iv) Pour tout ouvert U de $\mathbb{X}$,

$$\liminf_{n \rightarrow \infty} \mu_n(U) \geq \mu(U).$$

- (v) Pour tout fermé F de $\mathbb{X}$,

$$\limsup_{n \rightarrow \infty} \mu_n(F) \leq \mu(F).$$

- (vi) Pour tout borélien A de $\mathbb{X}$ tel que $\mu(\partial A) = 0$,

$$\lim_{n \rightarrow \infty} \mu_n(A) = \mu(A).$$

Démonstration. (i) $\Rightarrow$ (ii) Soit φ semi-continue inférieurement et minorée. Le lemme 1.19 donne alors une suite (φ_k) de fonctions continues et bornées telles que $\varphi = \sup_k \varphi_k$. Par convergence monotone,

$$\begin{aligned} \liminf_{n \rightarrow \infty} \int \varphi d\mu_n &= \liminf_{n \rightarrow \infty} \sup_k \int \varphi_k d\mu_n \\ &\geq \sup_k \liminf_{n \rightarrow \infty} \int \varphi_k d\mu_n \\ &\geq \sup_k \int \varphi_k d\mu_\infty = \int \varphi d\mu. \end{aligned}$$

Donc $\mathcal{H}$ est stable par convergence monotone bornée et on peut appliquer le théorème B.22.

(ii) $\Leftrightarrow$ (iii) Immédiat par la transformation φ devient $-\varphi$.

(ii) $\Rightarrow$ (iv) Immédiat : les indicatrices de fermés sont semi-continues supérieurement et majorées.

(iv) $\Leftrightarrow$ (v) Immédiat par passage au complémentaire.

(iv) et (v) $\Rightarrow$ (vi) On a alors $\mu(A) = \mu(\mathring{A}) = \mu(\bar{A})$ et donc $\liminf_n \mu_n(\mathring{A}) \geq \limsup_n \mu_n(\bar{A})$. Comme $\mathring{A} \subset \bar{A}$, on doit avoir égalité.

(vi) $\Rightarrow$ (i) Soit $\varphi \in \mathcal{C}_b(\mathbb{X})$. Quitte à décomposer en partie positive et négative, on peut supposer que $\varphi \geq 0$. Notons aussi $M := \|\varphi\|_\infty$. Comme $\varphi \geq 0$, on a

$$\varphi(x) = \int_0^{\varphi(x)} dt = \int_0^\infty 1_{t \leq \varphi(x)} dt = \int_0^M 1_{t \leq \varphi(x)} dt.$$

Par le théorème de Fubini–Tonelli, on trouve

$$\int \varphi d\mu_n = \int_0^M \mu_n(\varphi \geq t) dt.$$

Montrons maintenant que l'ensemble

$$T := \{t \in \mathbb{R} : \mu(\varphi = t) > 0\},$$

est dénombrable. Pour cela, supposons par l'absurde que T n'est pas dénombrable. On écrit $T = \bigcup_{n \in \mathbb{N}} T_n$ avec

$$T_n := \left\{t \in \mathbb{R} : \mu(\varphi = t) \geq \frac{1}{n}\right\}, \quad \forall n \in \mathbb{N}^*.$$

Comme T n'est pas dénombrable, il existe $N \in \mathbb{N}^*$ tel que T_N contient une infinité d'éléments. Prenons $t_0, \dots, t_N \in T_N$ tous distincts, alors

$$\sum_{i=0}^N \mu(\varphi = t_i) \geq 1 + \frac{1}{N}.$$

Absurde.

Puisque φ est continue, $\partial\{\varphi \geq t\} \subset \{\varphi = t\}$, et on trouve que pour tout $t \notin T$, et donc pour presque tout $t \in \mathbb{R}$ puisque T est dénombrable que

$$\mu_n(\varphi \geq t) \xrightarrow{n \rightarrow \infty} \mu_\infty(\varphi \geq t).$$

Puis par convergence dominée,

$$\int \varphi d\mu_n = \int_0^M \mu_n(\varphi \geq t) dt \xrightarrow{n \rightarrow \infty} \int_0^M \mu_\infty(\varphi \geq t) dt = \int \varphi d\mu.$$

□

Corollaire 1.23. Soient (X_n) une suite de variables aléatoires à valeurs dans $\mathbb{X}$ et $x_\infty \in \mathbb{X}$ tels que (X_n) converge en loi vers x_∞ . Alors, (X_n) converge vers x_∞ en probabilité.

Démonstration. Soit $\varepsilon > 0$, puisque $B(x_\infty, \varepsilon)$ est ouvert, le théorème du portemanteau (théorème 1.22) donne que

$$\limsup_{n \rightarrow \infty} \mathbf{P}[d(X_n, x_\infty) \geq \varepsilon] \leq \delta_{x_\infty}(\mathbb{X} \setminus B(x_\infty, \varepsilon)) = 0.$$

□

1.3.2 TENSION ET RÉGULARITÉ DES MESURES DE PROBABILITÉS

Dans cette section, nous verrons que pour $\mathbb{X}$ polonais, les mesures de probabilités vérifient des propriétés de compatibilités avec la topologie de $\mathbb{X}$.

Tension

Cette première propriété nous indique que du point de vue d'une mesure de probabilité fixée, $\mathbb{X}$ est presque compact.

Théorème 1.24 (Une mesure de probabilité est tendue). *Pour tout $\mu \in \mathcal{P}(\mathbb{X})$ et $\varepsilon > 0$, il existe un compact K tel que*

$$\mu(K) > 1 - \varepsilon. \quad (1.2)$$

En terme probabiliste, pour toute variable aléatoire X et $\varepsilon > 0$, il existe un compact K tel que

$$\mathbf{P}[X \in K] > 1 - \varepsilon.$$

Pour aller plus loin 5 (Tension des mesures). Si μ est une mesure de probabilité sur un espace topologique séparé et satisfait (1.2), on dit que μ est *tendue*. Nous avons montré que sur un espace polonais toutes les mesures de probabilités sont tendues, mais ce n'est plus forcément le cas sur un espace général.

Démonstration. Fixons $\mu \in \mathcal{P}(\mathbb{X})$ et $\varepsilon > 0$. Considérons une suite (x_n) dense dans $\mathbb{X}$. Soit $m \in \mathbb{N}^*$. Par densité des (x_n) et σ -additivité de μ ,

$$\lim_{N \rightarrow \infty} \mu\left(\bigcup_{n=1}^N \bar{B}\left(x_n, \frac{1}{m}\right)\right) = 1.$$

Il existe donc $N(m) \in \mathbb{N}$ tel que

$$\mu\left(\bigcup_{n=1}^{N(m)} \bar{B}\left(x_n, \frac{1}{m}\right)\right) > 1 - \frac{\varepsilon}{2^m}.$$

Posons

$$K := \bigcap_{m \in \mathbb{N}} \bigcup_{n=1}^{N(m)} \bar{B}\left(x_n, \frac{1}{m}\right).$$

Alors d'une part K est précompact, et d'autre part il est fermé comme intersection de fermés. Comme par hypothèse $\mathbb{X}$ est un espace métrique complet, K est compact. En passant au complémentaire on trouve que

$$\mu(\mathbb{X} \setminus K) \leq \sum_{m \in \mathbb{N}} \mu\left(\mathbb{X} \setminus \bigcup_{n=1}^{N(m)} \bar{B}\left(x_n, \frac{1}{m}\right)\right) < \varepsilon.$$

□

Régularité

Nous montrons maintenant qu'une mesure de probabilité est caractérisée par ses valeurs sur tous les compacts, ou sur tous les ouverts.

Théorème 1.25 (Ulam). *Soit $\mu \in \mathcal{P}(\mathbb{X})$. Alors pour tout $A \in \mathfrak{B}(\mathbb{X})$*

$$\mu(A) = \sup\{\mu(K) : K \in \mathfrak{K}(\mathbb{X}), K \subset A\} = \inf\{\mu(U) : U \in \mathfrak{O}(\mathbb{X}), A \subset U\}. \quad (1.3)$$

Remarque 1.26. La première égalité se réécrit : soit A un borélien et $\varepsilon > 0$, alors il existe un compact $K \subset A$ tel que $\mu(K) \geq \mu(A) - \varepsilon$, autrement dit $\mu(A \setminus K) < \varepsilon$. La deuxième égalité pourrait se réécrire de manière similaire.

Démonstration. Il suffit de montrer la première égalité ; la deuxième s'obtient par passage au complémentaire. Écrivons $\mathfrak{A}$ pour l'ensemble des boréliens A tels que la première égalité de (1.3) est vraie pour A et pour $\mathbb{X} \setminus A$.

Montrons que $\mathfrak{A}$ est une classe monotone, stable par passage au complémentaire. Puisque μ est tendue (théorème 1.24), $\mathbb{X} \in \mathfrak{A}$. Par construction $\mathfrak{A}$ est stable par passage au complémentaire. Prenons (A_n) une suite croissante d'éléments de $\mathfrak{A}$ et $A := \bigcup_{n=1}^{\infty} A_n$. Soit $\varepsilon > 0$. Par définition de $\mathfrak{A}$, prenons des compacts (K_n) et (L_n) tels que

$$\mu(A_n \setminus K_n) < \frac{\varepsilon}{2}, \quad \text{et} \quad \mu((\mathbb{X} \setminus A_n) \setminus L_n) < \frac{\varepsilon}{2^n}.$$

D'une part, par continuité des mesures, il existe $N \in \mathbb{N}$, tel que

$$\mu(A_N) = \mu\left(\bigcup_{n=1}^N A_n\right) > \mu(A) - \frac{\varepsilon}{2}.$$

D'où il vient

$$\mu(K_N) > \mu(A) - \varepsilon.$$

D'autre part, avec $L := \bigcap_{n=1}^{\infty} L_n$, comme $(\mathbb{X} \setminus A) \setminus L \subset \bigcup_{n=1}^{\infty} ((\mathbb{X} \setminus A_n) \setminus L_n)$, on trouve

$$\mu((\mathbb{X} \setminus A) \setminus L) \leq \sum_{n \in \mathbb{N}} \frac{\varepsilon}{2^n} = \varepsilon.$$

Nous avons montré que $A \in \mathfrak{A}$. Ainsi, $\mathfrak{A}$ est une classe monotone stable par passage au complémentaire.

Montrons que $\mathfrak{O}(\mathbb{X}) \subset \mathfrak{A}$, ce qui permettra de conclure par le lemme de classe monotone que $\mathfrak{A} = \mathfrak{B}(\mathbb{X})$. Soient U un ouvert et $F := \mathbb{X} \setminus U$. Fixons $\varepsilon > 0$. Par le théorème 1.24 (tension), choisissons un compact K tel que $\mu(\mathbb{X} \setminus K) < \varepsilon/2$. Premièrement, posons

$$K_n := \left\{x \in \mathbb{X} : d(x, F) \geq \frac{1}{n}\right\} \cap K, \quad \forall n \in \mathbb{N}^*,$$

avec

$$d(x, F) := \inf\{d(x, y) : y \in F\}, \quad \forall x \in \mathbb{X}.$$

Alors, (K_n) est une suite croissante de compacts dont l'union est $U \cap K$. Par continuité de μ , il existe $n_0 \in \mathbb{N}$ tel que

$$\mu(K_{n_0}) > \mu(U \cap K) - \frac{\varepsilon}{2} > \mu(U) - \varepsilon.$$

Ensuite, on a toujours $\mu(F \cap K) > \mu(F) - \varepsilon$ et $F \cap K$ est compact. Donc $U \in \mathfrak{A}$. □

Pour aller plus loin 6 (Mesures de Radon). Pour une mesure μ non nécessairement finie, la deuxième égalité dans (1.3) n'est plus forcément vraie. Une mesure qui satisfait la première approximation dans est *intérieurement régulière*, et si elle satisfait la deuxième, elle est *extérieurement régulière*. La mesure μ est *localement finie* si pour tout $x \in \mathbb{X}$ il existe un voisinage V tel que $\mu(V) < \infty$. Une mesure localement finie et intérieurement régulière est dite *de Radon*. Les mesures de Radon sont en fait « le bon cadre » pour la théorie de la mesure et permet d'obtenir :

- l'existence de limites projectives de mesures ;
- de bonnes propriétés de la topologie étroite pour les mesures bornées — que nous allons voir en partie plus loin dans ce cours, et pour la topologie *vague* pour les mesures non bornées — dont nous ne parlerons pas ;
- l'existence de lois conditionnelles et théorèmes de désintégration.

1.3.3 DISTANCE DE FORTET-MOURIER

Nous allons maintenant définir une topologie sur $\mathcal{P}(\mathbb{X})$ à partir d'une distance.

Définition 1.27 (Distance de Fortet–Mourier). La *distance de Fortet–Mourier* est

$$d_{BL}(\mu_1, \mu_2) := \sup \left\{ \left| \int \varphi d(\mu_1 - \mu_2) \right| : \varphi \in \mathcal{C}_b(\mathbb{X}), \|\varphi\|_\infty \leq 1, \text{Lip}(\varphi) \leq 1 \right\},$$

où nous notons

$$\text{Lip}(\varphi) := \sup_{x \neq y} \frac{|\varphi(x) - \varphi(y)|}{d(x, y)}.$$

Lemme 1.28. La quantité d_{BL} est une distance sur $\mathcal{P}(\mathbb{X})$.

Démonstration. Par construction, $d_{BL} \in [0, 2]$. La symétrie et l'inégalité triangulaire sont évidentes. Supposons que $d_{BL}(\mu_1, \mu_2) = 0$. Donc $\int \varphi d\mu_1 = \int \varphi d\mu_2$ pour tout $\varphi \in \mathcal{C}_b(\mathbb{X})$ bornée par 1 et de constante de Lipschitz bornée par 1. Par linéarité de l'intégrale, on trouve que c'est aussi vrai pour toute fonction φ continue, bornée et Lipschitz. Soit $\mathcal{C}$ l'espace vectoriel des fonctions Lipschitz et bornées, il est stable par infimum et contient 1. Soit $\mathcal{H}$ l'ensemble des fonctions boréliennes bornées φ telles que $\int \varphi d\mu_1 = \int \varphi d\mu_2$. L'espace $\mathcal{H}$ est stable par convergence monotone bornée et contient $\mathcal{C}$. Par un argument de classe monotone (théorème B.22), $\mathcal{H}$ contient toutes les fonctions $\sigma(\mathcal{C})$ -mesurable. Maintenant, pour un ouvert $U \in \mathfrak{D}(\mathbb{X})$ la suite de fonction $(nd(\cdot, U^C) \wedge 1)$ est bornée et Lipschitz et converge en croissant vers 1_U . Donc $\mathcal{H}$ contient toutes les fonctions boréliennes bornées et $\mu_1 = \mu_2$. $\square$

Remarque 1.29 (Reformulation avec des variables aléatoires). Quand on manipule des variables aléatoires, il est parfois plus simple de manipuler la quantité

$$\mathbf{BL}(X_1, X_2) := \sup \{ |\mathbf{E}[\varphi(X_1) - \varphi(X_2)]| : \varphi \in \mathcal{C}_b(\mathbb{X}), \text{Lip}(\varphi) \leq 1 \}.$$

Clairement,

$$\mathbf{BL}(X_1, X_2) = d_{BL}(\mathbf{loi}[X_1], \mathbf{loi}[X_2]),$$

mais on se gardera bien de penser que $\mathbf{BL}$ est une distance.

Distance de Fortet–Mourier et épaissis

Le lemme technique suivant montre que deux mesures proches, au sens de la distance de Fortet–Mourier, chargent presque les mêmes ensembles. Pour un borélien A , considérons ses épaissis, c'est-à-dire la famille d'ensembles ouverts

$$A^\varepsilon := \cup_{x \in A} B(x, \varepsilon), \quad \forall \varepsilon > 0.$$

Lemme 1.30. *Soient μ et $\nu \in \mathcal{P}(\mathbb{X})$. Alors*

$$\mu(A) \leq \nu(A^\varepsilon) + \left(1 + \frac{1}{\varepsilon}\right) d_{BL}(\mu, \nu), \quad \forall A \in \mathcal{B}(\mathbb{X}), \varepsilon > 0.$$

Démonstration. Posons

$$\varphi := \left(1 - \frac{d(\cdot, A)}{\varepsilon}\right)_+.$$

Alors, $\text{Lip}(\varphi) \leq \frac{1}{\varepsilon}$ et $\|\varphi\|_\infty \leq 1$ et φ satisfait $1_A \leq \varphi \leq 1_{A^\varepsilon}$. En intégrant et en utilisant la définition de d_{BL} , on trouve

$$\mu(A) \leq \int \varphi d\mu \leq \int \varphi d\nu + \left(1 + \frac{1}{\varepsilon}\right) d_{BL}(\mu, \nu) \leq \nu(A^\varepsilon) + \left(1 + \frac{1}{\varepsilon}\right) d_{BL}(\mu, \nu).$$

□

Compatibilité de la distance et de la convergence étroite

Le résultat suivant montre que la distance de Fortet–Mourier métrise la convergence étroite. Nous verrons plus tard que la distance est complète et que $\mathcal{P}(\mathbb{X})$ est séparable.

Théorème 1.31 (La distance de Fortet–Mourier induit la convergence étroite). *Soient (μ_n) et μ_∞ des éléments de $\mathcal{P}(\mathbb{X})$. Alors*

$$\mu_n \xrightarrow[n \rightarrow \infty]{\mathcal{P}(\mathbb{X})} \mu_\infty \Leftrightarrow d_{BL}(\mu_n, \mu_\infty) \xrightarrow[n \rightarrow \infty]{} 0.$$

Démonstration. $\Rightarrow$ Supposons la convergence étroite. Fixons $\varepsilon > 0$. Puisque μ_∞ est tendue (théorème 1.24), on peut trouver $K \subset \mathbb{X}$ tel que $\mu_\infty(K) > 1 - \varepsilon$. Définissons

$$\mathcal{BL} := \{\varphi \in \mathcal{C}_b(\mathbb{X}) : \|\varphi\|_\infty \leq 1, \text{Lip}(\varphi) \leq 1\},$$

Par le théorème d'Ascoli, l'ensemble

$$\mathcal{K} := \{\varphi|_K : \varphi \in \mathcal{BL}\},$$

est compact dans l'espace de Banach $\mathcal{C}_b(\mathbb{X})$. En particulier, il existe un recouvrement de $\mathcal{K}$ par un nombre fini de boules de rayon ε . Plus précisément, il existe $\varphi_1, \dots, \varphi_l \in \mathcal{BL}$ les centres de ces boules telles que pour toute $\varphi \in \mathcal{BL}$, il existe $j \in \{1, \dots, l\}$ tel que

$$\sup_{x \in K} |\varphi_j(x) - \varphi(x)| \leq \varepsilon.$$

En particulier, pour $x \in K$ et $y \in K^\varepsilon$

$$\begin{aligned} |\varphi(y) - \varphi_j(y)| &\leq |\varphi(y) - \varphi(x)| + |\varphi(x) - \varphi_j(x)| + |\varphi_j(x) - \varphi_j(y)| \\ &\leq d(x, y) + \varepsilon + d(x, y) < 3\varepsilon. \end{aligned} \tag{1.4}$$

Comme $d_{BL}(\mu_n, \mu_\infty) \rightarrow 0$, par le lemme 1.30 sur les épaisiss, pour n assez grand

$$\mu_n(K^\varepsilon) > 1 - 2\varepsilon. \quad (1.5)$$

Finalement, il vient

$$\left| \int \varphi d(\mu_n - \mu_\infty) \right| \leq \int |\varphi - \varphi_j| d(\mu_n + \mu_\infty) + \left| \int \varphi_j d(\mu_n - \mu_\infty) \right|.$$

Pour le premier terme, on trouve

$$\begin{aligned} \int_{K^\varepsilon} |\varphi - \varphi_j| d(\mu_n + \mu_\infty) &< 2 \times 3\varepsilon, \quad \text{par (1.4).} \\ \int_{\mathbb{X} \setminus K^\varepsilon} |\varphi - \varphi_j| d(\mu_n + \mu_\infty) &\leq 2(\mu_n + \mu_\infty)(\mathbb{X} \setminus K^\varepsilon) < 2(2\varepsilon + \varepsilon) = 6\varepsilon, \quad \text{par (1.5).} \end{aligned}$$

Nous avons donc montré que

$$d_{BL}(\mu_n, \mu_\infty) \leq 12\varepsilon + \max_{j=1, \dots, l} \left| \int \varphi_j d(\mu_n - \mu_\infty) \right|.$$

Le deuxième terme est un maximum, car il y a un nombre fini d'indice. Puisque, par hypothèse, chacune des intégrales avec φ_j converge vers 0, on peut passer à la limite supérieure et faire $\varepsilon \rightarrow 0$.

$\Leftarrow$ Si $d_{BL}(\mu_n, \mu) \rightarrow 0$, alors $\int \varphi d\mu_n \rightarrow \int \varphi d\mu$ pour toute fonction φ bornée et Lipschitz. L'ensemble des fonctions bornées et Lipschitz est dense dans $\mathcal{C}_b(\mathbb{X})$ pour la convergence uniforme et on conclut par le lemme 1.4.

□

Définition topologie de la convergence étroite

Définition 1.32 (Topologie de la convergence étroite). La *topologie de la convergence étroite* (certains auteurs disent *topologie faible* ou *topologie en loi*) sur $\mathcal{P}(\mathbb{X})$ est la topologie induite par d_{BL} .

Pour aller plus loin 7 (D'autres définitions de la topologie). Nous avons défini la convergence étroite en testant contre les fonctions continues bornées, ce qui ne nécessite pas de distance sur $\mathbb{X}$, mais uniquement une topologie, pourtant notre construction dépend de manière critique du choix d'une distance sur $\mathbb{X}$ à travers l'utilisation des fonctions Lipschitz.

1. On ne pourrait pas définir une distance pour la convergence étroite uniquement à partir des fonctions $\mathcal{C}_b(\mathbb{X})$. En effet, si on écrit

$$\rho(\mu_1, \mu_2) := \sup \left\{ \left| \int \varphi d(\mu_1 - \mu_2) \right| : \varphi \in \mathcal{C}_b(\mathbb{X}), \|\varphi\|_\infty \leq 1 \right\},$$

on vérifie facilement que ρ définit une distance sur $\mathcal{P}(\mathbb{X})$ (exercice!). Par contre, on peut voir (autre exercice!) que $\rho(\delta_{1/n}, \delta_0) = 1$ pour tout $n \in \mathbb{N}$ alors que $\delta_{1/n} \rightarrow \delta_0$ étroite.

2. Il existe d'autres distances sur $\mathcal{P}(\mathbb{X})$ qui définissent aussi la topologie de la convergence étroitement. La plus célèbre est la distance de *Lévy-Prokhorov*

$$d_{LP}(\mu, \nu) := \inf \{ \varepsilon > 0 : \mu(A) \leq \nu(A^\varepsilon) + \varepsilon, \forall A \in \mathfrak{B}(\mathbb{X}) \},$$

où $A^\varepsilon := \{x \in \mathbb{X} : d(x, A) \leq \varepsilon\}$. Ici, la distance sur $\mathbb{X}$ est utilisée dans la définition de d_{LP} à travers les épaissis A^ε . Il est possible de montrer que d_{LP} est bien une distance et qu'elle définit une topologie équivalente à celle définie par d_{BL} . Mieux, on peut montrer que d_{LP} est aussi une distance complète. Par contre, les deux distances ne sont pas équivalentes au sens bi-Lipschitz : le rapport $\frac{d_{BL}}{d_{LP}}$ n'est pas borné.

3. La topologie de la convergence étroite peut se définir uniquement à partir de la topologie sur $\mathbb{X}$ en disant que c'est la topologie initiale associée aux applications

$$e_\varphi : \mu \mapsto \int \varphi d\mu, \quad \varphi \in \mathcal{C}_b(\mathbb{X}),$$

où $\mathcal{C}_b(\mathbb{X})$ est l'ensemble des fonctions continues et bornées sur $\mathbb{X}$. Autrement dit, c'est la topologie engendrée par les $e_\varphi^{-1}(U)$ quand φ parcourt $\mathcal{C}_b(\mathbb{X})$ et U parcourt les ouverts de $\mathbb{R}$. Nous avons montré que si $(\mathbb{X}, d)$ est complet et séparable, alors cette topologie initiale est métrisable par d_{BL} (ou d_{LP}). Nous verrons plus loin que $\mathcal{P}(\mathbb{X})$ est encore un espace séparable et que la distance d_{BL} est complète. Notons toutefois que cette topologie initiale ne dépend pas d'un choix d'une distance sur $\mathbb{X}$. Il n'y a pas besoin d'équiper $\mathbb{X}$ d'une distance particulière, mais seulement de savoir que sa topologie peut être engendrée par une distance complète.

4. On peut considérer l'ensemble $\mathcal{M}_b(\mathbb{X})$ des mesures boréliennes signées bornées de $\mathbb{X}$, c'est-à-dire les mesure de la forme $\mu = \mu_+ - \mu_-$ avec μ_+ et μ_- des mesures finies. Tous les résultats et définitions de cette section se généralise à l'espace vectoriel $\mathcal{M}_b(\mathbb{X})$. La définition de la topologie de la convergence étroite comme topologie initiale se réinterprète comme la topologie d'espace vectoriel localement convexe engendrée par les semi-normes

$$N_\varphi := \mu \mapsto \left| \int \varphi d\mu \right|, \quad \varphi \in \mathcal{C}_b(\mathbb{X}).$$

La discussion précédente montre que lorsque $\mathbb{X}$ est polonais, alors $\mathcal{M}_b(\mathbb{X})$ est un espace de Fréchet séparable.

Séparabilité Nous montrons ici une première propriété importante de la topologie étroite sur $\mathcal{P}(\mathbb{X})$: la séparabilité. Dans la section suivante, à l'aide du critère de Prokhorov, nous montrons que la distance est aussi complète.

Proposition 1.33 (Séparabilité de la topologie étroite). *La topologie étroite est séparable. Plus précisément, soit $\mathbb{D}$ un ensemble dénombrable dense dans $\mathbb{X}$, alors l'ensemble dénombrable*

$$\mathcal{D} := \left\{ \sum q_i \delta_{x_i} : q_i \in \mathbb{Q}_+, \sum q_i = 1, x_i \in \mathbb{D} \right\}$$

est dense dans $\mathcal{P}(\mathbb{X})$.

Démonstration. Soient $\mu \in \mathcal{P}(\mathbb{X})$ et $\varepsilon > 0$. Il suffit de montrer qu'il existe $\nu \in \mathcal{D}$ telle que $d_{BL}(\mu, \nu) \leq \varepsilon$. Posons $\delta := \varepsilon/4$. Soit (x_n) une énumération des éléments de $\mathbb{D}$. On construit par récurrence

$$\begin{cases} A_0 := B(x_0, \delta), \\ A_{n+1} := B(x_{n+1}, \delta) \setminus \bigcup_{i=0}^n B(x_i, \delta). \end{cases}$$

Alors les (A_n) forment une partition de $\mathbb{X}$. Comme $\sum_{n \in \mathbb{N}} \mu(A_n) = 1$, on trouve $N \in \mathbb{N}$ tel que

$$\sum_{n > N} \mu(A_n) \leq \delta. \tag{1.6}$$

Posons, alors

$$\begin{cases} a_0 := (\mu(A_0) + \sum_{n>N} \mu(A_n)), \\ a_n := \mu(A_n), \quad \forall n \in \{1, \dots, N\} \end{cases}$$

Par densité de $\mathbb{Q}$ dans $\mathbb{R}$, on peut trouver des rationnels (q_n) tels que

$$\sum_{n=0}^N q_n = 1, \quad \text{et} \quad \sum_{n=0}^N |a_n - q_n| \leq \delta.$$

Alors,

$$\nu := \sum_{n=0}^N q_n \delta_{x_n},$$

convient. En effet, soit φ 1-bornée et 1-Lipschitz

$$\begin{aligned} \left| \int \varphi d(\mu - \nu) \right| &\leq \sum_{n=0}^N \int_{A_n} |\varphi(x) - \varphi(x_n)| \mu(dx) \\ &\quad + \int 1_{\mathbb{X} \setminus \cup_{n>N} A_n} |\varphi| d\mu + |\varphi(x_N)| \sum_{n>N} \mu(A_n) \\ &\quad + \sum_{n=0}^N |a_n - q_n| |\varphi(x_n)|. \end{aligned}$$

Par construction des (A_n) et la condition de Lipschitz, le premier terme est majoré par δ . Par (1.6) et la borne sur φ , le second terme est majoré par 2δ . Par la condition sur les (q_n) et la borne sur φ , le troisième terme est majoré par δ . $\square$

1.3.4 COMPACTITÉ POUR LA CONVERGENCE ÉTROITE

Cas général

Compacité de $\mathcal{P}(\mathbb{X})$ quand $\mathbb{X}$ est compact Si $\mathbb{X}$ est fini, alors $\mathcal{P}(\mathbb{X})$ s'identifie à un fermé borné de $\mathbb{R}^{\mathbb{X}}$, et donc $\mathcal{P}(\mathbb{X})$ est compact. Ce critère se généralise de plusieurs manières. Tout d'abord sur les espaces compacts généraux.

Théorème 1.34 (L'espace des mesures de probabilité sur un espace compact est compact). *Soit $\mathbb{K}$ un espace compact métrisable, alors $\mathcal{P}(\mathbb{K})$ est compact métrisable.*

Démonstration. Par le théorème de Stone–Weierstraß, il existe une suite de fonctions (f_n) dense dans $\mathcal{C}(\mathbb{K})$. En posant

$$d(\mu_1, \mu_2) := \sum_{k \in \mathbb{N}} 2^{-k} \left| \int f_k d(\mu_1 - \mu_2) \right|,$$

on vérifie aisément que d est une distance et qu'elle métrise la topologie de la convergence étroite. Donc, il suffit de démontrer de la compacité séquentielle. Prenons (μ_n) une suite dans $\mathcal{P}(\mathbb{K})$ et définissons la suite

$$a_n := \left(\int f_1 d\mu_n, \int f_2 d\mu_n, \dots \right), \quad n \in \mathbb{N}.$$

Notons que (a_n) est une suite à valeurs dans $\prod_{k \in \mathbb{N}} [\min f_k, \max f_k]$ qui est un espace compact par le théorème de Tychonoff. Quitte à extraire une sous-suite on a donc que $a_n \rightarrow a$. En utilisant la densité des (f_k) on trouve que

$$I(f) := \lim_{n \rightarrow \infty} \int f d\mu_n,$$

existe pour tout $f \in \mathcal{C}(\mathbb{K})$. En effet, pour tout $f \in \mathcal{C}(\mathbb{K})$ et $\varepsilon > 0$ on peut trouver un f_k tel que $\|f - f_k\|_\infty < \varepsilon$. D'où

$$-\varepsilon \leq \int f_k d\mu_n - \int f d\mu_n \leq \varepsilon, \quad \forall k \in \mathbb{N}.$$

En prenant la limite supérieure et la limite inférieure, on trouve

$$\limsup_{n \rightarrow \infty} \int f d\mu_n - \liminf_{n \rightarrow \infty} \int f d\mu_n \leq 2\varepsilon.$$

Le terme de gauche étant indépendant du choix de ε , on trouve qu'il est en fait nul. Notons que I est une application linéaire, positive et $I(1) = 1$. Par le théorème de Riesz–Markov (théorème B.24), il se représente donc comme une mesure de probabilité. $\square$

Donnons tout de suite une application probabiliste de ce théorème.

Corollaire 1.35. *Soit $\mathbb{X}$ un espace topologique métrisable et (X_n) une suite de variables aléatoires à valeurs dans $\mathbb{X}$. S'il existe un compact $K \subset \mathbb{X}$ tel que $\mathbf{P}[X_n \in K] = 1$ pour tout $n \in \mathbb{N}$. Alors, il existe une mesure de probabilité $\mu_\infty \in \mathcal{P}(K)$ et une extraction (n_j) telles que*

$$X_{n_j} \xrightarrow[j \rightarrow \infty]{\text{loi}} \mu_\infty.$$

Équitension et critère de Prokhorov Travaillons à nouveau sous l'hypothèse 1, c'est-à-dire que $\mathbb{X}$ est un espace polonais.

Définition 1.36 (Équitension). On dit qu'une partie $\mathcal{H}$ de $\mathcal{P}(\mathbb{X})$ est *équitendue* si pour tout $\varepsilon > 0$ il existe un compact $K \subset \mathbb{X}$ tel que pour tout $\mu \in \mathcal{H}$

$$\mu(K) > 1 - \varepsilon.$$

Par un abus de langage, on dira aussi qu'une famille de variables aléatoires est équitendue si l'ensemble des lois correspondantes l'est.

Remarque 1.37. En notant $\mathfrak{K}(\mathbb{X})$ les compacts de $\mathbb{X}$, on peut récrire la propriété de manière concise

$$\inf_{K \in \mathfrak{K}(\mathbb{X})} \sup_{\mu \in \mathcal{H}} \mu(\mathbb{X} \setminus K) = 0.$$

En termes d'une famille (X_i) de variables aléatoires cela s'écrit

$$\inf_{K \in \mathfrak{K}(\mathbb{X})} \sup_{i \in I} \mathbf{P}[X_i \notin K] = 0.$$

La propriété d'équitension signifie que les mesures de $\mathcal{H}$ sont uniformément presque supportées sur un compact.

Il n'est pas surprenant que l'on puisse récupérer de la compacité à partir de cette condition.

Théorème 1.38 (Critère de Prokhorov). *Soit $\mathcal{H}$ une partie de $\mathcal{P}(\mathbb{X})$. Il y a équivalence entre :*

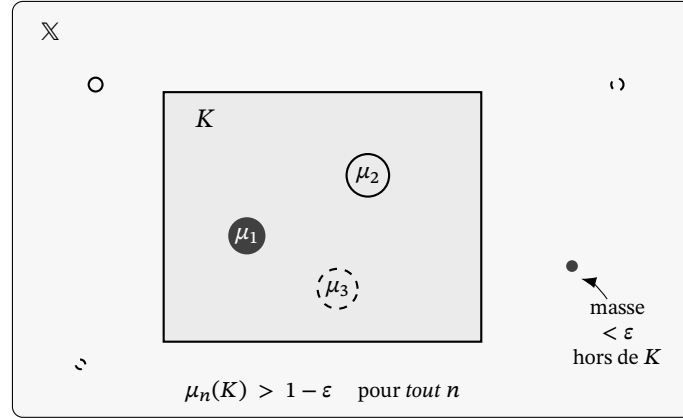


FIGURE 1.1 : Équitension : pour tout $\varepsilon > 0$, il existe *un même* compact K tel que $\mu_n(K) > 1 - \varepsilon$ pour tout n . Chaque mesure de la famille concentre au moins $1 - \varepsilon$ de sa masse dans K ; seule une fraction $< \varepsilon$, uniformément en n , peut s'échapper hors de K .

- (i) $\mathcal{H}$ est équitendue.
- (ii) $\mathcal{H}$ est séquentiellement relativement compacte : toute suite de $\mathcal{H}$ admet une sous-suite convergence dans $\mathcal{P}(\mathbb{X})$.
- (iii) $\mathcal{H}$ est relativement compacte pour la convergence étroite.
- (iv) $\mathcal{H}$ est précompacte.

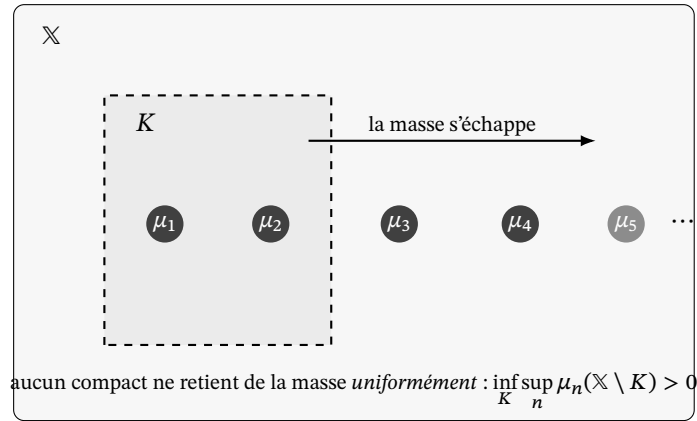


FIGURE 1.2 : Défaut de compacité : quel que soit le compact K , à partir d'un certain rang les μ_n chargent presque toute leur masse hors de K (ici $\mu_n = \delta_n$ sur $\mathbb{R}$). La famille n'est pas équitendue.

Démonstration. (i) $\Rightarrow$ (ii) Prenons (μ_n) une suite de $\mathcal{H}$. Par hypothèse, pour tout $j \in \mathbb{N}^*$, il existe un compact K_j tel que $\mu_n(K_j^C) < 1/j$ pour tout $n \in \mathbb{N}$. Sans perte de généralité, supposons la suite (K_j) croissante. Considérons

$$\mu_n^{(j)}(A) := \mu_n(A \cap K_j), \quad A \in \mathfrak{B}(\mathbb{X}).$$

La suite $(\mu_n^{(j)})_n$ s'identifie à une suite de mesures sur K_j de masse ≤ 1 . Une généralisation immédiate du théorème de compacité pour les mesures de probabilité sur les espaces compacts

(théorème 1.34) montre que c'est un espace compact. En particulier, il existe une extraction λ_j et une mesure $\mu^{(j)}$ sur K_j de masse ≤ 1 telles que

$$\mu_{\lambda_j(n)}^{(j)} \xrightarrow[n \rightarrow \infty]{\mathcal{P}(\mathbb{X})} \mu^{(j)}.$$

Comme (K_j) est croissante, on trouve que pour tout $j \in \mathbb{N}^*$ et toute φ continue, bornée et positive,

$$\int_{\mathbb{X}} \varphi d\mu^{(j)} = \lim_{n \rightarrow \infty} \int_{K_j} \varphi d\mu_{(\lambda_j \circ \lambda_{j+1})(n)} \leq \lim_{n \rightarrow \infty} \int_{K_{j+1}} \varphi d\mu_{(\lambda_j \circ \lambda_{j+1})(n)} = \int_{\mathbb{X}} \varphi d\mu^{(j+1)}.$$

Ainsi, la suite $(\mu^{(j)})$ est croissante. En posant $\mu_{\infty} := \sup_j \mu^{(j)}$ on obtient une mesure de masse ≤ 1 . Montrons que le long d'une sous-suite $\mu_n \rightarrow \mu_{\infty}$. Notons immédiatement que cela implique que $\mu_{\infty}(\mathbb{X})$ est une probabilité, car $1 \in \mathcal{C}_b(\mathbb{X})$.

Soit $\varphi \in \mathcal{C}_b(\mathbb{X})$. Fixons $\varepsilon > 0$ et prenons j assez grand pour que $\frac{1}{j} < \varepsilon$ et

$$\int \varphi d\mu^{(j)} \geq \int \varphi d\mu_{\infty} - \varepsilon.$$

Nous avons alors les majorations suivantes

$$\begin{aligned} \left| \int \varphi d(\mu_n - \mu_{\infty}) \right| &\leq \left| \int \varphi d(\mu_n^{(j)} - \mu^{(j)}) \right| + \left| \int \varphi d(\mu_n - \mu_n^{(j)}) \right| + \left| \int \varphi d(\mu_{\infty} - \mu^{(j)}) \right| \\ &\leq \left| \int \varphi d(\mu_n^{(j)} - \mu^{(j)}) \right| + \|\varphi\|_{\infty} \mu(\mathbb{X} \setminus K_j) + \varepsilon \\ &\leq \left| \int \varphi d(\mu_n^{(j)} - \mu^{(j)}) \right| + 2\varepsilon. \end{aligned}$$

Ainsi, nous avons montré que pour $j > \frac{1}{\varepsilon}$,

$$\limsup_{n \rightarrow \infty} \left| \int \varphi d(\mu_{\lambda_j(n)} - \mu_{\infty}) \right| \leq \varepsilon.$$

On conclut par un argument diagonal.

(ii) $\Leftrightarrow$ (iii) Puisque la topologie est définie par une distance, la compacité relative et la compacité relative séquentielle sont équivalentes.

(iii) $\Rightarrow$ (iv) Dans un espace métrique, tous les compacts sont précompacts.

(iv) $\Rightarrow$ (i) Soit $\varepsilon > 0$. On considère une famille $\mu_1, \dots, \mu_l \in \mathcal{P}(\mathbb{X})$ telle que

$$\mathcal{H} \subset \bigcup_{j=1}^l B_{BL}(\mu_j, \frac{\varepsilon^2}{4}).$$

Par le théorème de tension (théorème 1.24) la famille finie $\{\mu_1, \dots, \mu_l\}$ est équitendue. Soit donc K_{ε} un compact tel que $\mu_j(K_{\varepsilon}) > 1 - \varepsilon/2$ pour tout $j = 1, \dots, l$. Comme K_{ε} est compact, il existe une famille finie $F = F(\varepsilon) := \{x_1^{\varepsilon}, \dots, x_{l_{\varepsilon}}^{\varepsilon}\}$ telle que

$$K_{\varepsilon} \subset \bigcup_{k=1}^{l_{\varepsilon}} B(x_k^{\varepsilon}, \frac{\varepsilon}{2}) = F^{\frac{\varepsilon}{2}}.$$

En particulier $\mu_j(F^{\varepsilon/2}) \geq 1 - \frac{\varepsilon}{2}$ pour tout $j = 1, \dots, l$. Par le lemme sur les épaissis (lemme 1.30), on trouve finalement que

$$\mu(F^\varepsilon) \geq 1 - \varepsilon - \frac{\varepsilon^2}{4}, \quad \forall \mu \in \mathcal{H}.$$

Pour $\delta > 0$, posons

$$\varepsilon_m := \frac{\delta}{2^m}, \quad \forall m \in \mathbb{N}^*; \quad K := \bigcap_{m \in \mathbb{N}^*} \bigcup_{x \in F(\varepsilon_m)} \bar{B}(x, \varepsilon_m).$$

Alors K est fermé comme intersection de fermé et il est précompact. Donc K est compact et on trouve pour tout $\mu \in \mathcal{H}$

$$\mu(\mathbb{X} \setminus K) \leq \sum_{m \in \mathbb{N}^*} \delta 2^{-m} = \delta.$$

Cela montre que $\mathcal{H}$ est équitendue. □

La topologie étroite est polonaise Puisque les suites de Cauchy sont précompactes et que par théorème 1.38 les ensembles précompacts sont compacts, on est déduit qu'elles convergent. Ainsi $(\mathcal{P}(\mathbb{X}), d_{BL})$ est complet. Avec proposition 1.33, on trouve le résultat suivant.

Théorème 1.39 (L'espace des mesures de probabilités est polonais). *Si $\mathbb{X}$ est polonais, alors $\mathcal{P}(\mathbb{X})$ muni de la topologie étroite est également polonais.*

Compacité dans les espaces produits On en déduit ce théorème utile pour les mesures définies sur un espace produit $\mathbb{X} := \mathbb{X}_1 \times \mathbb{X}_2$, où $(\mathbb{X}_i, d_i)$ sont des espaces métriques complets et séparables. Dans ce cas, on peut équiper $\mathbb{X}$ de la distance

$$d((x_1, x_2), (y_1, y_2)) := d_1(x_1, y_1) + d_2(x_2, y_2),$$

pour laquelle c'est à nouveau un espace métrique complet et séparable. Notons également e_i la projection canonique de $\mathbb{X}$ vers $\mathbb{X}_i$ et pour $\mu \in \mathcal{P}(\mathbb{X})$, $\mu_i := (e_i)_\# \mu$ pour ses marginales.

Proposition 1.40. *Soit $\mathcal{H} \subset \mathcal{P}(\mathbb{X})$. Alors $\mathcal{H}$ est équitendu si et seulement si les ensembles*

$$\mathcal{H}_i := \{\mu_i : \mu \in \mathcal{H}\},$$

sont équitendus.

Démonstration. C'est nécessaire car comme les projections e_i sont continues, les applications $(e_i)_\# : \mathcal{P}(\mathbb{X}) \rightarrow \mathcal{P}(\mathbb{X}_i)$ sont aussi continues. Montrons que c'est suffisant. Pour tout $\varepsilon > 0$, on peut trouver des compacts K_i de $\mathbb{X}_i$ tels que

$$\mu_i(\mathbb{X} \setminus K_i) \leq \varepsilon, \quad \forall \mu \in \mathcal{H}.$$

Puisque

$$\mu(\mathbb{X} \setminus K_1 \times K_2) \leq \mu_1(\mathbb{X}_1 \setminus K_1) + \mu_2(\mathbb{X}_2 \setminus K_2),$$

le résultat suit. □

Corollaire 1.41. *Soient μ et ν deux mesures de probabilités sur des espaces polonais alors l'ensemble des variables aléatoires (X, Y) telles que $X \sim \mu$ et $Y \sim \nu$ est compact pour la convergence en loi.*

Démonstration. La compacité relative vient du théorème précédent et du fait que $\{\mu\}$ et $\{\nu\}$ sont tendus par théorème 1.24. La contrainte sur les marginales est fermée. □

Cas des variables aléatoires numériques

Étudions le critère abstrait sur l'exemple plus simple des espaces euclidiens. Commençons par récrire l'équitension dans ce cas.

Lemme 1.42. *Une suite (X_n) de variables aléatoires à valeurs dans $\mathbb{R}^d$ est équitendue si et seulement si*

$$\lim_{r \rightarrow \infty} \sup_{n \in \mathbb{N}} \mathbf{P}[\|X_n\| \geq r] = 0.$$

Démonstration. C'est une réécriture immédiate de la définition en utilisant que les compacts de $\mathbb{R}^d$ sont les fermés bornés. $\square$

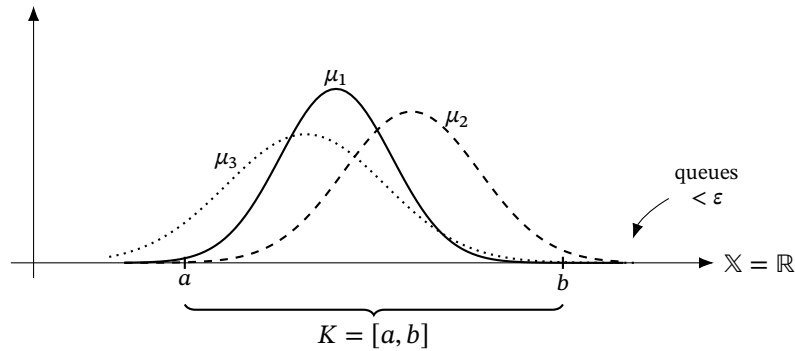


FIGURE 1.3 : Équitension, lecture par les densités : toutes les lois μ_n restent sous une même enveloppe concentrée sur $K = [a, b]$. La masse hors de K (les queues) est majorée par ε uniformément en n ; c'est exactement $\sup_n \mu_n(\mathbb{R} \setminus K) \leq \varepsilon$.

Passons à une condition suffisante de compacité sur $\mathcal{P}(\mathbb{R}^d)$.

Proposition 1.43. *Soit $\varphi : \mathbb{R}^d \rightarrow [0, \infty]$ une fonction semi-continue inférieurement telle que $\varphi(x) \rightarrow +\infty$ quand $\|x\| \rightarrow +\infty$. Alors les ensembles*

$$\left\{ \mu \in \mathcal{P}(\mathbb{R}^d) : \int \varphi d\mu \leq C \right\}, \quad \forall C \in \mathbb{R}_+,$$

sont compacts pour la topologie étroite.

Démonstration. Il suffit de le montrer pour $C := 1$. Pour tout $\varepsilon > 0$, posons

$$K_\varepsilon := \left\{ \varphi \leq \frac{1}{\varepsilon} \right\}.$$

Comme φ est semi-continue inférieurement, K_ε est fermé, et par coercivité K_ε est borné. Donc K_ε est compact par le théorème de Heine–Borel et pour tout μ dans l'ensemble en question

$$\mu(\mathbb{R}^d \setminus K_\varepsilon) \leq \varepsilon \int_{\mathbb{R}^d \setminus K_\varepsilon} \varphi d\mu \leq \varepsilon.$$

Donc l'ensemble est équitendu, et donc relativement compact par le théorème de compacité de Prokhorov (théorème 1.38). Par le lemme du portemanteau (théorème 1.22), comme φ est semi-continue inférieurement, les applications $\mu \mapsto \int \varphi d\mu$ sont aussi semi-continues inférieurement. L'ensemble est fermé comme sous-ensemble de niveau d'une fonction semi-continue inférieurement (lemme 1.19 (ii)). $\square$

Corollaire 1.44 (Les boules fermées de $\mathcal{L}^p$ sont étroitement compactes). Soit (X_n) une suite de variables aléatoires à valeurs dans $\mathbb{R}^d$ telle que pour un certain $1 \leq p < \infty$

$$\sup_n \mathbf{E}[\|X_n\|^p] < \infty,$$

alors il existe une extraction (n_j) et $\mu_\infty \in \mathcal{P}(\mathbb{R}^d)$ avec $\int \|x\|^p \mu_\infty(dx) < \infty$ telles que

$$X_{n_j} \xrightarrow[j \rightarrow \infty]{\text{loi}} \mu_\infty.$$

Théorème 1.45 (Lévy). Soient (X_n) une suite de variables aléatoires à valeurs dans $\mathbb{R}^d$ et leur fonction caractéristique

$$\hat{\mu}_n(t) := \mathbf{E}[e^{it \cdot X_n}], \quad \forall t \in \mathbb{R}^d.$$

Alors, (X_n) converge en loi si et seulement $(\hat{\mu}_n)$ converge simplement vers une fonction continue en 0.

Démonstration. Si (X_n) converge en loi vers μ_∞ alors on sait par le théorème 1.6 que $\hat{\mu}_n \rightarrow \hat{\mu}_\infty$ simplement et $\hat{\mu}_\infty$ est continue en 0 (exercice!). Maintenant supposons que $(\hat{\mu}_n)$ converge simplement vers une fonction f continue en 0. Je prétends alors que (μ_n) est équitendue et que c'est suffisant pour conclure. En effet, par équitension et le théorème de Prokhorov (théorème 1.38) on peut extraire une sous suite qui converge étroitement vers une certaine loi μ_∞ . Mais, alors par le théorème 1.6 et l'unicité de la limite simple $f = \hat{\mu}_\infty$, donc $\hat{\mu}_n \rightarrow \hat{\mu}_\infty$ et on conclut par le théorème 1.6. Commençons par le cas $d := 1$. Soit $u > 0$. Par le théorème de Fubini

$$\frac{1}{u} \int_{-u}^u (1 - \hat{\mu}_n(t)) dt = \int \frac{1}{u} \int_{-u}^u (1 - e^{itx}) dt \mu_n(dx) = 2 \int \left(1 - \frac{\sin(ux)}{ux}\right) \mu_n(dx).$$

Puisque l'intégrande est positive et $\geq \frac{1}{2}$ dès que $|ux| \geq 2$, on trouve que

$$\mu_n\left(|x| \geq \frac{2}{u}\right) \leq \frac{1}{u} \int_{-u}^u (1 - \hat{\mu}_n(t)) dt.$$

Comme $\mu_n(0) = 1$, on a aussi que $f(0) = 1$ donc par le théorème fondamental de l'analyse comme f est continue

$$\frac{1}{u} \int_{-u}^u (1 - f(t)) dt \xrightarrow{u \rightarrow 0} 0.$$

Par convergence dominée on a aussi que pour u fixé

$$\frac{1}{u} \int_{-u}^u (1 - \hat{\mu}_n(t)) dt \xrightarrow{n \rightarrow \infty} \frac{1}{u} \int_{-u}^u (1 - f(t)) dt.$$

On conclut avec le lemme 1.42. Pour la dimension quelconque, on se ramène à l'extension des marginales par la proposition 1.40, en observant que la fonction caractéristique des marginales s'obtient en restreignant la fonction caractéristique à des droites de $\mathbb{R}^d$. $\square$

Exercices fondamentaux

Les exercices ci-dessous, tirés du livret d'exercices, constituent l'entraînement minimal recommandé sur ce chapitre.

Rappels. Exercices [3](#), [4](#).

Convergence en loi. Exercices [5](#), [6](#), [7](#).

Propriétés de la convergence étroite. Exercices [12](#), [13](#), [14](#), [15](#).

Compacité et théorème de Prokhorov. Exercices [17](#), [18](#), [19](#).

Fonctions caractéristiques. Exercices [8](#), [9](#), [10](#).

Construction rigoureuse de l'espérance conditionnelle et des lois conditionnelles

Dans tout ce chapitre, nous travaillons sur un espace de probabilités $(\Omega, \mathfrak{B}, \mathbf{P})$ et nous formalisons rigoureusement les notions d'*espérance conditionnelle* et de *loi conditionnelle*. Tout d'abord, sans aucune hypothèse supplémentaire sur Ω nous définissons l'*espérance conditionnelle de F sachant $\mathfrak{A}$* $\mathbf{E}[F | \mathfrak{A}]$, pour une variable aléatoire F intégrable et une sous-tribu $\mathfrak{A}$ de $\mathfrak{B}$. Intuitivement, l'espérance conditionnelle $\mathbf{E}[F | \mathfrak{A}]$ représente la « meilleure approximation » de F sachant $\mathfrak{A}$. La sous-tribu $\mathfrak{A}$ représente une « information limitée » à certains événements. On s'attend à ce que $\mathbf{E}[F | \mathfrak{A}]$ soit une variable aléatoire $\mathfrak{A}$ -mesurable qui satisfasse certaines propriétés naturelles qui reflètent notre intuition.

2.1 Espérance conditionnelle par rapport à une tribu

2.1.1 DÉFINITION

Existence d'un objet universel

Commençons par construire un objet canonique partir de d'une sous-tribu et d'une variable aléatoire.

Théorème 2.1 (Propriété universelle). *Soit $\mathfrak{A}$ une sous-tribu de $\mathfrak{B}$. Soit F une variable aléatoire ($\mathfrak{B}$ -mesurable) à valeurs dans $[0, \infty]$. Alors, il existe une unique variable aléatoire G , mesurable par rapport à $\mathfrak{A}$ et à valeurs dans $[0, \infty]$, telle que*

$$\mathbf{E}[F 1_A] = \mathbf{E}[G 1_A], \quad \forall A \in \mathfrak{A}. \quad (2.1)$$

De plus, si $F \in \mathcal{L}^1(\mathfrak{B})$, non nécessairement positive, alors il existe une unique $G \in \mathcal{L}^1(\mathfrak{B})$ telle que (2.1) reste vraie.

Remarque 2.2. G étant une variable aléatoire, l'unicité s'entend au sens presque sûr.

L'exercice 24 détaille ce point de vue.

Démonstration. Montrons d'abord le théorème dans le cas où F est positive, le cas L^1 s'en déduisant.

Unicité. Soient G et G' deux variables aléatoires comme dans l'énoncé. En prenant $Z := 1_{G > G'}$ qui est $\mathfrak{A}$ -mesurable puisque G et G' le sont, on trouve

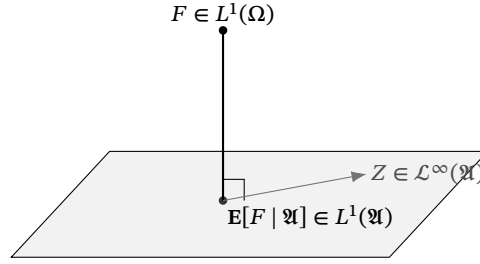
$$\mathbf{E}[(G - G') 1_{G > G'}] = 0.$$

Comme l'intégrand est positive, $(G - G') 1_{G > G'} = 0$ puis $1_{G > G'} = 0$. Donc $G' \geq G$. De manière similaire $G \geq G'$ et $G = G'$.

Existence. Posons

$$\mathbf{Q}(A) := \mathbf{E}[F 1_A], \quad \forall A \in \mathfrak{A}.$$

Alors $\mathbf{Q}$ est une mesure sur $(\Omega, \mathfrak{A})$ (exercice!). De plus, si $A \in \mathfrak{A}$ est tel que $\mathbf{P}(A) = 0$, alors $\mathbf{Q}(A) = 0$. Donc $\mathbf{Q} \ll \mathbf{P}$ (plus précisément $\mathbf{Q}$ est absolument continue par rapport à la



$$F - \mathbf{E}[F | \mathfrak{A}] \perp L^\infty(\mathfrak{A}) \quad \text{c.-à-d.} \quad \mathbf{E}[(F - \mathbf{E}[F | \mathfrak{A}])Z] = 0 \quad \text{pour tout } Z \text{ } \mathfrak{A}\text{-mesurable borné}$$

FIGURE 2.1 : $\mathbf{E}[F | \mathfrak{A}]$ est la « projection » de F sur les variables $\mathfrak{A}$ -mesurables. L'orthogonalité s'exprime par la relation $\mathbf{E}[FZ] = \mathbf{E}[\mathbf{E}[F | \mathfrak{A}]Z]$ pour tout Z $\mathfrak{A}$ -mesurable bornée.

restriction de $\mathbf{P}$ à la sous-tribu $\mathfrak{A}$), et par le théorème de Radon–Nikodym (théorème B.12), il existe $G \geq 0$ tel que

$$\mathbf{E}[F1_A] = \mathbf{E}[G1_A], \quad A \in \mathfrak{A}.$$

La propriété (2.1) est donc vérifiée.

Cas L^1 . Si maintenant $F \in \mathcal{L}^1(\mathfrak{B})$, on écrit $F = F_+ - F_-$ et on utilise le cas d'avant. De plus $\mathbf{Q}(\Omega) = \mathbf{E}[F] < \infty$ et donc la densité G est intégrable. $\square$

Définition à partir de la propriété universelle

Définition 2.3 (Espérance conditionnelle). Soit F une variable aléatoire à valeurs dans $[0, \infty]$, ou une variable aléatoire dans $\mathcal{L}^1(\mathfrak{B})$. On note $\mathbf{E}[F | \mathfrak{A}]$ l'unique variable aléatoire $\mathfrak{A}$ -mesurable qui satisfait (2.1). On appelle (2.1) la « propriété universelle » de l'espérance conditionnelle.

Remarque 2.4. Nous verrons qu'il est parfois plus pratique de vérifier (2.1) avec 1_A remplacé par $Z \in \mathcal{L}^\infty(\mathfrak{A})$ (éventuellement positif si on est dans le cas positif). Par un théorème de limite monotone, c'est équivalent.

2.1.2 PROPRIÉTÉS

Propriétés qui généralisent celles de l'espérance

L'espérance conditionnelle partage beaucoup de similarités avec l'espérance. On désignera les propriétés énoncées ci-dessous comme la version « conditionnelle » du résultat usuel pour l'espérance. On parlera par exemple du « lemme de Fatou conditionnel » ou d'« inégalité de Jensen conditionnelle », etc. Il faut cependant bien garder en tête que $\mathbf{E}[F | \mathfrak{A}]$ est une variable aléatoire ! En particulier, les inégalités ou les limites doivent être comprises au sens des variables aléatoires, c'est-à-dire au sens presque sûr.

Théorème 2.5 (Propriétés génériques de l'espérance conditionnelle). Soit $\mathfrak{A}$ une sous-tribu de $\mathfrak{B}$. Prenons deux variables aléatoires F et F' ou bien à valeurs dans $[0, \infty]$ ou bien dans $\mathcal{L}^1(\mathfrak{B})$.

(i) $\mathbf{E}[\mathbf{E}[F | \mathfrak{A}]] = \mathbf{E}[F]$.

(ii) Soient t et $s \in \mathbb{R}_+$ (cas positif) ou dans $\mathbb{R}$ (cas L^1), alors

$$\mathbf{E}[tF + sF' | \mathfrak{A}] = t \mathbf{E}[F | \mathfrak{A}] + s \mathbf{E}[F' | \mathfrak{A}].$$

(iii) Si $F \leq F'$, alors

$$\mathbf{E}[F | \mathfrak{A}] \leq \mathbf{E}[F' | \mathfrak{A}].$$

(iv) Soit φ une fonction convexe telle que $\mathbf{E}[\varphi(F) \mid \mathfrak{A}]$ est bien définie, on a

$$\varphi(\mathbf{E}[F \mid \mathfrak{A}]) \leq \mathbf{E}[\varphi(F) \mid \mathfrak{A}].$$

(v) Pour $p \in [1, \infty]$, l'application linéaire

$$\mathcal{L}^p(\mathfrak{B}) \ni F \mapsto \mathbf{E}[F \mid \mathfrak{A}] \in \mathcal{L}^p(\mathfrak{A}),$$

est une contraction, en particulier elle est continue.

(vi) Si (F_n) est une suite croissante de variables aléatoires positives, on a

$$\mathbf{E}\left[\lim_{n \rightarrow \infty} F_n \mid \mathfrak{A}\right] = \lim_{n \rightarrow \infty} \mathbf{E}[F_n \mid \mathfrak{A}].$$

(vii) Si (F_n) est une suite de variables aléatoires positives, on a

$$\mathbf{E}\left[\liminf_{n \rightarrow \infty} F_n \mid \mathfrak{A}\right] \leq \liminf_{n \rightarrow \infty} \mathbf{E}[F_n \mid \mathfrak{A}].$$

(viii) Si $(F_n) \subset \mathcal{L}^1$ converge presque sûrement vers F et $\sup_n |F_n| \leq Z$ pour Z intégrable, alors

$$\mathbf{E}[F_n \mid \mathfrak{A}] \xrightarrow[n \rightarrow \infty]{p.s./\mathcal{L}^1} \mathbf{E}[F \mid \mathfrak{A}].$$

Démonstration. (i) Par la propriété universelle avec $Z := 1 = 1_\Omega$, on a

$$\mathbf{E}[\mathbf{E}[F \mid \mathfrak{A}]] = \mathbf{E}[\mathbf{E}[F \mid \mathfrak{A}]Z] = \mathbf{E}[FZ] = \mathbf{E}[F].$$

(ii) Soit $Z := 1_A$ avec $A \in \mathfrak{A}$. Par linéarité de l'espérance et la propriété universelle (2.1)

$$\begin{aligned} \mathbf{E}[(tF + sF')Z] &= t \mathbf{E}[FZ] + s \mathbf{E}[F'Z] = t \mathbf{E}[\mathbf{E}[F \mid \mathfrak{A}]Z] + s \mathbf{E}[\mathbf{E}[F' \mid \mathfrak{A}]Z] \\ &= \mathbf{E}[(t \mathbf{E}[F \mid \mathfrak{A}] + s \mathbf{E}[F' \mid \mathfrak{A}])Z]. \end{aligned}$$

(iii) Si maintenant $F \geq 0$ et $Z \geq 0$, on trouve

$$\mathbf{E}[\mathbf{E}[F \mid \mathfrak{A}]Z] = \mathbf{E}[FZ] \geq 0.$$

Donc $\mathbf{E}[F \mid \mathfrak{A}] \geq 0$. Par la propriété précédente, c'est suffisant.

(iv) Posons

$$E_\varphi := \{(a, b) \in \mathbb{Q}^2 : \forall t \in \mathbb{R}, \varphi(t) \geq at + b\}.$$

Alors $\varphi(t) = \sup\{at + b : (a, b) \in E_\varphi\}$. Par monotonie et linéarité de l'espérance conditionnelle, pour $(a, b) \in E_\varphi$,

$$\mathbf{E}[\varphi(F) \mid \mathfrak{A}] \geq \mathbf{E}[aF + b \mid \mathfrak{A}] = a \mathbf{E}[F \mid \mathfrak{A}] + b.$$

Comme E_φ est dénombrable on peut passer au supremum dans l'inégalité presque sûre et on trouve le résultat.

(v) Supposons $p < \infty$. En appliquant l'inégalité de Jensen que l'on vient de démontrer à la fonction $t \mapsto |t|^p$, il vient

$$|\mathbf{E}[F \mid \mathfrak{A}]|^p \leq \mathbf{E}[|F|^p \mid \mathfrak{A}].$$

Puis on utilise que $\mathbf{E}[\mathbf{E}[|F|^p \mid \mathfrak{A}]] = \mathbf{E}[|F|^p]$ par le premier point. Le cas $p = \infty$ est obtenu en passant à la limite $p \rightarrow \infty$.

(vi) Par monotonie, $(\mathbf{E}[F_n | \mathfrak{A}])$ est aussi une suite croissante de variables aléatoires positives. On prend alors $Z = 1_A$ avec $A \in \mathfrak{A}$ et, par convergence monotone, on passe à la limite dans

$$\mathbf{E}[F_n Z] = \mathbf{E}[\mathbf{E}[F_n | \mathfrak{A}] Z].$$

(vii) Par la propriété de convergence monotone que l'on vient d'établir

$$\mathbf{E}\left[\liminf_{n \rightarrow \infty} F_n | \mathfrak{A}\right] = \lim_{n \rightarrow \infty} \mathbf{E}\left[\inf_{k \geq n} F_k | \mathfrak{A}\right].$$

Mais par monotonie,

$$\mathbf{E}\left[\inf_{k \geq n} F_k | \mathfrak{A}\right] \leq \inf_{k \geq n} \mathbf{E}[F_k | \mathfrak{A}].$$

(viii) Par le lemme de Fatou conditionnelle on a déjà

$$\mathbf{E}[F | \mathfrak{A}] \leq \liminf_{n \rightarrow \infty} \mathbf{E}[F_n | \mathfrak{A}].$$

Pour l'inégalité sur la limite supérieure, on utilise que $Z - F_n \geq 0$ et on invoque de nouveau le lemme de Fatou

$$\mathbf{E}[Z - F | \mathfrak{A}] = \mathbf{E}\left[\liminf_{n \rightarrow \infty} (Z - F_n) | \mathfrak{A}\right] \leq \mathbf{E}[Z | \mathfrak{A}] - \limsup_{n \rightarrow \infty} \mathbf{E}[F_n | \mathfrak{A}].$$

Ce qui donne

$$\limsup_{n \rightarrow \infty} \mathbf{E}[F_n | \mathfrak{A}] \leq \mathbf{E}[F | \mathfrak{A}],$$

et, par suite, la convergence presque sûre. La convergence $\mathcal{L}^1$ est obtenue par convergence dominée avec

$$\sup_n |\mathbf{E}[F_n | \mathfrak{A}]| \leq \sup_n \mathbf{E}[|F_n| | \mathfrak{A}] \leq \mathbf{E}[Z | \mathfrak{A}],$$

et $\mathbf{E}[\mathbf{E}[Z | \mathfrak{A}]] = \mathbf{E}[Z] < \infty$.

□

Propriétés spécifiques

Passons maintenant à des propriétés nouvelles. Par simplicité, nous nous restreignons au cas intégrable. Les propriétés restent vraies dans le cas positif à condition de remplacer les hypothèses d'intégrabilité par de la positivité. Nous laissons cette généralisation en exercice.

Théorème 2.6 (Propriétés spécifiques de l'espérance conditionnelle). *Soient $F \in \mathcal{L}^1(\mathfrak{B})$ et $\mathfrak{A}$ une sous-tribu de $\mathfrak{B}$. Alors :*

(i) *F est indépendante de $\mathfrak{A}$ si et seulement si pour toute fonction h borélienne tel que $h(F)$ intégrable*

$$\mathbf{E}[h(F) | \mathfrak{A}] = \mathbf{E}[h(F)].$$

(ii) *Si G est une variable aléatoire $\mathfrak{A}$ -mesurable telle que FG est intégrable, on a*

$$\mathbf{E}[FG | \mathfrak{A}] = G \mathbf{E}[F | \mathfrak{A}].$$

(iii) *Si $\mathfrak{A}'$ est une sous-tribu de $\mathfrak{A}$ alors*

$$\mathbf{E}[\mathbf{E}[F | \mathfrak{A}] | \mathfrak{A}'] = \mathbf{E}[F | \mathfrak{A}'].$$

Démonstration. (i) Supposons que F soit indépendante de $\mathfrak{A}$. Prenons $Z \in \mathcal{L}^\infty(\mathfrak{A})$ et h comme dans l'énoncé. Par indépendance,

$$\mathbf{E}[h(F)Z] = \mathbf{E}[h(F)] \mathbf{E}[Z] = \mathbf{E}[\mathbf{E}[h(F)]Z].$$

Ce qui montre que $\mathbf{E}[h(F)]$ satisfait la propriété universelle (2.1). Réciproquement, supposons la formule pour les espérances conditionnelles. Soient $B \in \mathfrak{B}(\mathbb{R})$ et $A \in \mathfrak{A}$,

$$\mathbf{P}[(F \in B) \cap A] = \mathbf{E}[1_B(F)1_A] = \mathbf{E}[1_B(F)] \mathbf{E}[1_A] = \mathbf{P}[F \in B] \mathbf{P}[A].$$

(ii) Au vu de la linéarité, quitte à décomposer avec les parties positives et négatives, supposons que $F \geq 0$ et $G \geq 0$. Prenons $A \in \mathfrak{A}$, $Z := 1_A$, et pour $n \in \mathbb{N}$, $Z_n := G 1_{G \leq n} Z \in \mathcal{L}^\infty(\mathfrak{A})$. Par la propriété universelle

$$\mathbf{E}[FZ_n] = \mathbf{E}[\mathbf{E}[F | \mathfrak{A}]Z_n].$$

Par convergence monotone, on passe à la limite dans les deux espérances et on trouve

$$\mathbf{E}[FGZ] = \mathbf{E}[\mathbf{E}[F | \mathfrak{A}]GZ].$$

Ce qui montre d'une part que $\mathbf{E}[\mathbf{E}[F | \mathfrak{A}]G] = \mathbf{E}[FG] < \infty$ et d'autre part que $\mathbf{E}[F | \mathfrak{A}]G$ satisfait la propriété universelle (2.1).

(iii) Posons $G := \mathbf{E}[F | \mathfrak{A}]$. Prenons $Z \in \mathcal{L}^\infty(\mathfrak{A}') \subset \mathcal{L}^\infty(\mathfrak{A})$. Par la propriété universelle,

$$\mathbf{E}[FZ] = \mathbf{E}[GZ].$$

Ce qui conclut. □

2.1.3 CONDITIONNEMENT PAR RAPPORT À UNE VARIABLE ALÉATOIRE

Dans cette partie $(\mathbb{X}, \mathfrak{X})$ est un espace mesurable.

Définition 2.7. Soient $F \in \mathcal{L}^1(\mathfrak{B})$ (donc à valeurs réelles) et X une variable aléatoire à valeurs dans $\mathbb{X}$. On pose

$$\mathbf{E}[F | X] := \mathbf{E}[F | \sigma(X)].$$

Lemme de Doob

Au vu du résultat suivant les variables aléatoires $\sigma(X)$ -mesurable, les espérances conditionnelles par rapport à X ont une forme close. Ce qui les rendra plus faciles à calculer en pratique.

Théorème 2.8 (Doob). *Soit X une variable aléatoire à valeurs dans $\mathbb{X}$. Soit $G : \Omega \rightarrow \mathbb{R}$. Alors G est $\sigma(X)$ -mesurable si et seulement s'il existe une fonction mesurable $h : \mathbb{X} \rightarrow \mathbb{R}$ telle que $G = h(X)$. De plus, cette application est unique $\mathbf{loi}[X]$ -presque sûrement.*

Démonstration. Il est clair que la condition est suffisante. Pour montrer qu'elle est aussi nécessaire, supposons d'abord que G est étagée : il existe une suite numérique (y_n) telle que $G = \sum_n y_n 1_{G=y_n}$. Comme G est supposée $\sigma(X)$ -mesurable, pour tout $n \in \mathbb{N}$, $\{G = y_n\}$ est de la forme $X^{-1}(A_n)$ pour un certain $A_n \in \mathfrak{X}$. Posons $C_n := A_n \setminus (\cup_{k < n} A_k)$. Alors $C_n \in \mathfrak{X}$, les ensembles sont disjoints et on a

$$f^{-1}(C_n) = \{G = y_n\} \setminus \cup_{k < n} \{G = y_k\} = \{G = y_n\}.$$

Définissons $h = \sum_{n \in \mathbb{N}} y_n 1_{C_n}$. Par construction, h est mesurable et $G = h(X)$.

Dans le cas général, il existe une suite (G_n) de variables aléatoires étagées qui converge vers G . Par ce qui précède, $G_n = h_n(X)$. Posons, $h := \limsup_n h_n$. Alors h est mesurable et $G = h(X)$. L'unicité est immédiate. □

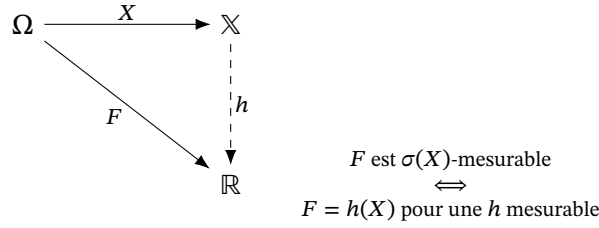


FIGURE 2.2 : Une variable aléatoire F est $\sigma(X)$ -mesurable si et seulement si elle se factorise par X , c'est-à-dire $F = h \circ X$ pour une fonction mesurable h (flèche en pointillé).

L'espérance conditionnelle comme une fonction

D'après le théorème de Doob (théorème 2.8), $\mathbf{E}[F | X]$ étant $\sigma(X)$ -mesurable par définition, il existe une fonction ψ , unique $\mathbf{loi}[X]$ -presque sûrement, telle que $\mathbf{E}[F | X] = \psi(X)$. En pratique, le calcul d'une espérance conditionnelle se ramène au calcul de cette fonction ψ . Nous allons donc introduire une notation pour ψ .

Définition 2.9. On note

$$\psi : E \ni x \mapsto \mathbf{E}[F | X = x],$$

l'unique (aux ensembles $\mathbf{loi}[X]$ -négligeables près) application telle que $\psi(X) = \mathbf{E}[F | X]$.

Remarque 2.10. Le lecteur prendra soin de noter que l'évènement $\{X = x\}$ est en général de mesure nulle. À part dans le cas discret décrit ci-dessous, l'expression $\mathbf{E}[F | X = x]$ ne s'interprète pas comme une espérance conditionnellement à un évènement.

Conditionnement discret

Ici, nous supposons que X est une variable aléatoire discrète, c'est-à-dire que $\mathbb{X}$ est au plus dénombrable et $\mathfrak{X} := \mathfrak{P}(\mathbb{X})$ et nous écrivons

$$\mu_x := \mathbf{P}[X = x], \quad \forall x \in \mathbb{X}.$$

Sans perte de généralité, nous supposons que $\mu_x > 0$ pour tout $x \in \mathbb{X}$. Nous avons la formule explicite suivante.

Théorème 2.11 (Conditionnement par rapport à une variable aléatoire discrète). *Pour $F \in \mathcal{L}^1(\mathfrak{B})$*

$$\mathbf{E}[F | X = x] = \frac{\mathbf{E}[F 1_{X=x}]}{\mu_x}.$$

Autrement dit

$$\mathbf{E}[F | X] = \sum_{x \in \mathbb{X}} \frac{1}{\mu_x} \mathbf{E}[F 1_{X=x}] 1_{X=x}.$$

Démonstration. Pour une fonction h , écrivons

$$\mathbf{E}[F h(X)] = \sum_{x \in \mathbb{X}} h(x) \mathbf{E}[F 1_{X=x}] = \mathbf{E}[h(X) \psi(X)],$$

avec

$$\psi(x) := \frac{\mathbf{E}[F 1_{X=x}]}{\mu_x}.$$

Par le théorème de Doob (théorème 2.8), cela montre que la variable aléatoire annoncée satisfait bien la propriété universelle (2.1). $\square$

Remarque 2.12. Dans les plus petites classes, vous aviez défini une nouvelle mesure de probabilités

$$\mathbf{P}[A | B] := \frac{\mathbf{P}[A \cap B]}{\mathbf{P}[B]},$$

dès que $\mathbf{P}[B] > 0$. L'espérance par rapport à cette mesure est donnée par

$$\mathbf{E}[F | B] = \frac{\mathbf{E}[F 1_B]}{\mathbf{P}[B]}.$$

En prenant $B := \{X = x\}$ pour un certain $x \in \mathbb{X}$, on retrouve exactement la formule du théorème. L'approche naïve des probabilités conditionnelles est consistante avec la théorie plus générale que nous développons ici.

2.2 Lois conditionnelles et noyaux de Markov

Motivation à partir de la loi jointe Considérons un autre espace mesuré $(\mathbb{Y}, \mathfrak{Y})$. Soient X et Y des variables aléatoires à valeurs dans $\mathbb{X}$ et $\mathbb{Y}$ respectivement, et $\varphi : \mathbb{X} \times \mathbb{Y} \rightarrow \mathbb{R}$ mesurable. Pour calculer en pratique $\mathbf{E}[\varphi(X, Y) | X]$, on cherche une fonction $\psi : \mathbb{X} \rightarrow \mathbb{R}$ mesurable telle que

$$\mathbf{E}[\varphi(X, Y)h(X)] = \mathbf{E}[\psi(X)h(X)], \quad \forall h : \mathbb{X} \rightarrow \mathbb{R} \text{ mesurable et borné.}$$

Dans ce cas, on a $\mathbf{E}[\varphi(X, Y) | X] = \psi(X)$, ou encore

$$\psi(x) = \mathbf{E}[\varphi(X, Y) | X = x].$$

Écrivons $\pi := \mathbf{loi}[(X, Y)] \in \mathcal{P}(\mathbb{X} \times \mathbb{Y})$ et $\mu := \mathbf{loi}[X] \in \mathcal{P}(\mathbb{X})$. La relation précédente s'écrit

$$\int \varphi(x, y)h(x)\pi(\mathrm{d}x\mathrm{d}y) = \int \psi(x)h(x)\mu(\mathrm{d}x).$$

2.2.1 NOYAUX DE MARKOV

Pour résoudre cette équation, nous cherchons à « factoriser » μ en dehors de π . Cela conduit à la notion de *noyau de probabilité*.

Définition 2.13 (Noyau de Markov). Soit $\mathbb{Y}$ un espace polonais et $(\mathbb{X}, \mathfrak{X})$ un espace mesurable. Un *noyau de Markov*, on dit aussi *noyau de transition*, de $\mathbb{X}$ dans $\mathbb{Y}$ est une application mesurable

$$p : \mathbb{X} \rightarrow \mathcal{P}(\mathbb{Y}),$$

l'espace $\mathcal{P}(\mathbb{Y})$ étant muni de la tribu borélienne associée à la topologie de la convergence étroite. De manière plus prosaïque, p est une application

$$p : \mathbb{X} \times \mathfrak{B}(\mathbb{Y}) \rightarrow [0, 1],$$

telle que

- (i) pour tout $x \in \mathbb{X}$, $p(x, \cdot)$ est une mesure de probabilités borélienne sur $\mathbb{Y}$;
- (ii) pour tout $A \in \mathfrak{B}(\mathbb{Y})$, $p(\cdot, A)$ est $\mathfrak{X}$ -mesurable.

Opérateurs associés

Nous allons faire agir le noyau q sur les fonctions et sur les mesures. Ce formalisme sera pratique pour étudier les chaînes de Markov.

Définition 2.14 (Opérateurs associés aux noyaux). Soit q un noyau de probabilité sur $\mathbb{X}$. L'opérateur associé, noté Q , agit :

(i) à gauche sur les fonctions φ mesurables et bornées par

$$(Q\varphi)(x) := \int \varphi(y)q(x, dy), \quad \forall x \in \mathbb{X};$$

(ii) à droite sur les mesures (positives non nécessairement finies) ν par

$$(\nu Q)(A) := \int q(x, A)\nu(dx), \quad \forall A \in \mathfrak{B}(\mathbb{Y}).$$

La vérification que $Q\varphi$ est bien une fonction mesurable bornée et que νQ est bien une mesure est laissé en exercice.

Nous définissons la multiplication des noyaux sur $\mathbb{X}$ et l'opérateur associé à la multiplication de deux noyaux est obtenus par composition (multiplication des opérateurs).

Lemme 2.15 (Propriété de semi-groupe des noyaux). Soient q et r deux noyaux de probabilité sur l'espace polonais $\mathbb{X}$. Alors,

$$qr(x, A) := \int r(y, A)q(x, dy),$$

est encore un noyau de probabilité. En notant QR l'opérateur associé on a bien $Q(R\varphi) = QR\varphi$ pour φ mesurable et bornée et $(\nu R)Q = \nu QR$.

Démonstration. Immédiat. □

2.2.2 LOI CONDITIONNELLES

Définition de la loi conditionnelle

Théorème 2.16 (Existence et unicité des lois conditionnelles). Soient $(\Omega, \mathbf{P})$ un espace de probabilité polonais, $(\mathbb{X}, \mathfrak{X})$ un espace mesurable et $X : \Omega \rightarrow \mathbb{X}$ une variable aléatoire. Écrivons $\mu := \mathbf{loi}[X] = X_{\#} \mathbf{P}$. Alors, il existe un noyau de probabilité p de $\mathbb{X}$ dans Ω tel que

$$\mathbf{P}[A] = \int p(x, A)\mu(dx), \quad \forall A \in \mathfrak{B}(\mathbb{X}), \quad (2.2)$$

De plus, p est essentiellement unique : si $\tilde{p}$ est un autre noyau qui satisfait (2.2) alors l'ensemble

$$\{x \in \mathbb{X} : p(x, \cdot) \neq \tilde{p}(x, \cdot)\},$$

est μ -négligeable.

Définition 2.17 (Loi conditionnelle). Avec les notations du théorème précédent on note

$$\mathbf{P}[A \mid X = x] := p(x, A), \quad \forall A \in \mathfrak{B}(\Omega), x \in \mathbb{X},$$

et on appelle cette probabilité la *loi conditionnelle sachant que $X = x$* . Lorsque $\mathbb{X} := \Omega$ et que $\mathfrak{X} := \mathfrak{F}$ est une sous tribu de $\mathfrak{B}(\Omega)$ et que X est l'identité, on parlera aussi de *loi conditionnelle sachant $\mathfrak{F}$* et on notera

$$\mathbf{P}[A \mid \mathfrak{F}](\omega) := p(\omega, A), \quad A \in \mathfrak{B}(\Omega), \omega \in \Omega.$$

Démonstration. Nous découpons la démonstration en plusieurs étapes. Le « gros morceau » qui est l'existence d'un noyau satisfaisant (2.2) se démontre d'une manière similaire à celle de l'espérance conditionnelle, via le théorème de Radon–Nikodym. C'est toutefois plus technique : il faut utiliser un théorème qui garantit l'existence des mesures de probabilité. Nous utilisons ici le théorème de Riesz–Markov, ce qui justifie de commencer par le cas compact.

Unicité Considérons une suite (ω_n) dense dans Ω et une énumération (q_n) des rationnels et

$$\mathfrak{B}_0(\Omega) := \{B(\omega_n, q_m) : (n, m)\}.$$

Soit p et $\tilde{p}$ deux noyaux qui satisfont (2.2), alors

$$\int (p(x, A) - \tilde{p}(x, A))\mu(dx) = 0, \quad \forall A \in \mathfrak{B}_0(\Omega).$$

Comme $\mathfrak{B}_0(\Omega)$ est dénombrable, on trouve que

$$\mu\{x \in \mathbb{X} : p(x, A) = \tilde{p}(x, A), \forall A \in \mathfrak{B}_0(\Omega)\} = 1.$$

Deux mesures de probabilités qui coïncident sur $\mathfrak{B}_0(\Omega)$ coïncident sur $\mathfrak{B}(\Omega)$ et on obtient l'unicité.

Existence dans le cas compact Commençons par montrer (2.2) dans le cas particulier où Ω est compact. Soit $G \in \mathcal{C}(\Omega)$ quitte à décomposer en partie négative et positive, prenons $G \geq 0$. Considérons la mesure $\nu := X_{\#}(G \mathbf{P})$. Soit $B \in \mathfrak{X}$ tel que $\mu(B) = \mathbf{P}[X \in B] = 0$. Alors

$$\nu(B) = \mathbf{E}[G 1_{X \in B}] = 0.$$

Donc, ν est absolument continue par rapport à μ et par le théorème de Radon–Nikodym (théorème B.12), on peut trouver une densité g telle que $\nu = g\mu$. La densité g est définie uniquement sur un $B_g \in \mathfrak{X}$ tel que $\mu(B_g) = 1$. Fixons un représentant particulier $g := 0$ en dehors de B_g .

Par compacité de Ω , il existe un ensemble dénombrable $\mathcal{H}$ dense dans $\mathcal{C}(\Omega)$. On peut supposer que $\mathcal{H}$ est un $\mathbb{Q}$ -espace vectoriel, qui contient les constantes et qui est stable par $\vee$ et $\wedge$. Considérons pour $x \in \mathbb{X}$, l'application

$$\Lambda_x : \mathcal{H} \ni G \mapsto g(x).$$

Pour tout $x \in B := \cap_{g \in \mathcal{H}} B_g$, qui est de mesure pleine puisque $\mathcal{H}$ est dénombrable, on a que Λ_x est une forme linéaire croissante et $\Lambda_x(1) = 1$. En particulier, Λ_x est continue avec $\|\Lambda_x\| \leq 1$. Par le théorème de prolongement des applications continues, pour tout $x \in B$, prolongeons Λ_x en une forme linéaire continue croissante et qui préserve 1. Par le théorème de représentation

de Riezs–Markov (théorème B.24), Λ_x se représente par une mesure de probabilité $\mathbf{P}_x$. Si $x \notin B$, alors on pose $\mathbf{P}_x = \mathbf{P}$ (ou n’importe quelle autre mesure de probabilité). Considérons la fonction

$$p(x, G) := \mathbf{E}_x[G], \quad \forall x \in \mathbb{X}, G \text{ borélienne bornée.}$$

Quand $G \in \mathcal{H}$, par construction $p(\cdot, G) = g$ est une application mesurable. Par convergence uniforme, $p(\cdot, G)$ donc mesurable lorsque $G \in \mathcal{C}(\Omega)$, puis par un argument de classe monotone, lorsque G est une variable aléatoire bornée. Ainsi la restriction de p aux indicatrices, c’est-à-dire

$$p(x, A) := \mathbf{P}_x[A], \quad \forall x \in \mathbb{X}, A \in \mathfrak{B}(\Omega),$$

définit bien un noyau de probabilité. Quand $G \in \mathcal{H}$,

$$\mathbf{E}[G] = \nu(\mathbb{X}) = \int g(x)\mu(dx) = \int \mathbf{E}_x[G]\mu(dx).$$

Cette identité s’étend de la même manière que précédemment à toute variable aléatoire bornée. En particulier pour $G := 1_A$ on a, comme attendu,

$$\mathbf{P}[A] = \int p(x, A)\mu(dx).$$

Existence dans le cas général Par le théorème de tension (théorème 1.24), considérons une suite croissante de compact (K_n) telle que $\mathbf{P}[\cup K_n] = 1$. Posons

$$\forall n \in \mathbb{N}, \begin{cases} \mathbf{P}_n := \mathbf{P}[\cdot \cap K_n]; \\ \mu_n := X_{\#} \mathbf{P}_n. \end{cases}$$

Alors $(\mathbf{P}_n)$ et (μ_n) sont des suites croissantes de mesures de masse ≤ 1 , de limite monotone respective $\mathbf{P}$ et μ . En particulier, par le théorème de Radon–Nikodym (théorème B.12), il existe une suite croissante de fonctions mesurables (g_n) telles que $\sup g_n = 1$ et

$$\mu_n = g_n \mu, \quad \forall n \in \mathbb{N}.$$

Par le cas précédent, ou plus exactement par une généralisation immédiate au cas des mesures de masse ≤ 1 , on trouve une famille (p_n) de noyaux de probabilité tels que

$$\mathbf{P}_n[A, X \in B] = \int_B p_n(x, A)g_n(x)\mu(dx), \quad \forall A \in \mathfrak{B}(\Omega), B \in \mathfrak{X}.$$

Par croissance de $(\mathbf{P}_n)$ et en utilisant de nouveau qu’une mesure sur Ω est caractérisée par ses valeurs sur un ensemble dénombrable de boréliens, il existe $B_0 \in \mathfrak{X}$ de mesure pleine tel que $(g_n(x)p_n(x, \cdot))_n$ est croissante pour tout $x \in B_0$. Posons

$$p(x, \cdot) := \begin{cases} \sup_n g_n(x)p_n(x, \cdot), & \text{si } x \in B_0 \\ \mathbf{P}, & \text{sinon.} \end{cases}$$

Fixons $A \in \mathfrak{B}(\Omega)$. D’une part, comme le supremum est sur un ensemble dénombrable, $x \mapsto p(x, A)$ est mesurable. D’autre part, par convergence monotone

$$\mathbf{P}[A] = \sup_n \mathbf{P}_n[A] = \int p(x, A)\mu(dx).$$

Ce qui montre que $p(x, \Omega) = 1$ μ -presque partout et que p est bien le noyau recherché (quitte à le modifier sur l’ensemble négligeable où ce n’est pas une probabilité). $\square$

La loi conditionnelle est portée par les fibres

Nous allons voir que si l'ensemble d'arrivée $\mathbb{X}$ satisfait notre hypothèse topologique usuelle, la loi conditionnelle admet la propriété naturelle : conditionnellement à $X = x$, alors $X = x$ presque sûrement.

Théorème 2.18. *Avec les notations du théorème 2.16, supposons de plus que $\mathbb{X}$ est polonais et que $\mathfrak{X} = \mathfrak{B}(\mathbb{X})$, alors, la loi conditionnelle satisfait*

$$\mathbf{P}[X = x \mid X = x] = 1, \quad \mu\text{-p.t. } x. \quad (2.3)$$

Démonstration. Considérons la variable aléatoire

$$\tilde{X} : \Omega \ni \omega \mapsto (\omega, X(\omega)) \in \Omega \times \mathbb{X}.$$

Alors, avec les notations de la démonstration précédente, en prenant $G := 1_A$ avec $A \in \mathfrak{B}(\Omega)$ et $B \in \mathfrak{B}(\mathbb{X})$

$$\mathbf{P}[\tilde{X} \in A \times B] = \mathbf{E}[1_A 1_{X \in B}] = \nu(B) = \int_B \mathbf{P}_x[A] \mu(dx).$$

Autrement dit $\mathbf{loi}[\tilde{X}] = \int \mathbf{P}_x \otimes \delta_x \mu(dx)$. Soit J une réunion dénombrable de compacts de Ω telle que $\mathbf{P}[J] = 1$. Si $\bar{J} := \tilde{X}(J)$ était borélien, on aurait alors que

$$1 = \mathbf{P}[\tilde{X} \in \bar{J}] = \int (\mathbf{P}_x \otimes \delta_x)[\bar{J}] \mu(dx),$$

et alors on aurait qu'il existe $B \in \mathfrak{B}(\mathbb{X})$ de mesure pleine tel que

$$(\mathbf{P}_x \otimes \delta_x)[\bar{J}] = 1, \quad \forall x \in B.$$

Or, la coupe selon $x \in \mathbb{X}$ de $\bar{J}$ est donné par

$$\bar{J}_x := \{X = x\}.$$

Puisque, par le théorème de Fubini,

$$(\mathbf{P}_x \otimes \delta_x)[\bar{J}] = \mathbf{P}_x[\bar{J}_x], \quad \forall x \in \mathbb{X},$$

cela permettrait de conclure. En général, on ne sait pas si $\bar{J}$ est borélien : c'est l'image d'un ensemble mesurable par une application mesurable. Toutefois, en utilisant que J est une réunion dénombrable de compacts, on peut montrer les deux points suivants :

- (i) Il existe une tribu $\mathfrak{B}_u(\Omega \times \mathbb{X})$ sur $\Omega \times \mathbb{X}$ telle que $\bar{J}$ est $\mathfrak{B}_u(\Omega \times \mathbb{X})$ -mesurable. Plus précisément, la tribu $\mathfrak{B}_u(\Omega \times \mathbb{X})$ est la plus petite tribu qui contient $\mathfrak{B}(\Omega \times \mathbb{X})$ et les ensembles Λ -négligeables pour toute mesure de probabilité borélienne Λ sur $\Omega \times \mathbb{X}$. On appelle $\mathfrak{B}_u(\Omega \times \mathbb{X})$, le complété universel de $\mathfrak{B}(\Omega \times \mathbb{X})$.
- (ii) Toute mesure de probabilité borélienne Λ sur $\Omega \times \mathbb{X}$ se prolonge de manière unique en une probabilité $\hat{\Lambda}$ sur $\mathfrak{B}_u(\Omega \times \mathbb{X})$. Plus précisément, l'application $\Lambda \mapsto \hat{\Lambda}$ est une bijection de l'ensemble des mesures de probabilité sur $\mathfrak{B}(\Omega \times \mathbb{X})$ vers celui de mesures de probabilité sur $\mathfrak{B}_u(\Omega \times \mathbb{X})$.

Ces deux points précédents sont des résultats fins de la théorie de la mesure et sont admis, on pourra consulter [DM75, Thm. 18] ou [Bog07, Chap. 6]. Ainsi quitte à relever $\mathbf{loi}[\tilde{X}]$ sur $\mathfrak{B}_u(\Omega \times \mathbb{X})$ on peut bien parler de $\mathbf{P}[\tilde{X} \in \bar{J}]$, et l'argument précédent est valide. $\square$

2.2.3 LIEN ENTRE LOI CONDITIONNELLE ET ESPÉRANCE CONDITIONNELLE

Comme on pouvait s'y attendre, tant par la construction que par la terminologie, il y a un lien fort entre loi conditionnelle et espérance conditionnelle.

Le théorème abstrait

Théorème 2.19 (L'espérance conditionnelle à partir de la loi conditionnelle). *Soit X une variable aléatoire défini sur un espace de probabilité polonais $(\Omega, \mathbf{P})$ et à valeurs dans un espace mesurable $(\mathbb{X}, \mathfrak{X})$. Alors, en notant, $\mathbf{P}_x := \mathbf{P}[\cdot | X = x]$ et $\mathbf{E}_x$ pour l'intégrale associée*

$$\mathbf{E}[F | X = x] = \mathbf{E}_x[F], \quad \forall F \in L^1.$$

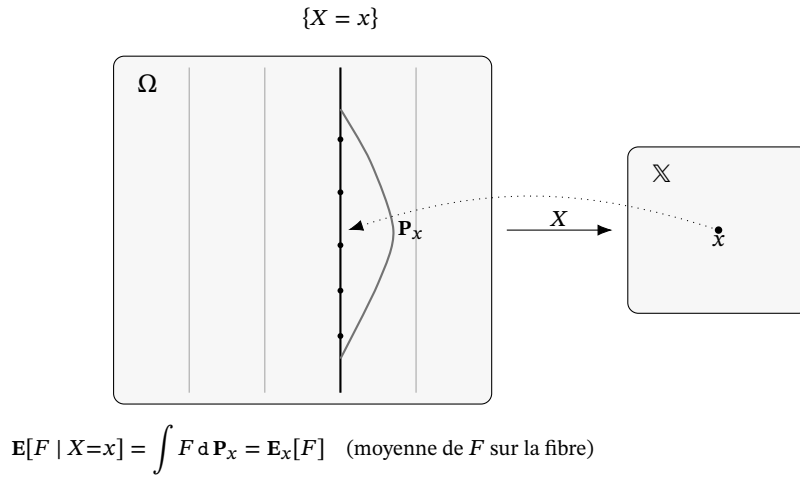


FIGURE 2.3 : Loi et espérance conditionnelles : X partitionne Ω en fibres $\{X = x\} = X^{-1}(x)$. La loi conditionnelle $\mathbf{P}_x$ est portée par la fibre et $\mathbf{E}[F | X = x]$ est la moyenne de F contre $\mathbf{P}_x$.

Remarque 2.20. Autrement dit, l'espérance conditionnelle est obtenue en intégrant contre la loi conditionnelle. Le lecteur gardera bien en tête que toutes ces identités sont comprises pour μ -presque tout x . En particulier, il serait plus correct de dire une version de l'espérance conditionnelle est obtenue en intégrant contre une version de la loi conditionnelle.

Remarque 2.21. Par construction l'application,

$$\mathbb{X} \ni x \mapsto \mathbf{P}_x \in \mathcal{P}(\Omega),$$

est mesurable. On peut donc considérer la variable aléatoire à valeurs dans $\mathcal{P}(\Omega)$, $\mathbf{P}_X$, définie par

$$\mathbf{P}_X(\omega) := \mathbf{P}_{X(\omega)}, \quad \forall \omega \in \Omega.$$

Le théorème précédent peut alors se lire dans sa version aléatoire

$$\mathbf{E}[F | X] = \mathbf{E}_X[F].$$

Remarque 2.22. Considérons une sous tribu $\mathfrak{F}$ de $\mathfrak{B}(\Omega)$ et X l'application identité de $(\Omega, \mathfrak{B}(\Omega))$ vers $(\Omega, \mathfrak{F})$. Alors $\mathbf{E}[F | X] = \mathbf{E}[F | \mathfrak{F}]$, et on voit que l'espérance conditionnelle par rapport à une tribu se retrouve également grâce à la probabilité conditionnelle par rapport à la tribu.

Pour aller plus loin 8. La construction de l'espérance conditionnelle ne demande aucune hypothèse topologique sur Ω . À l'inverse la construction de probabilité conditionnelle repose fortement sur le caractère polonais de Ω . Dans le cas où Ω n'est pas polonais, on peut montrer qu'il n'existe pas toujours de loi conditionnelle.

Démonstration. Par linéarité et le procédé d'approximation usuel, il suffit de montrer le résultat pour $F := 1_A$ et $A \in \mathfrak{B}(\Omega)$. Prenons $B \in \mathfrak{X}$. Par définition de la loi conditionnelle, avec $\mu := \mathbf{loi}[X]$,

$$\begin{aligned} \mathbf{E}[F 1_{X \in B}] &= \int_B \mathbf{E}_x[F] \mu(dx). \\ &= \mathbf{E}\left[1_B(X) \int F(\omega) p(X, d\omega)\right]. \end{aligned}$$

Par le lemme de Doob (théorème 2.8), la propriété universelle (2.1) est démontrée et cela conclut la démonstration. $\square$

Le théorème 2.19 admet une réciproque.

Théorème 2.23 (Noyau à partir de l'espérance conditionnelle). *Soient X une variable aléatoire définie sur un espace polonais $(\Omega, \mathbf{P})$ et à valeurs dans un espace mesurable $(\mathbb{X}, \mathfrak{X})$ et p un noyau de probabilité tel que*

$$\mathbf{E}[F | X = x] = \int F(\omega) p(x, d\omega), \quad \forall F \in L^1.$$

Alors, p est (une version de) la loi conditionnelle de $\mathbf{P}$ sachant X .

Démonstration. Soient $A \in \mathfrak{B}(\Omega)$ et $B \in \mathfrak{B}(\mathbb{X})$. Alors, en notant $\mu := \mathbf{loi}[X]$,

$$\mathbf{P}[A, X \in B] = \mathbf{E}[1_B(X) \mathbf{E}[1_A | X]] = \mathbf{E}[1_B(X) p(X, A)] = \int_B p(x, A) \mu(dx).$$

$\square$

Application à des couples de variables aléatoires

Nous allons maintenant spécifier les résultats généraux au cas des espaces produits. Ce cas particulier est suffisamment remarquable et important pour que l'on explicite complètement les résultats dans ce cas. Dans toute cette partie, $\mathbb{X}$ et $\mathbb{Y}$ sont des espaces polonais. On écrit $e_{\mathbb{X}}$ et $e_{\mathbb{Y}}$ pour les projections de $\mathbb{X} \times \mathbb{Y} \rightarrow \mathbb{X}$ et de $\mathbb{X} \times \mathbb{Y} \rightarrow \mathbb{Y}$.

Définition 2.24 (Marginale). Pour $\pi \in \mathcal{P}(\mathbb{X} \times \mathbb{Y})$, la *marginale en $\mathbb{X}$* de π est la mesure $\mu := (e_{\mathbb{X}})_\# \pi$. Autrement dit, si $(X, Y) \sim \pi$, alors $\mu := \mathbf{loi}[X]$. De manière plus élémentaire,

$$\mu(B) := \pi(B \times \mathbb{Y}), \quad \forall B \in \mathfrak{B}(\mathbb{X}).$$

Théorème 2.25 (Désintégration des lois jointes). *Soit $\pi \in \mathcal{P}(\mathbb{X} \times \mathbb{Y})$, de marginale en $\mathbb{X}$ donnée par μ , il existe un unique (au sens du théorème 2.16) noyau de probabilité q de $\mathbb{X}$ vers $\mathbb{Y}$ tel que*

$$\pi(A \times B) = \int_A q(x, B) \mu(dx), \quad \forall A \in \mathfrak{B}(\mathbb{X}), B \in \mathfrak{B}(\mathbb{Y}).$$

De plus

$$q(x, \{X = x\}) = 1, \quad \mu\text{-p.t. } x.$$

Démonstration. Il suffit d'appliquer les théorèmes 2.16 et 2.18 à l'espace de probabilité polonais $(\mathbb{X} \times \mathbb{Y}, \pi)$. Le noyau q est alors la loi conditionnelle de π sachant X , où X est la projection sur $\mathbb{X}$. $\square$

Définition 2.26 (Loi conditionnelle d'une variable aléatoire par rapport à une autre). Avec les notations du théorème 2.25, considérons un couple de variables aléatoires $(X, Y) \sim \pi$. On appelle q la loi conditionnelle de Y sachant X .

Réécrivons notre théorème 2.19 dans le cadre d'un couple de variables aléatoires.

Théorème 2.27 (Espérance conditionnelle pour un couple). Soit (X, Y) un couple de variables aléatoires à valeurs dans $\mathbb{X} \times \mathbb{Y}$. En notant q la loi conditionnelle de Y sachant X , on a pour tout $\varphi : \mathbb{X} \times \mathbb{Y} \rightarrow \mathbb{R}$ mesurable et π -intégrable,

$$\mathbf{E}[\varphi(X, Y) | X] = \int \varphi(X, y)q(X, dy).$$

Autrement dit

$$\mathbf{E}[\varphi(X, Y) | X = x] = \int \varphi(x, y)q(x, dy), \quad \mu\text{-p.t. } x.$$

En particulier, si on se restreint à $\varphi : \mathbb{Y} \rightarrow \mathbb{R}$, alors

$$\mathbf{E}[\varphi(Y) | X] = \int \varphi(y)q(X, dy).$$

2.2.4 CALCULS EXPLICITES D'ESPÉRANCES CONDITIONNELLES

Nous donnons le calcul explicite de la loi de Y sachant X dans quelques cas importants.

Cas des variables aléatoires indépendantes

Les variables aléatoires X et Y sont indépendantes si et seulement si la loi du couple se tensorise $\pi = \mu \otimes \nu$, avec $\mu \in \mathcal{P}(\mathbb{X})$ et $\nu \in \mathcal{P}(\mathbb{Y})$. Dans ce cas, on vérifie aisément que la loi conditionnelle est donné par $q(x, dy) = \nu(dy)$.

Le corollaire immédiat de cette observation avec le théorème 2.27 est le suivant.

Corollaire 2.28 (Conditionnement d'un couple de variables aléatoires indépendantes). Soient X et Y deux variables aléatoires à valeurs dans $\mathbb{X}$ et $\mathbb{Y}$ et indépendantes et $\varphi : \mathbb{X} \times \mathbb{Y} \rightarrow \mathbb{R}$ mesurable telle que $\varphi(X, Y)$ est intégrable. Alors,

$$\mathbf{E}[\varphi(X, Y) | X = x] = \mathbf{E}[\varphi(x, Y)].$$

Insistons sur l'utilisation de ce théorème pour calculer l'espérance conditionnelle de $F := \varphi(X, Y)$ sachant X dans le cas indépendant.

1. On fixe $x \in \mathbb{X}$ et on calcule

$$\psi(x) := \mathbf{E}[\varphi(x, Y)].$$

Cette quantité est « juste » une intégrale : il y a de bonnes chances de savoir la calculer !

2. Ensuite, on a $\mathbf{E}[F | X] = \psi(X)$.

Cas des variables aléatoires à densité

Passons maintenant à un autre cas important. Choisissons des variables aléatoires X et Y à valeurs dans $\mathbb{R}^d$ et $\mathbb{R}^m$ respectivement telles que le couple (X, Y) admet une densité ρ par rapport à la mesure de Lebesgue. Cela signifie que pour toute fonction borélienne bornée φ , on a

$$\mathbf{E}[\varphi(X, Y)] = \int \varphi(x, y) \rho(x, y) dx dy.$$

Dans ce cas, X admet une densité donnée par

$$m(x) := \int \rho(x, y) dy.$$

Ayant fait cette observation, on vérifie qu'un noyau de désintégration est donné par

$$p(x, dy) := \frac{1}{m(x)} \rho(x, y) dy.$$

Remarque 2.29. Pour être parfaitement rigoureux il faut faire attention à ce qu'il se passe sur l'ensemble $m = 0$. Comme cet ensemble est μ -négligeable ça n'a pas d'importance dans les calculs et nous laissons ces détails sous le tapis.

Comme avant, nous en déduisons un corollaire du théorème 2.27.

Corollaire 2.30 (Conditionnement des variables aléatoires à densité). *Soit $\varphi : \mathbb{R}^d \times \mathbb{R}^m \rightarrow \mathbb{R}$ borélienne telle que $\varphi(X, Y)$ soit intégrable. Alors,*

$$\mathbf{E}[\varphi(X, Y) | X = x] = \frac{1}{m(x)} \int \varphi(x, y) \rho(x, y) dy.$$

Cas des vecteurs gaussiens

Soit $(X_1, \dots, X_d, Y)$ un vecteur gaussien centré, avec le raccourci $X := (X_1, \dots, X_d)$. La loi jointe $\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R})$ est gaussienne, ainsi que la loi marginale $\mu \in \mathcal{P}(\mathbb{R}^d)$. Nous allons voir qu'un noyau de désintégration, et donc l'espérance conditionnelle se calcule à l'aide d'une projection orthogonale. Plus précisément, considérons $\mathbb{V}$ le sous-espace vectoriel de dimension fini de $\mathcal{L}^2(\mathbf{P})$ engendré par $X_1, \dots, X_d$. Ainsi, il existe des réels $t_1, \dots, t_d$ et $Z \in \mathbb{V}^\perp$ tels que

$$Y = Z + \sum_{i=1}^d t_i X_i.$$

Corollaire 2.31 (Conditionnement des vecteurs gaussiens). *Avec les notations précédentes, un noyau de désintégration est donné par*

$$p(x, \cdot) = \mathcal{N}\left(\sum_{i=1}^d t_i x_i, \mathbf{E}[Z^2]\right), \quad x = (x_1, \dots, x_d) \in \mathbb{R}^d.$$

Autrement dit,

$$\mathbf{E}[\varphi(Y) | X = x] = \mathbf{E}\left[\varphi\left(\sum_{i=1}^d t_i x_i + Z\right)\right].$$

En particulier $\mathbf{E}[Y | X]$ coïncide avec la projection orthogonale de Y sur $\mathbb{V}$.

Démonstration. Puisque $(X, Z) = (X, Y - \sum_{i=1}^d t_i X_i)$, (X, Z) est aussi un vecteur gaussien centré. Puisque $Z \perp \mathbb{V}$, on trouve que $\mathbf{E}[ZX_i] = 0$ pour $i = 1, \dots, d$, et donc, puisque nous sommes dans un cadre gaussien, Z est indépendant de X et donc aussi de $\sum_{i=1}^d t_i X_i$. Le reste vient de la formule pour les variables aléatoires indépendantes (corollaire 2.28). $\square$

Remarque 2.32. En pratique, il suffit de calculer les (t_i) qui s'obtiennent à partir de la matrice de covariance. En effet, par la propriété universelle

$$\mathbf{E}[YX_i] = \mathbf{E}[\mathbf{E}[Y | X]X_i] = \sum_{j=1}^d t_j \mathbf{E}[X_i X_j], \quad i = 1, \dots, d.$$

En posant $C := \mathbf{E}[X \otimes X]$ et $\vec{v} = (\mathbf{E}[X_i Y])$ il faut trouver une solution au système linéaire $\vec{v} = C\vec{t}$ où l'inconnu est $\vec{t}$. Nous verrons en TD des exemples concrets.

2.3 Applications : les théorèmes ergodiques de Birkhoff et de Choquet

Nous allons montrer que les outils probabilistes de l'espérance conditionnelle et des lois conditionnelles permettent d'obtenir des informations précises sur des systèmes dynamiques, qui ne sont pas forcément probabilistes (voir la feuille de TD pour des exemples). Plus précisément, on s'intéresse à l'évolution d'un système sous les itérations répétées d'une même transformation. Nous verrons également que ces théorèmes sont très utiles pour étudier les chaînes de Markov.

Définition 2.33 (Système dynamique). Un *système dynamique* est un espace polonais Ω muni d'une transformation borélienne $\theta : \Omega \rightarrow \Omega$.

Définition 2.34 (Mesure invariante). Pour un système dynamique (Ω, θ) , on dit que $\mathbf{P} \in \mathcal{P}(\Omega)$ est *invariante* si $\theta_{\#} \mathbf{P} = \mathbf{P}$. On note $\mathcal{I} = \mathcal{I}(\theta)$ l'ensemble des mesures invariantes.

Autrement dit si $X \sim \mathbf{P}$ alors $\theta(X) \sim \mathbf{P}$. La loi $\mathbf{P}$ est vue comme une mesure d'équilibre du système récurrent $X_{n+1} = \theta(X_n)$.

Définition 2.35 (Tribu invariante). La *tribu invariante* associée au système dynamique (Ω, θ) est la tribu

$$\mathfrak{i} := \{A \in \mathfrak{B}(\Omega) : \theta^{-1}A = A\}.$$

Une variable aléatoire F est *invariante* si $F \circ \theta = F$. Le lecteur vérifiera aisément que F est invariante si et seulement si elle est $\mathfrak{i}$ -mesurable.

Le théorème suivant caractérise le comportement statistique moyen des trajectoires d'un système dynamique invariant.

Théorème 2.36 (Théorème de Birkhoff). Soit $(\Omega, \mathbf{P}, \theta)$ un système dynamique invariant et $\mathfrak{i}$ la tribu invariante associée. Alors, pour tout $F \in L^1(\mathbf{P})$

$$\frac{1}{n} \sum_{k=0}^{n-1} F \circ \theta^k \xrightarrow[n \rightarrow \infty]{p.s.} \mathbf{E}[F | \mathfrak{i}]. \quad (2.4)$$

La démonstration repose sur l'intuition suivante : si la limite existait, elle devrait être invariante. Le candidat naturel pour la limite est « la meilleure approximation invariante de F », c'est-à-dire $\mathbf{E}[F | \mathfrak{i}]$. Nous formalisons cette intuition à la fin de cette section en donnant une démonstration complète du théorème. Nous allons d'abord étudier plusieurs conséquences et concepts liés à ce théorème.

2.3.1 MESURES ERGODIQUES

Tel que formulé, le théorème de Birkhoff (théorème 2.36) reste d'une utilité limitée. En toute généralité, on ne sait pas calculer $\mathbf{E}[F | \mathfrak{i}]$. Dans le cas particulier où $\mathfrak{i}$ est triviale, l'espérance conditionnelle devient alors une vraie espérance. Cela motive la définition suivante.

Définition 2.37 (Mesure ergodique). Une mesure invariante $\mathbf{P}$ est *ergodique* si $\mathbf{P}[I] \in \{0, 1\}$ pour tout $I \in \mathfrak{i}$. On note $\mathcal{E} \subset \mathcal{F}$ l'ensemble des mesures ergodiques.

Lemme 2.38. Soient $\mathbf{P}$ ergodique et $F \in L^1(\mathbf{P})$. Alors,

$$\mathbf{E}[F | \mathfrak{i}] = \mathbf{E}[F].$$

Démonstration. Posons $\bar{F} := \mathbf{E}[F | \mathfrak{i}]$. Définissons

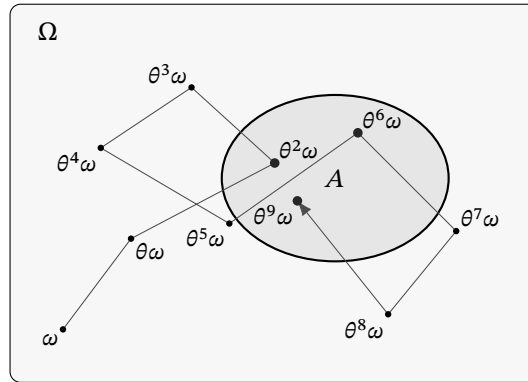
$$A_{>} := \{\bar{F} > \mathbf{E}[F]\}, \quad A_{<} := \{\bar{F} < \mathbf{E}[F]\}, \quad A_{=} := \{\bar{F} = \mathbf{E}[F]\}.$$

Ces trois ensembles forment une partition $\mathfrak{i}$ -mesurable de Ω . Donc par hypothèse d'ergodicité seulement un des trois ensembles est de mesure 1 et les autres sont de mesures 0. Si on avait $\mathbf{P}[A_{>}] = 1$, alors $\mathbf{E}[\bar{F}] = \mathbf{E}[F] > \mathbf{E}[F]$. Absurde. De la même manière, $\mathbf{P}[A_{<}] = 0$. $\square$

Une conséquence immédiate du lemme précédent avec le théorème de Birkhoff (théorème 2.36) est le raffinement suivant. Il est d'une importance pratique suffisante pour que nous l'écrivions explicitement.

Théorème 2.39 (Théorème ergodique de Birkhoff). Soit $(\Omega, \mathbf{P}, \theta)$ un système dynamique ergodique. Alors, pour tout $F \in L^1(\mathbf{P})$

$$\frac{1}{n} \sum_{k=0}^{n-1} F \circ \theta^k \xrightarrow[n \rightarrow \infty]{p.s.} \mathbf{E}[F]. \quad (2.5)$$



$$\underbrace{\frac{1}{n} \sum_{k=0}^{n-1} 1_A(\theta^k \omega)}_{\text{fraction du temps passé dans } A} \xrightarrow[n \rightarrow \infty]{} \underbrace{\mathbf{P}[A]}_{\text{masse de } A}$$

FIGURE 2.4 : Le long d'une trajectoire $(\theta^k \omega)$, la fraction du temps passée dans une région A converge presque sûrement vers sa masse $\mathbf{P}[A]$. Les gros points gras marquent les instants où la trajectoire visite A .

Parmi les mesures invariantes, nous allons caractériser les mesures ergodiques. Nous aurons besoin du résultat suivant : deux mesures ergodiques distinctes ne se voient pas.

Lemme 2.40. Deux mesures de probabilité ergodiques distinctes sont toujours étrangères.

Démonstration. Soient $\mathbf{P}_1 \neq \mathbf{P}_2 \in \mathcal{E}$. On peut trouver un borélien $A \subset \Omega$ tel que $\mathbf{P}_1(A) > \mathbf{P}_2(A)$. Par (2.5), il existe Ω_1 et Ω_2 avec $\mathbf{P}_i(\Omega_i) = 1$ tel que, pour tout $\omega \in \Omega_i$

$$\frac{1}{n} \sum_{k=0}^{n-1} 1_{\theta_k(\omega) \in A} \xrightarrow{n \rightarrow \infty} \mathbf{P}_i(A).$$

Donc $\Omega_1 \cap \Omega_2 = \emptyset$, et $\mathbf{P}_1$ et $\mathbf{P}_2$ sont étrangères. $\square$

Passons au résultat suivant qui caractérise complètement les mesures ergodiques.

Théorème 2.41 (Théorème de Choquet, caractérisation des mesures ergodiques). *Les points extrémaux de l'ensemble des mesures de probabilité invariantes sont exactement les mesures de probabilité ergodiques. De plus, toute mesure de probabilité invariante peut s'écrire comme une combinaison convexe de mesures ergodiques. Autrement dit, pour tout $\mathbf{Q} \in \mathcal{I}$, il existe une mesure de probabilité $\lambda := \lambda_{\mathbf{Q}}$ sur $\mathcal{E}$ tel que*

$$\mathbf{Q}(A) = \int \mathbf{P}(A) \lambda(d\mathbf{P}).$$

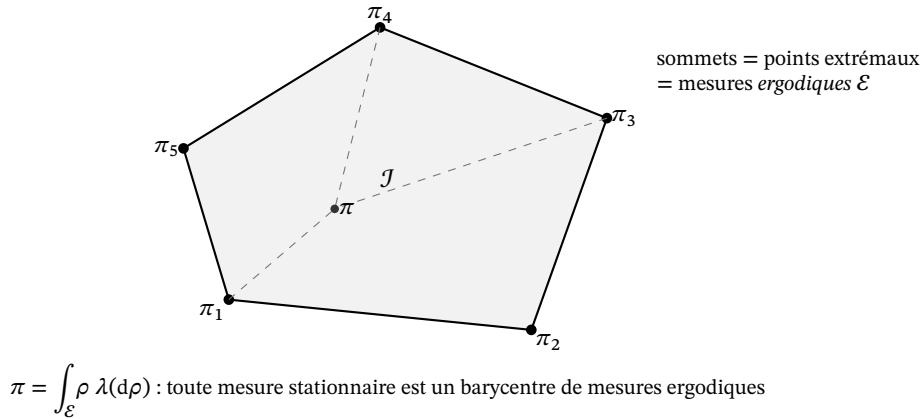


FIGURE 2.5 : L'ensemble $\mathcal{I}$ des mesures stationnaires est convexe ; ses points extrémaux (les sommets) sont exactement les mesures ergodiques $\mathcal{E}$, et tout $\pi \in \mathcal{I}$ se décompose comme barycentre de mesures ergodiques.

Remarque 2.42. Un point x d'un convexe K est *extrémal* s'il ne peut s'écrire comme une combinaison convexe non triviale d'autres éléments de K . C'est-à-dire, que si $x = (1-t)x_0 + tx_1$ avec $t \in (0, 1)$, alors $x_0 = x_1 = x$. Le théorème de Krein-Milman nous garantit que si K est un convexe compact, alors tout point de K peut s'écrire comme une combinaison convexe des points extrémaux de K . Ici, sauf si Ω est compact, nous n'avons en général pas de compacité pour $\mathcal{I}$ et nous utilisons donc une construction purement probabiliste.

Démonstration. Montrons d'abord que les mesures ergodiques sont extrémales. Prenons $\mathbf{P} \in \mathcal{E}$. Considérons une décomposition $\mathbf{P} = (1-t)\mathbf{P}_1 + t\mathbf{P}_2$ avec $t \in (0, 1)$. On a alors

$$\mathbf{P} = (1-t)\mathbf{P}_1 + t\mathbf{P}_2.$$

Pour tout $A \in \mathcal{I}$, $\mathbf{P}(A) \in \{0, 1\}$, et donc nécessairement $\mathbf{P}_1(A) = \mathbf{P}_2(A) = \mathbf{P}(A)$. Donc $\mathbf{P}_1$ et $\mathbf{P}_2$ sont ergodiques. Puisque $\mathbf{P}_1, \mathbf{P}_2 \ll \mathbf{P}$, des mesures ergodiques distinctes sont singulières (lemme 2.40) on a que $\mathbf{P} = \mathbf{P}_1 = \mathbf{P}_2$.

Montrons maintenant que les mesures extrémales sont ergodiques, par contraposée. Soit $\mathbf{P} \in \mathcal{F} \setminus \mathcal{E}$. Alors, il existe $A \in \mathfrak{i}$ tel que $\mathbf{P}(A) = t \in (0, 1)$. Posons

$$\mathbf{P}_1 := \frac{1}{t} \mathbf{P}(A \cap \cdot), \quad \mathbf{P}_2 = \frac{1}{1-t} \mathbf{P}(A^C \cap \cdot).$$

En utilisant que A et son complémentaire sont invariants, on trouve que $\mathbf{P}_1$ et $\mathbf{P}_2 \in \mathcal{F}$. Comme $\mathbf{P} = t \mathbf{P}_1 + (1-t) \mathbf{P}_2$, $\mathbf{P}$ n'est pas extrême.

Finissons par la décomposition ergodique d'une mesure invariante. Considérons le noyau conditionnel

$$q(\omega, A) := \mathbf{Q}(A \mid \mathfrak{i})(\omega).$$

Alors $q(\omega, \cdot)$ est invariante presque sûrement (exercice!) et satisfait pour $A \in \mathfrak{i}$

$$q(\omega, A) = \mathbf{Q}(A \mid \mathfrak{i})(\omega) = 1_A(\omega).$$

Donc $q(\omega, \cdot)$ est aussi presque sûrement ergodique. Considérons l'application

$$f : \Omega \ni \omega \mapsto q(\omega, \cdot) \in \mathcal{E},$$

qui est bien mesurable par définition des probabilités conditionnelles. Avec $\lambda := f_{\#} \mathbf{Q}$, on a bien

$$\mathbf{Q}(A) = \int q(\omega, A) \mathbf{Q}(d\omega) = \int \mathbf{P}(A) \lambda(d\mathbf{P}).$$

□

2.3.2 DÉMONSTRATION DU THÉORÈME DE BIRKHOFF

La démonstration repose sur le lemme maximal suivant, résultat intéressant en soi.

Théorème 2.43 (Théorème ergodique maximal). *Avec les mêmes notations que le théorème 2.36, posons*

$$S_n := \sum_{k=1}^n F \circ \theta_k, \quad \text{et} \quad S_n^* := \max\{0, S_1, S_2, \dots, S_n\}.$$

Alors, pour tout $n \in \mathbb{N}^*$

$$\mathbf{E}[F 1_{\{S_n^* > 0\}}] \geq 0.$$

Remarque 2.44. Plus généralement, on peut montrer que pour tout $n \in \mathbb{N}^*$

$$\mathbf{E}[F 1_{\{S_n^* > \lambda\}}] \geq \lambda \mathbf{P}[S_n^* > \lambda].$$

Démonstration. Prenons par convention $S_0 = S_0^* := 0$. Soient des entiers $n \geq k$, en utilisant que $S_N^* \geq S_k$ presque sûrement, on trouve que

$$F + S_n^* \circ \theta \geq F + S_k \circ \theta = S_{k+1}.$$

Puis en prenant le maximum en $k \geq n$, il vient

$$F \geq \max\{S_1, \dots, S_n\} - S_n^* \circ \theta.$$

Comme sur $\{S_n^* > 0\}$ le premier maximum coïncide avec S_n^* , on obtient

$$F 1_{\{S_n^* > 0\}} \geq 1_{\{S_n^* > 0\}} (S_n^* - S_n^* \circ \theta). \quad (2.6)$$

Puisque, d'un part

$$1_{\{S_n^* > 0\}} S_n^* \circ \theta \leq S_n^* \circ \theta,$$

et d'autre part

$$S_n^* = S_n^* 1_{\{S_n^* > 0\}},$$

en intégrant l'inégalité (2.6), on trouve que

$$\mathbf{E}[F 1_{\{S_n^* > 0\}}] \geq \mathbf{E}[S_n^* - S_n^* \circ \theta].$$

L'espérance de droite vaut 0 par invariance. $\square$

Démonstration du théorème 2.36. Montrons d'abord (2.4) quand $\pi \in \mathcal{F}$. Puisque $\mathbf{E}[F \mid \mathfrak{i}]$ est invariante, quitte à remplacer F par $F - \mathbf{E}[F \mid \mathfrak{i}]$, on peut supposer que $\mathbf{E}[F \mid \mathfrak{i}] = 0$. Posons

$$L := \limsup_{n \rightarrow \infty} \frac{1}{n} S_n.$$

Il suffit de montrer que $L \leq 0$, car en appliquant l'argument à $-F$ on trouvera que la limite inférieure est ≥ 0 ce qui donne que la limite existe et vaut 0. Fixons $\varepsilon > 0$ et posons

$$\begin{aligned} A_\varepsilon &:= \{L > \varepsilon\}, \\ F_\varepsilon &:= (F - \varepsilon) 1_{A_\varepsilon}, \\ S_{n,\varepsilon} &:= \sum_{k=0}^{n-1} F_\varepsilon \circ \theta^k, \\ S_{n,\varepsilon}^* &:= \max\{0, S_{1,\varepsilon}, \dots, S_{n,\varepsilon}\}. \end{aligned}$$

La suite $(\{S_{n,\varepsilon}^* > 0\})_n$ est croissante et satisfait donc

$$\bigcup_{n \in \mathbb{N}} \{S_{n,\varepsilon}^* > 0\} = \left\{ \sup_n S_{n,\varepsilon} > 0 \right\} =: B_\varepsilon.$$

Puisque $F_\varepsilon 1_{A_\varepsilon^C} = 0$, $S_{n,\varepsilon}^* = 0$ sur A_ε^C , on trouve que $B_\varepsilon \subset A_\varepsilon$. Ensuite, comme la borne supérieure est toujours plus grande que la limite supérieure, on trouve que $A_\varepsilon \subset B_\varepsilon$ et donc $A_\varepsilon = B_\varepsilon$. En appliquant le lemme maximal (théorème 2.43) et le théorème de convergence dominée, on trouve que

$$0 \leq \mathbf{E}[F_\varepsilon 1_{\{S_{n,\varepsilon}^* > 0\}}] \xrightarrow{n \rightarrow \infty} \mathbf{E}[F_\varepsilon 1_{A_\varepsilon}] = \mathbf{E}[F 1_{A_\varepsilon}] - \varepsilon \mathbf{P}[A_\varepsilon].$$

Par construction L est invariant, c'est-à-dire $L \circ \theta = L$, et donc $A_\varepsilon \in \mathfrak{i}$, par la propriété universelle de l'espérance conditionnelle (2.1), on trouve

$$\mathbf{E}[F 1_{A_\varepsilon}] = \mathbf{E}[\mathbf{E}[F \mid \mathfrak{i}] 1_{A_\varepsilon}] = 0.$$

On a donc montré que $0 \leq -\varepsilon \mathbf{P}[A_\varepsilon]$ et donc $\mathbf{P}[A_\varepsilon] = 0$. Comme c'est vrai pour tout $\varepsilon > 0$, cela donne que $L \leq 0$ et conclut la démonstration. $\square$

Exercices fondamentaux

Les exercices ci-dessous, tirés du livret d'exercices, constituent l'entraînement minimal recommandé sur ce chapitre.

Espérance conditionnelle : premiers calculs. Exercices 20, 22, 21.

Cas discret. Exercices 26, 38, 39.

Cas à densité. Exercices 46, 47.

Cas indépendant. Exercices 43, 42.

Cas gaussien. Exercices 44.

Noyaux. Exercices 31, 32, 33, 36.

Théorème de Birkhoff. Exercices 50, 51, 53.

Deuxième partie

Chaînes de Markov

Chaînes de Markov sur un espace quelconque

Nous arrivons maintenant au thème central de ce cours : les chaînes de Markov (à temps discret). Il s'agit d'une classe spéciale de processus indexé par $\mathbb{N}$, c'est-à-dire une suite de variables aléatoires (X_n) , que l'on interprète comme l'évolution d'un système aléatoire. On s'autorise de la dépendance *contrôlée* : l'état au temps $n + 1$ dépend éventuellement de X_n et d'un aléa indépendant, mais pas des $X_{n-1}, \dots, X_0$. On dit souvent que c'est un système sans mémoire : l'évolution de X_n vers X_{n+1} est déterminée par X_n , mais pas par ce qu'il s'est passé avant.

Dans tout ce chapitre on se place sur un espace métrique complet et séparable $(\mathbb{X}, d)$. Le cas le plus important est le cas des espaces dénombrables, mais nous donnons ici une définition générale. Les variables aléatoires (X_n) prendront leur valeur dans $\mathbb{X}$. On définit aussi

$$\mathfrak{F}_n := \sigma(X_0, \dots, X_n), \quad \forall n \in \mathbb{N}.$$

3.1 Définitions et premières propriétés

3.1.1 DÉFINITION

Formalisons cette intuition d'absence de mémoire à l'aide de l'espérance conditionnelle. Rappelons la définition d'un noyau de probabilité et son utilisation dans le calcul d'espérance conditionnelle vu en sections 2.2 et 2.2.4. Nous allons considérer des noyaux de probabilité de $\mathbb{X}$ vers $\mathbb{X}$.

Définition 3.1. Un *noyau markovien* est un noyau de probabilité q de $\mathbb{X}$ vers $\mathbb{X}$.

Définition 3.2 (Chaîne de Markov). Soient q un noyau markovien et (X_n) un processus à valeurs dans $\mathbb{X}$. Le processus (X_n) est une *chaîne de Markov de noyau de transition* q si, pour toute fonction φ mesurable bornée et $n \in \mathbb{N}$,

$$\mathbb{E}[\varphi(X_{n+1}) \mid X_n = x, X_{n-1} = x_{n-1}, \dots, X_0 = x_0] = \int \varphi(y)q(x, dy).$$

Autrement dit, on demande que

$$\mathbb{E}[\varphi(X_{n+1}) \mid \mathfrak{F}_n] = Q\varphi(X_n).$$

Remarque 3.3. En termes de noyau, on demande l'existence d'un noyau de désintégration de la loi de $(X_0, X_1, \dots, X_{n+1})$ le long de la loi de $(X_0, \dots, X_n)$ et que ce noyau qui a priori dépend a priori de $X_0, \dots, X_n$ ne dépendent en fait que de X_n .

Remarque 3.4. On pourrait aussi définir des chaînes de Markov non homogène en temps dans ce cas on voudrait une famille de noyaux de transition sur $\mathbb{X}$ (q_n) et que, pour toute fonction φ mesurable bornée et $n \in \mathbb{N}$,

$$\mathbb{E}[\varphi(X_{n+1}) \mid X_n = x, X_{n-1} = x_{n-1}, \dots, X_0 = x_0] = \int \varphi(y)q_n(x, dy).$$

Nous n'abordons pas plus le sujet durant ce cours.

Cas des espaces finis

On considère un espace fini $\mathbb{X} := \{y_1, \dots, y_N\}$. On identifie une fonction $\varphi : \mathbb{X} \rightarrow \mathbb{R}$ au vecteur colonne de $\mathbb{R}^{\mathbb{X}} = \mathbb{R}^N$ en l'identifiant à $(\varphi(y_1), \dots, \varphi(y_N))^T$. On identifie aussi une mesure ν au vecteur ligne $(\nu(y_1), \dots, \nu(y_N))$. En particulier l'espace des mesures de probabilités s'identifie à un sous-espace fermé de $[0, 1]^N$ qui est donc compact par Heine–Borel. C'est aussi une conséquence générale du théorème de Prokhorov (théorème 1.38). Avec ces identifications on trouve que

$$\int \varphi d\nu = \sum_{i=1}^N \varphi(y_i) \nu(y_i) = \nu \varphi.$$

Matrices stochastiques Dans ce cas les noyaux de $\mathbb{X}$ vers $\mathbb{X}$ prennent une forme simple.

Définition 3.5 (Matrices stochastiques). Une matrice de taille $N \times N$ est *stochastique* si chacune de ses lignes représentent une mesure de probabilités sur $\mathbb{X}$. Autrement dit $Q = (q(x, y))_{x, y \in \mathbb{X}}$ est une matrice stochastique si et seulement si

- (i) $q(x, y) \geq 0$, pour tout x et $y \in \mathbb{X}$;
- (ii) $\sum_{y \in \mathbb{X}} q(x, y) = 1$.

Une matrice stochastique telle que Q^T est aussi stochastique est dite *bistochastique*.

Remarque 3.6. Par définition, la fonction $\mathbf{1} := (1, \dots, 1)^T$ est un vecteur propre à droite de Q . La propriété suivante est évidente et connecte les matrices stochastiques au noyau de transition.

Proposition 3.7 (Les matrices stochastiques sont les noyaux markoviens). *Il y a identification entre matrices stochastiques et noyau markovien. Plus précisément, si Q est une matrice stochastique, alors*

$$q(x, A) := \sum_{y \in A} Q(x, y), \quad \forall x \in \mathbb{X},$$

définit un noyau markovien. Réciproquement, si q est un noyau markovien alors

$$Q(x, y) := q(x, \{y\}), \quad \forall x, y \in \mathbb{X},$$

définit une matrice stochastique. Il y a une correspondance entre le calcul d'opérateur abstrait développé pour les noyaux sur un espace général et le calcul matriciel. C'est-à-dire, l'opérateur Q associé au noyau q est simplement la matrice stochastique Q .

Graphe d'incidence Une représentation visuelle utile de la matrice stochastique Q et de la chaîne de Markov associée se fait à l'aide d'un graphe orienté.

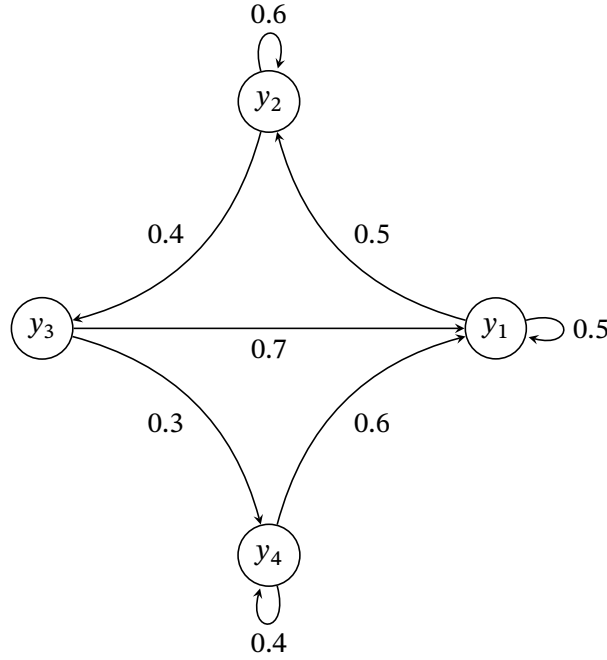
Définition 3.8 (Graphe d'incidence). Étant donné une matrice stochastique Q , le graphe d'incidence associé est le graphe orienté tel que :

- (i) les sommets sont les éléments de $\mathbb{X}$,
- (ii) il y a une arête orientée de x vers y si et seulement si $Q(x, y) > 0$,
- (iii) l'arête $x \rightarrow y$ du graphe est décorée de la valeur numérique $Q(x, y)$.

Exemple 6 (Le graphe d'incidence d'une matrice stochastique de taille 4). Considérons la matrice stochastique suivante

$$Q := \frac{1}{10} \begin{pmatrix} 5 & 5 & 0 & 0 \\ 0 & 6 & 4 & 0 \\ 7 & 0 & 0 & 3 \\ 6 & 0 & 0 & 4 \end{pmatrix}.$$

Alors le graphe d'incidence associé est donné par



3.1.2 UNICITÉ, EXISTENCE ET CARACTÉRISATION

Loi d'une chaîne de Markov

Le théorème suivant montre que l'on peut calculer explicitement la loi d'une chaîne de Markov unique à partir de son noyau de transition et de sa loi initiale. En particulier, à condition initiale donnée et noyau de transition donné il y a unicité en loi d'une chaîne de Markov.

Théorème 3.9 (Le noyau de Markov caractérise la loi). Soient (X_n) un processus et q de noyau de probabilité. Notons $\mu = \text{loi}[X_0]$. Le processus (X_n) est une chaîne de Markov de noyau de transition q si et seulement si pour tout $A_0, \dots, A_n \in \mathfrak{X}$

$$\mathbf{P}[X_0 \in A_0, \dots, X_n \in A_n] = \int_{\prod A_i} q(x_{n-1}, dx_n) q(x_{n-2}, dx_{n-1}) \dots q(x_0, dx_1) \mu(dx_0). \quad (3.1)$$

Dans ce cas, on a en particulier, pour tout $n \in \mathbb{N}$

$$\text{loi}[X_n] = \mu Q^n.$$

Remarque 3.10. (3.1) est équivalente à ce que pour toutes fonctions mesurables et bornées $\varphi_0, \dots, \varphi_n$

$$\mathbf{E} \left[\prod_{i=1}^n \varphi_i(X_i) \right] = \int \varphi_0(x_0) \dots \varphi_n(x_n) q(x_{n-1}, dx_n) q(x_{n-2}, dx_{n-1}) \dots q(x_0, dx_1) \mu(dx_0).$$

En particulier, une notation moins rigoureuse mais plus compréhensible de (3.1) est

$$\mathbf{P}[X_0 \in dx_0, \dots, X_n \in dx_n] = q(x_{n-1}, dx_n) q(x_{n-2}, dx_{n-1}) \dots q(x_0, dx_1) \mu(dx_0).$$

Démonstration. Soient $\varphi_0, \dots, \varphi_n$ des fonctions mesurables et bornées. Supposons (X_n) est markovien. On procède par récurrence. (3.1) est trivialement vraie pour $n = 0$. Si on l'a montrée pour un $n \in \mathbb{N}$ alors

$$\mathbf{E} \left[\prod_{i=0}^{n+1} \varphi_i(X_i) \right] = \mathbf{E} \left[\prod_{i=0}^n \varphi_i(X_i) \mathbf{E}[\varphi_{n+1}(X_{n+1}) \mid \mathfrak{F}_n] \right] = \mathbf{E} \left[\prod_{i=0}^n \varphi_i(X_i) Q \varphi_{n+1}(X_n) \right].$$

Puis, il suffit d'appliquer l'hypothèse de récurrence.

Inversement, par théorème 2.8, montrer que (X_n) est markovien revient à établir que pour toutes fonctions mesurables et bornées $h : \mathbb{X}^n \rightarrow \mathbb{R}$ et $\varphi : \mathbb{X} \rightarrow \mathbb{R}$

$$\mathbf{E}[h(X_0, \dots, X_{n-1})\varphi(X_n)] = \mathbf{E}[h(X_0, \dots, X_{n-1})Q\varphi(X_{n-1})].$$

Par un argument de classe monotone, c'est équivalent à montrer que

$$\mathbf{E}\left[\prod_{i=0}^{n-1} \varphi_i(X_i)\varphi(X_n)\right] = \mathbf{E}\left[\prod_{i=0}^{n-1} \varphi_i(X_i)Q\varphi(X_{n-1})\right],$$

qui est immédiat par (3.1). □

Les chaînes de Markov comme système récurrent

La définition des chaînes de Markov ou le critère (3.1) sont particulièrement difficiles à vérifier. Heureusement, il existe le critère suivant qui est celui que l'on utilisera toujours en pratique.

Théorème 3.11 (Les systèmes récurrents sont des chaînes de Markov). *Soient :*

- (i) X_0 une variable aléatoire de loi μ ;
- (ii) (U_n) une suite de variables aléatoires indépendantes et identiquement distribuées à valeurs dans un espace mesurable $(\mathbb{U}, \mathcal{U})$ indépendante de X_0 ;
- (iii) et $f : \mathbb{X} \times \mathbb{U} \rightarrow \mathbb{X}$ mesurable.

Alors le processus défini récursivement par $X_{n+1} = f(X_n, U_{n+1})$ est une chaîne de Markov de noyau de transition

$$q(x, A) = \mathbf{P}[f(x, U_0) \in A].$$

Remarque 3.12. Ce théorème est particulièrement important car il permet de montrer en pratique qu'un processus donné est une chaîne de Markov.

Démonstration. Prenons φ mesurable et bornée. Par corollaire 2.28, on trouve que

$$\mathbf{E}[\varphi(X_{n+1}) | \mathfrak{F}_n] = \psi(X_n),$$

avec $\psi(x) = \mathbf{E}[\varphi(f(x, U_0))]$. En posant $q(x, \cdot) := \mathbf{loi}[f(x, U_0)]$, on a bien $\psi = Q\varphi$. □

Exemples importants

Exemple 7 (Systèmes dynamiques déterministes). Soit $f : \mathbb{X} \rightarrow \mathbb{X}$ une fonction mesurable et X_0 une variable aléatoire. Alors le processus défini par $X_n = f^n(X_0)$ est une chaîne de Markov de noyau de transition

$$q(x, \cdot) = \delta_{f(x)}.$$

Dans ce cas, $Q\varphi = \varphi \circ f$ et $\mu Q = f_{\#}\mu$.

Exemple 8 (Variables aléatoires indépendantes et identiquement distribuées). Soit (X_n) une suite de variables aléatoires indépendantes et de même loi μ . C'est une chaîne de Markov de noyau de transition $q(x, \cdot) = \mu$.

Exemple 9 (Marches aléatoires). Soit μ une mesure de probabilité sur $\mathbb{R}^d$. Alors le processus défini par $X_n := \varepsilon_1 + \dots + \varepsilon_n$, où les (ε_n) sont des variables aléatoires indépendantes de même loi μ est une chaîne de Markov de noyau de transition

$$q(x, A) = \mu(A - x),$$

où $A - x := \{z - x : z \in A\}$. On l'appelle la *marche aléatoire* sur $\mathbb{R}^d$ de loi de saut μ (commencée en 0). Dans ce cas, Q est l'opérateur de convolution par μ . Le cas le plus élémentaire est celui de la *marche aléatoire simple et symétrique sur $\mathbb{Z}^d$* où l'on choisit les (ε_n) indépendants et uniforme sur $\{\pm e_i\}$ où (e_i) est la base canonique de $\mathbb{R}^d$.

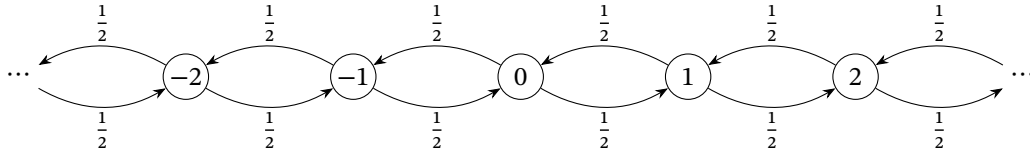


FIGURE 3.1 : Le graphe d'incidence de la marche aléatoire simple sur $\mathbb{Z}$

Exemple 10 (Marches aléatoires sur des groupes). Soit G un groupe topologique, muni de sa tribu borélienne. On fixe μ une mesure de probabilité sur G . Le processus défini par $X_n := \varepsilon_0 \varepsilon_1 \dots \varepsilon_n$ est une chaîne de Markov de noyau de transition

$$q(x, A) = \mu(x^{-1}A),$$

avec $x^{-1}A := \{x^{-1}z : z \in A\}$. On l'appelle la *marche aléatoire invariante à gauche* sur G de loi de saut μ (commencée en ε_0). On peut de la même manière définir les marches aléatoires invariantes à droite. Dans le cas abélien, comme à l'exemple précédent, il n'y a pas de différence.

Exemple 11 (Marche aléatoire simple sur un graphe). Un graphe non orienté avec sommets $\mathbb{X}$ est la donnée d'un sous ensemble d'arêtes $\mathbb{A} \subset \{\{x, y\} : x, y \in \mathbb{X}\}$. On dit que les sommets x et y sont connectés si $\{x, y\} \in \mathbb{A}$. Les voisins de x sont donnés par l'ensemble $\mathbb{A}_x := \{y \in \mathbb{X} : \{x, y\} \in \mathbb{A}\}$. On définit un processus (X_n) sur $\mathbb{X}$ en choisissant X_0 suivant une loi de probabilité donnée, puis à l'étape n on choisit pour X_{n+1} un voisin de X_n pris uniformément au hasard de X_n . Alors (X_n) est une chaîne de Markov de noyau de transition

$$q(x, \cdot) := \text{Uniforme}(\mathbb{A}_x), \quad x \in \mathbb{X}.$$

On l'appelle *marche aléatoire simple* sur le graphe. Dans ce cas le graphe d'incidence a les mêmes sommets que le graphe $\mathbb{X}$.

Existence

Le résultat suivant est important et montre que l'on peut toujours construire une chaîne de Markov. La preuve du cas général repose sur un théorème d'extension de Kolmogorov. La preuve dans le cas dénombrable repose sur une construction ad hoc d'un système récurrent. Nous la donnons car elle est intéressante.

Théorème 3.13 (Existence des chaînes de Markov). Soit $\mu \in \mathcal{P}(\mathbb{X})$ et q un noyau markovien. Il existe un espace de probabilité $(\Omega, \mathfrak{B}, \mathbf{P})$ et une suite (X_n) de variables aléatoires définies sur Ω et à valeurs dans $\mathbb{X}$ tels que (X_n) est une chaîne de Markov de noyau de transition q et de loi initiale μ .

Démonstration dans le cas dénombrable. On suppose que $\mathbb{X}$ est dénombrable et on considère $\{y_1, y_2, \dots\}$ une énumération de ses éléments. Il suffit de poser, avec $d := \text{Card}(\mathbb{X}) \in \mathbb{N} \cup \{\infty\}$,

$$f(x, u) := \sum_{k=1}^d y_k 1_{\{\sum_{1 \leq j < k} Q(x, y_j) < u \leq \sum_{1 \leq j \leq k} Q(x, y_j)\}}.$$

Cette fonction est mesurable. Considérons maintenant l'espace $\Omega := \mathbb{X} \otimes [0, 1]^{\mathbb{N}}$ équipé de la tribu produit et la mesure $\mathbf{P} := \mu \otimes \text{Leb}^{\mathbb{N}}$. Pour $\omega = (x, u_1, u_2, \dots) \in \Omega$, considérons les variables aléatoires

$$X_0(\omega) := x, \quad \text{et} \quad U_i(\omega) := u_i, \quad \forall i \in \mathbb{N}^*.$$

Définissons par récurrence

$$X_{n+1} := f(X_n, U_{n+1}), \quad \forall n \in \mathbb{N}.$$

Alors par le théorème 3.11, (X_n) est une chaîne de Markov dont la matrice de transition est donnée par

$$\mathbf{P}[f(x, U) = y_k] = \text{Leb}\left\{ \sum_{1 \leq j < k} Q(x, y_j) < u \leq \sum_{1 \leq j \leq k} Q(x, y_j) \right\} = Q(x, y_k).$$

□

Démonstration dans le cas général. Dans le cas général, on pose $\Omega := \mathbb{X}^{\mathbb{N}}$ muni de la tribu produit. D'abord, posons

$$X_n(\omega) := x_n, \quad \forall \omega = (x_1, x_2, \dots) \in \Omega.$$

Ensuite, pour tout $n \in \mathbb{N}$, on définit $\mathbf{P}_n$ la mesure de probabilité définie sur $\mathbb{X}^{n+1}$ à partir de (3.1), vue comme une mesure de probabilité sur $\mathbb{X}^{\mathbb{N}}$. Alors, $(\mathbf{P}_n)$ est un système projectif. Par le théorème d'extension de Kolmogorov (théorème B.29), on trouve $\mathbf{P} \in \mathcal{P}(\Omega)$ telle que sous $\mathbf{P}$, $(X_1, \dots, X_n) \sim \mathbf{P}_n$. En particulier, sous $\mathbf{P}$ $(X_1, \dots, X_n)$ satisfait (3.1). □

3.2 Analyse stochastique

3.2.1 PROCESSUS CANONIQUE

L'espace canonique

Par le théorème d'existence (théorème 3.13), nous savons associer à tout noyau q et condition initiale μ , une chaîne de Markov. Il peut y avoir plusieurs chaînes de Markov de même matrice de transition, mais d'après le théorème 3.9, elles ont toutes la même loi. Il suffit donc d'en étudier une en particulier que nous allons maintenant construire de manière canonique. Pour cela posons $\Omega := \mathbb{X}^{\mathbb{N}}$ l'espace des suites à valeurs dans $\mathbb{X}$, équipé de la tribu produit, c'est-à-dire de la tribu cylindrique.

Applications coordonnées

Pour tout $n \in \mathbb{N}$, on définit la variable aléatoire X_n à valeur dans $\mathbb{X}$ par

$$X_n(\omega) := \omega_n.$$

Chaîne de Markov canonique

Théorème 3.14 (Existence de la chaîne de Markov canonique). *Soit q un noyau sur $\mathbb{X}$. Pour tout $x \in \mathbb{X}$, il existe une unique mesure de probabilité $\mathbf{P}_x$ sur Ω tel que, sous $\mathbf{P}_x$, (X_n) est une chaîne de Markov de noyau de transition q et de condition initiale x .*

Démonstration. Par le théorème 3.13, on peut considérer une espace $(\Omega', \mathfrak{B}, \mathbf{P}')$ et une variable aléatoire $X' : \Omega' \rightarrow \Omega$ tels que sous $\mathbf{P}'$, X' est une chaîne de Markov commencée en x et de noyau de transition q . On pose $\mathbf{P}_x$ la mesure image de $\mathbf{P}'$ par X' . Par définition des mesures images $\mathbf{P}_x[X_0 = x] = \mathbf{P}'[X'_0 = x] = 1$ et, de la même manière,

$$\mathbf{P}_x[X_0 \in A_0, \dots, X_n \in A_n] = \mathbf{P}'[X' \in A_0 \times \dots \times A_n \times \mathbb{X} \times \mathbb{X} \times \dots].$$

Puisque le terme de droite est donné par (3.1), on conclut par le théorème 3.9 que $\mathbf{P}_x$ convient. $\square$

Autres conditions initiales

Pour tout $\mu \in \mathcal{P}(\mathbb{X})$, on pose

$$\mathbf{P}^\mu(A) := \int \mathbf{P}_x[A] \mu(dx).$$

On vérifie que $\mathbf{P}^\mu$ est encore une mesure de probabilité sur Ω et que sous $\mathbf{P}^\mu$, (X_n) est une chaîne de Markov de matrice de transition Q et de condition initiale μ . De manière plus générale si ν est une mesure σ -finie on définit de la même manière

$$\mathbf{P}^\nu := \int \mathbf{P}_x \nu(dx).$$

Notons que $\mathbf{P}^\nu$ est une probabilité (resp. finie) si et seulement si ν est une probabilité (resp. finie).

3.2.2 REFORMULATION DE LA PROPRIÉTÉ DE MARKOV

Nous allons maintenant interpréter un peu différemment la définition des chaînes de Markov. Informellement, on veut dire que conditionnellement à $(X_0, X_1, \dots, X_n)$, $(X_n, X_{n+1}, \dots)$ est encore une chaîne de Markov de noyau de transition q , mais commencée en X_n .

Opérateurs de translation

Pour tout $k \in \mathbb{N}^*$, on définit les opérateurs canoniques de translation $\theta_k : \Omega \rightarrow \Omega$, donnés par

$$\theta_k(\omega) = (\omega_k, \omega_{k+1}, \dots), \quad \omega \in \Omega.$$

Espérance aléatoire

Pour toute variable aléatoire F positive (ou intégrable), l'expression $\mathbf{E}_x[F]$ se calcule à l'aide de (3.1). Ceci montre que $x \mapsto \mathbf{E}_x[F]$ est mesurable, et on peut définir la variable aléatoire $\mathbf{E}_{X_n}[F]$.

Propriété de Markov faible

Théorème 3.15 (Propriété de Markov faible). Soient $n \in \mathbb{N}$, F une variable aléatoire positive $\mathfrak{F}_n$ -mesurable et G une variable aléatoire positive. Alors,

$$\mathbf{E}_x[F(G \circ \theta_n)] = \mathbf{E}_x[F \mathbf{E}_{X_n}[G]], \quad \forall x \in \mathbb{X}.$$

Autrement dit

$$\mathbf{E}_x[G \circ \theta_n | \mathfrak{F}_n] = \mathbf{E}_{X_n}[G], \quad \mathbf{P}_x - p.s., \quad \forall x \in \mathbb{X}.$$

Démonstration. Par un argument de classe monotone, il suffit de prouver le résultat pour

$$F := \varphi_0(X_0) \dots \varphi_n(X_n), \quad G := \psi_0(X_0) \dots \psi_p(X_p),$$

pour tout $p \in \mathbb{N}$ et (φ_i) et (ψ_j) des fonctions boréliennes. Alors

$$G \circ \theta_n = \psi_0(X_n) \dots \psi_p(X_{n+p}).$$

Notons

$$\mu_k^x(dx_1 \dots dx_k) := q(x_{k-1}, dx_k) \dots q(x_2, dx_1)q(x, dx_1), \quad \forall k \in \mathbb{N}^*.$$

Par la formule (3.1), on trouve d'abord

$$\mathbf{E}_x[F(G \circ \theta_n)] = \varphi_0(x) \int \varphi_1 \otimes \dots \otimes \varphi_{n-1} \otimes (\varphi_n \psi_0) \otimes \psi_1 \dots \otimes \psi_p d\mu_{n+p}^x.$$

De la même manière, on trouve que

$$\mathbf{E}_{X_n}[G] = \psi_0(X_n) \int \psi_1 \otimes \dots \otimes \psi_p d\mu_p^{X_n}.$$

Puis finalement

$$\mathbf{E}_x[F \mathbf{E}_{X_n}[G]] = \varphi_0(x) \int \varphi_1 \otimes \dots \otimes \varphi_{n-1} \otimes (\varphi_n \psi_0) \otimes \psi_1 \dots \otimes \psi_p d\nu_{n+p}^x,$$

avec

$$\nu_{n+p}^x(dx_1 \dots dx_{n+p}) := \mu_p^{X_n}(dx_{n+1} \dots dx_{n+p})\mu_n^x(dx_1 \dots dx_n).$$

Mais en développant l'expression de μ_n et μ_p on voit que $\nu_{n+p}^x = \mu_{n+p}^x$. □

Remarque 3.16. De la même manière on pourrait montrer un résultat avec $\mathbf{E}^\mu$ pour tout $\mu \in \mathcal{P}(\mathbb{X})$.

3.2.3 RENFORCEMENT DE LA PROPRIÉTÉ DE MARKOV

Nous allons maintenant donner un renforcement de la propriété précédente où on autorise des translations par un temps aléatoire. Dans le cas des chaînes de Markov la preuve de ce renforcement est en fait trivial. Cette partie va toutefois nous permettre

- (i) de se familiariser avec les temps d'arrêt qui jouent un rôle important dans l'étude des processus stochastiques généraux;
- (ii) de formaliser la propriété de renouvellement des chaînes de Markov.

Temps d'arrêt

Introduisons les temps aléatoires qui satisfont les bonnes propriétés de mesurabilité pour établir ce renforcement.

Définition 3.17 (Temps d'arrêt). Une variable aléatoire τ à valeurs dans $\mathbb{N} \cup \{\infty\}$ est un *temps d'arrêt* si pour tout $k \in \mathbb{N}$, $\{\tau = k\} \in \mathfrak{F}_k$.

Définition 3.18 (Événements antérieurs à un temps d'arrêt). Soit τ un temps d'arrêt, on pose

$$\mathfrak{F}_\tau := \{A \in \mathfrak{B} : A \cap \{\tau = k\} \in \mathfrak{F}_k, \forall k \in \mathbb{N}\}.$$

Proposition 3.19 (Tribu associée à un temps d'arrêt). Si τ est un temps d'arrêt, alors $\mathfrak{F}_\tau$ est une tribu.

Démonstration. Vérifions les trois axiomes des tribus.

- (i) Comme tous les $\mathfrak{F}_k$ sont des tribus, elles contiennent l'ensemble vide, donc $\mathfrak{F}_\tau$ aussi.
- (ii) Pour $A \in \mathfrak{B}$, $A^C \cap \{\tau = k\} = \{\tau = k\} \cap (\{\tau = k\} \cap A)^C$. Donc si $A \in \mathfrak{F}_\tau$, on a que $A \cap \{\tau = k\} \in \mathfrak{F}_k$ par définition et donc son complémentaire aussi. Comme τ est un temps d'arrêt, on trouve que $A^C \cap \{\tau = k\} \in \mathfrak{F}_k$, donc $A^C \in \mathfrak{F}_\tau$.
- (iii) Pour une suite d'événements (A_n) ,

$$\left(\bigcap_{n \in \mathbb{N}} A_n \right) \cap \{\tau = k\} = \bigcap_{n \in \mathbb{N}} [A_n \cap \{\tau = k\}],$$

ce qui montre que $\mathfrak{F}_\tau$ est stable par intersection dénombrable.

□

Temps d'atteinte

Définition 3.20 (Temps d'atteinte). Étant donné $A \in \mathfrak{B}(\mathbb{X})$ on pose

$$\tau_A := \inf\{n \in \mathbb{N} : X_n \in A\}.$$

Pour $x \in \mathbb{X}$, on écrit simplement $\tau_x := \tau_{\{x\}}$.

Lemme 3.21. La variable aléatoire τ_A est un temps d'arrêt.

Démonstration. Soit $n \in \mathbb{N}$,

$$\{\tau_A = n\} = \bigcap_{i=0}^{n-1} \{X_i \notin A\} \cap \{X_n \in A\},$$

qui est $\mathfrak{F}_n$ -mesurable.

□

Propriété de Markov forte

Les temps d'arrêt forment la bonne classe de temps aléatoire pour généraliser la propriété de Markov à des temps aléatoires.

Théorème 3.22 (Propriété de Markov forte). *Soient τ un temps d'arrêt, F et G deux variables aléatoires positives avec F qui est $\mathfrak{F}_\tau$ -mesurable. Alors,*

$$\mathbf{E}_x[1_{\{\tau < \infty\}} F(G \circ \theta_\tau)] = \mathbf{E}_x[1_{\{\tau < \infty\}} F \mathbf{E}_{X_\tau}[G]].$$

Démonstration. Soit $n \in \mathbb{N}$. Comme F est $\mathfrak{F}_\tau$ -mesurable, par définition, on trouve que $F 1_{\{\tau=n\}}$ est $\mathfrak{F}_n$ -mesurable. Donc, par la propriété de Markov faible du théorème 3.15, on trouve

$$\mathbf{E}_x[1_{\{\tau=n\}} F(G \circ \theta_\tau)] = \mathbf{E}_x[1_{\{\tau=n\}} F(G \circ \theta_n)] = \mathbf{E}_x[1_{\{\tau=n\}} F \mathbf{E}_{X_n}[G]].$$

Il suffit ensuite de sommer sur $n \in \mathbb{N}$. □

Donnons tout de suite un corollaire important de ce résultat qui généralise l'idée qu'une chaîne de Markov est indépendante de son passé.

Corollaire 3.23 (Propriété de renouvellement des chaînes de Markov). *Soient x et $y \in \mathbb{X}$, tels que $\mathbf{P}_x[\tau_y < \infty] > 0$. Sous $\mathbf{P}_x[\cdot \mid \tau_y < \infty]$, $\theta_{\tau_y} = (X_{\tau_y}, X_{\tau_y+1}, \dots)$ est indépendante de $\mathfrak{F}_{\tau_y}$ et de loi $\mathbf{P}_y$.*

Démonstration. Soit F une variable aléatoire positive et $\mathfrak{F}_{\tau_y}$ mesurable et G une variable aléatoire positive. Par le théorème 3.22,

$$\begin{aligned} \mathbf{E}_x[F(G \circ \theta_{\tau_y}) 1_{\{\tau_y < \infty\}}] &= \mathbf{E}_x[F \mathbf{E}_{X_{\tau_y}}[G] 1_{\{\tau_y < \infty\}}] = \mathbf{E}_x[F \mathbf{E}_y[G] 1_{\{\tau_y < \infty\}}] \\ &= \mathbf{E}_x[F 1_{\{\tau_y < \infty\}}] \mathbf{E}_y[G]. \end{aligned}$$

Ce qui donne le résultat annoncé. □

3.3 Analyse harmonique

Commençons par une partie plus analytique, où nous mettons en perspective des objets motivés par une étude probabiliste avec la terminologie venue la théorie des équations aux dérivées partielles linéaires. Ce cours introductif est trop court pour développer pleinement ce parallèle.

3.3.1 FONCTIONS HARMONIQUES

Fonctions harmoniques sur tout l'espace

Définition 3.24 (Fonction harmonique). On dit qu'une fonction mesurable $u : \mathbb{X} \rightarrow [0, \infty]$ est

- (a) *surharmonique* ou *excessive* si $u \geq Qu$;
- (b) *harmonique* si $u = Qu$.

Remarque 3.25. Le fait d'être positive mais possiblement non finie fait partie de la définition. On pourrait considérer des fonctions harmoniques de signe quelconques mais en faisant attention aux questions de finitude — pour éviter d'écrire $\infty - \infty$!

Remarque 3.26. On peut aussi de la même manière définir les fonctions *sousharmoniques*.

Pour aller plus loin 9 (Lien avec le laplacien). En analyse, on dit que u est *harmonique* si $\Delta u = 0$. Ici le rôle du laplacien est joué par l'opérateur « discret » $I - Q$, que l'on appelle le *générateur* de la chaîne. Si Q est la marche aléatoire simple sur $h\mathbb{Z}$ avec $h > 0$ alors u est harmonique si et seulement si

$$u(x + h) + u(x - h) - 2u(x) = 0,$$

qui est une approximation discrète de $u''(x)/2 = 0$. En utilisant les processus stochastiques à temps continu, on peut définir le *mouvement brownien* qui est un *processus de Markov* à valeur dans $\mathbb{R}^d$ et on peut montrer que son générateur est exactement l'opérateur différentiel

$$u \mapsto \frac{u''}{2}.$$

Dans ce cas, les fonctions harmoniques au sens des probabilités coïncident avec les fonctions harmoniques au sens de l'analyse.

Harmonicité sur un domaine

Demander à une fonction d'être harmonique sur tout $\mathbb{X}$ est parfois trop restrictif.

Définition 3.27 (Fonctions harmonique dans un domaine). Soient $A \in \mathfrak{B}(X)$. Une fonction mesurable $u : \mathbb{X} \rightarrow [0, \infty]$ est *harmonique dans A* si $u = Qu$ sur A .

Remarque 3.28. Même si l'équation $u = Qu$ ne doit être vérifiée que sur A , il faut que u soit définie partout. En effet, calculer $Qu(x)$ pour x dans A peut nécessiter de connaître les valeurs de u en dehors de A .

Pour aller plus loin 10 (Fonctions harmoniques et martingales). Si (X_n) est une chaîne de Markov et u est une fonction harmonique dans A alors $(u(X_{n \wedge \tau_A}))$ est une martingale où τ_A est le *temps d'atteinte de A* .

3.3.2 TEMPS D'ATTEINTE ET PROBLÈME DE DIRICHLET

Un problème classique étant donné une chaîne de Markov est de se demander quel est le temps moyen à attendre avant qu'un certain événement ne se produise.

Les problèmes de calcul de temps d'atteinte peuvent se ramener à des calculs de fonctions harmoniques qui sont des problèmes linéaires.

Théorème 3.29 (Les valeurs de sortie sont des solutions au problème de Dirichlet). Soit $A \in \mathfrak{B}(\mathbb{X})$ et $\varphi : A \rightarrow [0, \infty]$ mesurable. On pose

$$u(x) := \mathbf{E}_x[\varphi(X_{\tau_A})1_{\tau_A < \infty}], \quad \forall x \in \mathbb{X}.$$

Alors, u est la plus petite solution du problème de Dirichlet suivant

$$\begin{cases} (I - Q)u = 0, & \text{dans } \mathbb{X} \setminus A; \\ u = \varphi, & \text{dans } A. \end{cases} \quad (3.2)$$

Remarque 3.30. On dit que les fonctions solutions du problème de Dirichlet (3.2) sont harmoniques dans $\mathbb{X} \setminus A$. Autrement dit, le théorème donne une recette pour, étant donné une fonction sur A , construire une extension harmonique de cette fonction. De plus, l'extension ainsi construite est la plus petite.

Démonstration. D'abord, quitte à travailler avec $\varphi \wedge k$ puis faire $k \rightarrow \infty$, on peut supposer que φ est bornée. Montrons que u est solution. Si $x \in A$, alors sous $\mathbf{P}_x$, $\tau_A = 0$ et donc $u(x) = \varphi(x)$. Si $x \notin A$, posons $F := \varphi(X_{\tau_A})1_{\tau_A < \infty}$ de sorte que $u(x) = \mathbf{E}_x[F]$. Calculons

$$F \circ \theta_1 = \varphi(X_{\tau_A} \circ \theta_1)1_{\tau_A \circ \theta_1 < \infty} = \sum_{n \in \mathbb{N}} \varphi(X_n \circ \theta_1)1_{\tau_A \circ \theta_1 = n}.$$

Comme sous $\mathbf{P}_x$, $\tau_A > 0$, $\tau_A \circ \theta_1 = \tau_A - 1$. Donc, $\mathbf{P}_x$ -p.s.,

$$F \circ \theta_1 = \sum_{n \in \mathbb{N}} \varphi(X_{n+1})1_{\tau_A = n+1} = \sum_{n \in \mathbb{N}^*} \varphi(X_n)1_{\tau_A = n} = F,$$

car $\mathbf{P}_x[\tau_A = 0] = 0$. Par définition $u(X_1) = \mathbf{E}_{X_1}[F]$ donc par la propriété de Markov du théorème 3.22 (en fait ici on n'utilise que la propriété faible)

$$Qu(x) = \mathbf{E}_x[u(X_1)] = \mathbf{E}_x[\mathbf{E}_{X_1}[F]] = \mathbf{E}_x[F \circ \theta_1] = \mathbf{E}_x[F] = u(x).$$

Montrons que c'est la plus petite solution positive. La positivité est claire. Soit v une autre solution positive. Posons

$$u_n(x) := \mathbf{E}_x[\varphi(X_{\tau_A})1_{\tau_A \leq n}], \quad \forall n \in \mathbb{N}.$$

Par convergence monotone, on a que $u_n \rightarrow u$. On va montrer par récurrence que

$$v \geq u_n, \quad \forall n \in \mathbb{N}.$$

On a

$$u_0(x) = \varphi(x) \mathbf{P}_x[\tau_A = 0] = \varphi(x)1_{x \in A}.$$

Puisque $v = \varphi$ sur A et $v \geq 0$ on a bien $v \geq u_0$. Supposons $v \geq u_n$ pour un certain $n \in \mathbb{N}$ et montrons que $v \geq u_{n+1}$. Par construction

$$v(x) = \varphi(x) = u_{n+1}(x), \quad \forall x \in A.$$

Prenons $x \notin A$, un calcul similaire à celui précédent sur F montre que (exercice!)

$$\varphi(X_{\tau_A})1_{\tau_A \leq n+1} = (\varphi(X_{\tau_A})1_{\tau_A \leq n}) \circ \theta_1.$$

Par la propriété de Markov (théorème 3.22), on trouve donc $u_{n+1} = Qu_n$. En utilisant la croissance de Q et que v est harmonique sur $\mathbb{X} \setminus A$ on trouve donc

$$u_{n+1}(x) = Qu_n(x) \leq Qv(x) = v(x), \quad \forall x \in \mathbb{X} \setminus A.$$

□

Remarque 3.31. En fait une fois que l'on connaît la théorie des martingales, on peut montrer que si $\mathbf{P}_x[\tau_A < \infty]$, pour tout $x \in A$, et que φ est bornée, alors u est l'unique solution bornée du problème de Dirichlet.

On peut de la même manière calculer des espérances de temps d'arrêt.

Proposition 3.32. Avec les notations du théorème 3.29, posons $n(x) := \mathbf{E}_x[\tau_A]$. Alors, n est la plus petite solution positive de

$$\begin{cases} n(x) = 0, & \forall x \in A; \\ Qn(x) = n(x) - 1, & \forall x \notin A. \end{cases}$$

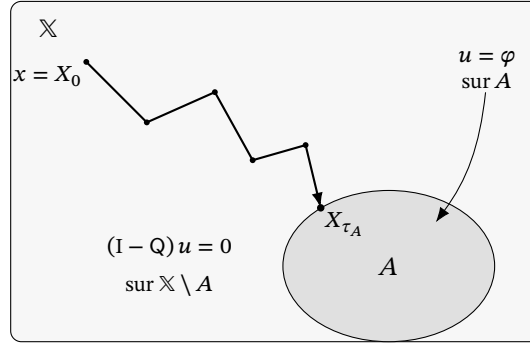


FIGURE 3.2 : La valeur de sortie $u(x) = \mathbf{E}_x[\varphi(X_{\tau_A})\mathbf{1}_{\tau_A < \infty}]$ est harmonique dans $X \setminus A$ et vaut φ sur A .

Démonstration. C'est essentiellement la même preuve. On écrit

$$Qn(x) = \mathbf{E}_x[n(X_1)] = \mathbf{E}_x[\mathbf{E}_{X_1}[\tau_A]] = \mathbf{E}_x[\tau_A \circ \theta_1] = \mathbf{E}_x[\tau_A] - 1.$$

Les détails sont laissés en exercice. □

Exemple 12. On considère le groupe de symétrie du triangle équilatéral. Écrivons r pour la rotation d'angle $2\pi/3$ et ρ_A, ρ_B et ρ_C pour les transformations qui laissent invariants le sommet A, B ou C respectivement. On peut prendre comme générateurs r et τ_C qui correspond au graphe de Cayley suivant.

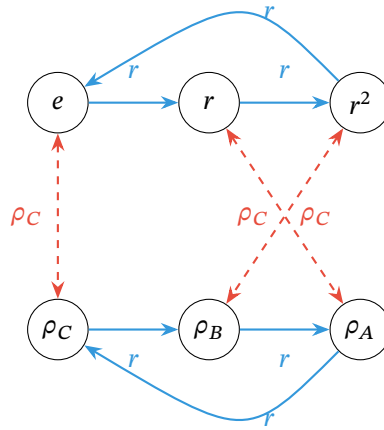


FIGURE 3.3 : Le graphe de transition de la chaîne de Markov

Posons

$$\tau := \inf\{n \in \mathbb{N} : X_n \in \{\rho_A, \rho_B, \rho_C\}\}.$$

On veut calculer la loi de X_τ sous $\mathbf{P}_x$ pour tout $x \in \Sigma_3$. Il est clair que X_τ est à valeurs dans $\{\rho_A, \rho_B, \rho_C\}$ donc il suffit de calculer les probabilités respectives. Avec $\varphi(y) := 1_{y=\rho_A}$, on a

$$u(x) := \mathbf{E}_x[\varphi(X_\tau)] = \mathbf{P}_x[X_\tau = \rho_A].$$

En appliquant le théorème 3.29, on trouve d'une part que $u(\rho_A) = 1$ et que $u(\rho_B) = u(\rho_C) = 0$ et d'autre part que

$$\begin{cases} Q(e) = \frac{u(r) + u(\rho_C)}{2} = u(e), \\ Qu(r) = \frac{u(r^{-1}) + u(\rho_A)}{2} = u(r), \\ Qu(r^{-1}) = \frac{u(e) + u(\rho_B)}{2} = u(r^{-1}). \end{cases}$$

En résolvant on trouve

$$u(r) = \frac{4}{7}, \quad u(e) = \frac{2}{7}, \quad u(r^{-1}) = \frac{1}{7}.$$

En résolvant de manière symétrique pour ρ_B (ou ρ_C) on trouve que sous $\mathbf{P}_e$, X_τ suit $\frac{4}{7}\delta_{\rho_C} + \frac{2}{7}\delta_{\rho_A} + \frac{1}{7}\delta_{\rho_B}$.

Maintenant, posons $n(x) := \mathbf{E}_x[\tau]$. Par la proposition 3.32, on a $n(\rho_A) = n(\rho_B) = n(\rho_C) = 0$. D'autre part, avec des calculs similaires

$$\begin{cases} \frac{n(r)}{2} = n(e) - 1, \\ \frac{n(r^{-1})}{2} = n(r) - 1, \\ \frac{n(e)}{2} = n(r^{-1}) - 1. \end{cases}$$

En utilisant que par symétrie tous doivent être égaux on trouve que $n(e) = n(r) = n(r^{-1}) = 2$.

3.3.3 POTENTIEL D'UNE CHAÎNE DE MARKOV

Informellement, trouver les fonctions harmoniques consiste à identifier $\ker(I - Q)$, une fois cet ensemble identifié on cherche à calculer des solutions non homogènes.

Noyau potentiel

Pour cela l'opérateur de Green joue un rôle important.

Définition 3.33 (Potentiel / opérateur de Green). On appelle l'opérateur de Green ou opérateur du potentiel l'opérateur

$$G := \sum_{n \in \mathbb{N}} Q^n.$$

Comme il y a une somme infinie et que l'on n'a pas parlé de topologie pour assurer la convergence précisons que l'on fait agir G uniquement sur les fonctions positives φ et que dans ce cas

$$G\varphi(x) = \sum_{n \in \mathbb{N}} \mathbf{E}_x[\varphi(X_n)].$$

Équation de Poisson

Définition 3.34 (Équation de Poisson). On dit qu'une fonction mesurable positive u est solution de l'équation de Poisson avec terme h si

$$(I - Q)u = h.$$

On vérifie immédiatement que $G - I = QG = GQ$. En appliquant G dans l'équation de Poisson on s'attend à ce que les solutions s'écrivent $u = Gh + u_0$ avec u_0 harmonique.

3.4 Mesures stationnaires

3.4.1 DÉFINITIONS ET PREMIERS RÉSULTATS

Stationnarité

Une fonction harmonique est un vecteur propre à droite de Q associé à la valeur propre 1. Puisque Q agit aussi à gauche sur les mesures on peut généraliser cette notion.

Définition 3.35 (Mesures stationnaires). Une mesure ν sur $\mathbb{X}$ est *stationnaire* si elle est σ -finie et que $\nu Q = \nu$.

Proposition 3.36 (Stationnarité des lois marginales). Si π est une mesure de probabilité stationnaire, alors sous $\mathbf{P}^\pi$, X_n a pour loi π , pour tout $n \in \mathbb{N}$.

Démonstration. On a vu au théorème 3.9 que si $X_0 \sim \pi$, alors $X_n \sim \pi Q^n = \pi$. $\square$

Les mesures stationnaires jouent donc un rôle important pour étudier le comportement en temps long des chaînes de Markov.

Définition 3.37 (Noyau fellerien). On dit que Q est *fellerien* si $Q\varphi$ est continue et bornée dès que φ est continue et bornée.

Proposition 3.38. Supposons le noyau fellerien. Soient μ et $\pi \in \mathcal{P}(\mathbb{X})$ telles que, sous $\mathbf{P}^\mu$, X_n converge en loi vers π , alors π est stationnaire.

Démonstration. Prenons $\varphi \in \mathcal{C}_b(\mathbb{X})$. Par l'hypothèse de convergence en loi et le théorème 3.9, on trouve que

$$\int \varphi d(Q^n \mu) \xrightarrow{n \rightarrow \infty} \int \varphi d\nu.$$

D'autre part, puisque $Q\varphi \in \mathcal{C}_b(\mathbb{X})$, on trouve que

$$\int \varphi d(Q^n \mu) = \int Q\varphi d(Q^{n-1} \mu) \xrightarrow{n \rightarrow \infty} \int Q\varphi d\nu.$$

Par unicité de la limite (dans $\mathbb{R}$), il vient que

$$\int Q\varphi d\nu = \int \varphi d\nu.$$

Puisque c'est vrai pour tout $\varphi \in \mathcal{C}_b(\mathbb{X})$. Par le théorème B.22 de classe monotone, on trouve que $\nu = \nu Q$. $\square$

Réversibilité

La notion de stationnarité peut être renforcée.

Définition 3.39 (Mesures réversibles). Une mesure ν sur $\mathbb{X}$ est *réversible* si elle est σ -finie et que

$$\int_A q(x, B) \nu(dx) = \int_B q(x, A) \nu(dx), \quad \forall A, B \in \mathfrak{X}.$$

Lemme 3.40. Toute mesure réversible est stationnaire.

Démonstration. Par définition de la réversibilité et en utilisant que q est markovien, on trouve

$$\int_{\mathbb{X}} q(x, A) \nu(dx) = \int_A q(x, \mathbb{X}) \nu(dx) = \nu(A), \quad \forall A \in \mathfrak{X}.$$

□

Pour aller plus loin 11 (Stationnarité et réversibilité sur $\mathbb{Z}$). En fait, quand ν est stationnaire, on peut construire une mesure $\mathbf{P}^\nu$ sur $\mathbb{X}^\mathbb{Z}$ muni des applications coordonnées

$$(\dots, X_{-n}, \dots, X_0, \dots, X_n, \dots).$$

Les opérateurs de translation θ_k pour $k \in \mathbb{Z}$ consistent à faire agir le groupe $\mathbb{Z}$ sur $\mathbb{X}^\mathbb{Z}$ de manière canonique. On peut ensuite envoyer faire agir ces translations sur $\mathcal{P}(\mathbb{X}^\mathbb{Z})$ via $(\theta_k)_\#$. La mesure ν est stationnaire si et seulement si $\mathbf{P}^\nu$ est invariante sous l'action des translations. De manière similaire on peut faire agir l'involution $n \mapsto -n$ sur $\mathbb{X}^\mathbb{Z}$ via l'opérateur

$$(\tau x)_k := x_{-k}, \quad x \in \mathbb{X}^\mathbb{Z}.$$

Alors, ν est réversible si et seulement si $\tau_\#(\mathbf{P}^\nu) = \mathbf{P}^\nu$.

3.4.2 MESURES STATIONNAIRES ET THÉORIE ERGODIQUE

Lien avec les systèmes dynamiques

En fait, les mesures stationnaires donnent lieu à une propriété d'invariance au niveau de Ω .

Théorème 3.41 (Stationnarité de la loi). *Soit ν une mesure stationnaire, alors*

$$\theta_\# \mathbf{P}^\nu = \mathbf{P}^\nu.$$

Autrement dit, la mesure $\mathbf{P}^\nu$ est invariante pour le système dynamique (Ω, θ) .

Démonstration. Il suffit de vérifier que pour tout $n \in \mathbb{N}$ et $\varphi : \mathbb{X}^{n+1} \rightarrow [0, \infty]$ mesurable,

$$\mathbf{E}_\nu[\varphi(X_0, \dots, X_n)] = \mathbf{E}_\nu[\varphi(X_1, \dots, X_{n+1})].$$

En appliquant (3.1) et en renommant les variables d'intégration on trouve que le terme de gauche vaut

$$\begin{aligned} & \int \varphi(x_0, \dots, x_n) q(x_{n-1}, dx_n) \dots q(x_0, dx_1) \nu(dx_0) \\ &= \int \varphi(x_1, \dots, x_{n+1}) q(x_n, dx_{n+1}) \dots q(x_1, dx_2) \nu(dx_1). \end{aligned}$$

Par invariance, $\nu(dx_1) = q(x_0, dx_1) \nu(dx_0)$. On conclut en remplaçant dans l'équation précédente et en utilisant de nouveau (3.1). □

Mesures ergodiques

Définition 3.42 (Mesures stationnaires ergodiques). Soit ν une mesure stationnaire pour la chaîne. On dit que ν est *ergodique* si $\mathbf{P}^\nu$ est ergodique pour le système dynamique (Ω, θ) .

Comme pour les systèmes dynamiques, nous avons une caractérisation des mesures stationnaires ergodiques. Avec un léger abus de notation, notons

$$\mathcal{F} = \mathcal{F}(Q) := \{\pi \in \mathcal{P}(\mathbb{X}) : \pi Q = \pi\}.$$

Proposition 3.43. *L'ensemble $\mathcal{F}$ est convexe. Si de plus Q est fellérien, alors $\mathcal{F}$ est fermé dans $\mathcal{P}(\mathbb{X})$.*

Démonstration. La convexité est immédiate. Si (π_n) est une suite de mesures stationnaire, alors comme Q est fellérien, on peut passer à la limite dans $\pi_n Q = \pi_n$. $\square$

Nous allons voir que l'on peut précisément étudier cet ensemble convexe. On travaille sur l'espace canonique Ω muni de l'opérateur de translation $\theta := \theta_1$. Rappelons que l'on note $\mathfrak{i}$ la tribu des ensembles θ -invariants, c'est-à-dire

$$\mathfrak{i} := \{A \in \mathfrak{B}(\Omega) : \theta^{-1}(A) = A\}.$$

Définition 3.44 (Mesures ergodiques). On dit que π est *ergodique* si pour tout $A \in \mathfrak{i}$, $\mathbf{P}^\pi(A) \in \{0, 1\}$.

Les ensembles invariants, et donc les mesures ergodiques, sont caractérisés à partir d'ensembles dans Ω ce qui les rend peu pratique à manipuler. On peut définir d'autres tribus directement sur $\mathbb{X}$ qui jouent un rôle similaire.

Définition 3.45. Soit $\pi \in \mathcal{F}$. On dit que $B \in \mathfrak{B}(\mathbb{X})$ est π -stationnaire si

$$q(x, B) = 1, \quad \pi\text{-presque tout } x \in B.$$

On pose $\mathfrak{i}^\pi$ la tribu des ensembles π -stationnaire.

Remarque 3.46. En terme d'opérateur, cela signifie que

$$\int 1_B Q 1_B d\pi = \pi(B).$$

Remarque 3.47. Un ensemble absorbant est toujours π -stationnaire pour tout $\pi \in \mathcal{F}$.

Proposition 3.48. *Soit $\pi \in \mathcal{F}$, alors les tribus $\mathfrak{i}^\pi$ et $\mathfrak{i}$ sont équivalentes. Autrement dit, pour tout $A \in \mathfrak{i}$, il existe un ensemble π -stationnaire B tel que*

$$\mathbf{P}^\pi(A \Delta B^\mathbb{N}) = 0.$$

En particulier, π est ergodique si et seulement si pour tout B π -stationnaire, $\pi(B) \in \{0, 1\}$.

Démonstration. On admet la première partie du résultat sur l'équivalence des tribus, qui est technique et peu intéressante. La partie sur la caractérisation de l'ergodicité s'obtient en remarquant que si B est π -stationnaire, alors $\pi(B) = \mathbf{P}^\pi(B^\mathbb{N})$. $\square$

Théorème 3.49. *L'ensemble des mesures de probabilité stationnaires forme un ensemble convexe de $\mathcal{P}(\mathbb{X})$ dont les points extrémaux sont exactement les mesures ergodiques. De plus, deux mesures ergodiques sont ou bien égales ou bien mutuellement singulières.*

Idée de démonstration. L'intuition principale consiste à utiliser que $\nu \mapsto \mathbf{P}^\nu$ est une bijection des mesures stationnaires de la chaîne vers les mesures invariantes pour (Ω, θ) puis à invoquer la caractérisation des mesures ergodiques pour (Θ, θ) comme points extrémaux théorème 2.41. La procédure qui permet de relier les ensembles de $\mathbb{X}$ avec les ensembles de Ω est exactement la proposition précédente. $\square$

Notons deux corollaires utiles de cette représentation.

Corollaire 3.50 (Ergodicité en cas d'unicité). *Supposons que la chaîne admette une unique mesure stationnaire, alors elle est ergodique.*

Corollaire 3.51 (Ergodicité des masses de Dirac). *Si la chaîne admet une masse de Dirac comme mesure de probabilité invariante, alors elle est ergodique.*

Pour les systèmes dynamiques ergodiques, nous avons énoncé le théorème de Birkhoff (théorème 2.39). Donnons une conséquence immédiate deux chaînes de Markov.

Théorème 3.52 (Loi des grands nombres pour les chaînes de Markov ergodiques stationnaires). *Soit π une mesure de probabilité stationnaire et ergodique. Alors*

$$\frac{1}{n} \sum_{k=0}^{n-1} \varphi(X_k) \xrightarrow[n \rightarrow \infty]{\mathbf{P}^\pi \text{-p.s.}} \int \varphi d\pi, \quad \forall \varphi \in L^1(\pi).$$

Démonstration. Il suffit d'appliquer le théorème 2.39 à $F := \varphi(X_0)$, on a alors $F \circ \theta^k = \varphi(X_k)$ et $\mathbf{E}^\pi[F] = \int \varphi d\pi$. $\square$

Remarque 3.53 (Un renforcement de la loi des grands nombres). Nous avons obtenu le théorème précédent en appliquant le théorème de Birkhoff à une variable aléatoire F qui ne dépend que d'une coordonnée. En choisissant $F := \varphi(X_0, \dots, X_L)$ avec L fixé et $\varphi : \mathbb{X}^{L+1} \rightarrow \mathbb{R}$ disons mesurable et bornée, on obtient

$$\frac{1}{n} \sum_{k=0}^{n-1} \varphi(X_0, \dots, X_L) \xrightarrow[n \rightarrow \infty]{\mathbf{P}^\pi \text{-p.s.}} \mathbf{E}^\pi[\varphi(X_0, \dots, X_L)].$$

Dans ce cas, l'expression pour l'espérance dans le terme de droite de la limite ne se simplifie pas plus.

Pour aller plus loin 12 (Monte-Carlo par chaîne de Markov). Imaginons que l'on souhaite simuler uniquement une loi de probabilité π que l'on sait être l'unique mesure stationnaire d'une chaîne de Markov. Le théorème précédent nous dit que si l'on sait bien approcher numériquement la chaîne de Markov stationnaire, alors on peut utiliser les sommes partielles $\frac{1}{n} \sum_{k=0}^{n-1} \delta_{X_k}$ pour approcher π . Mise en œuvre : exercices 74 et 87 ainsi que dans les TPs.

3.4.3 PROPRIÉTÉ DE LIOUVILLE ET NON-UNICITÉ

Finissons cette partie avec un critère simple pour montrer la non-unicité des probabilités stationnaires.

Définition 3.54 (Propriété de Liouville). On dit qu'un noyau markovien Q a la *propriété de Liouville* si les seules fonctions harmoniques bornées sont les fonctions constantes.

Remarque 3.55. La terminologie vient de la propriété de Liouville : les fonctions holomorphes bornées sont constantes. On remarquera que h est holomorphe alors par les équations de Cauchy–Riemann $\Delta h = 0$.

Lemme 3.56. *Soit Q un noyau markovien qui admet une mesure réversible m (possiblement infinie). Si h est harmonique, alors hm est également une mesure stationnaire.*

Démonstration. Il suffit de calculer pour un borélien B

$$(hm)Q(B) = \int hQ1_B dm = \int 1_B Qh dm = \int 1_B h dm = (hm)(B).$$

□

Proposition 3.57 (Critère de non-unicité). *Si Q est un noyau markovien qui admet une mesure de probabilité réversible μ et qui ne satisfait pas la propriété de Liouville, alors μ n'est pas l'unique mesure de probabilité stationnaire.*

Exercices fondamentaux

Les exercices ci-dessous, tirés du livret d'exercices, constituent l'entraînement minimal recommandé sur ce chapitre.

Espace canonique et temps d'arrêt. Exercices [56](#), [57](#), [58](#), [59](#), [60](#).

Propriété de Markov. Exercices [55](#).

Fonctions harmoniques et problème de Dirichlet. Exercices [61](#), [62](#).

Marches aléatoires. Exercices [63](#), [64](#), [65](#).

Mesures stationnaires. Exercices [67](#), [68](#).

Étude des mesures stationnaires et comportement en temps long

Au vu des résultats précédents, plusieurs questions naturelles se posent :

- (a) A-t-on des critères simples d'existence ou d'unicité des mesures stationnaires ?
- (b) Que se passe-t-il quand une mesure stationnaire est de masse infinie ?
- (c) Est-ce que partant de n'importe quel état, la distribution de la chaîne fini par s'équilibrer suivant une mesure stationnaire ?

Nous allons donner des éléments de réponse dans un cadre de plus en plus général. Nous commençons par les espaces finis, où, au vu de l'analogie entre les noyaux de Markov et les matrices stochastiques, les résultats s'expliquent particulièrement bien à l'aide de l'algèbre linéaire. Nous passons ensuite à un cadre intermédiaire : les chaînes de Markov à valeurs dans les espaces dénombrables. Ici nous verrons que des phénomènes d'échappée à l'infini se produisent et rendent les résultats plus fins. Nous concluons avec les chaînes de Markov à valeurs dans les espaces polonais quelque conquies. Il n'existe pas de théorie aussi définitive que dans les cas précédents et nous nous contentons de donner un aperçu de quelques résultats généraux.

4.1 Cas des espaces finis

4.1.1 MATRICES IRRÉDUCTIBLES

On rappelle que dans le cas des espaces finis, il y a correspondance entre les chaînes de Markov et les matrices stochastiques. Nous allons commencer par étudier le cas d'une classe « sympathique » de matrices stochastiques.

Irréductibilité

Une matrice est irréductible s'il est possible d'aller de n'importe quel point du graphe vers n'importe quel point en suivant les flèches orientées. D'un point de vue probabiliste cela signifie qu'étant donné n'importe quel point de départ la chaîne finira par visiter tous les points. Intuitivement si la matrice est irréductible la chaîne « bouge » dans tous les sens et cela aide pour l'existence et l'unicité de la mesure stationnaire.

Définition 4.1 (Irréductibilité). Une matrice stochastique Q est *irréductible* si pour tout $x, y \in \mathbb{X}$, il existe $n \in \mathbb{N}$ tel que $Q^n(x, y) > 0$. Autrement dit, le chaîne est irréductible si pour tout $x, y \in \mathbb{X}$, $G(x, y) > 0$.

On appelle un *chemin* de x vers y dans le graphe d'incidence tout tuple $(z_1, \dots, z_l)$ de sommets du graphe tels que

$$x \rightarrow z_1 \rightarrow \dots \rightarrow z_l \rightarrow y.$$

Alors, Q est irréductible si et seulement si pour tout x et y il existe un chemin de x vers y dans le graphe d'incidence.

Remarque 4.2. On n'a pas forcément $G(y, x) > 0$ même si $G(x, y) > 0$. Autrement dit, l'existence d'un chemin de x vers y n'implique pas l'existence d'un chemin de y vers x .

Périodicité

Une autre obstruction à la convergence est si la chaîne « tourne » sur elle-même.

Définition 4.3 (Temps de retours). L'ensemble des temps de retour possibles en $x \in \mathbb{X}$ sont donnés par

$$\mathfrak{R}(x) := \{n \in \mathbb{N}^* : Q^n(x, x) > 0\}$$

Définition 4.4 (Période). La période de $x \in \mathbb{X}$ est le plus grand diviseur commun de $\mathfrak{R}(x)$.

Lemme 4.5. Soit Q une matrice stochastique irréductible. Alors, tous les états ont la même période.

Démonstration. Soient $x, y \in \mathbb{X}$ de période respective k_x et k_y . Par irréductibilité, on peut trouver n et $m \in \mathbb{N}$ tels que $Q^n(x, y)Q^m(y, x) > 0$. On a alors $n + m \in \mathfrak{R}(x) \cap \mathfrak{R}(y)$. En particulier k_x et k_y divisent $n + m$. Comme on a aussi $n + m + k_y \in \mathfrak{R}(x)$, on trouve que k_x divise $n + m + k_y$ et donc k_x divise k_y . Par un argument symétrique, k_y divise k_x et donc $k_x = k_y$. $\square$

Définition 4.6 (Périodicité). On appelle *période* d'une chaîne irréductible, la période commune de tous les états. Si c'est 1 on dit que la chaîne est *apériodique*.

Lemme 4.7. Une matrice stochastique irréductible Q est de période $d \in \mathbb{N}^*$ si et seulement si, il existe une partition $A_1, \dots, A_d$ de $\mathbb{X}$ qui satisfait la propriété suivante : pour tout x et $y \in \mathbb{X}$ tels que $Q(x, y) > 0$, si $x \in A_k$ alors $y \in A_{k+1 \bmod d}$.

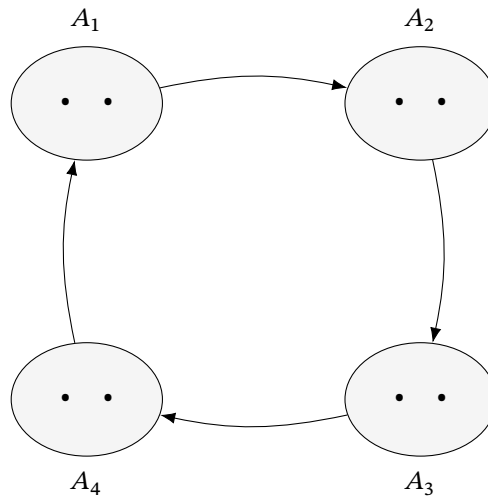


FIGURE 4.1 : Structure cyclique d'une chaîne irréductible de période d (ici $d = 4$) : si $Q(x, y) > 0$ et $x \in A_k$, alors $y \in A_{k+1 \bmod d}$.

Démonstration. Choisissons un $x_* \in \mathbb{X}$ arbitraire. Posons

$$A_k := \{y \in \mathbb{X} : \exists j \in \mathbb{Z}, Q^{k+jd}(x_*, y) > 0\}, \quad k \in \{1, \dots, d\}.$$

Par irréductibilité, $\cup A_k = \mathbb{X}$. Soit $y \in \mathbb{X}$, également par irréductibilité, il existe $m \in \mathbb{N}$ tel que $Q^m(y, x_*) > 0$. Donc pour tout k_1 et k_2 tels que $Q^{k_i}(x_*, y) > 0$ on a que d divise $m + k_i$, d'où $k_1 - k_2 = 0 \pmod{d}$. Ce qui montre que les (A_k) sont disjoints. Finalement, la propriété est immédiate. $\square$

Remarque 4.8. En fait on peut étendre ce résultat à des matrices non irréductibles.

Théorème de Perron–Frobenius

Commençons par un résultat d'algèbre linéaire que nous interpréterons d'un point de vue probabiliste. Dans l'énoncé suivant, les vecteurs propres sont compris à gauche

Théorème 4.9 (Perron–Frobenius). *Soit Q une matrice stochastique irréductible de période d .*

- (a) *Toutes les valeurs propres sont de module ≤ 1 .*
- (b) *1 est valeur propre simple associée à un vecteur μ dont on peut choisir toutes les entrées strictement positives.*
- (c) *Les seules valeurs propres sur le cercle unité sont les racines d -ième de l'unité. Ce sont des valeurs propres simples. L'espace propre associé à $\lambda_j := e^{2i\pi j/d}$ ($j = 0, \dots, d-1$) est engendré par le vecteur μ_j qui satisfait*

$$\mu_j(x) = e^{-2i\pi \frac{jk}{d}} \mu(x), \quad x \in A_k,$$

où (A_k) est la partition obtenue en lemme 4.7.

Démonstration. (a) Pour tout vecteur ν

$$\sum_x |(\nu Q)(x)| \leq \sum_{x,y} |\nu(y)| q(y, x) \leq \|\nu\|_{\ell^1}.$$

Donc si ν est un vecteur propre de valeur propre associée λ , on trouve que

$$|\lambda| \|\nu\|_{\ell^1} \leq \|\nu\|_{\ell^1}.$$

Comme les vecteurs propres sont $\neq 0$, on trouve que $|\lambda| \leq 1$.

(b) Puisque $Q\mathbf{1} = \mathbf{1}$, 1 est une valeur propre (à gauche) de Q^T et donc 1 est également une valeur propre de Q . Soit μ un vecteur propre associé.

Commençons par montrer que l'on peut prendre les entrées positives. Par irréductibilité, il existe $n \in \mathbb{N}^*$ et $\delta > 0$ tels que les coefficients de la matrice

$$R := \frac{1}{n} \sum_{k=1}^n Q^k,$$

sont tous $\geq \delta$. Écrivons $\mu = \mu_+ - \mu_-$ avec $\mu_{\pm}$ des vecteurs dont toutes les entrées sont positives. On a

$$(\mu_+ R)(x) = \sum_y \mu_+(y) R(x, y) \geq \delta \|\mu_+\|_{\ell^1},$$

et une inégalité similaire pour μ_- . En posant $\alpha := \min\{\|\mu_+\|_{\ell^1}, \|\mu_-\|_{\ell^1}\}$, il vient

$$\|\mu\| = \|\mu R\| \leq \|\mu_+ R - \delta \alpha \mathbf{1}\| + \|\mu_- R - \delta \alpha \mathbf{1}\| \leq \|\mu_+\| + \|\mu_-\| - 2\delta \alpha = \|\mu\| - 2\delta \alpha,$$

où toutes les normes sont au sens ℓ^1 . Ce qui donne que $\alpha = 0$. Donc toutes les entrées de μ ont le même signe et, quitte à prendre $-\mu$ au lieu de μ et à normaliser, on peut supposer que $\mu \in \mathcal{P}(\mathbb{X})$. Finalement $\mu = \mu R$ qui est à entrées strictement positives.

Maintenant prenons μ et $\mu' \in \mathcal{P}(\mathbb{X})$ deux mesures stationnaires et posons $\tilde{\mu} := \mu - \mu'$. D'où $\tilde{\mu}Q = \tilde{\mu}$ donc les entrées de $\tilde{\mu}$ ont toutes le même signe mais puisque $\tilde{\mu}(\mathbb{X}) = 0$, on trouve que $\tilde{\mu} = 0$.

(c) Soit ν un vecteur propre associé à la valeur propre $e^{i\theta}$ avec $\theta \neq 0 \bmod 2\pi$. Écrivons les coordonnées en polaire

$$\nu(x) = r(x)e^{i\theta(x)}, \quad x \in \mathbb{X}.$$

Le vecteur r est à entrées positives et, quitte à normaliser, on peut supposer que $r \in \mathcal{P}(\mathbb{X})$. La relation $\nu Q = \nu$ se réécrit

$$\sum_{y \in \mathbb{X}} e^{i\theta(y)} Q(x, y) r(y) = e^{i(\theta + \theta(x))} r(x). \quad (4.1)$$

D'où il vient

$$\sum_{x, y \in \mathbb{X}} e^{i(\theta(y) - \theta - \theta(x))} Q(x, y) r(y) = 1.$$

Par hypothèse on a que $(Q(x, y)r(y))_{x, y}$ est une suite à double index de nombres ≥ 0 qui somme à 1, donc

$$\theta(y) - \theta(x) - \theta = 0 \bmod 2\pi, \quad \forall x, y : Q(x, y) \neq 0. \quad (4.2)$$

(4.1) se réécrit $rQ = r$ et donc par le point précédent $r = \mu$. Par irréductibilité, cette famille d'égalité détermine complètement les $(\theta(x))_{x \in \mathbb{X}}$ à condition de connaître θ et disons $\theta(x_1)$. Quitte à tout multiplier par un nombre complexe (ce qui ne change rien), on va supposer que $\theta(x_1) = 0$.

Soit $z \in \mathbb{X}$ et $N \in \mathcal{R}(z)$. Soit (A_k) la partition du lemme 4.7. Alors, on trouve un i_0 tel que $z \in A_{i_0}$, puis on peut trouver un chemin $z = z_1, z_2, \dots, z_N = z$ tel que $z_i \in A_{i+i_0 \bmod d}$ et $Q(z_i, z_{i+1}) > 0$ pour tout $i = 1, \dots, d$. En écrivant les d équations correspondants à (4.2) pour les $y := z_i$ et $x := z_{i+1}$ puis en sommant on trouve que

$$0 = N\theta \bmod 2\pi.$$

Par les propriétés du pgcd et de la division euclidienne, on doit avoir que $kd \in \mathcal{R}(x)$ pour tout k assez grand. Ainsi $e^{i\theta}$ est bien une racine d -ième de l'unité. Par des mêmes arguments on trouve que si $x \in A_k$ alors

$$e^{i\theta(x)} = e^{i\theta k}.$$

□

La conséquence probabiliste du théorème de Perron–Frobenius est la suivante.

Théorème 4.10 (Convergence : cas des matrices irréductibles et apériodiques). *Soit (X_n) une chaîne de Markov sur un espace fini, irréductible et apériodique. Alors elle admet une unique mesure de probabilité stationnaire μ_∞ , et de plus pour tout $x \in \mathbb{X}$, sous $\mathbf{P}_x$*

$$X_n \xrightarrow[n \rightarrow \infty]{\text{loi}} \mu_\infty.$$

Démonstration. Par le théorème de Perron–Frobenius (théorème 4.9), il existe une unique mesure de probabilité stationnaire. Si Q est diagonalisable on écrit $Q = P\Lambda P^{-1}$, avec Λ la matrice des valeurs propres. Par Perron–Frobenius, toutes les entrées de Λ^n convergent vers 0 quand $n \rightarrow \infty$, sauf celle qui correspond à la valeur propre 1 qui est constante. Donc Q^n converge vers une matrice de rang 1 dont l'image est l'espace vectoriel engendré par μ_∞ . Donc pour tout $\nu \in \mathcal{P}(\mathbb{X})$, $\nu Q^n \rightarrow t\mu_\infty$ pour un certain $t \in \mathbb{R}$. Mais comme on sait que $\nu Q^n \in \mathcal{P}(\mathbb{X})$ pour tout $n \in \mathbb{N}^*$, on a $t = 1$. Même si la matrice n'est pas diagonalisable, le même argument marche quitte à passer en forme de Jordan. $\square$

4.2 Cas des espaces dénombrables

Toutes les matrices stochastiques ne sont pas irréductibles. Nous allons maintenant étudier le cas plus général des chaînes de Markov sur un espace dénombrable $\mathbb{X}$, qui permet en particulier de traiter le cas des matrices stochastiques non irréductibles. Comme précédemment, on note (y_i) pour une énumération des éléments de $\mathbb{X}$. Les notations matricielles et vectorielles restent valides à condition d'accepter de travailler avec des « matrices » et des « vecteurs » infinis. De la même manière, les définitions d'irréductibilité et période continuent à faire sens. Par contre, le théorème de Perron–Frobenius, basé sur l'algèbre linéaire, ne fonctionne plus dans ce cas général.

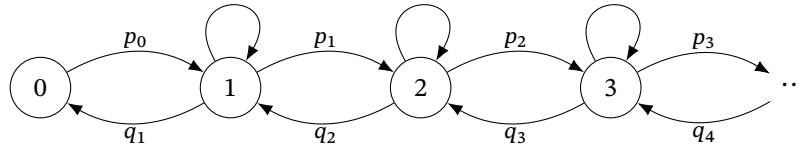


FIGURE 4.2 : Chaîne de naissance et de mort sur $\mathbb{N}$: montée p_x , descente q_x , sur-place $1 - p_x - q_x$ (boucles), avec $q_0 = 0$. L'espace d'états est infini : le comportement (réurrence positive, nulle, ou transience) dépend des suites (p_x) et (q_x) .

4.2.1 CLASSIFICATION DES ÉTATS

Classes de communication

Avec x et $y \in \mathbb{X}$, si $Q^n(x, y) > 0$ pour un certain n alors il y a une probabilité non nulle que la chaîne aille de x à y en n étapes. Si on veut juste savoir si on peut aller de x à y cela revient à demander que $G(x, y) > 0$.

Définition 4.11 (Chemin et communication). Soient x et $y \in \mathbb{X}$. On dit qu'il existe un *chemin* de x vers y et on note $x \rightarrow y$, si $G(x, y) > 0$. On dit que x et y *communiquent* et on note $x \leftrightarrow y$, si $x \rightarrow y$ et $y \rightarrow x$.

Remarque 4.12. En termes du graphe associé, $x \rightarrow y$ si et seulement si, il existe $p \in \mathbb{N}$ et une suite de sommets $x =: x_0, x_1, \dots, x_p =: y$ telle que les arêtes orientées $x_{i-1} \rightarrow x_i$, $i \in \{1, \dots, p\}$ sont présentes dans le graphe.

Proposition 4.13 (La communication est une relation d'équivalence). La relation $\leftrightarrow$ définit une relation d'équivalence sur $\mathbb{X}$.

Démonstration. La réflexivité et la symétrie sont immédiates. La transitivité s'obtient en composant les chemins sur le graphe. D'un point de vue matriciel cela vient de l'inégalité $Q^{n+p}(x, z) \geq Q^n(x, y)Q^p(y, z)$ pour tout $x, y, z \in \mathbb{X}$ et $k, n \in \mathbb{N}$. $\square$

États transitoires et récurrents

Pour une matrice non irréductible, il est possible que certains états ne soient visités qu'un nombre fini de fois. Dans le cas infini dénombrable il est même possible que tous les états ne soient visités qu'un nombre fini de fois (fuite à l'infini). Nous allons voir une notion qui permet de contrôler ce phénomène. On travaille sur l'espace canonique.

Définition 4.14 (Temps de retour et visites). Pour $x \in \mathbb{X}$, on pose le *temps de retour*

$$T_x := \inf\{n \in \mathbb{N}^* : X_n = x\},$$

et le *nombre de visites*

$$N_x := \sum_{n \in \mathbb{N}} 1_{X_n = x}.$$

Définition 4.15 (Récurrence et transience). Un état $x \in \mathbb{X}$ est *récurrent* si $G(x, x) = +\infty$, et *transitoire* sinon.

Par définition

$$G(x, y) = \sum_{n \in \mathbb{N}} Q^n(x, y) = \sum_{n \in \mathbb{N}} \mathbf{P}_x[X_n = y] = \mathbf{E}_x[N_y].$$

Dire qu'un état récurrent signifie qu'en moyenne en partant de x on revient une infinité de fois en x . Nous allons voir que ce comportement en moyenne peut en fait être caractérisé par des propriétés presque sûres a priori beaucoup plus faibles.

Théorème 4.16 (Dichotomie des états). Soit un état $x \in \mathbb{X}$. Sous $\mathbf{P}_x$, N_x suit une loi géométrique de paramètre $\mathbf{P}_x[T_x = \infty]$. En particulier, on a

$$G(x, x) = \frac{1}{\mathbf{P}_x[T_x = \infty]} \in [0, \infty],$$

et aussi

- x est récurrent si et seulement si $\mathbf{P}_x[T_x < \infty] = 1$ si et seulement si $\mathbf{P}_x[N_x = \infty] = 1$.
- x est transitoire si et seulement si $\mathbf{P}_x[T_x < \infty] < 1$ si et seulement si $\mathbf{P}_x[N_x < \infty] = 1$.

Démonstration. Pour $k \in \mathbb{N}^*$, on a

$$\{N_x \geq k + 1\} = \{T_x < \infty\} \cap \{N_x \circ \theta_{T_x} \geq k\},$$

En effet on visite au moins $k + 1$ fois x si et seulement si on visite x une première fois x en un temps fini et qu'après cette première visite x au moins k fois. En intégrant et en utilisant la propriété de Markov (théorème 3.22), on trouve que

$$\mathbf{P}_x[N_x \geq k + 1] = \mathbf{P}_x[T_x < \infty] \mathbf{P}_x[N_x \geq k].$$

Par récurrence, on trouve que $\mathbf{P}_x[N_x \geq k] = \mathbf{P}_x[T_x < \infty]^{k-1}$. □

Classe de récurrence

Soit x un état récurrent, alors si on peut aller de x vers y on s'attend à ce que y soit également visité une infinité de fois, donc également récurrent. De plus, on s'attend à pouvoir revenir de y vers x sinon x ne serait pas récurrent. Le résultat suivant nous dit que c'est effectivement le cas.

Lemme 4.17. *Soit x un état récurrent, et $y \in \mathbb{X}$ tel que $G(x, y) > 0$. Alors, y est récurrent, $\mathbf{P}_y[T_x < \infty] = 1$ et $G(y, x) > 0$.*

Démonstration. On écrit

$$\{N_x < \infty\} \supset \{T_y < \infty\} \cap \{T_x \circ \theta_{T_y} = \infty\}.$$

En effet, si on atteint y en temps fini et que l'on n'atteint plus jamais x à partir de y alors $N_x < \infty$. En intégrant et en utilisant d'une part le théorème 4.16 pour l'état récurrent x et d'autre part la propriété de Markov du théorème 3.22, on trouve que

$$0 = \mathbf{P}_x[N_x < \infty] \geq \mathbf{P}_x[T_y < \infty] \mathbf{P}_y[T_x = \infty].$$

Comme on a $G(x, y) > 0$, le premier terme du membre de droite est positif et donc le deuxième terme doit être nul. Ce qui donne $\mathbf{P}_y[T_x < \infty] = 1$. En adaptant la preuve du théorème 4.16, on trouve que $G(y, x) = \mathbf{P}_y[T_x < \infty]G(x, x)$, d'où $G(y, x) > 0$.

Finalement, par définition du produit matriciel et en utilisant que les entrées de Q sont positives :

$$Q^{n+p+k}(y, y) \geq Q^n(y, x)Q^p(x, x)Q^k(x, y), \quad n, p, k \in \mathbb{N}.$$

En sommant sur p , on obtient

$$G(y, y) \geq Q^n(y, x)G(x, x)Q^k(x, y), \quad k, n \in \mathbb{N}.$$

Comme $G(x, y)G(y, x) > 0$ on peut trouver k et n tels que les deux termes non diagonaux sont strictement positifs. Puisque x est récurrent, on trouve donc que $G(y, y) = \infty$. Ainsi y est récurrent. $\square$

Théorème 4.18 (Classification des états). *La relation d'équivalence $\leftrightarrow$ partitionne les états récurrents en classes d'équivalence qui satisfont les propriétés suivantes.*

(i) (On ne sort pas d'une classe de récurrence) Soit x un état récurrent alors, pour tout $y \in \mathbb{X}$, $\mathbf{P}_x$ -presque sûrement

$$\begin{cases} N_y = \infty, & \text{si } x \leftrightarrow y; \\ N_y = 0, & \text{sinon,} \end{cases}$$

(ii) (Les états transitoires s'échappent ou sont absorbés) Soit x un état transitoire. Notons τ le temps d'entrée dans la chaîne dans un état récurrent, autrement dit

$$\tau := \bigwedge_{y \text{ récurrent}} \tau_y.$$

Alors, $\mathbf{P}_x$ -presque sûrement

- (échappée) sur l'évènement $\{\tau = \infty\}$, pour tout $y \in \mathbb{X}$, $N_y < \infty$;
- (absorption) sur l'évènement $\{\tau < \infty\}$, pour tout $n \geq \tau$, $X_n \leftrightarrow X_\tau$.

Démonstration. On sait que $\leftrightarrow$ est une relation d'équivalence sur $\mathbb{X}$. Vérifions que $\leftrightarrow$ est une relation d'équivalence sur les états récurrents. Par le lemme 4.17 si x est récurrent et $x \leftrightarrow y$ alors y est aussi récurrent, donc on peut restreindre la relation $\leftrightarrow$ aux états récurrents, et on a $x \leftrightarrow y$ si et seulement si $G(x, y) > 0$.

Maintenant, pour x récurrent. Par le lemme 4.17, si $x \nleftrightarrow y$, $G(x, y) = \mathbf{E}_x[N_y] = 0$ et donc $N_y = 0$ $\mathbf{P}_x$ -presque sûrement. Pour $x \rightarrow y$, on a par contre que $\mathbf{P}_x[T_y < \infty] = 1$ et en utilisant la propriété de Markov du théorème 3.22, on trouve que

$$\mathbf{P}_x[N_y = \infty] = \mathbf{P}_x[\{T_y < \infty\} \cap \{N_y \circ \theta_{T_y} = \infty\}] = \mathbf{P}_x[T_y < \infty] \mathbf{P}_y[N_y = \infty] = 1. \quad (4.3)$$

Soit maintenant x transitoire. On va d'abord montrer que si $\mathbf{P}_x[\tau = \infty] > 0$, alors

$$\mathbf{P}_x[\forall y \in \mathbb{X}, N_y < \infty \mid \tau = \infty] = 0.$$

Quitte à renormaliser, on suppose que $\mathbf{P}_x[\tau = \infty] = 1$. Si $x \nleftrightarrow y$, alors $\mathbf{P}_x[N_y = 0] = 1$, et si $x \rightarrow y$ on utilise (4.3).

Ensuite, par le même argument de normalisation supposons que $\mathbf{P}_x[\tau < \infty] = 1$. Par la propriété de Markov $(X_{n+\tau})$ est de nouveau une chaîne de Markov, commencée en X_τ qui est un état récurrent. On applique alors la première partie sur les états récurrents. $\square$

Une conséquence immédiate du théorème 4.18 est l'alternative suivante.

Corollaire 4.19 (Unicité des états pour les chaînes irréductibles). *Soit Q une chaîne irréductible.*

- Ou bien tous les états sont récurrents dans la même classe de récurrence et

$$\mathbf{P}_x[\forall y \in \mathbb{X}, N_y = \infty] = 1, \quad x \in \mathbb{X}.$$

- Ou bien tous les états sont transitoires et

$$\mathbf{P}_x[\forall y \in \mathbb{X}, N_y < \infty] = 1.$$

De plus, si $\mathbb{X}$ est fini tous les états sont récurrents.

Démonstration. Tout est immédiat avec le théorème de classification (théorème 4.18) sauf la partie avec $\mathbb{X}$ fini. Dans ce cas, la suite (X_n) prend ses valeurs dans un ensemble fini. Il y a donc au moins une valeur qui est prise une infinité de fois. $\square$

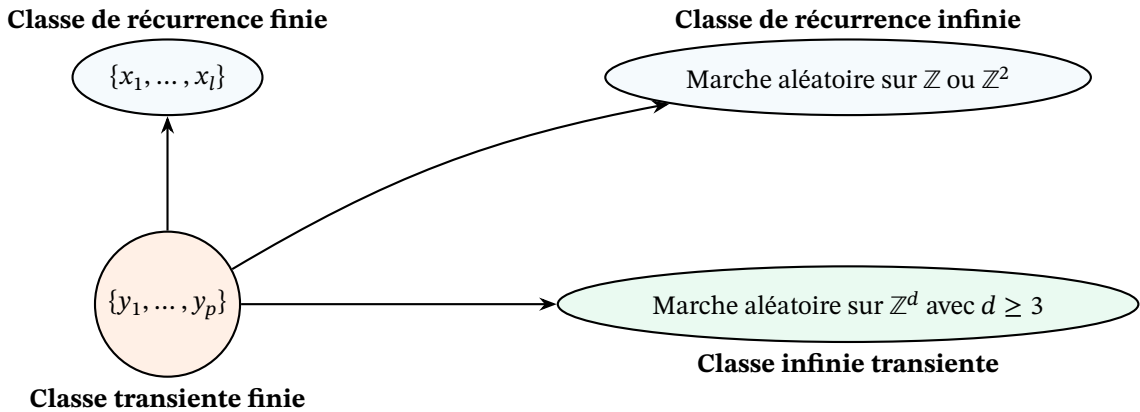


FIGURE 4.3 : Exemple des différentes classes d'irréductibilité

Récurrance et transience de la marche aléatoire simple et symétrique sur $\mathbb{Z}^d$

Dans cette partie on considère une marche aléatoire simple et symétrique sur $\mathbb{Z}^d$. Sur $\mathbb{Z}^d$, on dit que deux points x et y sont *voisins*, et on note $x \sim y$, s'il existe un élément de la base canonique e_i tel que $y = \pm e_i$. Cela induit la structure de graphe (réseau) usuelle sur $\mathbb{Z}^d$ et on considère la marche aléatoire simple. C'est la chaîne de Markov associée à la matrice

$$Q(x, y) := \frac{1}{2d} 1_{x \sim y}.$$

Théorème 4.20 (Pólya). *La marche aléatoire simple sur $\mathbb{Z}^d$ est récurrente pour $d \in \{1, 2\}$ et transiente sinon.*

Démonstration. Puisque la chaîne est irréductible, il suffit d'étudier la transience ou la récurrence d'un seul point, disons 0. Pour cela on va estimer $Q^n(0, 0)$ pour tout n . Déjà il est clair qu'on ne peut revenir en 0 depuis 0 qu'en un nombre pair d'étapes donc $Q^{2n+1}(0, 0) = 0$ pour tout $n \in \mathbb{N}$. Pour le cas pair, c'est un calcul direct en fonction de d .

Cas $d := 1$ Pour retourner en 0 depuis 0 en $2n$ pas, la marche doit faire exactement n pas à droite et exactement n pas à gauche, dans un ordre quelconque. Ainsi

$$Q^{(2n)}(0, 0) = 2^{-2n} \binom{2n}{n} \simeq c_1 n^{-1/2}.$$

Ce n'est pas sommable et donc $G(0, 0) = +\infty$.

Cas $d := 2$ On considère deux marches aléatoires indépendantes sur $\mathbb{Z}$ une pour les abscisses et l'autre pour l'ordonnée. Cela donne une marche aléatoire sur $\mathbb{Z}^2$ mais où l'on se déplace en diagonale. Par chance, ce réseau est isomorphe au réseau qui nous intéressent et on trouve

$$Q^{2n}(0, 0) = \left(2^{-2n} \binom{2n}{n} \right)^2 \simeq c_2 n^{-1},$$

qui n'est toujours pas sommable.

Cas général On ne peut plus identifier la marche aléatoire simple avec d marches aléatoires indépendantes sur $\mathbb{Z}$. Pour $d := 3$, on pourrait faire un argument combinatoire difficile et trouver que $Q^{2n}(0, 0) \simeq c_3 n^{-3/2}$. On va plutôt utiliser la transformée de Fourier qui comme bien souvent nous sauve. Soit ν la loi des sauts et $\xi \sim \nu$, on a que

$$\hat{\nu}(\theta) = \mathbf{E}[e^{i\xi \cdot \theta}] = \frac{1}{d} \sum_{j=1}^d \cos \theta_j =: \varphi(\theta).$$

Puisque ν_n la loi de X_n est donnée par une convolution (somme de variables aléatoires indépendantes et identiquement distribuées) on trouve que

$$\hat{\nu}_n(\theta) = \varphi(\theta)^n.$$

Par inversion de Fourier, on trouve donc que

$$\mu_n(0) = \frac{1}{(2\pi)^d} \int_{[0, 2\pi]^d} \varphi(\theta)^n d\theta.$$

Comme φ est à valeurs dans $[-1, 1]$, en sommant on trouve que

$$G(0, 0) = \frac{1}{(2\pi)^d} \int_{[0, 2\pi]^d} \frac{1}{1 - \varphi(\theta)} d\theta.$$

La seule singularité possible et quand $\varphi(\theta)$ est proche de 1, c'est à dire $\theta \simeq 0$. En faisant un développement limité de cos près de 0, on trouve

$$1 - \varphi(\theta) \simeq \frac{1}{2d} \|\theta\|^2 + o(\|\theta\|^2).$$

Maintenant

$$\int_{B(0, t)} \frac{d\theta}{\|\theta\|^2} = \int_0^t |\partial B(0, r)| \frac{1}{r^2} dr.$$

Comme la sphère de rayon r en dimension d à une aire surface proportionnelle à r^{d-1} , on intègre proche de 0, r^{d-1-2} qui est intégrable si et seulement si $d \geq 3$. $\square$

4.2.2 COMPORTEMENT EN TEMPS LONG

Mesures stationnaires

Pour un espace d'état fini, les mesures stationnaires sont des vecteurs propres et peuvent se calculer avec de l'algèbre linéaire. Dans le cas général, l'équation $\mu Q = \mu$ est toujours linéaire mais les outils pour résoudre cette équation sont plus avancés — analyse fonctionnelle, théorie des opérateurs, théorie du potentiel. Dans le cas infini dénombrable, on a la représentation probabiliste suivante.

Théorème 4.21 (Construction probabiliste de mesures stationnaires). *Soit x un état récurrent, alors la mesure*

$$\nu_x := \mathbf{E}_x \left[\sum_{k=0}^{T_x-1} \delta_{X_k} \right],$$

est l'unique (à constante multiplicative près) mesure stationnaire qui charge la classe de récurrence de x . De plus, $\nu_x(y) < \infty$, pour tout $y \in \mathbb{X}$, et $\nu_x(y) > 0$ pour tout y dans la classe de récurrence.

Démonstration. Tout d'abord si y n'est pas dans la classe de communication de x on a par la classification du théorème 4.18 que

$$0 = \mathbf{E}_x[N_y] \geq \nu_x(y).$$

Ce qui montre que ν_x s'annule en dehors de la classe de récurrence de x .

Ensuite pour tout $y \in \mathbb{X}$, on trouve que

$$\nu_x(y) = \sum_{z \in \mathbb{X}} \sum_{k \in \mathbb{N}} \mathbf{E}_x[1_{k \leq T_x} 1_{X_{k-1}=z} 1_{X_k=y}].$$

Comme T_x est un temps d'arrêt, $\{k \leq T_x\} = \{T_x \leq k-1\}^C$ est $\mathfrak{F}_{k-1}$ -mesurable, et $\{X_{k-1} = z\}$ est aussi trivialement $\mathfrak{F}_{k-1}$ -mesurable. D'où par la propriété de Markov, on trouve que

$$\nu_x(y) = \sum_{z \in \mathbb{X}} \mathbf{E}_x \left[\sum_{k=1}^{T_x} 1_{X_{k-1}=z} Q(z, y) \right].$$

Ce qui montre que $\nu_x = \nu_x Q$.

Si $x \leftrightarrow y$, alors il existe n et $m \in \mathbb{N}$ tels que $Q^n(x, y)Q^m(y, x) > 0$. En notant que $\nu_x(x) = 1$ et en itérant, l'invariance on trouve que

$$1 = \sum_{z \in \mathbb{X}} \nu_x(z)Q^n(z, x),$$

donc nécessairement $\nu_x(y) < \infty$. On a aussi

$$\nu_x(y) = \sum_{z \in \mathbb{X}} \nu_x(z)Q^m(z, y) \geq Q^m(x, y) > 0.$$

Montrons que ν_x est unique. Pour cela on va montrer que, si ν est une autre mesure stationnaire supportée sur la classe de récurrence, alors

$$\nu(z) \geq \nu(x)\nu_x(z), \quad \forall z \in \mathbb{X}. \quad (4.4)$$

En utilisant alors que ν et ν_x sont stationnaires et que $\nu_x(x) = 1$, on trouve à partir de (4.4) que pour tout $n \in \mathbb{N}$

$$\nu(x) = \sum_{z \in \mathbb{X}} \nu(z)Q^n(z, x) \geq \sum_{z \in \mathbb{X}} \nu(x)\nu_x(z)Q^n(z, x) = \nu(x)\nu_x(x) = \nu(x).$$

Ce qui montre que

$$\nu(z) = \nu(x)\nu_x(z), \quad \forall z \leftrightarrow x.$$

Mais comme $\nu(z) = \nu_x(z) = 0$ pour $z \not\leftrightarrow x$, cela conclura la preuve. Montrons (4.4). Déjà pour $x = z$, (4.4) est vraie et donc on suppose que $x \neq z$. Posons

$$\nu_{x,p} := \mathbf{E}_x \left[\sum_{k=0}^{p \wedge (T_x-1)} \delta_{X_k} \right], \quad \forall p \in \mathbb{N},$$

et montrons par récurrence que

$$\nu(z) \geq \nu(x)\nu_{x,p}(z), \quad \forall z \in \mathbb{X}, \quad \forall p \in \mathbb{N},$$

ce qui donne (4.4) en faisant $p \rightarrow \infty$. Pour $p = 0$, comme $z \neq x$, l'inégalité est immédiate. Si c'est vrai à l'ordre p , avec un argument analogue au calcul précédent on trouve que

$$\begin{aligned} \nu(z) &\geq \nu(x) \sum_{w \in \mathbb{X}} \nu_{x,p}(w)Q(w, z) = \nu(x) \sum_{w \in \mathbb{X}} \sum_{k=0}^p \mathbf{E}_x[1_{X_k=w} 1_{T_x-1 \geq k}]Q(w, z) \\ &= \nu(x) \sum_{w \in \mathbb{X}} \sum_{k=0}^p \mathbf{E}_x[1_{X_k=w} 1_{X_{k+1}=z} 1_{k \leq T_x-1}] = \nu(x)\nu_{x,p+1}(z). \end{aligned}$$

□

Définition 4.22 (Récurrence positive et nulle). Une classe récurrente est *positive* si toutes les mesures stationnaires sont de masse totale finie et *nulle* si leurs masses totales sont toutes infinies. En particulier, dans le cas positive, il existe une unique mesure de probabilité stationnaire supportée sur la classe de récurrence.

Corollaire 4.23 (Construction de la mesure stationnaire). Soit x un état récurrent. On a l'alternative suivante :

- (i) Ou bien la classe de x est positive, et cela se produit si et seulement si $\mathbf{E}_x[T_x] < \infty$, auquel cas la probabilité stationnaire associée π est ergodique et satisfait

$$\pi(x) = \frac{1}{\mathbf{E}_x[T_x]}.$$

(ii) Ou bien la classe de x est nulle, et cela se produit si et seulement si $\mathbf{E}_x[T_x] = \infty$.

De plus toute classe récurrente contenant un nombre fini d'état est nécessairement positive.

Démonstration. Par le théorème 4.21 toutes les mesures stationnaires sont colinéaires, donc ou bien toutes finies ou bien toutes infinies. Par construction,

$$\nu_x(\mathbb{X}) = \sum_{y \in \mathbb{X}} \mathbf{E}_x \left[\sum_{k=0}^{T_x-1} 1_{X_k=y} \right] = \mathbf{E}_x[T_x].$$

Ce qui donne la caractérisation de la finitude en termes de $\mathbf{E}_x[T_x]$. Ensuite, si π est l'unique mesure de probabilité stationnaire, en utilisant de nouveau toutes les mesures stationnaires sont colinéaires (théorème 4.21), on trouve que

$$\pi(x) = \frac{\nu_x(x)}{\nu_x(\mathbb{X})} = \frac{1}{\nu_x(\mathbb{X})} = \frac{1}{\mathbf{E}_x[T_x]}.$$

Soient π_1 et π_2 deux mesures ergodiques telles que $\pi = (1-t)\pi_1 + t\pi_2$ avec $0 < t < 1$. Par la caractérisation des mesures ergodiques du théorème 3.49, π_1 et π_2 sont mutuellement étrangères, donc à supports disjoints. Mais par théorème 4.21 π est l'unique mesure de probabilité stationnaire supportée par la classe de x . Donc $\pi_1 = \pi_2$ et π est ergodique. Finalement, si la classe de récurrence contient un nombre fini d'états, les mesures sont nécessairement finies par σ -additivité et le fait que $\nu_x(y) < \infty$ pour tout $y \in \mathbb{X}$. $\square$

Loi des grands nombres

Nous savons par le théorème 3.52 qu'une chaîne de Markov commencée en une probabilité stationnaire ergodique satisfait une loi des grands nombres. Dans le cas des espaces dénombrables, nous pouvons raffiner ce résultat.

Définition 4.24 (Temps de retours successifs). Pour $x \in \mathbb{X}$ on appelle *temps de retour successifs* en x la suite de variables aléatoires

$$\tau_x^{(0)} := \tau_x, \quad \text{et} \quad \tau_x^{(n+1)} := \inf \{k > \tau_x^{(n)} : X_k = x\}, \quad n \in \mathbb{N}.$$

On vérifie aisément que les $(\tau_x^{(n)})_{n \in \mathbb{N}}$ forment une suite de temps d'arrêt.

Définition 4.25 (Excursion). Les *excursions* de la chaîne de Markov sont les variables aléatoires

$$E_x^{(k)} := (X_{\tau_x^{(k)}}, X_{\tau_x^{(k)}+1}, \dots, X_{\tau_x^{(k+1)}-1}), \quad k \in \mathbb{N}.$$

Pour une fonction mesurable f on considère la *somme de f sur l'excursion*

$$Z_x^{(k)}(f) := \sum_{j=\tau_x^{(k)}}^{\tau_x^{(k+1)}-1} f(X_j).$$

La *durée* d'une excursion est

$$l_x^{(k)} := \tau_x^{(k)} - \tau_x^{(k-1)}, \quad \forall \quad k \in \mathbb{N}^*.$$

Lemme 4.26 (Asymptotique des temps de retour). *On considère une chaîne récurrente irréductible et une loi initiale $\mu \in \mathcal{P}(\mathbb{X})$. Sous $\mathbf{P}^\mu$, pour tout $x \in \mathbb{X}$, les durées d'excursion $(l_x^{(n)})_{n \in \mathbb{N}}$ forment une suite de variables aléatoires positives indépendantes et identiquement distribuées. En particulier, pour tout $x \in \mathbb{X}$,*

$$\frac{\tau_x^{(n)}}{n} \xrightarrow[n \rightarrow \infty]{\mathbf{P}^\mu - p.s.} \mathbf{E}_x[T_x].$$

Démonstration. On commence par montrer le résultat pour $\mu := \delta_x$, c'est-à-dire sous $\mathbf{P}_x$. En particulier, $\mathbf{P}_x$ -presque sûrement $\tau_x^{(0)} = 0$ et $\tau_x^{(1)} = T_x$. On montre par récurrence sur $k \in \mathbb{N}^*$ que les variables aléatoires

$$l_x^{(1)}, \dots, l_x^{(k)}, \quad \forall k \in \mathbb{N}^*,$$

sont indépendantes et identiquement distribuées. Pour $k := 1$, il n'y a rien à faire. Ensuite, si on a montré le résultat pour un certain $k \in \mathbb{N}^*$, on écrit

$$(l_x^{(2)}, \dots, l_x^{(k+1)}) = (l_x^{(1)}, \dots, l_x^{(k)}) \circ \theta_{T_x}.$$

D'une part, $l_x^{(1)}$ est $\mathcal{F}_{T_x}$ -mesurable, d'autre part par la propriété de Markov forte théorème 3.22, θ_{T_x} est une chaîne de Markov indépendante de $\mathcal{F}_{T_x}$ et de même loi que $(X_0, X_1, \dots)$. Ce qui montre bien le caractère indépendant et identiquement distribuées.

Pour le cas général, il suffit d'utiliser la propriété de Markov. En effet, par récurrence $\tau_x < \infty$ et donc θ_{τ_x} est de loi $\mathbf{P}_x$. Il suffit alors d'appliquer le cas particulier.

Pour le théorème limite, on applique la loi des grands nombres à

$$\tau_x^{(n)} = \sum_{k=1}^n l_x^{(k)}.$$

Puisque $\mathbf{E}_x[l_x^{(1)}] = \mathbf{E}_x[T_x - 0]$, on conclut. $\square$

Théorème 4.27 (Convergence des mesures empiriques). *On considère une chaîne récurrente irréductible. Alors pour toute condition initiale μ*

$$\frac{1}{n} \sum_{k=0}^{n-1} 1_{X_k=x} \xrightarrow[n \rightarrow \infty]{\mathbf{P}^\mu - p.s.} \frac{1}{\mathbf{E}_x[T_x]}.$$

En particulier, si la chaîne est récurrente positive on a

$$\frac{1}{n} \sum_{k=0}^{n-1} \delta_{X_k} \xrightarrow[n \rightarrow \infty]{\mathbf{P}^\mu - p.s.} \pi,$$

où π est l'unique mesure de probabilité stationnaire. Si elle est nulle, alors

$$\frac{1}{n} \sum_{k=0}^{n-1} \delta_{X_k} \xrightarrow[n \rightarrow \infty]{\mathbf{P}^\mu - p.s.} 0.$$

Démonstration. Écrivons, pour tout $n \in \mathbb{N}$, $N_x(n) := \sum_{k=0}^{n-1} 1_{X_k=x}$. Alors la variable aléatoire $N_x(n)$ représente le nombre de visites en x avant l'instant n (strictement), Ce qui donne $\tau_x^{N_x(n)} \leq n \leq \tau_x^{N_x(n)+1}$. En particulier on trouve

$$\frac{N_x(n)}{\tau_x^{N_x(n)+1}} \leq \frac{N_x(n)}{n} \leq \frac{N_x(n)}{\tau_x^{N_x(n)}}.$$

Comme $N_x(n) \rightarrow \infty$ presque sûrement par récurrence, on peut passer à la limite et utiliser la loi des grands nombres précédente. $\square$

Théorème de Birkhoff

Un corollaire immédiat du théorème 4.27 et que dans le cas irréductible récurrent positif, si f et $g \in L^1(\pi)$ avec $g > 0$ alors

$$\frac{\sum_{k=0}^n f(X_k)}{\sum_{k=0}^n g(X_k)} \xrightarrow[n \rightarrow \infty]{p.s.} \frac{\int f d\pi}{\int g d\pi}.$$

Dans le cas récurrent nul, on ne peut a priori rien dire car on a une limite de la forme 0/0. Nous allons voir qu'en travaillant un tout petit peu plus on peut en fait obtenir un résultat pour toutes les chaînes récurrentes.

Lemme 4.28 (Asymptotique des moyennes sur les excursions). *On considère une chaîne récurrente irréductible et une condition initiale $x \in \mathbb{X}$. Soit $f : \mathbb{X} \rightarrow [0, \infty]$ mesurable. Sous $\mathbf{P}_x$, pour tout $x \in \mathbb{X}$, les moyennes sur les excursions $(Z_x^{(n)})_{n \in \mathbb{N}}$ forment une suite de variables aléatoires positives indépendantes et identiquement distribuées. Si ν est une mesure stationnaire, alors*

$$\frac{1}{n} \sum_{k=1}^n Z_x^{(k)}(f) \xrightarrow[n \rightarrow \infty]{\mathbf{P}_x - p.s.} \frac{1}{\nu(x)} \int f d\nu.$$

Démonstration. La preuve est très proche de la preuve précédente. On obtient l'indépendance et la même loi par récurrence : $Z_x^{(0)}(f), \dots, Z_x^{(k)}(f)$ sont tous $\mathfrak{F}_{\tau_x^{(k+1)}}$ mesurable et $Z_x^{(k+1)}(f) = Z_x^{(0)}(f) \circ \theta_{\tau_x^{(k+1)}}$ puis on utilise la propriété de Markov. Pour le cas général, il suffit de translater par $\theta_{\tau_x^{(0)}}$.

Finalement pour le théorème limite on utilise la loi des grands nombres avec

$$\mathbf{E}_x[Z_x^{(0)}(f)] = \mathbf{E}_x\left[\sum_{k=0}^{T_x-1} \sum_{y \in \mathbb{X}} f(y) 1_{X_k=y}\right] = \sum_{y \in \mathbb{X}} f(y) \nu_x(y) = \frac{1}{\nu(x)} \int f d\nu.$$

□

Théorème 4.29 (Théorème de Birkhoff, version infinie dénombrable). *On considère une chaîne récurrente et irréductible. Soit ν une mesure stationnaire. Soient f et $g \in L^1(\nu)$ des fonctions strictement positives telles que $\int f g d\nu > 0$. Alors, pour toute condition initiale μ*

$$\frac{\sum_{k=1}^n f(X_k)}{\sum_{k=1}^n g(X_k)} \xrightarrow[n \rightarrow \infty]{\mathbf{P}^\mu - p.s.} \frac{\int f d\nu}{\int g d\nu}.$$

Pour aller plus loin 13 (Comparaison avec le théorème ergodique de Birkhoff). Notons deux différences majeures avec le théorème de Birkhoff (théorème 3.52) obtenu via des arguments de théorie ergodique :

- (i) On peut traiter le cas des mesures stationnaires non finies. En effet, nous avons défini un système dynamique comme étant équipé d'une probabilité invariante. On peut définir une théorie ergodique pour des systèmes dynamiques équipés d'une mesure invariante infinie. Il existe une généralisation à ce cadre du théorème de Birkhoff (théorème 2.39).
- (ii) On peut traiter n'importe quel condition initiale. En effet, la condition que $\mathbf{P}$ soit invariante pour le système dynamique (Ω, θ) impose de travailler sous la chaîne invariante. Ici, on obtient une information beaucoup plus précise : on peut commencer de n'importe quelle condition initiale.

Démonstration. C'est exactement comme avant. Commençons par le cas $\mu := \delta_x$ pour $x \in \mathbb{X}$. On écrit

$$\frac{1}{N_x(n)} \sum_{k=0}^{\tau_x^{N_x(n)}} f(X_k) \leq \frac{1}{N_x(n)} \sum_{k=0}^n f(X_k) \leq \frac{1}{N_x(n)} \sum_{k=0}^{\tau_x^{N_x(n)+1}} f(X_k).$$

En utilisant le lemme précédent, on trouve donc que

$$\frac{1}{N_x(N)} \sum_{k=0}^n f(X_k) \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \frac{1}{\nu(x)} \int f d\nu.$$

Avec le même raisonnement pour g on trouve le théorème dans le cas $\mu := \delta_x$. Le cas général vient en remarquant que la limite ne dépend pas de x et que l'on a montré le résultat pour tout $x \in \mathbb{X}$. $\square$

Convergence en loi

Nous avons, à l'aide du théorème de Perron–Frobenius, que les chaînes de Markov irréductibles et apériodiques sur un espace d'état fini convergent vers leur mesure d'équilibre. Nous généralisons ici au cas infini dénombrable.

Théorème 4.30 (Convergence des chaînes récurrentes positives et apériodiques). *On considère une chaîne irréductible, récurrente positive et apériodique. On écrit π pour l'unique mesure de probabilité stationnaire. Alors pour toute condition initiale $\mu \in \mathcal{P}(\mathbb{X})$,*

$$\sum_{y \in \mathbb{X}} |\mu Q^n(y) - \pi(y)| \xrightarrow[n \rightarrow \infty]{} 0.$$

Pour aller plus loin 14 (Convergence en variation totale). En d'autres termes, la loi de X_n converge vers l'unique mesure de probabilité au sens de la convergence L^1 des densités.

Démonstration. Nous utilisons une méthode de couplage. On considère la matrice stochastique sur $\mathbb{X} \times \mathbb{X}$

$$\hat{Q}((x_1, x_2), (y_1, y_2)) := Q(x_1, y_1)Q(x_2, y_2), \quad \forall x_1, x_2, y_1, y_2.$$

On note avec un chapeau les objets sur l'espace canonique associé. En particulier,

$$\hat{X} = (X_n^1, X_n^2) : \hat{\Omega} \rightarrow \mathbb{X}^2.$$

Soient (x_1, x_2) et $(y_1, y_2) \in \mathbb{X}^2$. Comme Q est récurrente apériodique, on peut trouver $n_i \in \mathbb{N}$ tels que

$$Q^n(x_i, y_i) > 0, \quad \forall n \geq n_i.$$

Donc en prenant les $n \geq \max\{n_1, n_2\}$, on voit que $\hat{Q}$ est aussi irréductible. Comme $\pi \otimes \pi$ est une mesure de probabilité stationnaire pour $\hat{Q}$, la chaîne est nécessairement récurrente positive. Fixons $n \in \mathbb{N}$. On peut écrire

$$(\mu Q^n - \pi)(y) = \hat{\mathbf{P}}^{\mu \otimes \pi}[X_n^1 = y] - \hat{\mathbf{P}}^{\mu \otimes \pi}[X_n^2 = y].$$

Définissons le temps d'arrêt

$$S := \inf\{n \in \mathbb{N} : X_n^1 = X_n^2\}.$$

Ce temps d'arrêt est fini car $S \leq \hat{\tau}_{(x,x)}$ qui est fini par récurrence. Pour tout $k = 0, \dots, n$, on trouve que

$$\begin{aligned} \hat{\mathbf{E}}^{\mu \otimes \pi} [1_{S=k} 1_{X_n^2=y}] &= \sum_{z \in \mathbb{X}} \hat{\mathbf{E}}^{\mu \otimes \pi} [1_{S=k} 1_{X_n^2=y} 1_{X_n^1=X_n^2=z}] \\ &= \sum_{z \in \mathbb{X}} \hat{\mathbf{E}}^{\mu \otimes \pi} [1_{S=k} 1_{X_n^1=X_n^2=z}] Q^{n-k}(z, y). \\ &= \sum_{z \in \mathbb{X}} \hat{\mathbf{E}}^{\mu \otimes \pi} [1_{S=k} 1_{X_n^1=y} 1_{X_n^1=X_n^2=z}] \\ &= \hat{\mathbf{E}}^{\mu \otimes \pi} [1_{S=k} 1_{X_n^1=y}] \end{aligned}$$

Donc on peut réécrire

$$\begin{aligned} (\mu Q^n - \pi)(y) &= \hat{\mathbf{E}}^{\mu \otimes \pi} [1_{S>n} (1_{X_n^1=y} - 1_{X_n^2=y})] + \hat{\mathbf{E}}^{\mu \otimes \pi} [1_{S \leq n} (1_{X_n^1=y} - 1_{X_n^2=y})] \\ &= \hat{\mathbf{E}}^{\mu \otimes \pi} [1_{S>n} (1_{X_n^1=y} - 1_{X_n^2=y})]. \end{aligned}$$

D'où il vient finalement,

$$\sum_{y \in \mathbb{X}} |(\mu Q^n - \pi)(y)| \leq 2 \hat{\mathbf{P}}^{\mu \otimes \pi}(S > n).$$

Cette dernière quantité tend vers 0, comme on a déjà remarqué que $S < \infty$. □

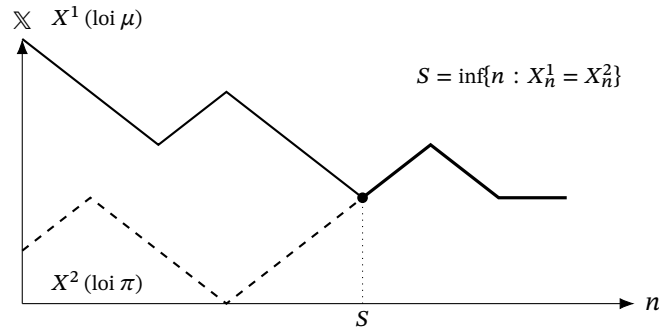


FIGURE 4.4 : Couplage de deux copies de la chaîne : avant S elles évoluent indépendamment, après S elles coïncident. On contrôle $\sum_y |\mu Q^n(y) - \pi(y)| \leq 2 \hat{\mathbf{P}}(S > n)$.

4.3 Cas général

Dans le cas dénombrable il existe des critères raisonnables pour l'existence et l'unicité de la mesure stationnaire ainsi que pour la convergence grâce à la classification des états. Dans le cas non dénombrables on a typiquement que $\mathbf{P}[X_n = x] = 0$ pour tout x et donc la classification des états est plus compliquée. Dans ce dernier chapitre, nous donnons quelques approches pour étudier le comportement asymptotique des chaînes de Markov. Dans toute cette partie $\mathbb{X}$ est un espace polonais, muni de sa tribu borélienne, et q est un noyau markovien.

4.3.1 EXISTENCE DE MESURES STATIONNAIRES

L'idée générale est d'utiliser le critère de compacité de Prokhorov (théorème 1.38). Cela demande une certaine compatibilité entre la chaîne et la topologie. Rappelons la condition de régularité de Feller de la définition 3.37.

En pratique, on peut lire la propriété de Feller sur la fonction de pas.

Théorème 4.31. Soient $\mathbb{U}$ un espace mesurable, (U_n) une suite de variables aléatoires indépendantes et identiquement distribuées à valeurs dans $\mathbb{U}$ et $f : \mathbb{X} \times \mathbb{U} \rightarrow \mathbb{X}$ une fonction mesurable. On suppose que $f(\cdot, U) : \mathbb{X} \rightarrow \mathbb{X}$ est presque sûrement continue. Alors, toute chaîne de Markov définie par la récurrence

$$X_{n+1} = f(X_n, U_{n+1}), \quad \forall n \in \mathbb{N},$$

a la propriété de Feller.

Démonstration. Soit $\varphi \in \mathcal{C}_b(\mathbb{X})$ et (x_n) une suite de $\mathbb{X}$ convergeant vers x_∞ . Alors par convergence dominée, on peut passer à la limite sous le signe intégral dans

$$Q\varphi(x_n) = \mathbf{E}[\varphi(f(x_n, U))].$$

En utilisant la continuité de φ et de f , on a montré que

$$Q\varphi(x_n) \xrightarrow{n \rightarrow \infty} Q\varphi(x_\infty).$$

Donc $Q\varphi$ est continue. □

Critère par compacité

Théorème 4.32 (Critère Krylov–Bogolioubov). Soit q un noyau fellérien tel que pour un certain $x \in \mathbb{X}$, la famille $(q^n(x, \cdot))$ est équitendue. Alors, Q admet une mesure de probabilité stationnaire.

Démonstration. Définissons,

$$\mu_n := \frac{1}{n} \sum_{k=1}^n q^k(x, \cdot).$$

Alors, (μ_n) est aussi équitendue — il suffit de prendre le même compact que pour la suite $(q^n(x, \cdot))$. Par théorème 1.38, à extraction près, cette suite converge vers μ_∞ . Soit $\varphi \in \mathcal{C}_b(\mathbb{X})$, on a

$$|(\mu_\infty Q - \mu_\infty)(\varphi)| \leq |(\mu_\infty - \mu_n)(Q\varphi)| + |(\mu_n Q - \mu_n)(\varphi)| + |(\mu_n - \mu_\infty)(\varphi)|.$$

Par convergence d'une sous-suite vers μ_∞ , comme φ et $Q\varphi \in \mathcal{C}_b(\mathbb{X})$ on peut faire tendre le premier et le dernier terme vers 0. Par construction,

$$\mu_n(Q - I) = \frac{1}{n}(q^{n+1}(x, \cdot) - q(x, \cdot)).$$

Ainsi le second terme se contrôle par $\frac{2}{n}\|\varphi\|_\infty$. □

Corollaire 4.33. Soit q un noyau fellérien sur un espace compact, alors il existe une mesure de probabilité stationnaire.

Démonstration. Par le théorème 1.34, toute famille de probabilité sur un espace compact est équitendue. □

L'hypothèse de compacité est en général trop restrictive. Heureusement on peut se limiter à un « bon » ensemble.

Définition 4.34 (Ensembles absorbants). Un ensemble fermé $B \subset \mathbb{X}$ tel que

$$q(x, B) = 1, \quad \forall x \in B,$$

est dit *absorbant*.

Remarque 4.35. En termes de la chaîne (X_n) , cela signifie que, $\mathbf{P}^\mu[X_{n+\tau_B} \in B] = 1$ pour tout $n \in \mathbb{N}$ et tout $\mu \in \mathcal{P}(\mathbb{X})$.

Corollaire 4.36. *Soit q un noyau fellerien qui admet un ensemble absorbant compact. Alors, il existe une mesure de probabilité stationnaire.*

Démonstration. Soit $x \in B$, alors la suite $(q^n(x, \cdot))_n$ s'identifie à une suite de $\mathcal{P}(B)$ qui est compact par le théorème 1.34. $\square$

Critère par fonctions de Lyapunov

Nous allons maintenant donner un critère fonctionnel pour l'existence d'une mesure stationnaire.

Définition 4.37 (Fonctions de Lyapunov). Une fonction mesurable $V : \mathbb{X} \rightarrow [0, \infty]$ est une *fonction de Lyapunov* pour le noyau markovien q si

- (i) il existe $x \in \mathbb{X}$ tel que $V(x) < \infty$;
- (ii) les sous-ensembles de niveau $\{V \leq c\}$ sont compacts pour tout $c \in \mathbb{R}_+$;
- (iii) il existe $\gamma \in (0, 1)$ et $C \in \mathbb{R}$ telles que

$$QV(x) \leq \gamma V(x) + C,$$

dès que $V(x) \neq \infty$.

Théorème 4.38 (Critère de Lyapunov). *Tout noyau fellerien qui admet une fonction de Lyapunov a une mesure stationnaire.*

Démonstration. Prenons $x \in \mathbb{X}$ tel que $V(x) \neq \infty$. Posons

$$v_n := \mathbf{E}_x[V(X_n)] = Q^n V(x), \quad \forall n \in \mathbb{N}.$$

Par récurrence sur la condition de Lyapunov, on trouve que

$$v_n \leq \gamma v_{n-1} + C \leq \gamma^2 v_{n-2} + C(1 + \gamma) \leq \dots \leq \gamma^n v_0 + \frac{C}{1 - \gamma}.$$

Donc la suite (v_n) est bornée, disons, par v^* . En utilisant l'inégalité de Tchebychev,

$$\mathbf{P}_x[V(X_n) > c] \leq \frac{v^*}{c}.$$

Soient $\varepsilon > 0$, en considérant le compact $K := \{V \leq \varepsilon/v^*\}$, on a montré que $q^n(x, K) \geq 1 - \varepsilon$. On applique ensuite le critère de Krylov–Bogolioubov théorème 4.32. $\square$

Théorème 4.39 (Intégrabilité des fonctions de Lyapunov). *Soit v une fonction de Lyapunov pour un noyau markovien q . Alors, si μ est une mesure stationnaire pour q on a*

$$\int v d\mu \leq \frac{C}{1 - \gamma}.$$

Démonstration. On prend $M \in \mathbb{R}_+$ et on écrit

$$\mu(V \wedge M) = \mu Q(V \wedge M) = \mu(Q(V \wedge M)) \leq \mu((\gamma V + C) \wedge M).$$

D'où on obtient par récurrence

$$\mu(V \wedge M) \leq \int \left(\gamma^n V + \frac{C}{1-\gamma} \right) \wedge M d\mu, \quad \forall n \in \mathbb{N}.$$

Par convergence dominée, on peut passer à la limite en n ce qui donne

$$\int V \wedge M d\mu \leq \frac{C}{1-\gamma}.$$

On conclut en faisant $M \rightarrow \infty$ et en utilisant la convergence monotone. $\square$

4.3.2 UNICITÉ DE LA MESURE STATIONNAIRE

Phénomène de contraction déterministe

On sait qu'un système récurrent $x_{n+1} = f(x_n)$ avec f strictement contractante admet un unique point fixe. Pour une chaîne de Markov donnée comme un système récurrent aléatoire $X_{n+1} = f(X_n, U_{n+1})$, on peut espérer que si f est suffisamment contractante en moyenne alors on peut toujours conclure. C'est effectivement le cas comme montré par ce critère simple.

Théorème 4.40 (Critère de contraction). *Supposons qu'il existe $\gamma \in (0, 1)$ tel que*

$$\mathbf{E}[d(f(x, U), f(y, U))] \leq \gamma d(x, y), \quad \forall x, y \in \mathbb{X}. \quad (4.5)$$

Alors, toute chaîne de Markov définie par la récurrence $X_{n+1} = f(X_n, U_{n+1})$ admet au plus une mesure de probabilité stationnaire.

Démonstration. Soit (X_n) et (Y_n) deux chaînes de Markov définies par récurrence avec le même aléa (U_n) et de condition initiale indépendante de loi donnée par une mesure stationnaire. Par construction de (X_n, Y_n) et par (4.5), on trouve

$$\mathbf{E}[1 \wedge d(X_{n+1}, Y_{n+1})] = \mathbf{E}[1 \wedge d(f(X_n, U_n), f(Y_n, U_n))] \leq \mathbf{E}[1 \wedge \gamma d(X_n, Y_n)].$$

D'où par récurrence,

$$\mathbf{E}[1 \wedge d(X_n, Y_n)] \leq \mathbf{E}[1 \wedge \gamma^n d(X_0, Y_0)].$$

En utilisant la proposition 1.40, on voit que (X_n, Y_n) est équitendue. En passant à la limite le long d'une sous-suite convergente et en utilisant le théorème de convergence dominée, on trouve que

$$\mathbf{E}[1 \wedge d(X_\infty, Y_\infty)] = 0.$$

On trouve donc que

$$\mathbf{P}[X_\infty = Y_\infty] = 1,$$

donc en particulier X_∞ et Y_∞ ont même loi mais comme les chaînes (X_n) et (Y_n) sont stationnaires les conditions initiales sont les mêmes. $\square$

Vérifier (4.5) sur tout $\mathbb{X}$ est en général trop restrictif. Comme pour les critères de compacité, on aimerait bien se limiter à un « bon » ensemble. Dans ce cas les ensembles absorbants ne suffisent plus.

Définition 4.41 (Ensembles attracteurs). Pour tout ensemble mesurable B on associe la suite récurrente $B_0 := B$ puis

$$B_{n+1} := \{x \in \mathbb{X} : q(x, B_n) > 0\}, \quad n \in \mathbb{N}.$$

L'ensemble B est *attracteur* si $\mathbb{X} = \bigcup_{n \in \mathbb{N}} B_n$.

Théorème 4.42 (Réduction à un ensemble absorbant et attracteur). Soit B un ensemble absorbant et attracteur, alors toute mesure de probabilité stationnaire est supportée sur B .

Démonstration. Montrons par récurrence que $q^n(x, B) > 0$ pour tout $n \in \mathbb{N}$ et $x \in B_n$. C'est vrai pas hypothèse pour $n := 0$. Si c'est vrai au rang $n \in \mathbb{N}$, alors

$$q^{n+1}(x, B) = \int_{\mathbb{X}} q^n(y, B) q(x, dy) \geq \int_{B_n} q^n(y, B) q(x, dy), \quad \forall x \in \mathbb{X}.$$

Par hypothèse de récurrence $q^n(\cdot, B) > 0$ sur B_n , et par définition de B_{n+1} , $q^n(x, B_n) > 0$ pour tout $x \in B_{n+1}$.

Prenons μ une mesure de probabilité stationnaire et supposons que $\mu(B) < 1$. Comme B est attracteur et que $\mu(\mathbb{X}) = 1$, il existe $n \in \mathbb{N}^*$ tel que $\mu(B_n \setminus B) > 0$. Par invariance

$$\mu(B) = \mu Q(B) = \int q^n(x, B) \mu(dx) \geq \int_B q^n(x, B) \mu(dx) + \int_{B_n \setminus B} q^n(x, B) \mu(dx).$$

Par invariance de B , $q^n(x, B) = 1$ pour tout $x \in B$. De plus, par construction $\mu(B_n \setminus B) > 0$ et on a montré que $q^n(\cdot, B) > 0$ sur B_n . Donc l'inégalité précédente implique que $\mu(B) > \mu(B)$. Contradiction. $\square$

Théorème 4.43 (Critère de contraction sur les ensembles absorbants et attracteurs). Si la chaîne admet un ensemble absorbant et attracteur B , alors pour que les conclusions du théorème précédent restent vrai, il suffit de vérifier (4.5) pour tout $x, y \in B$.

Démonstration. C'est la même preuve. Puisque par le théorème 4.42, les mesures stationnaires sont supportées dans B , on trouve que presque sûrement $X_0 \in B$ et $Y_0 \in B$. Par invariance, on a alors que $X_n \in B$ et $Y_n \in B$ et on voit que l'on a besoin de (4.5) uniquement sur B . $\square$

Phénomènes probabilistes

S'il n'y a pas unicité de la mesure stationnaire, alors il existe deux mesures ergodiques donc étrangères. Si cela implique que les supports sont disjoints, on aura un méta-théorème que si la chaîne « visite suffisamment » $\mathbb{X}$ alors il devrait y avoir unicité. Ce n'est toutefois pas vrai, même si on suppose la propriété de Feller, que nous devons donc renforcer.

Définition 4.44 (Propriété de Feller forte). L'opérateur Q est *fortement fellérien* si $Q\varphi \in \mathcal{C}_b(\mathbb{X})$ dès que φ est mesurable et bornée.

Lemme 4.45. Soit un opérateur de Markov fortement fellérien. Alors, deux mesures de probabilité stationnaires étrangères sont toujours à support disjoint.

Démonstration. Soient μ et $\nu \in \mathcal{F}$ étrangères. Choisissons un borélien B tel que $\mu(B) = 1$ et $\nu(B) = 0$. Puisque

$$1 = \mu(B) = \int q(x, B) \mu(dx),$$

on trouve que $q(x, B) = 1$ pour μ -presque tout $x \in \mathbb{X}$. De la même manière, $\nu(x, B) = 0$ pour ν -presque tout x .

Considérons $\varphi := Q1_B$. Par le calcul précédent $\varphi = 1$ μ -presque sûrement et $\varphi = 0$ ν -presque sûrement. Comme φ est continue, on trouve aussi que $\varphi = 1$ sur le support de μ et 0 sur le support de ν . $\square$

Définition 4.46 (Point accessible). Le point $x \in \mathbb{X}$ est *accessible* si pour tout $y \in \mathbb{X}$ et voisinage U de x , il existe $k \in \mathbb{N}^*$ tel que

$$q^k(y, U) > 0.$$

Lemme 4.47. Soit x un point accessible, alors il est dans le support de toute mesure de probabilité stationnaire.

Démonstration. Soient π une probabilité stationnaire, x un point accessible et U un voisinage de x . Posons

$$R := \sum_{n \in \mathbb{N}^*} 2^{-n} Q^n.$$

Par invariance

$$\pi(U) = (\pi R)(U) = \int r(y, U) \pi(dy).$$

La dernière quantité est strictement positive par hypothèse. $\square$

Remarque 4.48. L'accessibilité est une propriété topologique. Toutefois, la propriété de Feller est trop faible et l'accessibilité est trop faible pour garantir l'unicité.

Théorème 4.49. Si une chaîne de Markov fortement fellerienne admet un point accessible alors elle a au plus une mesure de probabilité stationnaire.

Démonstration. Par l'absurde, supposons que la chaîne admette deux mesures de probabilité stationnaires. Alors par théorème 2.41, elle admet deux mesures ergodiques. Par lemme 2.40 elles sont étrangères et par lemme 4.45, elles sont à support disjoint, mais par lemme 4.47 leur support contient le point accessible. $\square$

4.3.3 OUVERTURE : VITESSE DE CONVERGENCE

Les théorèmes 4.10 et 4.30 garantissent que μQ^n converge vers l'unique mesure stationnaire π , mais ils restent *qualitatifs* : ils ne disent rien de la *vitesse* de cette convergence. Pour les applications numériques, la question centrale est précisément : au bout de combien d'étapes μQ^n est-elle *proche* de π ? Nous esquissons ici, sans démonstration, un cadre qui permet de répondre à cette question ; les preuves font l'objet d'exercices de la feuille de TD.

Distance en variation totale.

Une façon de quantifier l'écart entre deux lois est la distance en *variation totale*. Pour μ et $\nu \in \mathcal{P}(\mathbb{X})$,

$$\|\mu - \nu\|_{TV} := \sup\{|\mu(A) - \nu(A)| : A \subset \mathbb{X}\} = \inf\{\mathbf{P}[X \neq Y] : X \sim \mu, Y \sim \nu\}.$$

Remarque 4.50. Si $\mu - \nu$ est remplacée par une mesure signée, on peut vérifier que l'expression définit une norme sur l'espace des mesures signées. C'est pourquoi nous adoptons la notation de norme bien que quand on se limite aux mesures de probabilités, il serait plus correct d'écrire $d_{TV}(\mu, \nu)$.

On peut vérifier que si μ et ν ont toutes les deux une densité par rapport à une mesure commune λ (non nécessairement finie), alors

$$\|\mu - \nu\|_{TV} = \frac{1}{2} \left\| \frac{d\mu}{d\lambda} - \frac{d\nu}{d\lambda} \right\|_{L^1(\lambda)},$$

et que cette quantité ne dépend pas du choix de λ .

La convergence du théorème 4.30, écrite avec la somme $\sum_y |\mu Q^n(y) - \pi(y)|$, n'est donc rien d'autre que

$$\|\mu Q^n - \pi\|_{TV} \rightarrow 0.$$

La démonstration de ce théorème fournit d'ailleurs *déjà* une borne quantitative : avec le couplage (X^1, X^2) et le temps de rencontre S qui y sont construits,

$$\|\mu Q^n - \pi\|_{TV} \leq \hat{\mathbf{P}}^{\mu \otimes \pi}[S > n].$$

Obtenir une borne sur la vitesse de convergence revient ainsi à contrôler la queue du temps de couplage S : plus les deux copies se rencontrent vite, plus la chaîne « oubliée » vite sa condition initiale.

Contraction et coefficient de Dobrushin

Sur un espace *fini*, cet oubli est géométrique. La quantité qui en donne le taux est le *coefficient de contraction de Dobrushin*

$$\delta(Q) := \max_{x, y \in \mathbb{X}} \|Q(x, \cdot) - Q(y, \cdot)\|_{TV} \in [0, 1],$$

qui mesure à quel point deux lignes de Q peuvent différer. On montre que Q *contracte* la variation totale au taux $\delta(Q)$:

$$\|\mu Q - \nu Q\|_{TV} \leq \delta(Q) \|\mu - \nu\|_{TV}, \quad \forall \mu, \nu \in \mathcal{P}(\mathbb{X}),$$

et que δ est sous-multiplicative, $\delta(Q^n) \leq \delta(Q)^n$. En itérant à partir de $\nu := \pi$, on obtient la *convergence géométrique* : dès que $\delta(Q^{n_0}) < 1$ pour un certain n_0 (ce qui a lieu si Q est irréductible et apériodique), il existe $\kappa \in (0, 1)$ et $C > 0$ tels que

$$\|\mu Q^n - \pi\|_{TV} \leq C \kappa^n, \quad \forall \mu \in \mathcal{P}(\mathbb{X}), n \in \mathbb{N}.$$

Ce résultat et le calcul explicite de la constante κ font l'objet de l'exercice 73 de la feuille de TD.

Pour aller plus loin 15 (Temps de mélange et cutoff). La convergence géométrique ci-dessus invite à définir le *temps de mélange*

$$t_{\text{mix}}(\varepsilon) := \inf \left\{ n \in \mathbb{N} : \max_{x \in \mathbb{X}} \|\delta_x Q^n - \pi\|_{TV} \leq \varepsilon \right\},$$

c'est-à-dire le nombre d'étapes nécessaire pour être ε -proche de l'équilibre, quelle que soit la condition initiale. Pour de nombreuses familles de chaînes indexées par un paramètre de taille N (mélanges de cartes, marches sur des graphes), on observe un phénomène remarquable de *cutoff* : la distance à l'équilibre reste proche de 1 jusqu'à un temps critique t_N , puis chute brutalement vers 0 sur une fenêtre négligeable devant t_N . La convergence n'est alors pas seulement géométrique : elle est *abrupte*. L'exercice 84 en donne l'exemple emblématique du mélange d'un jeu de n cartes, où $t_n \sim n \log n$. On pourra consulter [Sal] pour un exposé systématique.

Trou spectral et approche L^2

Le coefficient de Dobrushin contrôle la vitesse par un argument de couplage, de nature probabiliste et adapté à la variation totale. Lorsque la chaîne est *réversible* par rapport à une unique mesure de probabilités stationnaire π (définition 3.39) une seconde approche, de nature analytique, éclaire différemment le phénomène. La réversibilité par rapport à π signifie que le générateur $L := Q - I$ est *autoadjoint* sur l'espace de Hilbert $L^2(\pi)$. Vu les propriétés de Q , L est aussi négatif. Par le théorème spectral, $-L$ est alors diagonalisable le long d'une résolution spectrale $(\mathcal{H}_\lambda)_{\lambda \geq 0}$. Puisque $1 \in L^2(\pi)$ et que $L1 = 0$, on a que $\mathcal{H}_0$ contient les constantes. De plus si $\mathcal{H}_0$ ne contient pas que les constantes, il existe des fonctions harmoniques intégrables non constantes et cela contredirait l'unicité de π . On dit que la chaîne admet un *trou spectral* λ_* si $\mathcal{H}_\lambda = \mathbb{R}$ pour tout $\lambda < \lambda_*$. Dans le cas d'un spectre discret (par exemple pour les espaces d'états finis), le trou spectral est simplement la seconde plus petite valeur propre de $-L$. En utilisant le théorème min-max (quotient de Rayleigh), on montre que la chaîne admet un trou spectral λ_* si et seulement si on a l'*inégalité de Poincaré*

$$\lambda_* \mathbf{Var}_\pi[\varphi] \leq - \int \varphi L \varphi d\pi,$$

pour tout φ dans une classe suffisamment large. On peut alors démontrer que l'inégalité de Poincaré implique également une vitesse de convergence géométrique (en norme L^2). C'est un des sujets abordés dans [BGL14].

Exercices fondamentaux

Les exercices ci-dessous, tirés du livret d'exercices, constituent l'entraînement minimal recommandé sur ce chapitre.

Mesures stationnaires et convergence. Exercices 67, 68, 70, 72.

Périodicité. Exercices 69, 75.

Modélisation. Quatre exercices parmi 76, 77, 78, 79, 80, 81, 82, 83.

Modèles de population. Exercices 89, 88.

Temps de retour et transformée de Doob. Exercices 91, 92.

Espaces d'états généraux. Exercices 94, 96, 97.

Troisième partie

Rappels

Définition A.1 (Equivalence topologique). Deux espaces topologiques sont *homéomorphes* s'il existe entre eux une bijection continue d'inverse continue.

Du point de vue de la topologie, deux espaces homéomorphes sont les mêmes.

Définition A.2 (Compacité). Un espace topologique est *compact* si de tout recouvrement ouvert, on peut extraire un sous-recouvrement fini.

Théorème A.3 (Tychonoff). *Un produit d'espaces compacts est compact.*

Définition A.4 (Compacité séquentielle). Un espace topologique est *séquentiellement compact* si de toute suite on peut extraire une sous-suite convergente.

En toute généralité ces deux notions ne sont pas comparables. Par exemple, par le théorème de Tychonoff l'espace des fonctions de $\mathbb{R}$ dans $\{0, 1\}$ muni de la topologie de la convergence simple, c'est-à-dire l'espace produit $\{0, 1\}^{\mathbb{R}}$ est compact. Par contre, la suite de fonctions donnée par $\varphi_n(x)$ = le n -ième bit dans la décomposition dyadique de x ne peut pas admettre de sous-suite convergente (exercice!) et l'espace est non séquentiellement compact.

Réciproquement, en notant ω_1 le premier ordinal non dénombrable et $[0, \omega_1[$ l'espace topologique associé (pour la topologie de l'ordre), c'est un espace non compact mais séquentiellement compact. En effet, la famille d'ouverts

$$\{[0, \alpha[: \alpha < \omega_1\}$$

est un recouvrement de $[0, \omega_1[$ dont on ne peut pas extraire de recouvrement fini (exercice!). Par contre, $[0, \omega_1[$ est séquentiellement compact. Soit (x_n) une suite. Quitte à extraire on peut supposer que (x_n) est monotone (exercice!). Maintenant

(a) ou bien (x_n) est décroissante, auquel cas elle est constante à partir d'un certain rang (car 0 est un minorant), et en particulier convergente;

(b) ou bien (x_n) est strictement croissante et alors elle converge vers $s := \sup_n x_n = \cup_n x_n$ qui est dénombrable (axiome du choix dénombrable) donc $s < \omega_1$.

La notion d'espace métrique permet de réconcilier les deux notions.

Définition A.5 (Distance). Une distance sur $\mathbb{X}$ est une fonction

$$d : \mathbb{X} \times \mathbb{X} \rightarrow \mathbb{R}_+,$$

qui satisfait, pour tout $x, y, z \in \mathbb{X}$

- (définie) $d(x, y) = 0 \Leftrightarrow x = y$;
- (symétrie) $d(x, y) = d(y, x)$;
- (inégalité triangulaire) $d(x, y) \leq d(x, z) + d(z, y)$.

Définition A.6 (Boules). La boule ouverte (resp. fermée) de centre $x \in \mathbb{X}$ et de rayon $r \geq 0$ est l'ensemble $B(x, r)$ (resp. $\bar{B}(x, r)$) de tous les points $y \in \mathbb{X}$ tels que $d(x, y) < r$ (resp. $d(x, y) \leq r$).

Définition A.7 (Espace métrique). Un espace métrique $(\mathbb{X}, d)$ est un espace muni d'une distance. Toute espace métrique est un espace topologique dont les ouverts sont les unions quelconques de boules ouvertes.

Définition A.8 (Espace métrisable). Un espace topologique est *métrisable* s'il existe une distance qui induit sa topologie. Autrement dit, il est métrisable s'il est homéomorphe à un espace métrique.

Théorème A.9 (Bolzano–Weierstrass). *Dans un espace métrisable, les compacts coïncident avec les compacts séquentiels.*

Définition A.10 (Espace complet). Un espace métrique est *complet* si de toute suite de Cauchy on peut extraire une sous-suite convergente.

Définition A.11 (Espace précompact). Un métrique est *précompact* si pour tout $\varepsilon > 0$ il admet un recouvrement par des boules ouvertes de rayon ε . Autrement dit, un espace est précompact si de toute suite on peut extraire une sous-suite de Cauchy.

Un ensemble compact est toujours précompact (exercice!). La condition pour l'implication inverse est exactement la complétude.

Théorème A.12. *Un espace métrique est complet si et seulement si les ensembles relativement compacts coïncident avec les ensembles précompacts.*

Démonstration. Sens direct. Toute suite d'un ensemble précompact est de Cauchy (exercice!) donc convergente par complétude. Donc les précompacts sont séquentiellement compacts et donc compacts par Bolzano–Weierstrass. Pour le sens indirect, on utilise que les suites de Cauchy sont précompacts, donc relativement compacts, donc convergentes. $\square$

Définition A.13 (Espace séparable). Un espace est séparable s'il contient un ensemble dénombrable dense.

Théorème A.14. *Un espace compact métrisable est séparable.*

Démonstration. Il suffit d'utiliser Bolzano–Weierstrass. Pour tout n il existe une famille finie de boule de rayon $1/(n+1)$ qui recouvre tout l'espace. L'union du centre de toutes ces boules est un espace dénombrable dense. $\square$

Théorème A.15 (Stone–Weierstrass). *Soit $\mathbb{X}$ un espace compact. Une sous-algèbre de $\mathcal{C}(\mathbb{X})$ est dense (pour la norme uniforme) si et seulement si*

- (séparation) pour tout $x \neq y \in \mathbb{X}$, il existe $f \in A$ tel que $f(x) \neq f(y)$;
- (non-annulation) pour tout $x \in \mathbb{X}$, il existe $f \in A$ tel que $f(x) \neq 0$.

Corollaire A.16. *Soit $\mathbb{X}$ un compact métrisable, alors $\mathcal{C}(\mathbb{X})$ est séparable.*

Démonstration. On sait que $\mathbb{X}$ est séparable. Soit (x_n) une suite dense. Il suffit de prendre pour A la $\mathbb{Q}$ -algèbre engendrée par les fonctions $\mathbb{X} \ni y \mapsto d(x_n, y)$ et la fonction constante égale à 1. $\square$

B.1 Rappels d'intégration

Nous rappelons ici, sans preuve, quelques résultats utiles en intégration sur un espace de probabilités $(\Omega, \mathfrak{B}, \mathbf{P})$. La plupart des résultats restent vrais si on travaille plutôt sur un espace mesuré $(E, \mathfrak{G}, \mu)$, mais pas tous ! Comme on ne s'intéresse ici qu'aux probabilités on fait le choix de facilité de ne présenter les résultats que dans leur version probabiliste.

Théorème B.1 (Inégalité de Jensen). *Soient X une variable aléatoire et $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ une fonction convexe telles que $\varphi(X)$ soit intégrable, alors*

$$\varphi(\mathbf{E}[X]) \leq \mathbf{E}[\varphi(X)].$$

Définition B.2 (Normes et espaces $\mathcal{L}^p$). Soit X une de variable aléatoire. Pour $1 \leq p < \infty$, la norme $\mathcal{L}^p$ de X est

$$\|X\|_p := \mathbf{E}[|X|^p]^{1/p}.$$

Quand $p = \infty$, on pose

$$\|X\|_\infty := \inf\{t \in \mathbb{R} : \mathbf{P}[X > t] = 0\} = \lim_{p \rightarrow \infty} \|X\|_p.$$

On note $\mathcal{L}^p(\mathbf{P})$, ou tout simplement $\mathcal{L}^p$, l'espace des variables aléatoires X telles que $\|X\|_p < \infty$.

Théorème B.3 (Topologie des espaces $\mathcal{L}^p$). *Pour tout $p \in [1, \infty]$, l'espace normé $(\mathcal{L}^p, \|\cdot\|_p)$ est un espace de Banach. Quand $p = 2$, c'est même un espace de Hilbert.*

Théorème B.4 (Injection continue des espaces $\mathcal{L}^p$). *Pour toute variable aléatoire X et pour tout $1 \leq p \leq q \leq \infty$,*

$$\|X\|_p \leq \|X\|_q.$$

Définition B.5 (Conjugué de Hölder). Pour $p \in [1, \infty]$ on note $p^\star$ l'unique élément de $[1, \infty]$ tel que

$$\frac{1}{p} + \frac{1}{p^\star} = 1.$$

Théorème B.6 (Inégalité de Hölder). *On a l'inégalité de Hölder*

$$\|XY\|_1 \leq \|X\|_p \|Y\|_{p^\star}.$$

Pour aller plus loin 16 (Dualité des espaces $\mathcal{L}^p$). En fait, l'inégalité de Hölder peut s'interpréter comme une inégalité de dualité. Pour tout $Y \in \mathcal{L}^{p^\star}$, considérons la forme linéaire

$$\mathbf{T}_Y : \mathcal{L}^p \ni X \mapsto \mathbf{E}[XY].$$

L'inégalité de Hölder montre que cette application est bien définie et continue, avec $\|\mathbf{T}_Y\| \leq \|Y\|_{p^\star}$. En posant, $X := Y^{p^\star-1}$, on voit que c'est en fait une isométrie $\|\mathbf{T}_Y\| = \|Y\|_{p^\star}$. Autrement

dit, on a montré que $Y \mapsto \mathbf{T}_Y$ est une injection isométrique de $\mathcal{L}^{p^*}$ dans le dual de $\mathcal{L}^p$. Quand $p < \infty$, on peut en fait montrer que cette injection est un isomorphisme. En particulier, le dual de $\mathcal{L}^2$ s'identifie à $\mathcal{L}^2$, ce qui n'est rien d'autre qu'une reformulation au cadre $\mathcal{L}^2$ du théorème de Riesz sur l'identification du dual d'un espace de Hilbert avec l'espace lui-même. Plus généralement, comme $p^{**} = p$, cela montre que pour $p \in (1, \infty)$, le bidual de $\mathcal{L}^p$ s'identifie à $\mathcal{L}^p$, c'est-à-dire que $\mathcal{L}^p$ est un espace réflexif. Par contre, même si le dual de $\mathcal{L}^1$ est $\mathcal{L}^\infty$, on n'a pas en général que le dual de $\mathcal{L}^\infty$ est l'espace $\mathcal{L}^1$.

On peut montrer que le dual de $\mathcal{L}^\infty$ s'identifie à l'ensemble des applications $m : \mathfrak{B} \rightarrow \mathbb{R}$ telles que $m(A) = 0$ si $\mathbf{P}(A) = 0$ et $m(\cup_{i=1}^n A_i) = \sum_{i=1}^n m(A_i)$ pour toute famille finie d'ensembles disjoints de $\mathfrak{B}$. À la différence des mesures on demande ici à m d'être finiment additive et on autorise m à prendre des valeurs négatives. En particulier, toute mesure signée μ absolument continue par rapport à $\mathbf{P}$ est un élément du dual. Par le théorème de Radon–Nikodym (théorème B.12) une telle mesure s'identifie avec un élément de $\mathcal{L}^1$. Pour construire un élément du dual qui n'est pas une mesure absolument continue on a besoin du lemme de Zorn.

Théorème B.7 (Convergence dominée). *Soit (X_n) une suite de variables aléatoires qui converge presque sûrement vers X . S'il existe $Y \in L^1(\mathbf{P})$ telle que, pour tout $n \in \mathbb{N}$, $|X_n| \leq Y$. Alors $X \in L^1(\mathbf{P})$ et*

$$\mathbf{E}[|X_n - X|] \xrightarrow{n \rightarrow \infty} 0.$$

Théorème B.8 (Fatou). *Soit (X_n) une suite de variables aléatoires positives, alors*

$$\mathbf{E}\left[\liminf_{n \rightarrow \infty} X_n\right] \leq \liminf_n \mathbf{E}[X_n].$$

Théorème B.9 (Beppo–Levi). *Soit (X_n) une suite croissante de variables aléatoires positives. Alors,*

$$\lim_{n \rightarrow \infty} \mathbf{E}[X_n] = \mathbf{E}\left[\lim_{n \rightarrow \infty} X_n\right].$$

B.2 Rappels et compléments de théorie de la mesure

Ici on va travailler sur un espace mesurable $(\mathbb{X}, \mathfrak{X})$. On va souvent le munir d'une mesure λ σ -fini. Cela signifie qu'il existe une suite $(\mathbb{X}_n)$ d'ensembles mesurables et de mesure finie tels que $\lambda(\mathbb{X} \setminus \cup_{n \in \mathbb{N}} \mathbb{X}_n) = 0$. En posant

$$p := \sum_{n \in \mathbb{N}} \frac{1}{2^n(1 + \lambda(\mathbb{X}_n))} 1_{\mathbb{X}_n},$$

on trouve que $0 < p < 1$. En particulier en posant $\tilde{\lambda} := p\lambda$ on voit que l'on peut essentiellement se ramener au cas d'une mesure finie. En effet, $\tilde{\lambda}$ a exactement les mêmes ensembles négligeables que λ .

Exemple 13. (a) Toute mesure de probabilité est σ -finie.

(b) La mesure de Lebesgue est σ -finie mais pas finie.

(c) Plus généralement la mesure de Haar sur un groupe localement compact est σ -finie, mais en général pas finie.

Lemme B.10. *Soient μ une mesure non nécessairement σ -finie et λ une mesure finie. Alors, il existe une partition de E en deux ensembles mesurables P et N tels que, pour tout $A \in \mathfrak{E}$:*

$$(i) \quad \mu(A \cap N) \leq \lambda(A \cap N).$$

$$(ii) \quad \mu(A \cap P) \geq \lambda(A \cap P).$$

Pour aller plus loin 17. La mesure $\mu - \lambda$ est un exemple de mesure *signée*. La décomposition (N, P) s'appelle la décomposition de Hahn de $\mu - \lambda$.

Démonstration. Définissons

$$\mathfrak{N} := \{F \in \mathfrak{G} : \forall A \in \mathfrak{G}, \mu(A \cap F) \leq \lambda(A \cap F)\}.$$

Considérons une suite (N_n) de $\mathfrak{N}$ telle que

$$\mu(N_n) - \lambda(N_n) \xrightarrow{n \rightarrow \infty} c := \inf\{\mu(F) - \lambda(F) : F \in \mathfrak{N}\}.$$

Alors, par convergence monotone, $N := \bigcup_{n \in \mathbb{N}} N_n$ est un élément de $\mathfrak{N}$ qui réalise l'infimum. Posons $P := X \setminus N$ et vérifions qu'il a la propriété attendue. Par l'absurde supposons qu'il existe un ensemble mesurable $G_0 \subset P$ tel que $\mu(G_0) < \lambda(G_0)$. Par minimalité de N , $G_0 \notin \mathfrak{N}$, car sinon on aurait $N \cup G_0 \in \mathfrak{N}$ et $(\mu - \lambda)(N \cup G_0) < (\mu - \lambda)(N)$. Cela implique que G_0 contient un ensemble mesurable A tel que $\mu(A) > \lambda(A)$. Posons

$$k_1 := \inf\left\{k \in \mathbb{N} : \exists A \subset G_0, \mu(A) \geq \lambda(A) + \frac{1}{k}\right\}.$$

Soit G_1 un ensemble mesurable associé à k_1 . Comme

$$\mu(G_0 \setminus G_1) = \mu(G_0) - \mu(G_1) \leq \mu(G_0) - \lambda(G_1) - \frac{1}{k_1} < \lambda(G_0) - \lambda(G_1),$$

on construit, en appliquant la construction précédente à $G_0 \setminus G_1$, un ensemble mesurable $G_2 \subset G_0 \setminus G_1$ et un entier k_2 tels que $\mu(G_2) \geq \lambda(G_2) + \frac{1}{k_2}$. Puis par récurrence on construit une suite (G_n, k_n) . Puisque $\mu(G_0) \leq \lambda(G_0) < \infty$, on doit avoir que $\lim_{n \rightarrow \infty} \frac{1}{k_n} = 0$. Posons $N_0 := G_0 \setminus \bigcup_{n \in \mathbb{N}} G_n$. Par construction, N_0 ne contient plus aucun A mesurable tel que $\mu(A) > \lambda(A)$. Donc $N_0 \in \mathfrak{N}$. Mais, on sait que $N_0 \cap N = \emptyset$, et

$$(\mu - \lambda)(N_0) \leq (\mu - \lambda)(G_0) < 0,$$

ce qui contredit la minimalité de N . Donc $\mu(G_0) \geq \lambda(G_0)$, et cela conclut la démonstration. $\square$

Définition B.11 (Absolue continuité, singularité). Soit μ une mesure de probabilité (ou une mesure finie) sur E .

(i) On dit que μ est *absolument continue* par rapport à λ et on note $\mu \ll \lambda$ si pour tout $A \in \mathfrak{G}$ tel que $\lambda(A) = 0$, on a aussi $\mu(A) = 0$.

(ii) On dit que μ et λ sont *étrangères*, et on note $\lambda \perp \mu$ s'il existe $A_\mu, A_\lambda \in \mathfrak{G}$ tels que $A_\mu \cap A_\lambda = \emptyset$, et pour tout $A \in \mathfrak{G}$,

$$\mu(A) = \mu(A \cap A_\mu), \quad \text{et} \quad \lambda(A \cap A_\lambda) = \lambda(A).$$

La mesure $\mu := f\lambda$ avec $f \in \mathcal{L}^1(\lambda)$ est une mesure absolument continue par rapport à λ . On peut se demander s'il existe des mesures absolument continues qui n'ont pas cette forme. Par le théorème suivant, ce n'est pas possible tant que λ est « suffisamment raisonnable ».

Théorème B.12 (Radon–Nikodym). Soient λ une mesure σ -finie et μ une mesure.

(i) Il existe une unique paire de mesures (μ_a, μ_s) telles que $\mu_a \ll \lambda$, $\mu_s \perp \lambda$ et $\mu = \mu_a + \mu_s$.

(ii) Il existe une unique (à ensembles λ -négligeables près) f mesurable et à valeurs dans $[0, \infty]$ telle que

$$\mu_a = f\lambda.$$

Remarque B.13. La fonction $f \in \mathcal{L}^1(\lambda)$ si et seulement si μ est une mesure finie ; $f \in \mathcal{L}_{loc}^1(\lambda)$ si et seulement si μ est σ -finie.

Démonstration. Comme on a vu que l'on peut construire une mesure finie $\tilde{\lambda} = p\lambda$ avec $0 < p < \infty$, il suffit de montrer le théorème pour λ une mesure finie.

On suppose d'abord que μ est finie. Pour l'unicité, s'il existait deux telles paires (μ_a, μ_s) et (μ'_a, μ'_s) , on aurait $\mu_a - \mu'_a = \mu_s - \mu'_s$. Comme le terme de gauche est absolument continue par rapport à λ et celui de droite singulier, les deux côtés doivent faire 0. De la même manière si $\mu_a = h\lambda = h'\lambda$ pour $h, h' \in \mathcal{L}^1(\lambda)$, alors pour tout $A \in \mathfrak{E}$, on trouve que

$$\int_A (h - h') d\mu = 0,$$

et donc $h = h'$. Pour l'existence, l'idée est de montrer les deux propriétés en même temps. Considérons la mesure finie $\nu := \mu + \lambda$. Par Jensen, puis Cauchy-Schwarz, on a, pour $\varphi \in \mathcal{L}^2(\nu)$

$$\left| \int \varphi d\mu \right| \leq \int |\varphi| d\mu \leq \int |\varphi| d\nu \leq \|\varphi\|_{\mathcal{L}^2(\nu)} \nu(E)^{1/2}.$$

D'où, l'application linéaire $\varphi \mapsto \int \varphi d\mu$ est continue sur $\mathcal{L}^2(\nu)$. En utilisant que $\mathcal{L}^2(\nu)$ est son propre dual, on trouve donc un $g \in \mathcal{L}^2(\nu)$ tel que

$$\int \varphi d\mu = \int \varphi g d\nu, \quad \varphi \in \mathcal{L}^2(\nu).$$

En particulier, pour tout $A \in \mathfrak{E}$,

$$0 \leq \frac{\mu(A)}{\nu(A)} \leq \frac{1}{\nu(A)} \int g d\nu \leq 1.$$

Ce qui montre que $g \in [0, 1]$, ν -presque sûrement. En utilisant que $\nu = \mu + \lambda$, on trouve

$$\int (1 - g) \varphi d\mu = \int \varphi g d\lambda, \quad \varphi \in \mathcal{L}^2(\nu). \quad (\text{B.1})$$

Définissons les ensembles $A := \{0 \leq g < 1\}$ et $S := \{g = 1\}$, ainsi que les mesures

$$\mu_a := \mu(\cdot \cap A), \quad \mu_s := \mu(\cdot \cap S).$$

Vérifions que (μ_a, μ_s) satisfait les propriétés attendues. Tout d'abord on a

$$0 = \nu(g \notin [0, 1]) = \nu(\mathbb{X} \setminus (A \cap S)) \geq \mu(\mathbb{X} \setminus (A \cap S)).$$

Donc $\mu_a + \mu_s = \mu$. Deuxièmement, on écrit.

$$\lambda(S) = \int_S g d\lambda = \int_S (1 - g) d\mu = 0.$$

Ce qui donne $\lambda(S) = 0$ et ainsi $\mu_s \perp \lambda$. Finalement, comme $g \in [0, 1]$, g ainsi que toutes ses puissances sont dans $\mathcal{L}^2(\nu)$. On peut donc choisir dans (B.1), $\varphi := (1 + g + \dots + g^n)1_B$ pour $n \in \mathbb{N}$ et $B \in \mathfrak{E}$, ce qui donne

$$\int_B (1 - g^{n+1}) d\mu = \int_B g(1 + g + \dots + g^n) d\lambda.$$

Notons que sur S , $1 - g^{n+1} = 0$, tandis que sur A , g^{n+1} décroît vers 0 quand $n \rightarrow \infty$. Par convergence monotone, le terme de gauche converge vers $\mu(A \cap B) = \mu_a(B)$. D'autre part, $g(1 + g + \dots + g^n)$ converge en croissant vers une limite h , donc par convergence monotone, le terme de droite converge vers $\int_B h d\lambda$. On a donc montré que

$$\mu_a(B) = \int_B h d\lambda.$$

Cela établit que $\mu_a = h\lambda$ et en particulier que $\mu_a \ll \lambda$.

Supposons maintenant seulement μ σ -finie et écrivons (X_n) pour une suite d'ensembles disjoints et mesurables telle que $X = \bigcup_{n \in \mathbb{N}} X_n$ et $\mu(X \setminus \bigcup_{n \in \mathbb{N}} X_n) = 0$. L'unicité du couple (μ_a, μ_s) se prouve exactement de la même manière. Pour l'unicité de h , c'est également la même méthode à la différence près qu'il faut travailler sur des ensembles de μ -mesures finies et écrire pour tout $A \in \mathcal{E}$:

$$\int_{A \cap X_n} (h - h') d\lambda = 0,$$

et conclure par convergence monotone. Pour l'existence, on considère les mesures finies

$$\mu^{(n)} := \mu(\cdot \cap X_n), \quad n \in \mathbb{N}.$$

Par le cas précédent, il existe une suite $(h^{(n)})$ et une suite de mesures singulières $(\mu_s^{(n)})$, telles que

$$\mu^{(n)} = h^{(n)} d\lambda + \mu_s^{(n)}.$$

En posant

$$h := \sum_{n \in \mathbb{N}} h^{(n)}, \quad \text{et} \quad \mu_s = \sum_{n \in \mathbb{N}} \mu_s^{(n)},$$

on trouve une décomposition qui convient.

Finalement passons au cas de μ général. L'unicité du couple (μ_a, μ_s) se prouve à nouveau de la même manière. L'unicité pour h est un peu plus compliquée et comme pour l'existence nous devons effectuer quelques calculs préparatoires. L'idée est de trouver un $X_0 \in \mathcal{E}$ tel que

$$\mu_0 := \mu(\cdot \cap X_0), \quad \text{et} \quad \mu_\infty := \mu(\cdot \cap (X \setminus X_0)),$$

satisfont les propriétés suivantes :

1. μ_0 est σ -finie;
2. pour tout $A \in \mathcal{E}$, $\mu_\infty(A) = \infty$ si $\lambda(A) > 0$ et $\mu_\infty(A) = 0$ si $\lambda(A) = 0$.

Supposons construit un tel ensemble et les mesures associées et montrons comment conclure. Pour l'existence, par le cas précédent on peut trouver h_0 et μ_s telles que $\mu_0 = h_0 \lambda + \mu_s$. En posant

$$h := h_0 1_{X_0} + \infty 1_{X \setminus X_0},$$

on trouve la décomposition attendue. Pour l'unicité de h , supposons que l'on ait $\mu_a = \tilde{h} \lambda$. Tout d'abord en travaillant sur X_0 et en argumentant comme dans le cas σ -fini, on trouve que $\tilde{h} 1_{X_0} = h_0$. Montrons que $\tilde{h} 1_{X \setminus X_0} = +\infty$ ce qui finira la démonstration. Calculons

$$\mu_\infty(\{\tilde{h} \leq n\}) = \int_{X \setminus X_0} \tilde{h} 1_{\tilde{h} \leq n} d\lambda \leq n \lambda(E).$$

Comme μ_∞ ne peut prendre que deux valeurs 0 ou $+\infty$ on trouve donc que le terme de gauche est fini et nécessairement $\lambda(\{\tilde{h} \leq n\}) = 0$, ce qui donne par limite monotone $\lambda(\{\tilde{h} < \infty\}) = 0$.

Pour finir montrons que l'on peut bien décomposer μ comme annoncé. Notons (N_n) les ensembles obtenus dans la décomposition lemme B.10. C'est-à-dire que pour tout $A \in \mathfrak{C}$, $\mu(A \cap N_n) \leq n\lambda(A \cap N_n)$, puis que $\mu(A \cap (X \setminus N_n)) \geq n\lambda(A \cap (X \setminus N_n))$. Avec $\mathbb{X}_0 := \bigcup_{n \in \mathbb{N}} N_n$, on vérifie facilement les propriétés annoncées. $\square$

Remarque B.14. La condition de σ -finitude de λ est cruciale même si μ est finie. Si on prend, sur $\mathbb{R}$ muni de la tribu borélienne, pour λ la mesure de comptage sur les boréliens et pour μ la mesure gaussienne. Il est clair que $\mu \ll \lambda$. Mais on ne peut pas avoir $\mu = h\lambda$ (exercice!).

Les théorèmes suivants permettent de montrer des propriétés sur les mesures à partir de propriétés sur des classes plus petites respectant certaines propriétés.

Définition B.15 (Classe monotone). Un ensemble $\mathfrak{M}$ de parties de $\mathbb{X}$ est une *classe monotone* si

- (i) $\mathbb{X} \in \mathfrak{M}$;
- (ii) elle stable par unions dénombrables croissantes, c'est-à-dire : si (A_n) est une suite croissante d'éléments de $\mathfrak{M}$, alors $\bigcup A_n \in \mathfrak{M}$;
- (iii) et si $A, B \in \mathfrak{M}$ avec $B \subset A$, alors $A \setminus B \in \mathfrak{M}$.

Lemme B.16. Une intersection quelconque de classes monotones est une classe monotone.

Démonstration. Soit $(\mathfrak{M}_i)_{i \in I}$ une famille de classes monotones indexées par un ensemble quelconque. Puisque $\mathbb{X} \in \mathfrak{M}_i$ pour tout $i \in I$, on trouve que $\mathbb{X} \in \bigcap \mathfrak{M}_i$. Si (A_n) est une suite croissante dans $\mathfrak{M}$, alors c'est aussi une suite croissante dans $\mathfrak{M}_i$ pour tout $i \in I$ et donc $\bigcup A_n \in \mathfrak{M}_i$ pour tout $i \in I$. Si A et $B \in \mathfrak{M}$ avec $B \subset A$, en utilisant que l'on a aussi A et $B \in \mathfrak{M}_i$ pour tout i on trouve que $A \setminus B \in \mathfrak{M}_i$ pour tout $i \in I$. $\square$

Définition B.17 (Classe monotone engendrée). Pour un ensemble de parties $\mathfrak{C}$, on pose

$$\mathfrak{M}(\mathfrak{C}) := \bigcap \{ \mathfrak{M} : \mathfrak{M} \text{ est une classe monotone qui contient } \mathfrak{C} \}.$$

Remarque B.18. Comme pour les tribus, la classe monotone engendrée par $\mathfrak{C}$ est la plus petite classe monotone qui contient $\mathfrak{C}$.

Lemme B.19. Une classe monotone stable par intersection finie est une tribu.

Démonstration. Si $\mathfrak{m}$ est stable par intersection finie, elle est aussi stable par union finie en passant au complémentaire. Puis si (A_n) est une suite quelconque d'éléments de $\mathfrak{M}$, par la propriété d'unions croissantes on trouve que

$$\bigcup_{n \in \mathbb{N}} A_n = \bigcup_{n \in \mathbb{N}} \left(\bigcup_{i \leq n} A_i \right) \in \mathfrak{M}.$$

$\square$

Théorème B.20 (Lemme de classe monotone ensembliste). Soit $\mathfrak{C}$ un ensemble de parties de $\mathbb{X}$ stable par intersections finies. Alors, $\mathfrak{M}(\mathfrak{C}) = \sigma(\mathfrak{C})$.

Démonstration. Comme $\sigma(\mathfrak{C})$ est une classe monotone, on a bien que $\mathfrak{M}(\mathfrak{C}) \subset \sigma(\mathfrak{C})$. Pour montrer l'inclusion inverse, on va montrer que $\mathfrak{M}(\mathfrak{C})$ est une tribu. En vertu du lemme précédent, c'est suffisant de montrer que $\mathfrak{M}(\mathfrak{C})$ est stable par intersections finies. Fixons $A \in \mathfrak{M}(\mathfrak{C})$ et posons

$$\mathfrak{M}_0 := \{B \in \mathfrak{M}(\mathfrak{C}) : A \cap B \in \mathfrak{M}(\mathfrak{C})\}.$$

Si on montre que $\mathfrak{M}_0$ contient $\mathfrak{M}(\mathfrak{C})$ alors on aura fini. D'abord montrons que $\mathfrak{M}_0$ est une classe monotone. Il est clair que $\mathbb{X} \in \mathfrak{M}_0$. Si B et $B' \in \mathfrak{M}_0$ avec $B \subset B'$ alors

$$A \cap (B' \setminus B) = (A \cap B') \setminus (A \cap B) \in \mathfrak{M}(\mathfrak{C}),$$

et donc $B' \setminus B \in \mathfrak{M}_0$. Finalement, si (B_n) est une suite croissante de $\mathfrak{M}_0$, alors

$$A \cap (\cup B_n) = \cup (A \cap B_n) \in \mathfrak{M}(\mathfrak{C}).$$

Supposons d'abord que $A \in \mathfrak{C}$, comme $\mathfrak{C}$ est par intersections finies, on trouve alors que $\mathfrak{C} \subset \mathfrak{M}_0$. Ce qui permet de conclure. Dans le cas général, le cas particulier montre que $\mathfrak{C} \subset \mathfrak{M}_0$ et on conclut également. $\square$

L'exemple canonique d'utilisation du lemme de classe monotone est le suivant.

Proposition B.21. *Soit μ et ν deux mesures de probabilité sur $\mathbb{X}$ qui coïncident sur un ensemble de parties $\mathfrak{C}$, stable par intersections finies. Alors, μ et ν coïncident sur $\sigma(\mathfrak{C})$.*

Démonstration. Par le lemme de classe monotone, il suffit de montrer que l'ensemble

$$\mathfrak{M} := \{A \in \mathbb{X} : \mu(A) = \nu(A)\},$$

est une classe monotone. On aura alors $\mathfrak{M}(\mathfrak{C}) = \sigma(\mathfrak{C}) \subset \mathfrak{M}$. Nous laissons les détails en exercice. $\square$

Exemple 14. Deux mesures de probabilité qui ont les mêmes fonctions de répartition sont les mêmes.

Passons maintenant à une version fonctionnelle du lemme de classe monotone. Il est souvent plus utile en pratique, car il est plus facile de manipuler des fonctions plutôt que des ensembles.

Théorème B.22 (Lemme de classe monotone fonctionnel). *Soit $\mathcal{H}$ un ensemble de fonctions $f : \mathbb{X} \rightarrow \mathbb{R}$ bornées, tel que $\mathcal{H}$ est stable par convergence monotone bornée : si $(f_n) \subset \mathcal{H}$ est une suite croissante et telle qu'on ait $0 \leq f_n \leq c$ pour tout $n \in \mathbb{N}$ pour un certain $c > 0$, alors $\lim f_n \in \mathcal{H}$. Soit $\mathcal{C} \subset \mathcal{H}$ un espace vectoriel qui contient 1 et stable par $\wedge$. Alors, $\mathcal{H}$ contient toutes les fonctions bornées et mesurables par rapport à $\sigma(\mathcal{C})$.*

Démonstration. Considérons $\mathcal{E}$ l'ensemble des espaces vectoriels inclus dans $\mathcal{H}$, stables par $\wedge$ et qui contiennent $\mathcal{C}$. Soit $(\mathcal{C}_i)_{i \in I}$ une famille totalement ordonnée de $\mathcal{H}$. Alors je prétends que $\mathcal{C}' := \cup_{i \in I} \mathcal{C}_i$ est un majorant de cette famille. Il est clair que $\mathcal{C} \subset \mathcal{C}' \subset \mathcal{H}$. Maintenant soit $x, y \in \mathcal{C}'$ et $\lambda \in \mathbb{R}$. Il existe i et $j \in I$ tels que $x \in \mathcal{C}_i$ et $y \in \mathcal{C}_j$. Comme l'ensemble est complètement ordonné, on suppose sans perte de généralité que $\mathcal{C}_i \subset \mathcal{C}_j$. Ainsi, $x \in \mathcal{C}_j$. Comme $\mathcal{C}_j$ est un espace vectoriel stable par $\wedge$ on trouve que $\lambda x + y \in \mathcal{C}_j \subset \mathcal{C}'$ et $x \wedge y \in \mathcal{C}_j \subset \mathcal{C}'$. On a montré que $\mathcal{C}' \in \mathcal{E}$ et $\mathcal{C}'$ est clairement un majorant de la famille. En particulier, l'ensemble $\mathcal{E}$ muni de l'inclusion satisfait les hypothèses du lemme de Zorn et il existe un élément maximal $\mathcal{C}_0$. Par maximalité de $\mathcal{C}_0$, cet espace vectoriel est aussi stable par passage à la limite monotone positive.

Posons

$$\mathfrak{A} := \{A \in \mathfrak{P}(E) : 1_A \in \mathcal{C}_0\}.$$

Comme $\mathcal{C}_0$ est $\wedge$ -stable et que $1_A \wedge 1_B = 1_{A \cap B}$, on trouve que $\mathfrak{A}$ est stable par intersections finies. Comme $1 \in \mathcal{C}_0$ et que c'est un espace vectoriel, en écrivant $1_{A^c} = 1 - 1_A$, on voit que $\mathfrak{A}$ est stable par passage au complémentaire. Pour une famille dénombrable (A_n) de $\mathfrak{A}$ on a que

$$1_{\cup_{n \in \mathbb{N}} A_n} = \lim_{n \rightarrow \infty} 1_{\cup_{k \leq n} A_k}.$$

Comme la limite est monotone et que $\mathcal{C}_0$ est stable par une telle limite, on trouve que $\mathfrak{A}$ est une tribu. Puisque toutes les fonctions $\mathfrak{A}$ -mesurables sont des limites monotones de fonctions étagées sur $\mathfrak{A}$, on voit que $\mathcal{C}_0$ contient toutes les fonctions $\mathfrak{A}$ -mesurable.

Montrons maintenant que $\sigma(\mathcal{C}) \subset \mathfrak{A}$. Comme la tribu borélienne est engendrée par les intervalles, il suffit de montrer que, pour tout $X \in \mathcal{C}_0$, les ensembles $\{a \leq X \leq b\}$ pour $a, b \in \mathbb{R}$ sont dans $\mathfrak{A}$. Par les propriétés de $\mathcal{C}_0$, il suffit de le faire pour $\{X \geq 1\}$. Posons $Y := (X \wedge 1)_+ \in \mathcal{C}_0$. Soit $n \in \mathbb{N}$. Posons pour $s \in \mathbb{R}$, $\varphi_n(s) := c_n s^{n/(n-1)}$, avec $c_n := n^{-1/(n-1)} - n^{-n/(n-1)}$. Alors,

$$Y^n = \sup\{sY - \varphi_n(s) : s \in \mathbb{R}\}.$$

Comme $\mathcal{C}_0$ est un espace vectoriel qui contient les constantes, les fonctions affines en Y sont aussi dans $\mathcal{C}_0$, et comme elle est stable par limite monotone on trouve aussi que $Y^n \in \mathcal{C}_0$. Finalement $1_{X \geq 1}$ est la limite décroissante de Y^n . $\square$

Une conséquence immédiate de ce résultat est que les lois sont bien caractérisées par les fonctions continues bornées.

Proposition B.23. Soient μ et ν deux mesures de probabilité définies sur un espace métrique, muni de sa tribu borélienne, telles que $\int h d\mu = \int h d\nu$ pour toute fonction h continue et bornée, alors $\mu = \nu$.

Démonstration. On applique le résultat précédent avec $\mathcal{C}$ l'ensemble des fonctions continues et bornées et $\mathcal{H}$ celui des fonctions boréliennes bornées. $\square$

Théorème B.24 (Riesz–Markov). Soient $\mathbb{X}$ un espace polonais compact, et ι une forme linéaire continue croissante sur $\mathcal{C}(\mathbb{X})$ avec $\iota(1) = 1$. Alors, il existe une unique $\mu \in \mathcal{P}(\mathbb{X})$ telle que

$$\iota(\varphi) = \int \varphi d\mu, \quad \forall \varphi \in \mathcal{C}(\mathbb{X}).$$

Remarque B.25. Croissante signifie que si $\varphi \geq \psi$, alors $\iota(\varphi) \geq \iota(\psi)$. Par linéarité, il suffit d'avoir $\iota(\varphi) \geq 0$ dès que $\varphi \geq 0$.

Démonstration. Voir [Rud87, Thm. 2.14] pour une preuve ad hoc, ou [DM75, § 32-35] pour une approche par théorie des capacités de Choquet. $\square$

Définition B.26 (Mesure extérieure). Une mesure extérieure sur un ensemble $\mathbb{X}$ est une application $\mu : 2^{\mathbb{X}} \rightarrow [0, +\infty]$ qui satisfait

- (i) $\mu(\emptyset) = 0$;
- (ii) si $A \subset B$, alors $\mu(A) \leq \mu(B)$;
- (iii) pour toute famille dénombrable (A_k)

$$\mu(\cup_{k \in \mathbb{N}} A_k) \leq \sum_{k \in \mathbb{N}} \mu(A_k).$$

Théorème B.27 (Carthéodory). *Soit μ une mesure extérieure, alors il existe une tribu $\mathfrak{A}$ qui contient tous les ensembles μ -négligeable et telle que la restriction de μ à la tribu $\mathfrak{A}$ est une mesure.*

Démonstration. Il suffit de poser

$$\mathfrak{A} := \{B \subset \mathbb{X} : \forall A \subset \mathbb{X}, \mu(A) = \mu(A \cap B) + \mu(A \cap B^C)\}.$$

Les détails sont laissés en exercice. □

Grâce à ce théorème d'extension on montre facilement le résultat suivant sur les systèmes projectifs de mesure. Commençons par quelques notations.

Définition B.28 (Système projectif). Soient $(\mathbb{X}, \mathfrak{A})$ un espace mesurable et $\Omega := \mathbb{X}^{\mathbb{N}}$. Notons $P_{(n+1) \rightarrow n}$ et $P_{\mathbb{N} \rightarrow n}$ pour les projections canoniques $\mathbb{X}^{n+1} \rightarrow \mathbb{X}^n$ et $\Omega \rightarrow \mathbb{X}^n$. Un *système projectif* de mesures de probabilité sur Ω est une famille (μ_n) telle que $\mu_n \in \mathcal{P}(\mathbb{X}^n)$ pour tout $n \in \mathbb{N}^*$ et

$$(P_{(n+1) \rightarrow n})_{\#} \mu_{n+1} = \mu_n.$$

Théorème B.29 (Kolmogorov). *Soit (μ_n) un système projectif de mesures de probabilité sur $\Omega := \mathbb{X}^{\mathbb{N}}$. Alors, il existe $\mu \in \mathcal{P}(\Omega)$ telle que*

$$(P_{\mathbb{N} \rightarrow n})_{\#} \mu = \mu_n, \quad \forall n \in \mathbb{N}^*.$$

Remarque B.30. À ce niveau de généralité, on n'a pas l'unicité. Toutefois, si $\mathbb{X}$ est polonais alors on a l'unicité.

Vecteurs gaussiens et théorème central de la limite

C.1 Mesures gaussiennes sur la droite réelle**C.1.1 LOI GAUSSIENNE STANDARD**

Commençons par le cas crucial de la gaussienne standard.

Définition C.1. On note $\mathcal{N}(0, 1)$, ou γ , la mesure sur $\mathbb{R}$ donnée par

$$\gamma(dx) := e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}}.$$

Lemme C.2. La mesure γ est bien une loi de probabilité.

Démonstration. Par un argument de croissance, il est clair que

$$I := \int_{\mathbb{R}} e^{-\frac{x^2}{2}} < \infty.$$

Montrons que $I^2 = 2\pi$. En développant le carré et en faisant un changement en coordonnées radiales, on trouve bien que

$$I^2 = 2\pi \int_0^\infty r e^{-\frac{r^2}{2}} dr = 2\pi \left[e^{-\frac{r^2}{2}} \right]_{r=0}^{r=\infty} = 2\pi.$$

□

Rappelons maintenant plusieurs caractérisations importantes de la gaussienne standard.

Théorème C.3. Soit X une variable aléatoire à valeurs réelles. Les points suivants sont équivalents.

- (i) X suit une gaussienne standard.
- (ii) Pour toute fonction $h \in \mathcal{C}^1(\mathbb{R}) \cap L^2(\gamma)$ telle que $h' \in L^2(\gamma)$:

$$\mathbf{E}[Xh(X)] = \mathbf{E}[h'(X)].$$

- (iii) Pour tout $t \in \mathbb{R}$

$$\mathbf{E}[e^{itX}] = e^{-\frac{t^2}{2}}.$$

- (iv) Pour tout $k \in \mathbb{N}$,

$$\mathbf{E}[X^{2k+1}] = 0, \quad \text{et} \quad \mathbf{E}[X^{2k}] = \prod_{j=1}^k (2j-1).$$

C.1.2 LOI GAUSSIENNE AVEC MOYENNE ET VARIANCE

Définition C.4. Pour m et $\sigma \in \mathbb{R}$, on écrit $\mathcal{N}(m, \sigma^2)$ la loi de $m + \sigma X$ où X est une gaussienne standard.

Remarque C.5. À partir de la transformée de Fourier, X et $-X$ ont la même loi. Ainsi, la loi de $m + \sigma X$ dépend bien uniquement de σ^2 et non pas de σ .

En pratique toutes les propriétés des $\mathcal{N}(m, \sigma^2)$ se déduisent de celles de $\mathcal{N}(0, 1)$: on fait tous les calculs avec $m + \sigma X$ où X est une gaussienne standard. Par exemple, à partir de théorème C.3 on obtient le résultat suivant.

Proposition C.6. Soit $X \sim \mathcal{N}(m, \sigma^2)$, alors

- (i) $\mathbf{E}[e^{itX}] = e^{itm} e^{-\frac{t^2 \sigma^2}{2}}$.
- (ii) $\mathbf{E}[X] = m$ et $\mathbf{Var}[X] = \sigma^2$.

Lorsque $\sigma \neq 0$, la transformation affine $x \mapsto m + \sigma x$ est un $\mathcal{C}^1$ -difféomorphisme et on obtient par la formule de changement de variable une formule pour la densité de $\mathcal{N}(m, \sigma^2)$. Par contre, $\mathcal{N}(m, 0) = \delta_m$ n'a jamais de densité.

Proposition C.7. La loi $\mathcal{N}(m, \sigma^2)$ admet une densité si et seulement si $\sigma \neq 0$ et dans ce cas elle est donnée par

$$x \mapsto \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) \frac{1}{\sqrt{2\pi\sigma^2}}.$$

C.2 Vecteurs gaussiens

L'idée est la même que pour le cas scalaire. On commence avec le cas standard et on obtient le cas général par transformation affine. Comme on travaille sur $\mathbb{R}^d$ avec $d > 1$, il y a un petit peu d'arguments géométriques qui apparaissent.

C.2.1 VECTEUR GAUSSIEN STANDARD

Définition C.8. On note $\mathcal{N}(0, \mathbf{I}_d)$, ou γ_d , la mesure sur $\mathbb{R}^d$ donnée par

$$\gamma(dx) := \exp\left(-\frac{\|x\|^2}{2}\right) \frac{dx}{(2\pi)^{d/2}}.$$

Il est clair que $\gamma_d = \gamma^{\otimes d}$ et que donc γ_d définit bien une mesure de probabilité. Comme dans le cas scalaire, il existe de nombreuses caractérisations des vecteurs gaussiens standards.

Théorème C.9. Soit $X = (X_1, \dots, X_d)$ une variable aléatoire à valeurs dans $\mathbb{R}^d$. Alors les points suivants sont équivalents.

- (i) $X \sim \gamma_d$.
- (ii) Pour toute fonction $h \in \mathcal{C}^2(\mathbb{R}^d) \cap \mathcal{L}^2(\gamma_d)$ telle que ∇h et $\nabla^2 h \in \mathcal{L}^2(\gamma_d)$:

$$\mathbf{E}[\langle X, \nabla h(X) \rangle] = \mathbf{E}[\Delta h(X)].$$

- (iii) Pour tout $t \in \mathbb{R}^d$,

$$\mathbf{E}[\exp(i\langle t, X \rangle)] = \exp\left(-\frac{\|t\|^2}{2}\right).$$

- (iv) Les X_i sont des gaussiennes réelles standard indépendantes.

- (v) Avec $R := \|X\|$ et $U := \frac{X}{\|X\|}$, on a $R \perp U$, U est uniforme sur la sphère unité et R^2 suit une loi χ^2 à d degré de liberté, c'est-à-dire

$$\mathbf{E}[\exp(itR^2)] = (1 - 2it)^{-\frac{d}{2}}, \quad \forall t \in \mathbb{R}.$$

- (vi) La loi de X est invariante par transformation du groupe orthogonale et les coordonnées de X sont indépendantes.

C.2.2 VECTEURS GAUSSIENS AVEC MOYENNE ET COVARIANCE

Comme dans le cas scalaire, les lois gaussiennes générales sont obtenues par transformations affines.

Définition C.10. Pour une matrice A et un vecteur m , on écrit $\mathcal{N}(m, AA^t)$, pour la loi de $m + AX$ avec X un vecteur gaussien standard.

Remarque C.11. Comme dans le cas scalaire, grâce à l'expression de la transformée de Fourier, on voit que la loi de $m + AX$ ne dépend effectivement que de AA^t .

Définition C.12. On dit qu'une matrice C est une matrice de *covariance* si elle symétrique positive (pas forcément défini).

Pour une matrice A , a matrice AA^t est bien entendu une matrice de covariance. Inversement toute matrice de covariance C peut s'écrire sous la forme AA^t . En effet, on peut diagonaliser C en base orthonormée $C = P\Lambda P^{-1}$ avec P dans le groupe orthogonal et Λ une matrice diagonale dont toutes les entrées sont positives. On peut ensuite prendre $A := P\Lambda^{1/2}P^{-1}$. Ainsi on voit que $\mathcal{N}(m, C)$ existe si et seulement si C est une matrice de covariance.

Lorsque $\det A \neq 0$, alors la transformation affine est un $\mathcal{C}^1$ -difféomorphisme de $\mathbb{R}^d$ et la formule du changement de variable donne une densité pour la loi de $m + AX$. Par contre, si $\det A = 0$, alors $\text{rang}(A) < d$. Puisque trivialement, $\mathbf{P}[AX \in \text{Im}(A)] = 1$, on voit que $m + AX$ ne peut pas avoir de densité. On en déduit le résultat suivant.

Proposition C.13. Soient m un vecteur et C une matrice de covariance. La loi $\mathcal{N}(m, C)$ admet une densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$ si et seulement si $\det C > 0$. Dans le cas la densité est donné par

$$x \mapsto \exp\left(-\frac{\langle x - m, C^{-1}x - m \rangle}{2}\right) \frac{1}{(2\pi)^{d/2} \det C^{1/2}}.$$

En travaillant avec la transformée de Fourier de la gaussienne standard on trouve le résultat suivant.

Proposition C.14. Soit X une variable aléatoire à valeurs dans $\mathbb{R}^d$, alors X suit une loi $\mathcal{N}(m, C)$ si et seulement si pour tout $t \in \mathbb{R}^d$

$$\langle t, X \rangle \sim \mathcal{N}(\langle t, m \rangle, \langle t, Ct \rangle).$$

Finissons avec l'équivalence entre l'indépendance et la décorrélation qui n'est vraie que pour les vecteurs gaussiens.

Théorème C.15. Soit X un vecteur gaussien. Alors $X_i \perp X_j$ si et seulement si $\mathbf{Cov}[X_i, X_j] = 0$.

C.3 Le théorème central limite

Passons maintenant au résultat principal de la théorie de la probabilité. Vous en avez vu une démonstration, au moins dans le cas scalaire, en L3. Voir les exercices 9 et 10, pour se rafraîchir la mémoire.

Théorème C.16 (Théorème central limite). *Soit (X_n) une suite de variables aléatoires indépendantes et identiquement distribuées à valeurs dans $\mathbb{R}^d$ et de second moment fini. Alors,*

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n (X_k - \mathbf{E}[X_k]) \xrightarrow[n \rightarrow \infty]{\text{loi}} \mathcal{N}(0, C),$$

où C est la matrice de covariance de X_1 .

Tous les polycopiés sont disponibles sur la page du cours ; tous les livres mentionnés devraient être trouvables à la BU ou en ligne. Si vous avez du mal à trouver une référence, vous pouvez m'écrire. Il existe de nombreux textes sur les probabilités et les chaînes de Markov. Le lecteur qui souhaiterait approfondir certains sujets, ou avoir un point de vue différent de celui adopté ici, trouvera dans cette liste quelques recommandations. Pour les étudiants ayant besoin de rappel en théorie de la mesure et en probabilités, je recommande les excellentes notes de cours suivantes.

- [Féj] J. FÉJOZ. *Chapitres d'intégration et de probabilités*. URL : <https://www.ceremade.dauphine.fr/~fejoz/Integration/integration-probabilites.pdf> qui contient des suppléments intéressants en lien avec le cours.
- [Gal] J.-F. le GALL. *Intégration, probabilités et processus aléatoires*. URL : <https://www.imo.universite-paris-saclay.fr/~jean-francois.le-gall/IPPA2.pdf> qui couvre les essentiels au niveau L3 avec des compléments en chaînes de Markov sur les espaces dénombrables et en théorie des martingales à temps discret. Il a aussi été publié (en anglais) [Gal22].
- [Tao11] T. TAO. *An introduction to measure theory*. Grad. Stud. Math. 126 se concentre sur la théorie de la mesure, mais est très pédagogique et très direct.

Pour des références plus systématiques sur la théorie de la mesure, on pourra consulter les ouvrages suivants.

- [GG25] H. GEISS & S. GEISS. *Measure, probability and functional analysis*. Universitext. DOI : [10.1007/978-3-031-84067-8](https://doi.org/10.1007/978-3-031-84067-8) est bien conçue. Le livre couvre tous les sujets fondamentaux et propose des ouvertures vers des sujets plus avancés.
- [Rud87] W. RUDIN. *Real and complex analysis*. est une référence célèbre et complète.

Pour s'initier aux arcanes de la théorie de la mesure sur les espaces topologiques, on pourra consulter, avec précaution, certaines parties des ouvrages suivants. Attention, ces références ne sont pas faites pour être lues linéairement. Contacter un enseignant si vous souhaitez approfondir un sujet dedans.

- [Bou69] N. BOURBAKI. *Éléments de mathématique. Intégration. Chapitre 9*. DOI : [10.1007/978-3-540-34391-2](https://doi.org/10.1007/978-3-540-34391-2).
- [DM75] C. DELLACHERIE & P.-A. MEYER. *Probabilités et potentiel. Chap. I à IV*. Publ. Inst. Math. Univ. Strasbourg XV.
- [Bog07] V. I. BOGACHEV. *Measure theory. Vol. I and II*.

Pour des références spécialisées en probabilités et chaînes de Markov, on pourra se référer aux entrées suivantes.

- [Bre] J.-C. BRETON. *Processus discrets*. URL : https://jcbreton.pages.math.cnrs.fr/jcb/Lecture_notes/processus_discrets.pdf, le célèbre « JCB » local. Couvre les chaînes de Markov et les martingales à temps discrets.

- [Nev70] J. NEVEU. *Bases mathématiques du calcul des probabilités*, p. 213 est la référence classique de probabilité.
- [BL07] P. BARBE & M. LEDOUX. *Probabilité*. Enseign. SUP-Maths constitue aussi une excellente synthèse, en particulier en vue de l'agrégation.
- [Dud02] R. M. DUDLEY. *Real analysis and probability*. Camb. Stud. Adv. Math. 74. DOI : [10.1017/CB09780511755347](https://doi.org/10.1017/CB09780511755347) que j'ai moi-même beaucoup consulté pour les chapitres 1 et 2. On pourra y approfondir de nombreuses questions et trouver des exercices complémentaires (certains se sont d'ailleurs retrouvés dans la feuille de TD).
- [Dur19] R. DURRETT. *Probability. Theory and examples*. Camb. Ser. Stat. Probab. Math. 49. DOI : [10.1017/9781108591034](https://doi.org/10.1017/9781108591034) est l'ouvrage anglophone de référence.
- [Nor97] J. R. NORRIS. *Markov chains*. Camb. Ser. Stat. Probab. Math. 2. DOI : [10.1017/CB09780511810633](https://doi.org/10.1017/CB09780511810633) est une monographie sur les processus de Markov à temps continu. Le premier chapitre couvre le cas des chaînes de Markov sur les espaces dénombrables.
- [DM83] C. DELLACHERIE & P.-A. MEYER. *Probabilités et potentiel. Chapitres IX à XI : Théorie discrète du potentiel. Edition entièrement refondue*. Publ. Inst. Math. Univ. Strasbourg XVIII pour les chaînes de Markov à espaces d'états quelconques du point de vue notamment de la théorie du potentiel.
- [Hai] M. HAIRER. *Ergodic Properties of Markov Processes*. URL : <https://hairer.org/notes/Markov.pdf> pour les aspects de convergence en temps long pour les espaces d'états quelconques.
- [MT09] S. MEYN & R. L. TWEEDIE. *Markov chains and stochastic stability. Prologue by Peter W. Glynn*. Camb. Math. Libr. DOI : [10.1017/CB09780511626630](https://doi.org/10.1017/CB09780511626630) est la référence anglophone classique sur le sujet.

Pour approfondir des points spécifiques du cours ou des TDs, je recommande les ouvrages suivants.

- [Cou12] Y. COUDÈNE. *Théorie ergodique et systèmes dynamiques*. Savoirs Actuels pour un panorama de la théorie ergodique.
- [Sal] J. SALEZ. *Mixing times of Markov chains*. URL : <https://www.ceremade.dauphine.fr/~salez/mix.pdf> pour les questions de temps de mélange et de vitesse de convergence.
- [CM16] D. CHAFAÏ & F. MALRIEU. *Recueil de modèles aléatoires*. Math. Appl. (Berl.) 78. DOI : [10.1007/978-3-662-49768-5](https://doi.org/10.1007/978-3-662-49768-5) pour des exemples explicites de chaînes de Markov.
- [LL10] G. F. LAWLER & V. LIMIC. *Random walk : A modern introduction*. Camb. Stud. Adv. Math. 123. DOI : [10.1017/CB09780511750854](https://doi.org/10.1017/CB09780511750854) pour une revue détaillée et moderne des marches aléatoires sur $\mathbb{Z}^d$.
- [Sch65] L. SCHWARTZ. *Théorie des distributions* pour la théorie des distributions, une généralisation des mesures sur $\mathbb{R}^d$.
- [Bré94] H. BRÉZIS. *Analyse fonctionnelle. Théorie et applications* pour la théorie des opérateurs en lien avec les chaînes de Markov.
- [Woe08] W. WOESS. *Random walks on infinite graphs and groups*. Camb. Tracts Math. 138. DOI : [10.1017/CB09780511470967](https://doi.org/10.1017/CB09780511470967) pour les marches aléatoires en lien avec la géométrie des groupes.

Pour découvrir ce qui se cache après ce cours, voici une liste de livres plus avancés.

- [RY05] D. REVUZ & M. YOR. *Continuous martingales and Brownian motion*. Grundlehren Math. Wiss. 293 est la référence absolue pour la théorie des processus à temps continus, en particulier à travers le prisme du mouvement brownien.

- [DM87] C. DELLACHERIE & P.-A. MEYER. *Probabilités et potentiel. Chapitres XII à XVI : Théorie du potentiel associée à une résolvante. Théorie des processus de Markov*. Publ. Inst. Math. Univ. Strasbourg XIX pour une présentation générale de la théorie des processus de Markov.
- [Str03] D. W. STROOCK. *Markov processes from K. Itô's perspective*. T. 155. Ann. Math. Stud. DOI : [10.1515/9781400835577](https://doi.org/10.1515/9781400835577) donne un acompte de la théorie des processus stochastique d'un point de vue géométrique.
- [BGL14] D. BAKRY, I. GENTIL & M. LEDOUX. *Analysis and geometry of Markov diffusion operators*. Grundlehren Math. Wiss. 348. DOI : [10.1007/978-3-319-00227-9](https://doi.org/10.1007/978-3-319-00227-9) pour les liens avec les inégalités fonctionnelles.
- [BH91] N. BOULEAU & F. HIRSCH. *Dirichlet forms and analysis on Wiener space*. De Gruyter Stud. Math. 14. DOI : [10.1515/9783110858389](https://doi.org/10.1515/9783110858389) pour une introduction aux formes de Dirichlet.

Quelques textes historiques et fondateurs viennent compléter cette liste : d'abord les textes en français.

- [Poi90] H. POINCARÉ. « Sur le problème des trois corps et les équations de la dynamique ». *Acta Math.* 13, p. 1-270 est l'ancêtre commun de tous les problèmes de système dynamique, en lien avec la section [2.3](#).
- [Leb02] H. LEBESGUE. « Intégrale, longueur, aire. » *Annali di Mat. (3)* 7, p. 231-359. DOI : [10.1007/BF02420592](https://doi.org/10.1007/BF02420592) est le texte fondateur de la théorie de la mesure et de l'intégrale de Lebesgue.
- [Lév37] P. LÉVY. *Théorie de l'addition des variables aléatoires*. Monographies des Probabilités; calcul des probabilités et ses applications 1, p. 328+xvii étudie la convergence en loi et énonce le théorème qui porte son nom (remarque [1.7](#)).
- [Doe40] W. DOEBLIN. « Éléments d'une théorie générale des chaînes simples constantes de Markoff ». *Ann. Sci. Éc. Norm. Supér. (3)* 57, p. 61-111. DOI : [10.24033/asens.883](https://doi.org/10.24033/asens.883) étudie les questions de convergence des chaînes de Markov.
- [KB37] N. KRYLOV & N. BOGOLIOUBOV. « La théorie générale de la mesure dans son application à l'étude des systèmes dynamiques de la mécanique non linéaire ». *Ann. Math. (2)* 38, p. 65-113. DOI : [10.2307/1968511](https://doi.org/10.2307/1968511) établissent le critère démontré dans ce cours (théorème [4.32](#)).

L'histoire du développement de la théorie des probabilités et des chaînes de Markov contient aussi beaucoup de textes en langue étrangère.

- [Mar06] A. A. MARKOV. « Rasprostranenie zakona bol'shikh chisel na velichiny, zavisyaschie drug ot druga ». *Kasan Ges. (2)* 15.4, p. 135-156 est le texte fondateur des chaînes de Markov.
- [Per07] O. PERRON. « Zur Theorie der Matrices. » *Math. Ann.* 64, p. 248-263. DOI : [10.1007/BF01449896](https://doi.org/10.1007/BF01449896) donne une preuve du théorème de Perron–Frobenius (théorème [4.9](#)) dans le cas strictement positif.
- [Fro12] G. FROBENIUS. « Über Matrizen aus nicht negativen Elementen. » *Berl. Ber.* 1912, p. 456-477 donne la preuve générale.
- [Pól21] G. PÓLYA. « Über eine Aufgabe der Wahrscheinlichkeitsrechnung betreffend die Irrfahrt im Straßennetz. » *Math. Ann.* 84, p. 149-160. DOI : [10.1007/BF01458701](https://doi.org/10.1007/BF01458701) établit la récurrence ou la transience de la marche aléatoire simple symétrique sur $\mathbb{Z}^d$ pour $d = 1, 2, 3$ (théorème [4.20](#)).

- [Kol31] A. KOLMOGOROFF. « Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung ». *Math. Ann.* 104, p. 415-458. DOI : [10.1007/BF01457949](https://doi.org/10.1007/BF01457949) marque le début de la théorie moderne des processus de Markov et établit les équations de Chapman–Kolmogorov ((3.1)).
- [Bir31] G. D. BIRKHOFF. « Proof of the ergodic theorem. » *Proc. Natl. Acad. Sci. USA* 17, p. 656-660. DOI : [10.1073/pnas.17.12.656](https://doi.org/10.1073/pnas.17.12.656) établit le théorème ergodique presque sûr (théorème 2.36).
- [Neu32] J. von NEUMANN. « Proof of the quasi-ergodic hypothesis ». *Proc. Natl. Acad. Sci. USA* 18, p. 70-82. DOI : [10.1073/pnas.18.1.70](https://doi.org/10.1073/pnas.18.1.70), soumis à quelques mois d'écart, établit la version L^2 du théorème ergodique.
- [Kol33] A. KOLMOGOROV. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. *Ergeb. Math. Grenzgeb.* 2 est le texte fondateur sur la théorie des probabilités. Il contient en particulier la construction de l'espérance conditionnelle abstraite que nous verrons au chapitre 2. Pour les non-germanophiles, on pourra consulter la traduction anglaise des travaux de Kolmogorov [Kol92].
- [Kac47] M. KAC. « On the notion of recurrence in discrete stochastic processes ». *Bull. Am. Math. Soc.* 53, p. 1002-1010. DOI : [10.1090/S0002-9904-1947-08927-8](https://doi.org/10.1090/S0002-9904-1947-08927-8) s'intéresse au lien entre les mesures stationnaires et les temps de retour (théorème 4.21).
- [Pro56] Y. V. PROKHOROV. « Convergence of Random Processes and Limit Theorems in Probability Theory ». *Theory of Probability & Its Applications* 1.2, p. 157-214. DOI : [10.1137/1101016](https://doi.org/10.1137/1101016) introduit l'équation de Kolmogorov et établit le théorème qui porte aujourd'hui son nom (théorème 1.38).

Enfin, finissons cette bibliographie par quelques réflexions plus historiques ou philosophiques sur la théorie.

- [Lev70] P. LEVY. *Quelques aspects de la pensée d'un mathématicien*, p. 221 est l'autobiographie mathématique de Paul Lévy : il y retrace, à la première personne, la genèse de ses principaux outils probabilistes.
- [Cho86] G. CHOQUET. « La naissance de la théorie des capacités : réflexion sur une expérience personnelle ». *C. R. Acad. Sci., Paris, Sér. Gén., Vie Sci.* 3, p. 385-397 rassemble les réflexions personnelles de Choquet sur la naissance de la théorie des capacités, dans le même esprit rétrospectif.
- [Dyn89] E. B. DYNKIN. « Kolmogorov and the theory of Markov processes ». *Ann. Probab.* 17.3, p. 822-832. DOI : [10.1214/aop/1176991248](https://doi.org/10.1214/aop/1176991248) présente les contributions principales de Kolmogorov à la théorie des processus de Markov, dans un numéro spécial consacré aux contributions probabilistes de Kolmogorov.
- [Bru03] B. BRU. « Souvenirs de Bologne ». *Journal de la Société française de statistique* 144.1-2, p. 135-226. URL : https://www.numdam.org/item/JSFS_2003__144_1-2_135_0/ est un essai historique, construit à partir d'archives et de correspondances, sur le congrès de Bologne de 1928 et la naissance de la théorie moderne des probabilités.
- [BY00] B. BRU & M. YOR. « Sur l'équation de Kolmogoroff, par W. Doeblin ». *C. R. Acad. Sci., Paris, Sér. I, Math.* 331.12 (numéro spécial), p. ii + 1033-1187. URL : <https://www.sciencedirect.com/journal/comptes-rendus-de-lacademie-des-sciences-series-i-mathematics/vol/331/issue/12/part/P2> est un numéro spécial des Comptes Rendus de l'Académie des sciences consacré au pli cacheté de Doeblin, retrouvé et publié soixante ans après son dépôt.

Liste des théorèmes importants

1.22	Théorème du portemanteau : caractérisation de la convergence étroite avec des fonctions semi-continues inférieurement ou des ensembles	20
1.31	La distance de Fortet–Mourier métrise la convergence étroite	25
1.34	L'espace des mesures de probabilité sur un espace compact métrisable est compact métrisable	28
1.38	Critère de Prokhorov pour la compacité étroite : l'équitension est équivalente à la compacité relative	30
1.39	L'espace des mesures de probabilité sur un espace polonais est polonais	32
2.1	Existence et unicité de l'espérance conditionnelle par la propriété universelle .	37
2.8	Lemme de Doob : une variable aléatoire est $\sigma(X)$ -mesurable si et seulement si elle s'écrit $h(X)$	41
2.16	Existence et unicité des lois conditionnelles	44
2.19	On peut calculer l'espérance conditionnelle à partir de la loi conditionnelle . . .	48
2.28	Calcul d'espérance conditionnelle pour un couple de variables aléatoires indépendantes	50
2.30	Calcul d'espérance conditionnelle pour les variables aléatoires à densité	51
2.30	Calcul d'espérance conditionnelle pour les variables aléatoires à densité	51
2.39	Théorème de Birkhoff : les moyennes temporelles converges presque sûrement vers les moyennes spatiales	53
2.41	Théorème de Choquet : les mesures de probabilité ergodiques sont les points extrémaux des mesures de probabilité invariantes	54
3.9	La loi d'une chaîne de Markov est déterminée par son noyau	63
3.11	Les systèmes récurrents aléatoires sont des chaînes de Markov et l'on connaît leur noyau	64
3.13	Les chaînes de Markov existent	66
3.14	Existence et définition de la chaîne de Markov canonique et des probabilités canoniques $(\mathbf{P}_x)$	67
3.22	La propriété de Markov forte : les chaînes de Markov se renouvellent après les temps d'arrêt	70
3.29	Analyse sur un pas de temps : les valeurs moyennes de sortie sont harmoniques	71
3.32	Analyse sur une pas de temps : espérance des temps de sortie	73
3.41	Les mesures stationnaires sur l'espace d'état $\mathbb{X}$ induit une mesure invariant sur l'espace canonique Ω	76
3.49	Théorème de Choquet pour les chaînes de Markov : les points extrémaux des mesures de probabilité stationnaires sont les mesures de probabilité ergodiques	78
3.52	Théorème de Birkhoff pour les chaînes de Markov : une chaîne de Markov stationnaire ergodique satisfait une loi des grands nombres	78
4.9	Théorème de Perron–Frobenius : caractérisation des valeurs propres des matrices stochastiques	83

4.10	Une chaîne de Markov irréductible et apériodique sur un espace fini converge vers son unique mesure stationnaire	84
4.16	Dichotomie des états récurrents / transients	86
4.18	Classification des états dans le cas dénombrable	87
4.20	Théorème de Pólya : la marche aléatoire simple symétrique sur $\mathbb{Z}^d$ est récurrente si et seulement si $d \in \{1, 2\}$	89
4.23	Sur une classe récurrente positive l'unique mesure ergodique se construit à partir des espérances des temps de retour	92
4.29	Théorème de Birkhoff infini : les ratios des moyennes temporelles convergent vers les ratios des moyennes spatiales	94
4.30	Sur un espace d'état dénombrable, les chaînes de Markov irréductible, récurrente positive et apériodique converge vers leur unique mesure ergodique	95
4.32	Théorème de Krylov–Bogolioubov : les noyaux fellériens équitendus admettent une mesure de probabilité stationnaire	97
4.38	Théorème de Lyapunov : tout noyau fellérien qui admet une fonction de Lyapunov admet une mesure stationnaire	98
4.40	Théorème d'unicité par contraction déterministe : les chaînes de Markov strictement contractantes en moyenne admettent au plus une mesure de probabilité stationnaire	99
4.49	Théorème d'unicité probabiliste : un noyau fellérien avec un point accessible a au plus une mesure de probabilité stationnaire	101

Table des figures

0.1	Transition d'une urne avec force de rappel	5
0.2	Les premiers (en rouge) et deuxièmes (en bleu) coups possibles pour un cavalier	5
1.1	Équitension : pour tout $\varepsilon > 0$, il existe <i>un même</i> compact K tel que $\mu_n(K) > 1 - \varepsilon$ pour tout n . Chaque mesure de la famille concentre au moins $1 - \varepsilon$ de sa masse dans K ; seule une fraction $< \varepsilon$, uniformément en n , peut s'échapper hors de K	30
1.2	Défaut de compacité : quel que soit le compact K , à partir d'un certain rang les μ_n chargent presque toute leur masse hors de K (ici $\mu_n = \delta_n$ sur $\mathbb{R}$). La famille n'est pas équitendue.	30
1.3	Équitension, lecture par les densités : toutes les lois μ_n restent sous une même enveloppe concentrée sur $K = [a, b]$. La masse hors de K (les queues) est majorée par ε uniformément en n ; c'est exactement $\sup_n \mu_n(\mathbb{R} \setminus K) \leq \varepsilon$	33
2.1	$\mathbf{E}[F \mathfrak{A}]$ est la « projection » de F sur les variables $\mathfrak{A}$ -mesurables. L'orthogonalité s'exprime par la relation $\mathbf{E}[FZ] = \mathbf{E}[\mathbf{E}[F \mathfrak{A}]Z]$ pour tout Z $\mathfrak{A}$ -mesurable bornée.	38
2.2	Une variable aléatoire F est $\sigma(X)$ -mesurable si et seulement si elle se factorise par X , c'est-à-dire $F = h \circ X$ pour une fonction mesurable h (flèche en pointillé).	42
2.3	Loi et espérance conditionnelles : X partitionne Ω en fibres $\{X = x\} = X^{-1}(x)$. La loi conditionnelle $\mathbf{P}_x$ est portée par la fibre et $\mathbf{E}[F X = x]$ est la moyenne de F contre $\mathbf{P}_x$	48

2.4	Le long d'une trajectoire $(\theta^k \omega)$, la fraction du temps passée dans une région A converge presque sûrement vers sa masse $\mathbf{P}[A]$. Les gros points gras marquent les instants où la trajectoire visite A	53
2.5	L'ensemble $\mathcal{I}$ des mesures stationnaires est convexe; ses points extrémaux (les sommets) sont exactement les mesures ergodiques $\mathcal{E}$, et tout $\pi \in \mathcal{I}$ se décompose comme barycentre de mesures ergodiques.	54
3.1	Le graphe d'incidence de la marche aléatoire simple sur $\mathbb{Z}$	65
3.2	La valeur de sortie $u(x) = \mathbf{E}_x[\varphi(X_{\tau_A})\mathbf{1}_{\tau_A < \infty}]$ est harmonique dans $\mathbb{X} \setminus A$ et vaut φ sur A	73
3.3	Le graphe de transition de la chaîne de Markov	73
4.1	Structure cyclique d'une chaîne irréductible de période d (ici $d = 4$) : si $Q(x, y) > 0$ et $x \in A_k$, alors $y \in A_{k+1 \bmod d}$	82
4.2	Chaîne de naissance et de mort sur $\mathbb{N}$: montée p_x , descente q_x , sur-place $1 - p_x - q_x$ (boucles), avec $q_0 = 0$. L'espace d'états est infini : le comportement (réurrence positive, nulle, ou transience) dépend des suites (p_x) et (q_x)	85
4.3	Exemple des différentes classes d'irréductibilité	88
4.4	Couplage de deux copies de la chaîne : avant S elles évoluent indépendamment, après S elles coïncident. On contrôle $\sum_y \mu Q^n(y) - \pi(y) \leq 2 \hat{\mathbf{P}}(S > n)$	96

Espaces et topologie

$\mathbb{X}$	espace d'états, polonais sauf mention contraire
$\mathfrak{B}(\mathbb{X})$	tribu borélienne de $\mathbb{X}$
$\mathfrak{K}(\mathbb{X}), \mathfrak{O}(\mathbb{X})$	compacts, ouverts de $\mathbb{X}$
$\mathcal{C}_b(\mathbb{X})$	fonctions continues bornées $\mathbb{X} \rightarrow \mathbb{R}$
$\mathcal{P}(\mathbb{X})$	mesures de probabilité boréliennes sur $\mathbb{X}$
d_{BL}	distance de Fortet–Mourier (définition 1.27)

Noyaux et opérateurs

q, Q	noyau markovien et l'opérateur associé (matrice stochastique)
$Q\varphi, \nu Q$	action à droite sur les fonctions, à gauche sur les mesures
G	opérateur de Green / potentiel, $G := \sum_n Q^n$ (définition 3.33)

Chaîne canonique

Ω	l'espace de probabilité canonique $\Omega := \mathbb{X}^{\mathbb{N}}$
X_n	l'application coordonnée
$\mathbf{P}_x, \mathbf{E}_x$	loi et espérance canonique partant de x (théorème 3.14)
$\mathbf{P}^\mu$	chaîne de loi initiale μ
θ_k	opérateur de translation, $\theta_k = (X_k, X_{k+1}, \dots)$
τ_A	temps d'atteinte de A (définition 3.20)
T_x, N_x	temps de retour et nombre de visites en x (définition 4.14)
$\mathfrak{R}(x)$	temps de retour possibles en x (définition 4.3)

Ergodicité et stationnarité

$\mathcal{I}$	probas invariantes pour une transformation (définition 2.34) ou stationnaires pour une chaîne (définition 3.35)
$\mathcal{E}$	probas ergodiques pour une transformation (définition 2.37) ou une chaîne (définition 3.44)
$\mathfrak{i}$	tribu invariante sur Ω (définition 2.35)
$\mathfrak{i}_\pi$	tribu π -stationnaire sur $\mathbb{X}$ (définition 3.45)

- [BGL14] D. BAKRY, I. GENTIL & M. LEDOUX. *Analysis and geometry of Markov diffusion operators*. Grundlehren Math. Wiss. 348. Springer, 2014. DOI : [10.1007/978-3-319-00227-9](https://doi.org/10.1007/978-3-319-00227-9).
- [BH91] N. BOULEAU & F. HIRSCH. *Dirichlet forms and analysis on Wiener space*. De Gruyter Stud. Math. 14. de Gruyter, 1991. DOI : [10.1515/9783110858389](https://doi.org/10.1515/9783110858389).
- [Bir31] G. D. BIRKHOFF. « Proof of the ergodic theorem. » *Proc. Natl. Acad. Sci. USA* 17 (1931), p. 656-660. DOI : [10.1073/pnas.17.12.656](https://doi.org/10.1073/pnas.17.12.656).
- [BL07] P. BARBE & M. LEDOUX. *Probabilité*. Enseign. SUP-Maths. EDP Sciences, 2007.
- [Bog07] V. I. BOGACHEV. *Measure theory. Vol. I and II*. Springer, 2007.
- [Bou69] N. BOURBAKI. *Éléments de mathématique. Intégration. Chapitre 9*. Springer, 1969. DOI : [10.1007/978-3-540-34391-2](https://doi.org/10.1007/978-3-540-34391-2).
- [Bre] J.-C. BRETON. *Processus discrets*. URL : https://jcbreton.pages.math.cnrs.fr/jcb/Lecture_notes/processus_discrets.pdf.
- [Bré94] H. BRÉZIS. *Analyse fonctionnelle. Théorie et applications*. Masson, 1994.
- [Bru03] B. BRU. « Souvenirs de Bologne ». *Journal de la Société française de statistique* 144.1-2 (2003), p. 135-226. URL : https://www.numdam.org/item/JSFS_2003__144_1-2_135_0/.
- [BY00] B. BRU & M. YOR. « Sur l'équation de Kolmogoroff, par W. Doeblin ». *C. R. Acad. Sci., Paris, Sér. I, Math.* 331.12 (numéro spécial 2000), p. ii + 1033-1187. URL : <https://www.sciencedirect.com/journal/comptes-rendus-de-lacademie-des-sciences-series-i-mathematics/vol/331/issue/12/part/P2>.
- [Cho86] G. CHOQUET. « La naissance de la théorie des capacités : réflexion sur une expérience personnelle ». *C. R. Acad. Sci., Paris, Sér. Gén., Vie Sci.* 3 (1986), p. 385-397.
- [CM16] D. CHAFAÏ & F. MALRIEU. *Recueil de modèles aléatoires*. Math. Appl. (Berl.) 78. Springer, 2016. DOI : [10.1007/978-3-662-49768-5](https://doi.org/10.1007/978-3-662-49768-5).
- [Cou12] Y. COUDÈNE. *Théorie ergodique et systèmes dynamiques*. Savoirs Actuels. EDP Sciences; CNRS Éditions, 2012.
- [DM75] C. DELLACHERIE & P.-A. MEYER. *Probabilités et potentiel. Chap. I à IV*. Publ. Inst. Math. Univ. Strasbourg XV. Hermann, 1975.
- [DM83] C. DELLACHERIE & P.-A. MEYER. *Probabilités et potentiel. Chapitres IX à XI : Théorie discrète du potentiel. Edition entièrement refondue*. Publ. Inst. Math. Univ. Strasbourg XVIII. Hermann, Paris, 1983.
- [DM87] C. DELLACHERIE & P.-A. MEYER. *Probabilités et potentiel. Chapitres XII à XVI : Théorie du potentiel associée à une résolvante. Théorie des processus de Markov*. Publ. Inst. Math. Univ. Strasbourg XIX. Hermann, Paris, 1987.

- [Doe40] W. DOEBLIN. « Éléments d'une théorie générale des chaînes simples constantes de Markoff ». *Ann. Sci. Éc. Norm. Supér.* (3) 57 (1940), p. 61-111. DOI : [10.24033/asens.883](https://doi.org/10.24033/asens.883).
- [Dud02] R. M. DUDLEY. *Real analysis and probability*. Camb. Stud. Adv. Math. 74. Cambridge University Press, 2002. DOI : [10.1017/CB09780511755347](https://doi.org/10.1017/CB09780511755347).
- [Dur19] R. DURRETT. *Probability. Theory and examples*. Camb. Ser. Stat. Probab. Math. 49. Cambridge University Press, 2019. DOI : [10.1017/9781108591034](https://doi.org/10.1017/9781108591034).
- [Dyn89] E. B. DYNKIN. « Kolmogorov and the theory of Markov processes ». *Ann. Probab.* 17.3 (1989), p. 822-832. DOI : [10.1214/aop/1176991248](https://doi.org/10.1214/aop/1176991248).
- [Féj] J. FÉJOZ. *Chapitres d'intégration et de probabilités*. URL : <https://www.ceremade.dauphine.fr/~fejoz/Integration/integration-probabilites.pdf>.
- [Fro12] G. FROBENIUS. « Über Matrizen aus nicht negativen Elementen. » *Berl. Ber.* 1912 (1912), p. 456-477.
- [Gal] J.-F. le GALL. *Intégration, probabilités et processus aléatoires*. URL : <https://www.imo.universite-paris-saclay.fr/~jean-francois.le-gall/IPPA2.pdf>.
- [Gal22] J.-F. le GALL. *Measure theory, probability, and stochastic processes*. Grad. Texts Math. 295. Springer, 2022. DOI : [10.1007/978-3-031-14205-5](https://doi.org/10.1007/978-3-031-14205-5).
- [GG25] H. GEISS & S. GEISS. *Measure, probability and functional analysis*. Universitext. Springer, 2025. DOI : [10.1007/978-3-031-84067-8](https://doi.org/10.1007/978-3-031-84067-8).
- [Hai] M. HAIRER. *Ergodic Properties of Markov Processes*. URL : <https://hairer.org/notes/Markov.pdf>.
- [Kac47] M. KAC. « On the notion of recurrence in discrete stochastic processes ». *Bull. Am. Math. Soc.* 53 (1947), p. 1002-1010. DOI : [10.1090/S0002-9904-1947-08927-8](https://doi.org/10.1090/S0002-9904-1947-08927-8).
- [KB37] N. KRYLOV & N. BOGOLIOUBOV. « La théorie générale de la mesure dans son application à l'étude des systèmes dynamiques de la mécanique non linéaire ». *Ann. Math.* (2) 38 (1937), p. 65-113. DOI : [10.2307/1968511](https://doi.org/10.2307/1968511).
- [Kol31] A. KOLMOGOROFF. « Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung ». *Math. Ann.* 104 (1931), p. 415-458. DOI : [10.1007/BF01457949](https://doi.org/10.1007/BF01457949).
- [Kol33] A. KOLMOGOROV. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Ergeb. Math. Grenzgeb. 2. Springer-Verlag, 1933.
- [Kol92] A. N. KOLMOGOROV. *Selected works II. Probability theory and mathematical statistics*. Sous la dir. d'A. N. SHIRYAEV. Springer Collect. Works Math. Dordrecht : Springer, 1992.
- [Leb02] H. LEBESGUE. « Intégrale, longueur, aire. » *Annali di Mat.* (3) 7 (1902), p. 231-359. DOI : [10.1007/BF02420592](https://doi.org/10.1007/BF02420592).
- [Lév37] P. LÉVY. *Théorie de l'addition des variables aléatoires*. Monographies des Probabilités ; calcul des probabilités et ses applications 1. Gauthier-Villars, 1937, p. 328+xvii.
- [Lev70] P. LEVY. *Quelques aspects de la pensée d'un mathématicien*. Albert Blanchard, 1970, p. 221.
- [LL10] G. F. LAWLER & V. LIMIC. *Random walk : A modern introduction*. Camb. Stud. Adv. Math. 123. Cambridge University Press, 2010. DOI : [10.1017/CB09780511750854](https://doi.org/10.1017/CB09780511750854).
- [Mar06] A. A. MARKOV. « Rasprostranenie zakona bol'shikh chisel na velichiny, zavisyashchie drug ot druga ». *Kasan Ges.* (2) 15.4 (1906), p. 135-156.

- [MT09] S. MEYN & R. L. TWEEDIE. *Markov chains and stochastic stability. Prologue by Peter W. Glynn*. Camb. Math. Libr. Cambridge University Press, 2009. DOI : [10.1017/CB09780511626630](https://doi.org/10.1017/CB09780511626630).
- [Neu32] J. von NEUMANN. « Proof of the quasi-ergodic hypothesis ». *Proc. Natl. Acad. Sci. USA* 18 (1932), p. 70-82. DOI : [10.1073/pnas.18.1.70](https://doi.org/10.1073/pnas.18.1.70).
- [Nev70] J. NEVEU. *Bases mathématiques du calcul des probabilités*. Masson et Cie, 1970, p. 213.
- [Nor97] J. R. NORRIS. *Markov chains*. Camb. Ser. Stat. Probab. Math. 2. Cambridge University Press, 1997. DOI : [10.1017/CB09780511810633](https://doi.org/10.1017/CB09780511810633).
- [Per07] O. PERRON. « Zur Theorie der Matrices. » *Math. Ann.* 64 (1907), p. 248-263. DOI : [10.1007/BF01449896](https://doi.org/10.1007/BF01449896).
- [Poi90] H. POINCARÉ. « Sur le problème des trois corps et les équations de la dynamique ». *Acta Math.* 13 (1890), p. 1-270.
- [Pól21] G. PÓLYA. « Über eine Aufgabe der Wahrscheinlichkeitsrechnung betreffend die Irrfahrt im Straßennetz. » *Math. Ann.* 84 (1921), p. 149-160. DOI : [10.1007/BF01458701](https://doi.org/10.1007/BF01458701).
- [Pro56] Y. V. PROKHOROV. « Convergence of Random Processes and Limit Theorems in Probability Theory ». *Theory of Probability & Its Applications* 1.2 (1956), p. 157-214. DOI : [10.1137/1101016](https://doi.org/10.1137/1101016).
- [Rud87] W. RUDIN. *Real and complex analysis*. McGraw-Hill, 1987.
- [RY05] D. REVUZ & M. YOR. *Continuous martingales and Brownian motion*. Grundlehren Math. Wiss. 293. Springer, 2005.
- [Sal] J. SALEZ. *Mixing times of Markov chains*. URL : <https://www.ceremade.dauphine.fr/~salez/mix.pdf>.
- [Sch65] L. SCHWARTZ. *Théorie des distributions*. Hermann, 1965.
- [Str03] D. W. STROOCK. *Markov processes from K. Itô's perspective*. T. 155. Ann. Math. Stud. Princeton University Press, 2003. DOI : [10.1515/9781400835577](https://doi.org/10.1515/9781400835577).
- [Tao11] T. TAO. *An introduction to measure theory*. Grad. Stud. Math. 126. American Mathematical Society (AMS), 2011.
- [Woe08] W. WOESS. *Random walks on infinite graphs and groups*. Camb. Tracts Math. 138. Cambridge University Press, 2008. DOI : [10.1017/CB09780511470967](https://doi.org/10.1017/CB09780511470967).