

Processus stochastiques

Ébauche de notes de cours de M1

Département de mathématiques et applications (DMA)
École normale supérieure – PSL, 45 rue d'Ulm, Paris 5^e



2024 – 2025

Version corrigée du 11 mai 2025

$$\mathbb{E}(M_{n+1} | \mathcal{F}_n) = M_n$$

$$\mathbb{E}(M_0) = \mathbb{E}(M_{n \wedge T})$$

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

$$\mathbb{E}(f(N_{T+t} - N_T) | \mathcal{F}_T) = \mathbb{E}(f(N_t - N_0))$$

$$\sup_{s \in [0, t]} B_s \stackrel{d}{=} |B_t|$$

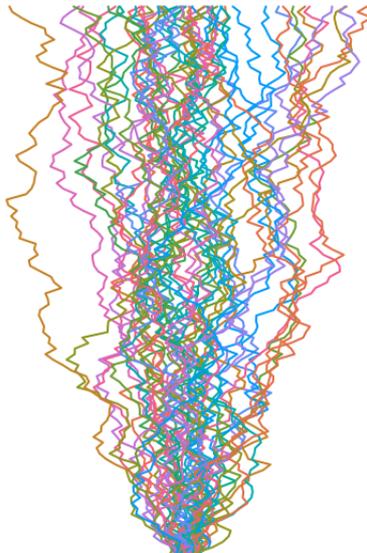


Table des matières

Table des matières	2
1 Rappels et compléments d'intégration et probabilité	6
1.1 Intégrale de Lebesgue	6
1.2 Variables aléatoires et lois	6
1.3 Inégalités de Markov, Cauchy–Schwarz, Hölder, Jensen	7
1.4 Convergences	7
1.5 Fonction caractéristique	9
1.6 Lemme de Borel–Cantelli	10
1.7 Loi des grands nombres, théorème limite central, loi du logarithme itéré	10
1.8 Autour de la convergence en loi : lemme de Slutsky et méthode delta	10
1.9 Intégrabilité uniforme ou équi-intégrabilité	11
2 Espérance conditionnelle	16
2.1 Introduction	16
2.2 Construction lorsque la tribu est discrète	17
2.3 Construction pour les variables de carré intégrable	17
2.4 Construction pour les variables positives	18
2.5 Construction pour les variables intégrables	18
2.6 Propriétés de l'espérance conditionnelle	18
2.7 Espérance conditionnelle par rapport à une variable aléatoire	20
2.8 Noyau conditionnel et loi conditionnelle	21
2.9 Motivation pour la suite : processus de Galton–Watson, processus auto-regressif	23
3 Vecteurs aléatoires gaussiens	24
3.1 Définition et simulation	25
3.2 Indépendance des composantes et densité	25
3.3 Cas isotrope : théorème de Cochran, théorème de Maxwell, équation de la chaleur	26
3.4 Espérance conditionnelle et loi conditionnelle	28
4 Martingales à temps discret	31
4.1 Processus, filtrations, martingales	31
4.2 Exemples de martingales	33
4.3 Martingales comme modélisation de stratégies	35
4.4 Temps d'arrêt et théorème d'arrêt de Doob	36
4.5 Arrêt optimal et enveloppe de Snell	39
4.6 Tribu d'arrêt et théorème d'arrêt généralisé	42
4.7 Décomposition de Doob et inégalité maximale de Doob	42
4.8 Convergence par la voie de carré intégrable (vallée de Doob)	44
4.9 Convergence par la voie des montées et descentes (arête de Doob)	48
4.10 Martingales fermées ou uniformément intégrables	49
4.11 Martingales rétrogrades ou renversées et application à la loi des grands nombres	50
4.12 Théorème limite central pour les martingales de carré intégrable	52
5 Chaînes de Markov à temps et espace discrets	54
5.1 Suites récurrentes aléatoires et noyaux de transition	55
5.2 Exemples de chaînes de Markov	57
5.3 Propriété de Markov forte	58
5.4 Classification des états, récurrence et transience	60
5.5 Simulation des trajectoires : suites récurrentes aléatoires, générateur, puissances	62
5.6 Atteinte, absorption, harmonicité, martingales	63
5.7 Chaînes irréductibles, mesures invariantes, récurrence positive, théorème ergodique	66
5.8 Critère de Foster et fonction de Liapounov	69
5.9 Loi des grands nombres ou théorème ergodique	70
5.10 Convergence en loi, période, couplage	72

5.11	Estimation du noyau et de la mesure invariante	75
5.12	Réversibilité et algorithme de Metropolis–Hastings	75
5.13	Cas des espaces d'état finis	78
5.14	Quelques exemples classiques	82
5.14.1	Processus de vie ou de mort sur \mathbb{N}	82
5.14.2	Marche aléatoire simple sur \mathbb{Z}^d	83
5.14.3	Urne d'Ehrenfest	84
5.14.4	Modèle de Wright–Fisher	85
5.14.5	Processus de Galton–Watson	86
6	Processus de Poisson et chaînes de Markov à temps continu et espace discret	89
6.1	Processus de Bernoulli et lois géométriques	89
6.2	Lois géométriques et lois exponentielles	90
6.3	Lois exponentielles et processus de Poisson	91
6.4	Propriété de Markov du processus de Poisson	91
6.5	Structure des temps de saut et des accroissements du processus de Poisson	93
6.6	Processus de Poisson comme limite de processus de Bernoulli	95
6.7	Paradoxe de l'autobus	96
6.8	Poissonisation d'une chaîne de Markov à temps discret	97
6.9	Chaînes de Markov à temps continu : construction	98
6.10	Chaînes de Markov à temps continu : non-explosion	98
6.11	Chaînes de Markov à temps continu : générateur	99
6.12	Chaînes de Markov à temps continu : semi-groupe et équations de Kolmogorov	102
6.13	Chaînes de Markov à temps continu : synthèse	102
6.14	Files d'attente et quelques autres exemples	103
7	Mouvement brownien	105
7.1	Caractérisations, martingales, invariance d'échelle	109
7.2	Variation des trajectoires et variation quadratique	110
7.3	Loi du zéro-un de Blumenthal et ses conséquences, loi de Bachelier–Lévy	111
7.4	Loi forte des grands nombres, invariance par inversion du temps	113
7.5	Propriété de Markov, principe de réflexion, temps d'atteinte	113
7.6	Construction avec le théorème de continuité de Kolmogorov, loi du logarithme itéré	116
7.7	Mesure de Wiener et mouvement brownien canonique	120
7.8	Intégrale de Wiener	120
7.9	Vers les intégrales de Itô et de Stratonovich	122
A	Variation totale sur espace au plus dénombrable	124
A.1	Formulations alternatives	124
A.2	Convergence étroite	125
B	Modèles emblématiques issus de l'informatique, de la physique, et de la biologie	126
C	Lexique bilingue	127
D	Chronologie	128
	Bibliographie	129

Équipe pédagogique associée à ce cours :

— 2024 – 2025 : Djalil Chafaï (cours), Alexis Metz–Donnadieu (travaux dirigés)

Ces notes ont bénéficié de relectures par Laëtitia Comminges et Arnaud Guyader

Ces notes sont une introduction à l'étude mathématique de processus stochastiques, c'est-à-dire à l'évolution aléatoire au cours du temps d'un système au travers d'une succession d'étapes dans l'espace. La nature au sens large est une grande pourvoyeuse de processus stochastiques, pour ceux qui savent les voir. Ces notes introductives de niveau master se focalisent sur les processus à temps discret, avec quelques incursions dans l'univers des processus à temps continu, en évitant les complexions liées à la dimension infinie. Le modèle le plus emblématique est peut-être celui des marches aléatoires, lié à une façon de faire des pas au hasard dans un espace donné. Il est possible de marcher sur un groupe, un graphe, etc. Les cas additifs les plus élémentaires sont connectés aux théorèmes limites standards sur les sommes de variables aléatoires indépendantes. Le cœur de ces notes est consacré à deux structures d'évolution importantes : les martingales et les chaînes de Markov, à temps discret, qui vont au-delà des sommes de variables indépendantes. Elles permettent de répondre aux questions naturelles du comportement asymptotique en temps long, et du comportement trajectorien comme les temps d'atteinte par exemple. Les processus stochastiques sont liés à des équations d'évolutions déterministes de lois, et donc aux systèmes dynamiques. Grâce au phénomène limite central, la limite d'échelle ou vision de loin d'une marche aléatoire est un mouvement brownien, objet incontournable des probabilités abordé dans un dernier chapitre. Ces notes de cours abordent également au passage les notions d'uniforme intégrabilité, d'espérance conditionnelle, de vecteurs gaussiens, de processus de Poisson, ainsi qu'une variété de modèles illustratifs issus de préoccupations concrètes.

Les noms propres qui figurent dans ces notes sont avant tout des conventions sociales de nommage, souvent prises en l'honneur de contributeurs majeurs au domaine. Cependant, de nombreux contributeurs à la création, mise en forme et diffusion des concepts ne sont pas apparents, comme par exemple Richard Edler von Mises.

Le prérequis incontournable est le cours d'intégration et probabilité de première année premier semestre.

Ces notes de cours sont librement inspirées d'une variété de références, données en bibliographie, ainsi que de considérations personnelles développées et accumulées au cours du temps, au fil du hasard et de la nécessité. . . Le nombre de pages est volontairement limité. Les dessins qu'elles ne contiennent pas (encore?) sont faits au tableau en cours. Le cours oral comprend également des commentaires improvisés qui ne figurent pas dans ces notes.

« *They say that Understanding ought to work by the rules of right reason. These rules are, or ought to be, contained in Logic; but the actual science of Logic is conversant at present only with things either certain, impossible, or entirely doubtful, none of which (fortunately) we have to reason on. Therefore the true Logic for this world is the Calculus of Probabilities, which takes account of the magnitude of the probability (which is, or which ought to be in a reasonable man's mind). This branch of Math., which is generally thought to favour gambling, dicing, and wagering, and therefore highly immoral, is the only "Mathematics for Practical Men," as we ought to be.* »

James Clerk Maxwell (1831 — 1879), Letter to Lewis Campbell (1850)

Extrait de *The Life of James Clerk Maxwell*, par L. Campbell et W. Garrett, Macmillan (1881) p. 80

« *If you ask me about my innermost conviction whether our century will be called the century of iron or the century of steam or electricity, I answer without hesitation : It will be called the century of the mechanical view of Nature, the century of Darwin.* »

Ludwig Boltzmann (1844 – 1906), Vienne (1886) Der zweite Hauptsatz der mechanischen Wärmetheorie

Tiré de *Lecture at the Festive Session*, Imperial Academy of Sciences

« *Uhlenbeck's attitude to Wiener's work was brutally pragmatic and it is summarized at the end of footnote 9 in his paper (written jointly with Ming Chen Wang) "On the Theory of Brownian Motion II" (1945) : the authors are aware of the fact that in the mathematical literature, especially in papers by N. Wiener, J. L. Doob, and others [cf. for instance Doob (Annals of Mathematics 43, 351 1942) also for further references], the notion of a random (or stochastic) process has been defined in a much more refined way. This allows [us], for instance, to determine in certain cases the probability that the random function $y(t)$ is of bounded variation or continuous or differentiable, etc. However it seems to us that these investigations have not helped in the solution of problems of direct physical interest and we will therefore not try to give an account of them.* »

Mark Kac (1914 – 1984) à propos de George Uhlenbeck (1900 – 1988)

Tiré de *Enigmas of Chance : an autobiography* (1984)

Organisation de l'année 2024-2025

Créneaux horaires

- 3h=2x1.5h de CM et 2h de TD par semaine
- Les CMs ont lieu le mercredi 10h45–12h15 (Borel, au 29) et le jeudi 10h45–12h15 (Cartan, au 45)
- Deux créneaux de TDs : le vendredi 15h15–17h15 (Cartan)

Ressources pédagogiques

Les notes de cours, feuilles de TD, DM, et annales se trouvent sur <https://moodle.psl.eu/>

Calendrier prévisionnel des cours

- **Semaine 1**
 - Me 11/09 : Rappels et compléments d'intégration et probabilités
 - Je 12/09 : Rappels et compléments d'intégration et probabilités
- **Semaine 2**
 - Me 18/09 : Conditionnement
 - Je 19/09 : Conditionnement
- **Semaine 3**
 - Me 25/09 : Conditionnement, Vecteurs aléatoires gaussiens
 - Je 26/09 : Vecteurs aléatoires gaussiens
- **Semaine 4**
 - Me 02/10 : Martingales
 - Je 03/10 : Martingales
- **Semaine 5**
 - Me 09/10 : Martingales
 - Je 10/10 : Martingales
- **Semaine 6**
 - Me 16/10 : Martingales
 - Je 17/10 : Martingales
- **Semaine 7**
 - Me 23/10 : Martingales, Chaînes de Markov
 - Je 24/10 : Chaînes de Markov
- **Vacances (semaine 28/10 – 01/11)**
- **Partiel (semaine 04/10 – 08/11) Durée : 2h, exercices**
- **Semaine 8**
 - Me 13/11 : Chaînes de Markov
 - Je 14/11 : Chaînes de Markov
- **Semaine 9**
 - Me 20/11 : Chaînes de Markov
 - Je 21/11 : Chaînes de Markov
- **PSL-Week (semaine 25/11 – 29/11)**
- **Semaine 10**
 - Me 04/12 : Chaînes de Markov
 - Je 05/12 : Chaînes de Markov
- **Semaine 11**
 - Me 11/12 : Processus de Poisson et chaînes de Markov à temps continu
 - Je 12/12 : Processus de Poisson et chaînes de Markov à temps continu
- **Semaine 12**
 - Me 18/12 : Processus de Poisson et chaînes de Markov à temps continu
 - Je 19/12 : Processus de Poisson et chaînes de Markov à temps continu
- **Vacances (semaines du 25/12 et du 01/01)**
- **Semaine 13**
 - Me 08/01 : Mouvement brownien
 - Je 09/01 : Mouvement brownien
- **Semaine 14**
 - Me 15/01 : Mouvement brownien
 - Je 16/01 : Mouvement brownien
- **Examen (semaine 23/01 – 27/01) Durée : 3h, exercices, problème**

Chapitre 1

Rappels et compléments d'intégration et probabilité

Prérequis : cours d'intégration et probabilités de L3, à base d'intégrale de Lebesgue.

Dans toutes ces notes, les espaces L^p sont, sauf mention explicite du contraire, relatifs à un espace de probabilité, par défaut noté $(\Omega, \mathcal{A}, \mathbb{P})$ si nécessaire. En particulier, ils contiennent les constantes, et sont emboîtés.

1.1 Intégrale de Lebesgue

L'intégrale de Lebesgue a du succès car elle possède des théorèmes de convergence simples et efficaces : théorème de convergence monotone, lemme de Fatou, et théorème de convergence dominée. Elle est de plus bien adaptée aux probabilités car elle permet de mathématiser efficacement l'information avec la notion d'événement, de tribu, et d'indépendance (théorie de la mesure de Borel et axiomatique de Kolmogorov), mais aussi d'englober dans le même formalisme les mesures discrètes $\mu(A) = \sum_{i: a_i \in A} m_i$ et les mesures à densité $\mu(A) = \int_A f(x) dx$, et leur combinaison, essentielle à la modélisation probabiliste. Le théorème de décomposition de Lebesgue affirme que si μ est une mesure sur \mathbb{R} muni de sa tribu borélienne, alors

$$\mu = \mu_{\text{dens}} + \mu_{\text{sing}} + \mu_{\text{disc}},$$

où μ_{dens} est absolument continue par rapport à la mesure de Lebesgue (à densité), μ_{sing} est sans atomes et singulière par rapport à la mesure de Lebesgue, et μ_{disc} est discrète (purement atomique).

Sur un groupe topologique localement compact, muni de sa tribu borélienne, il existe à dilatation près une unique mesure invariante par translation, appelée mesure de Haar. Sur \mathbb{R}^d il s'agit de la mesure de Lebesgue.

1.2 Variables aléatoires et lois

En général, les variables aléatoires interviennent à travers leur loi. Cela conduit aux problèmes suivants :

1. Partant d'une mesure de probabilité μ sur un espace mesurable (E, \mathcal{E}) , construire un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$, et une variable aléatoire $X : \Omega \rightarrow E$ de loi μ , les plus simples ou naturels possibles. La construction canonique consiste à prendre $(\Omega, \mathcal{A}, \mathbb{P}) = (E, \mathcal{E}, \mu)$ et $X = \text{Id}$.
2. Même problème mais avec deux mesures de probabilités μ et ν sur (E, \mathcal{E}) et (F, \mathcal{F}) , et la construction d'un couple (X, Y) de variables aléatoires sur un même espace, de lois marginales μ et ν . On parle de couplage. Lorsque dans une rédaction, on introduit une variable aléatoire, puis une autre, indépendante, l'espace de probabilité souvent implicite de définition de la première variable aléatoire ne suffit plus forcément, et doit être élargi pour inclure la seconde, un couplage a ainsi lieu, parfois à l'insu du lecteur, voire de l'auteur! La construction canonique consiste à considérer l'espace produit $(\Omega, \mathcal{A}, \mathbb{P}) := (E \times F, \mathcal{E} \otimes \mathcal{F}, \mu \otimes \nu)$ ce qui réalise X et Y comme les deux projections canoniques, à marges indépendantes.
3. Même problème mais avec la contrainte $X = f(U)$ où U est une variable aléatoire de loi uniforme sur $[0, 1]$. Il s'agit de préciser à la fois f et l'espace sur lequel U est définie, et de minimiser la complexité algorithmique d'évaluation de f . On parle de problème de la simulation. De manière équivalente, on peut remplacer U par une suite de v.a.i.i.d. de loi de Bernoulli symétriques sur $\{0, 1\}$ (bits).

L'indépendance des variables est équivalente à la structure produit de leur loi jointe.

Fondamental : si $X : \Omega \rightarrow E$ est une variable aléatoire définie sur l'espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ et à valeurs dans l'espace mesurable (E, \mathcal{E}) alors pour toute fonction $h : E \rightarrow \mathbb{R}$ mesurable, positive ou bornée,

$$\mathbb{E}(h(X)) = \int_{\Omega} h(X(\omega)) \mathbb{P}(d\omega) = \int_E h(x) \mu(dx)$$

où μ est la loi de X , mesure image de \mathbb{P} par X . C'est une mesure de probabilité sur (E, \mathcal{E}) . La formule consiste à réécrire la somme en regroupant par paquets correspondants aux valeurs prises par X . On parle de théorème ou lemme du transfert ou du transport. La formule reste valable pour h tel que $h(X)$ est intégrable.

Dans toute la suite, les variables aléatoires (v.a.) sont par défaut réelles : $E = \mathbb{R}$, \mathcal{E} est la tribu borélienne.

1.3 Inégalités de Markov, Cauchy–Schwarz, Hölder, Jensen

Inégalité de Markov. Si X est une v.a., $U(X) \geq 0$, U croissante, alors pour tout $r > 0$ tel que $U(r) > 0$,

$$\mathbb{1}_{\{X \geq r\}} \leq \frac{U(X)}{U(r)} \quad \text{d'où} \quad \mathbb{P}(X \geq r) \leq \frac{\mathbb{E}(U(X))}{U(r)}.$$

Cela permet de contrôler la queue de distribution avec les moments, tandis que réciproquement on a

$$\mathbb{E}(U(|X|)) = U(0) + \int_0^\infty U'(t) \mathbb{P}(|X| \geq t) dt.$$

En particulier, si $X \geq 0$ alors $\mathbb{E}(X) = \int_0^\infty \mathbb{P}(X > t) dt$, et si $\mathbb{P}(X \in \mathbb{N}) = 1$ alors $\mathbb{E}(X) = \sum_{n=0}^\infty \mathbb{P}(X > n)$.

Inégalité de Cauchy–Schwarz. Si $X, Y \in L^2$ alors

$$\mathbb{E}(XY) \leq \mathbb{E}(|X|^2)^{1/2} \mathbb{E}(|Y|^2)^{1/2}, \quad \text{dans } [0, \infty].$$

Égalité ssi X et Y sont colinéaires. Il s'agit d'une propriété typique des espaces de Hilbert, ici L^2 .

Inégalité de Hölder. Si $r, p, q \in [1, +\infty]$ et $1/r = 1/p + 1/q$ alors (donne CS pour $r = 1$ et $p = 2$)

$$\mathbb{E}(|XY|^r)^{1/r} \leq \mathbb{E}(|X|^p)^{1/p} \mathbb{E}(|Y|^q)^{1/q}, \quad \text{dans } [0, \infty].$$

Comme \mathbb{P} mesure finie, les constantes sont dans tout L^p , et donc $L^{p_2} \subset L^{p_1}$ si $p_1 \leq p_2$. En particulier $L^\infty \subset L^2 \subset L^1$.

Inégalité de Jensen. Si $U : \mathbb{R}^d \rightarrow \mathbb{R}$ est convexe et $X \in L^1$ avec $U(X) \in L^1$ alors

$$U(\mathbb{E}(X)) \leq \mathbb{E}(U(X)).$$

Égalité, quand U est strictement convexe, ssi X est déterministe. Ex. : $U(x) = x^p$, $p > 1$, $U(x) = e^{cx}$, $c \in \mathbb{R}$.

1.4 Convergences

Soit $(X_n)_{n \geq 1}$ des v.a. définies sur un même espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$, de lois μ_n, μ respectivement.

Convergence presque sûre. On dit que $X_n \xrightarrow{\text{p.s.}} X$ lorsque

$$\mathbb{P}(\lim_{n \rightarrow \infty} X_n = X) = 1$$

c'est-à-dire¹ que $\mathbb{P}(\{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\}) = 1$. Utilisé dans la loi forte des grands nombres.

Convergence en probabilité. On dit que $X_n \xrightarrow{\mathbb{P}} X$ lorsque

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| \geq \varepsilon) = 0,$$

c.-à-d. que $\lim_{n \rightarrow \infty} \mathbb{P}(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| \geq \varepsilon\}) = 0$. Utilisé dans la loi faible des grands nombres.

Convergence en moyenne. Pour tout $p \in [1, \infty)$, on dit que $X_n \xrightarrow{L^p} X$ lorsque

$$X \in L^p \quad \text{et} \quad \lim_{n \rightarrow \infty} \mathbb{E}(|X_n - X|^p) = 0,$$

c'est-à-dire que $X_n \rightarrow X$ dans L^p . En pratique, dans la plupart des cas, $p \in \{1, 2, 4\}$.

Convergence en loi. On dit que $X_n \xrightarrow{\text{loi}} X$, ou $X_n \xrightarrow{d} \mu$ (convergence en distribution), ou $\mu_n \xrightarrow{\text{étr.}} \mu$ (convergence étroite) lorsque pour tout $f : \mathbb{R} \rightarrow \mathbb{R}$ continue et bornée,

$$\lim_{n \rightarrow \infty} \mathbb{E}(f(X_n)) = \mathbb{E}(f(X)).$$

Utilisé dans le théorème limite central. Il est possible de remplacer les fonctions test continues et bornées par diverses classes de fonctions tests, paramétriques ou pas, et on parle alors parfois de théorème de Portemanteau². Dans le cas paramétrique, cela correspond alors à une convergence ponctuelle de transformée de la mesure.

- fonctions test \mathcal{C}^∞ et à support compact.
- fonctions trigonométriques e^{it} , $t \in \mathbb{R}$ (fonction caractéristique ou transformée de Fourier).
- fonctions indicatrices $\mathbb{1}_{(-\infty, x]}$, x point de continuité de F_X (fonction de répartition).
- fonctions exponentielles e^{-t} , $t \in \mathbb{R}_+$ (transformée de Laplace, v.a. positives).
- fonctions géométriques s^r , $s \in [0, 1]$ (fonction génératrice, v.a. discrètes sur \mathbb{N}).

1. L'événement considéré est bien mesurable car l'espace topologique \mathbb{R} est à base dénombrable de voisinages.
2. Il s'agit d'un canular de Billingsley dans [7], car ce théorème est en réalité dû à Alexandrov semble-t-il.

En revanche, les classes suivantes ne conviennent pas pour la convergence en loi :

- fonctions mesurables et bornées (classe trop grande).
- polynômes (sauf cas spécial comme support compact par exemple).

La convergence en loi ne nécessite pas que les variables possèdent des moments. En revanche, la topologie de la convergence en loi renforcée par la convergence des moments jusqu'à l'ordre p correspond exactement à la convergence pour la distance de couplage de Kantorovich – Monge – Wasserstein, liée au transport optimal.

Attention, la convergence en moyenne d'ordre p entraîne la convergence des moments jusqu'à l'ordre p . En revanche, ce n'est pas le cas des convergence presque sûre, en probabilité, et en loi en général. En particulier, la convergence de tous les moments est un mode de convergence spécial à distinguer des autres.

Au contraire des autres types de convergence, la convergence en loi ne dépend pas de la loi du couple (X_n, X) et n'utilise que les lois marginales. Le couplage n'est donc pas vu par ce type de convergence.

Théorème de convergence monotone. Si $(X_n)_{n \geq 1}$ prend ses valeurs dans $[0, \infty]$ et \nearrow alors

$$\mathbb{E}(\lim_{n \rightarrow \infty} X_n) = \lim_{n \rightarrow \infty} \mathbb{E}(X_n) \in [0, \infty].$$

Lemme de Fatou. Si $(X_n)_{n \geq 1}$ prend ses valeurs dans $[0, \infty]$ alors

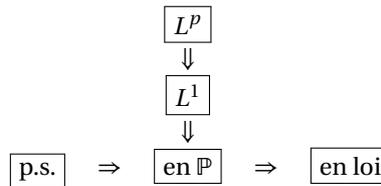
$$\mathbb{E}(\liminf_{n \rightarrow \infty} X_n) \leq \liminf_{n \rightarrow \infty} \mathbb{E}(X_n) \in [0, \infty].$$

En particulier si $0 \leq X_n \rightarrow X$ p.s. et (X_n) bornée dans L^1 alors $X \in L^1$.

Théorème de convergence dominée. Si $X_n \xrightarrow{\text{p.s.}} X$ et $\sup_n |X_n| \leq Y$, $\mathbb{E}(Y) < \infty$, alors

$$X \in L^1 \quad \text{et} \quad X_n \xrightarrow{L^1} X, \quad \text{et en particulier} \quad \lim_{n \rightarrow \infty} \mathbb{E}(X_n) = \mathbb{E}(\lim_{n \rightarrow \infty} X_n) = \mathbb{E}(X).$$

Théorème 1.4.1. Relations entre convergences.



Démonstration. L'inégalité de Hölder permet d'établir que la convergence en moyenne d'ordre $q \geq 1$ implique la convergence en moyenne d'ordre $1 \leq p \leq q$ car en notant $1/r = 1/p - 1/q$:

$$\mathbb{E}(|X_n - X|^p)^{1/p} \leq \mathbb{E}(|X_n - X|^q)^{1/q} \mathbb{E}(1^r)^{1/r} = \mathbb{E}(|X_n - X|^q)^{1/q}.$$

L'inégalité de Markov fait passer de la convergence en moyenne à la convergence en probabilités :

$$\mathbb{P}(|X_n - X| \geq \varepsilon) \leq \frac{\mathbb{E}(|X_n - X|^p)}{\varepsilon^p}.$$

On passe de la convergence presque sûre à la convergence en probabilité en traduisant tout d'abord les quantificateurs \forall et \exists par \cap et \cup comme suit : $\{\lim_{n \rightarrow \infty} X_n = X\} = \cap_{k=1}^{\infty} \cup_{m=1}^{\infty} \cap_{n \geq m} \{|X_n - X| \leq 1/k\}$, puis en observant que par monotonie, pour tout k fixé :

$$1 = \mathbb{P}(\lim_{n \rightarrow \infty} X_n = X) \leq \mathbb{P}(\cup_{m=1}^{\infty} \cap_{n \geq m} \{|X_n - X| \leq 1/k\}) = \lim_{m \rightarrow \infty} \mathbb{P}(\cap_{n \geq m} \{|X_n - X| \leq 1/k\}) \leq \lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| \leq 1/k).$$

Si $X_n \xrightarrow{\text{p.s.}} X$ alors $X_n \xrightarrow{\text{loi}} X$ car pour toute f continue et périodique ou à support compact, f est uniformément continue par le théorème de Heine, et donc pour tout $\eta > 0$, il existe $\varepsilon > 0$ tel que $|f(x) - f(y)| \leq \eta$ si $|x - y| \leq \varepsilon$, d'où

$$\begin{aligned}
 |\mathbb{E}(f(X_n)) - \mathbb{E}(f(X))| &\leq \mathbb{E}(|f(X_n) - f(X)|) \\
 &= \mathbb{E}(|f(X_n) - f(X)| \mathbb{1}_{|X_n - X| < \varepsilon}) + \mathbb{E}(|f(X_n) - f(X)| \mathbb{1}_{|X_n - X| \geq \varepsilon}) \\
 &\leq \eta + 2 \|f\|_{\infty} \mathbb{P}(|X_n - X| \geq \varepsilon).
 \end{aligned}$$

On peut donc choisir η assez petit, ce qui donne ε , puis n assez grand.

Raccourci. Le théorème de convergence dominée permet de déduire la convergence en loi de la convergence presque sûre sans passer par la convergence en probabilité. ■

Remarque 1.4.2. Réciproques.

1. Si X est constante (et égale p.s. à c) alors la convergence en loi entraîne la convergence en probabilité : si $f_{c,\varepsilon} : \mathbb{R} \rightarrow \mathbb{R}_+$ est continue bornée avec $f(c) = 0$ et $\mathbb{1}_{(c-\varepsilon, c+\varepsilon)^c} \leq f_{c,\varepsilon}$ alors

$$\mathbb{P}(|X_n - X| \geq \varepsilon) = \mathbb{E}(\mathbb{1}_{(c-\varepsilon, c+\varepsilon)^c}(X_n)) \leq \mathbb{E}(f_{c,\varepsilon}(X_n)) \rightarrow \mathbb{E}(f_{c,\varepsilon}(c)) = 0.$$

2. Si une suite converge en probabilité vers une variable aléatoire intégrable, alors elle converge en moyenne d'ordre 1 ssi elle est uniformément intégrable (u.i.) cf. section dédiée.

Exemple 1.4.3. Contrexemples.

1. *Convergence en probabilité $\not\Rightarrow$ presque sûrement* : suite tournante sur $[0, 1]$. Posons $X_{m+k} = \mathbb{1}_{[k, k+1)/m}(U)$ pour $0 \leq k < m$ et $m \geq 1$, où U est uniforme sur $[0, 1]$. On a alors $X_n \xrightarrow{\mathbb{P}} 0$ mais $(X_n)_{n \geq 1}$ ne converge pas p.s. car « tous les ω voient passer des 0 et des 1 une infinité de fois ». Autre exemple moins artificiel : $X_n = 1/(1 + |S_n|)$ avec S_n marche aléatoire simple, qui converge en probabilité vers 0 car S_n s'étale de plus en plus, mais ne converge pas p.s. en raison des retours à 0 (réurrence).
2. *Convergence en loi $\not\Rightarrow$ en probabilité*. Soit G une v.a. de loi gaussienne $\mathcal{N}(0, 1)$. Pour tout $n \geq 1$, posons $X_n := G$ et soit $X := -G$. La suite $(X_n)_{n \geq 1}$ converge en loi vers X . En revanche elle ne peut pas converger en probabilité, car $\mathbb{P}(|X_n - X| \geq \varepsilon) = \mathbb{P}(2|G| \geq \varepsilon) \not\rightarrow 0$ quand $n \rightarrow \infty$, pour tout $\varepsilon > 0$.
3. *Convergence en probabilité $\not\Rightarrow$ en moyenne*. On pose $X_n = n\mathbb{1}_{[0, 1/n]}(U)$ où U uniforme sur $[0, 1]$. On a $\mathbb{P}(|X_n - 0| \geq \varepsilon) = \mathbb{P}(U \leq 1/n) = 1/n \rightarrow 0$ pour tout $\varepsilon > 0$ et donc $X_n \xrightarrow{\mathbb{P}} 0$. Cependant, $\mathbb{E}(|X_n - 0|) = n \int_0^1 \mathbb{1}_{[0, 1/n]}(u) du = 1 \not\rightarrow 0$ et donc $(X_n)_{n \geq 1}$ ne converge pas en moyenne vers 0.

Remarque 1.4.4. Stabilités des convergences.

Les convergences p.s., en probabilité, et en moyenne sont stables par combinaisons linéaires finies. Les convergences p.s., en probabilité et en loi sont stables par composition avec une fonction continue.

Remarque 1.4.5. Extension aux vecteurs aléatoires.

Les trois premiers types de convergence s'étendent naturellement aux vecteurs aléatoires et au-delà, avec une distance ou une norme par exemple.

Remarque 1.4.6. Métriques.

Les convergences en moyenne d'ordre p , en probabilité, et en loi sont métriques ou métrisables, tandis que la convergence p.s. ne l'est pas.

1.5 Fonction caractéristique

La fonction caractéristique d'un vecteur aléatoire X de \mathbb{R}^d de loi μ est définie par

$$t \in \mathbb{R}^d \mapsto \varphi_X(t) = \mathbb{E}(e^{i\langle t, X \rangle}) = \int e^{i\langle t, x \rangle} \mu(dx).$$

1. Elle ne dépend que de la loi de X , et constitue sa transformée de Fourier.
 2. Elle est continue, bornée par 1, et $\varphi(0) = 1$.
 3. Elle caractérise la loi : X et Y de même loi ssi $\varphi_X = \varphi_Y$ (tomographie, théorème de Cramér–Wold³).
 4. Elle caractérise la convergence en loi : $X_n \rightarrow X$ en loi ssi $\varphi_{X_n}(t) \rightarrow \varphi_X(t)$ pour tout t .
De plus, si $\varphi_{X_n}(t) \rightarrow \varphi(t)$ pour tout t , avec φ continue en 0, alors il existe une variable aléatoire X telle que $\varphi_X = \varphi$ et $X_n \rightarrow X$ en loi (théorème de continuité de Cramér–Lévy).
 5. Elle caractérise l'indépendance : X et Y indépendants ssi $\varphi_{(X, Y)}(s, t) = \varphi_X(s)\varphi_Y(t)$ pour tous s, t .
 6. Elle est très commode pour les sommes de v.a.i.i.d. : $\varphi_{X_1 + \dots + X_n}(t) = \varphi_{X_1}(t)^n$.
-
3. Si X est un vecteur aléatoire de \mathbb{R}^n alors $\text{proj}_{\mathbb{R}^t}(X) = \langle t, X \rangle t$ pour tout $t \in \mathbb{R}^n$ tel que $|t| = 1$, et $\varphi_X(t) = \varphi_{\langle \frac{t}{|t|}, X \rangle}(|t|)$ pour tout $t \in \mathbb{R}^n$.

7. Si X a un moment d'ordre n , alors φ_X est $\mathcal{C}^n(\mathbb{R}, \mathbb{C})$, et $\varphi_X^{(k)}(0) = i^k \mathbb{E}(X^k)$ pour tout $k \leq n$. Réciproquement, si φ_X est dérivable n fois en 0, alors X possède un moment d'ordre n si n est pair et d'ordre $n-1$ si n est impair.
8. C'est en fait le dual de \mathbb{R}^n qui intervient pour transformer X en la v.a. $\langle t, X \rangle$ via $\ell(x) := \langle t, x \rangle$.
9. Elle ne fait que tester la loi de X sur les fonctions test trigonométriques $x \mapsto e^{i\langle t, x \rangle}$. En fonction du contexte, d'autres classes de fonctions tests conduisent à d'autres transformées qui peuvent rendre des services similaires : fonction de répartition et fonction de masse (indicatrices), fonction génératrice (puissance), transformée de Laplace (exponentielle), transformée de Cauchy–Stieltjes (inversion), etc.

1.6 Lemme de Borel–Cantelli

Pour une suite $(A_n)_n$ d'événements d'un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$, on pose

$$\begin{aligned} \underline{\lim}_n A_n &:= \cup_n \cap_{m \geq n} A_m = \{\omega \in \Omega : \omega \in A_n \text{ à partir d'un certain rang sur } n\}, \\ \overline{\lim}_n A_n &:= \cap_n \cup_{m \geq n} A_m = \{\omega \in \Omega : \omega \in A_n \text{ pour une infinité de valeurs de } n\}. \end{aligned}$$

On a $(\underline{\lim}_n A_n^c)^c = \overline{\lim}_n A_n$. D'autre part $\overline{\lim}_n \mathbb{1}_{A_n} = \mathbb{1}_{\overline{\lim}_n A_n}$ et $\underline{\lim}_n \mathbb{1}_{A_n} = \mathbb{1}_{\underline{\lim}_n A_n}$.

1. (Cantelli) si $\sum_n \mathbb{P}(A_n) < \infty$ alors $\mathbb{P}(\overline{\lim}_n A_n) = 0$.
2. (Loi du zéro-un de Borel) si $\sum_n \mathbb{P}(A_n) = \infty$ et les $(A_n)_n$ sont indépendants alors $\mathbb{P}(\overline{\lim}_n A_n) = 1$.

Le lemme de Borel–Cantelli est un grand pourvoyeur de presque sûr. Mieux : si X prend ses valeurs dans $[0, \infty]$ alors $\mathbb{E}(X) < \infty$ implique $X < \infty$ p.s., et ceci permet d'établir la partie Cantelli :

$$\sum_n \mathbb{P}(A_n) = \sum_n \mathbb{E} \mathbb{1}_{A_n} = \mathbb{E} \sum_n \mathbb{1}_{A_n} \quad \text{et} \quad \left\{ \sum_n \mathbb{1}_{A_n} = \infty \right\} = \overline{\lim}_n A_n.$$

Le lemme de Borel–Cantelli indique que la convergence en probabilité sommable implique la convergence presque sûre : si $\sum_n \mathbb{P}(|X_n - X| \geq \varepsilon) < \infty$ pour tout $\varepsilon > 0$, alors presque sûrement $|X_n - X| < \varepsilon$ pour n assez grand, et donc, en prenant $\varepsilon = 1/k$, il vient que presque sûrement, $X_n \rightarrow X$. En particulier, de toute suite convergeant en probabilité on peut extraire une sous-suite convergeant presque sûrement.

Remarque 1.6.1. Presque sûr.

Beaucoup de comportements p.s. passent au bout du compte par les propriétés élémentaires suivantes : si X prend ses valeurs dans $[0, \infty]$ alors $\mathbb{E}(X) = 0$ implique $X = 0$ p.s. tandis que $\mathbb{E}(X) < \infty$ implique $X < \infty$ p.s.

1.7 Loi des grands nombres, théorème limite central, loi du logarithme itéré

Loi forte des grands nombres. Si $X \in L^1$ et X_1, X_2, \dots sont des copies i.i.d. de X alors, avec $m := \mathbb{E}(X)$,

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} m \quad \text{et} \quad \frac{X_1 + \dots + X_n}{n} \xrightarrow[n \rightarrow \infty]{L^1} m.$$

Théorème limite central. Si de plus $X \in L^2$, alors, avec $\sigma^2 = \text{Var}(X) = \mathbb{E}((X - m)^2) = \mathbb{E}(X^2) - m^2$,

$$\frac{\sqrt{n}}{\sigma} \left(\frac{X_1 + \dots + X_n}{n} - m \right) = \frac{X_1 - m + \dots + X_n - m}{\sqrt{n}\sigma} \xrightarrow[n \rightarrow \infty]{\text{loi}} \mathcal{N}(0, 1).$$

Ceci implique une loi faible des grand nombres : $\frac{X_1 + \dots + X_n}{n} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} m$, par le lemme de Slutsky ci-dessous.

Loi du logarithme itéré. Sous les hypothèses et avec les notations du TLC, presque sûrement,

$$\overline{\lim}_{n \rightarrow \infty} \left(\frac{\sqrt{n}}{\sigma \sqrt{2 \log \log(n)}} \left(\frac{X_1 + \dots + X_n}{n} - m \right) \right) = \overline{\lim}_{n \rightarrow \infty} \left(\frac{X_1 - m + \dots + X_n - m}{\sqrt{2n \log \log(n)} \sigma} \right) = 1$$

et

$$\underline{\lim}_{n \rightarrow \infty} \left(\frac{\sqrt{n}}{\sigma \sqrt{2 \log \log(n)}} \left(\frac{X_1 + \dots + X_n}{n} - m \right) \right) = \underline{\lim}_{n \rightarrow \infty} \left(\frac{X_1 - m + \dots + X_n - m}{\sqrt{2n \log \log(n)} \sigma} \right) = -1.$$

D'autres théorèmes limites, comme les principes de grandes déviations, sont présentés dans [12].

1.8 Autour de la convergence en loi : lemme de Slutsky et méthode delta

La convergence en loi est stable par déformation continue⁴ : si $X_n \rightarrow X$ en loi alors $f(X_n) \rightarrow f(X)$ en loi, pour toute fonction continue f . En effet, cela découle du fait que $g \circ f \in \mathcal{C}_b$ pour tout $g \in \mathcal{C}_b$.

4. En anglais on parle de *continuous mapping theorem*.

Théorème 1.8.1. Lemme de Slutsky.

Si $X_n \rightarrow X$ en loi et $Y_n \rightarrow Y$ en loi avec Y constante alors $(X_n, Y_n) \rightarrow (X, Y)$ en loi.

En particulier, $X_n Y_n \xrightarrow{\text{loi}} XY$; $X_n + Y_n \xrightarrow{\text{loi}} X + Y$; $X_n / Y_n \xrightarrow{\text{loi}} X / Y$ si $Y \neq 0$. Cela permet notamment de remplacer une moyenne ou une variance par un estimateur empirique dans un théorème de convergence en loi comme le TLC, ce qui s'avère pratique pour fabriquer une région de confiance ou un test d'hypothèse statistique :

$$\frac{\sqrt{n}}{\hat{\sigma}_n} \left(\frac{X_1 + \dots + X_n}{n} - m \right) \xrightarrow[n \rightarrow \infty]{\text{loi}} \mathcal{N}(0, 1) \quad \text{pour tout estimateur } \hat{\sigma}_n \xrightarrow[n \rightarrow \infty]{\text{loi}} \sigma.$$

Démonstration. Comme Y est constante, $Y_n \rightarrow Y$ en probabilité, d'où, en utilisant également la continuité uniforme de $u \mapsto e^{iu}$ (théorème de Heine), il vient, pour tous $s, t \in \mathbb{R}$ et tout $\varepsilon > 0$, avec le paramètre $\eta = \eta(\varepsilon) > 0$ donné par la continuité uniforme et pour $n \geq N(\eta, \varepsilon)$ où le seuil $N(\eta, \varepsilon)$ est donné par la convergence en probabilité,

$$|\mathbb{E}(e^{itX_n + isY_n}) - \mathbb{E}(e^{itX_n + isY})| \leq 2\mathbb{P}(|Y_n - Y| \geq \eta) + \mathbb{E}(|e^{isY_n} - e^{isY}| \mathbb{1}_{|Y_n - Y| < \eta}) \leq 2\varepsilon + \varepsilon,$$

tandis que d'autre part, comme Y est constante et $X_n \rightarrow X$ en loi,

$$\mathbb{E}(e^{itX_n + isY}) = \mathbb{E}(e^{itX_n})\mathbb{E}(e^{isY}) \xrightarrow[n \rightarrow \infty]{} \mathbb{E}(e^{isY})\mathbb{E}(e^{itX}) = \mathbb{E}(e^{itX + isY}).$$

Cette preuve indique que le lemme de Slutsky se généralise au cas où $Y_n \rightarrow Y$ en loi avec Y indépendante de X . ■

Théorème 1.8.2. Méthode delta.

Si $a_n(X_n - b_n) \xrightarrow[n \rightarrow \infty]{\text{loi}} L$ avec $a_n \rightarrow \infty$ et $b_n \rightarrow b$ déterministes, alors pour toute $f \in \mathcal{C}^1(\mathbb{R}, \mathbb{R})$,

$$a_n(f(X_n) - f(b_n)) \xrightarrow[n \rightarrow \infty]{\text{loi}} \text{Loi}(f'(b)Z) \quad \text{où } Z \sim L.$$

Plus généralement, si $f \in \mathcal{C}^m(\mathbb{R}, \mathbb{R})$, $m \geq 1$, et $f^{(k)}(b) = 0$ pour tout $1 \leq k \leq m-1$, alors

$$a_n^m(f(X_n) - f(b_n)) \xrightarrow[n \rightarrow \infty]{\text{loi}} \text{Loi}\left(\frac{f^{(m)}(b)}{m!} Z^m\right) \quad \text{où } Z \sim L.$$

Cela permet par exemple de déduire un résultat de fluctuation asymptotique à partir d'un autre, comme le TLC.

Démonstration. Le théorème des accroissements finis sous forme d'égalité donne⁵ $f(X_n) - f(b_n) = f'(Y_n)(X_n - b_n)$ avec Y_n v.a. à valeurs entre X_n et b_n . Comme $a_n \rightarrow \infty$ et $a_n(X_n - b_n) \rightarrow L$ en loi, le lemme de Slutsky donne $(X_n - b_n) \rightarrow 0$ en loi, donc en probabilité, donc $Y_n \rightarrow b$ en probabilité, donc $f'(Y_n) \rightarrow f'(b)$ en probabilité, et le lemme de Slutsky à nouveau donne le résultat pour $m = 1$. Même méthode pour $m \geq 1$ avec une formule de Taylor. ■

1.9 Intégrabilité uniforme ou équi-intégrabilité

Pour toute variable aléatoire X , le théorème de convergence dominée indique que

$$X \in L^1 \quad \text{ssi} \quad \lim_{R \rightarrow +\infty} \mathbb{E}(|X| \mathbb{1}_{|X| \geq R}) = 0.$$

C'est une caractérisation de l'intégrabilité par troncurabilité. Voici une version uniforme pour les familles de v.a. :

Théodéf 1.9.1. Familles uniformément intégrables (u.i.).

Pour une famille $(X_i)_{i \in I} \subset L^1$, les propriétés suivantes sont équivalentes :

- (i) **(définition)** $\lim_{R \rightarrow +\infty} \sup_{i \in I} \mathbb{E}(|X_i| \mathbb{1}_{|X_i| \geq R}) = 0$.
- (ii) **(critère epsilon-delta)** bornitude dans L^1 : $\sup_{i \in I} \mathbb{E}(|X_i|) < \infty$, et $\forall \varepsilon > 0, \exists \delta > 0, \forall A \in \mathcal{A}$,

$$\mathbb{P}(A) \leq \delta \Rightarrow \sup_{i \in I} \mathbb{E}(|X_i| \mathbb{1}_A) \leq \varepsilon.$$

5. Pour rendre $f'(Y_n)$ mesurable, considérer par exemple le plus petit Y_n réalisant l'égalité des accroissements finis.

(iii) (**critère de de la Vallée Poussin**) il existe $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ mesurable telle que

$$\lim_{x \rightarrow +\infty} \frac{\varphi(x)}{x} = +\infty \quad \text{et} \quad \sup_{i \in I} \mathbb{E}(\varphi(|X_i|)) < \infty.$$

On dit alors que la famille $(X_i)_{i \in I}$ est **uniformément intégrable (u.i.)** ou **équi-intégrable**.

De plus, dans ce cas, la propriété (iii) est réalisable avec φ croissante, convexe, et à croissance modérée :

$$\varphi(x) \leq x^2 \quad \text{pour tout } x \geq 0.$$

Soit Φ l'ensemble des $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ croissantes telles que $\lim_{x \rightarrow +\infty} \frac{\varphi(x)}{x} = +\infty$. Exemple : $x \mapsto x^p$, $p > 1$.
Si $(X_i)_{i \in I}$ vérifie $\sup_{i \in I} \mathbb{E}(\varphi(|X_i|)) < \infty$ pour un $\varphi \in \Phi$, alors $(X_i)_{i \in I}$ est bornée dans L^1 .

Démonstration. Preuve de (i) \Rightarrow (ii). Pour la bornitude dans L^1 , pour tout $i \in I$, et $R \geq 0$ assez grand,

$$\mathbb{E}(|X_i|) = \mathbb{E}(|X_i| \mathbb{1}_{|X_i| < R}) + \mathbb{E}(|X_i| \mathbb{1}_{|X_i| \geq R}) \leq R + \sup_{i \in I} \mathbb{E}(|X_i| \mathbb{1}_{|X_i| \geq R}) < \infty.$$

Ensuite, par hypothèse, pour tout $\varepsilon > 0$, il existe R_ε tel que $\sup_{i \in I} \mathbb{E}(|X_i| \mathbb{1}_{|X_i| \geq R}) \leq \varepsilon$ pour tout $R \geq R_\varepsilon$. Par conséquent, si $A \in \mathcal{A}$ vérifie $\mathbb{P}(A) \leq \delta_\varepsilon := \varepsilon / R_\varepsilon$, alors pour tout $i \in I$,

$$\mathbb{E}(|X_i| \mathbb{1}_A) = \mathbb{E}(|X_i| \mathbb{1}_{|X_i| < R_\varepsilon} \mathbb{1}_A) + \mathbb{E}(|X_i| \mathbb{1}_{|X_i| \geq R_\varepsilon} \mathbb{1}_A) \leq R_\varepsilon \mathbb{P}(A) + \varepsilon \leq 2\varepsilon.$$

Preuve de (ii) \Rightarrow (i). Fixons $\varepsilon > 0$, et soit $\delta > 0$ comme dans (ii). Comme $(X_i)_{i \in I}$ est bornée dans L^1 , pour tout $j \in I$, si $A_j := \{|X_j| \geq R\}$, alors, par l'inégalité de Markov, $\mathbb{P}(A_j) \leq R^{-1} \sup_{i \in I} \mathbb{E}(|X_i|) \leq \delta$ pour R assez grand, uniformément en $j \in I$, et l'hypothèse pour $A = A_j$ donne $\lim_{R \rightarrow +\infty} \sup_{i, j \in I} \mathbb{E}(|X_i| \mathbb{1}_{|X_j| \geq R}) = 0$.

Preuve de (iii) \Rightarrow (i). Pour tout $\varepsilon > 0$, comme $\varphi \in \Phi$, par définition, il existe $R_\varepsilon \geq 0$ tel que $|x| \leq \varepsilon \varphi(|x|)$ si $|x| \geq R_\varepsilon$, d'où

$$\sup_{i \in I} \mathbb{E}(|X_i| \mathbb{1}_{|X_i| \geq R_\varepsilon}) \leq \varepsilon \sup_{i \in I} \mathbb{E}(\varphi(|X_i|) \mathbb{1}_{|X_i| \geq R_\varepsilon}) \leq \varepsilon \sup_{i \in I} \mathbb{E}(\varphi(|X_i|)).$$

Preuve de (i) \Rightarrow (iii). Pour une suite $(x_m)_{m \geq 1} \nearrow +\infty$ à déterminer, considérons la fonction positive

$$\varphi(x) = \sum_{m \geq 1} (x - x_m)^+.$$

Elle est croissante et convexe. De plus, par convergence monotone (pour la mesure de comptage),

$$\frac{\varphi(x)}{x} = \sum_{m \geq 1} \left(1 - \frac{x_m}{x}\right)^+ \xrightarrow{x \rightarrow +\infty} \sum_{m \geq 1} 1 = \infty.$$

Il reste à observer qu'en choisissant x_m grâce à l'hypothèse tel que $\mathbb{E}(|X_i| \mathbb{1}_{|X_i| \geq x_m}) \leq 2^{-m}$, il vient

$$\mathbb{E}(\varphi(|X_i|)) \leq \sum_{m \geq 1} \mathbb{E}(|X_i| \mathbb{1}_{|X_i| \geq x_m}) \leq \sum_{m \geq 1} 2^{-m} = 1 < \infty.$$

Ceci n'est pas abordé lors du cours oral.

Il est possible de remplacer 2^{-m} par toute autre quantité sommable en m . Alternativement, il est possible de rechercher φ dans la famille des fonctions linéaires par morceaux de la forme

$$\varphi(x) = \int_0^x \varphi'(t) dt \quad \text{avec} \quad \varphi' = \sum_{n=1}^{\infty} u_n \mathbb{1}_{[n, n+1)} \quad \text{et} \quad u_n = \sum_{m \geq 1} \mathbb{1}_{x_m \leq n}.$$

On a $\varphi \in \Phi$ car $\lim_{n \rightarrow \infty} u_n = +\infty$. De plus, comme $\varphi \leq \sum_{n=1}^{\infty} (u_1 + \dots + u_n) \mathbb{1}_{[n, n+1)}$, on a, pour tout $i \in I$,

$$\begin{aligned} \mathbb{E}(\varphi(|X_i|)) &\leq \sum_{n=1}^{\infty} (u_1 + \dots + u_n) \mathbb{P}(n \leq |X_i| < n+1) \\ &= \sum_{n=1}^{\infty} u_n \mathbb{P}(|X_i| \geq n) \\ &= \sum_{m \geq 1} \sum_{n \geq x_m} \mathbb{P}(|X_i| \geq n) \\ &\leq \sum_{m \geq 1} \sum_{n \geq x_m} n \mathbb{P}(n \leq |X_i| < n+1) \end{aligned}$$

$$\begin{aligned} &\leq \sum_{m \geq 1} \mathbb{E}(|X_i| \mathbb{1}_{|X_i| \geq x_m}) \\ &\leq \sum_{m \geq 1} 2^{-m} < \infty \end{aligned}$$

où la dernière inégalité large a lieu car $\sup_{i \in I} \mathbb{E}(|X_i| \mathbb{1}_{|X_i| \geq x_m}) \leq 2^{-m}$. La fonction φ est croissante et convexe. Reste à faire en sorte d'obtenir une croissance modérée. On observe tout d'abord que pour tout $x \geq 0$,

$$\varphi'(2x) \leq c\varphi'(x),$$

avec $c = 2$. En effet, on peut prendre x_m arbitrairement grand, par exemple tel que pour tout $n \geq 1$,

$$u_{2n} \leq u_n + \sum_{m \geq 1} \mathbb{1}_{n < x_m \leq 2n} \leq 2u_n.$$

Par conséquent, la fonction φ est à croissance modérée car avec $C := 2c$,

$$\varphi(2x) = \int_0^{2x} g(u) du = \int_0^x g(2t) 2dt \leq 2c\varphi(x) = C\varphi(x).$$

À présent $\varphi(x) \leq C^{1+k} \varphi(2^{-k}x)$ pour tout $k \geq 1$, et avec $k = k_x = \lceil \log_2(x) \rceil$, on obtient

$$\varphi(x) \leq C^2 C^{\log_2(x)} \varphi(1) = C^2 \varphi(1) 2^{\log_2(C) \log_2(x)} = C^2 \varphi(1) x^{\log_2(C)}.$$

Comme on peut prendre $c = 2$, on obtient $C = 4$, ce qui permet $\varphi(x) \leq x^2$ par dilatation. ■

Exemple 1.9.2. Exemples de familles u.i.

- Toute partie finie de L^1 est u.i. par convergence dominée.
- Plus généralement, si $|X_i| \leq X$ pour tout $i \in I$ avec $X \in L^1$ (domination ^a) alors $(X_i)_{i \in I}$ est u.i. En effet, pour tout $i \in I$ et tout $R > 0$, $|X_i| \mathbb{1}_{|X_i| \geq R} \leq |X| \mathbb{1}_{|X| \geq R}$ donc $\sup_{i \in I} \mathbb{E}(|X_i| \mathbb{1}_{|X_i| \geq R}) \leq \mathbb{E}(|X| \mathbb{1}_{|X| \geq R}) \rightarrow 0$ quand $R \rightarrow +\infty$ par convergence dominée car $X \in L^1$.
- Le caractère u.i. est stable par réunion finie de familles u.i., également par prise d'enveloppe convexe, mais n'est pas stable par prise d'espace vectoriel engendré en général.
- $(|X_i|)_{i \in I}$ est u.i. ssi $(X_i)_{i \in I}$ est u.i.
- Pour tout $p \geq 1$, si $X \in L^p$ et $X_n \xrightarrow{L^p} X$, alors $(|X_n|^p)_{n \geq 1} \cup \{|X|^p\}$ et $(|X_n - X|^p)_{n \geq 1}$ sont u.i. En effet, pour tout $\varepsilon > 0$, soit n tel que $\sup_{k \geq n} \mathbb{E}(|X_k - X|^p) \leq \varepsilon$. Soit $\delta > 0$ du critère epsilon-delta pour la famille finie $\{|X_1 - X|^p, \dots, |X_n - X|^p\}$. Pour tout $A \in \mathcal{A}$ t.q. $\mathbb{P}(A) \leq \delta$,

$$\sup_{n \geq 1} \mathbb{E}(|X_n - X|^p \mathbb{1}_A) \leq \max \left(\max_{k \leq n} \mathbb{E}(|X_k - X|^p \mathbb{1}_A); \sup_{k \geq n} \mathbb{E}(|X_k - X|^p) \right) \leq \varepsilon.$$

- Une famille bornée dans L^p pour un réel $p > 1$ est toujours u.i., conséquence du critère de de la Vallée Poussin. Ce critère est souvent utilisé avec $p = 2$ (second moment des v.a.). En revanche, une famille bornée dans L^1 n'est pas forcément u.i., il manque quelque chose! Les martingales étudiées dans le chapitre 4 fournissent de multiples contrexemples en la matière.

a. Cette domination pourrait s'écrire $\sup_{i \in I} |X_i| \in L^1$ mais ce sup n'est pas forcément mesurable si I n'est pas dénombrable.

Remarque 1.9.3. Intégrabilité améliorée.

Le critère de de la Vallée Poussin pour la famille u.i. singleton $\{X\}$ indique que $X \in L^1$ implique que $\varphi(|X|) \in L^1$ pour un $\varphi \in \Phi$. En d'autres termes, pour toute v.a., l'intégrabilité peut être améliorée. Il n'y a pas de paradoxe car φ dépend de X . Topologiquement, l'intégrabilité est une condition ouverte plutôt que fermée, et une instance élémentaire de ce phénomène est visible dans les séries de Riemann : si $\sum_{n \geq 1} n^{-s} < \infty$ pour un $s > 0$ alors $\sum_{n \geq 1} n^{-s'} < \infty$ pour un $s' < s$, car la condition de convergence $s > 1$ est ouverte.

Théorème 1.9.4. Convergence dominée améliorée.

Pour toute suite de v.a. $(X_n)_{n \geq 1}$ dans L^1 et toute variable aléatoire X ,

$$X \in L^1 \text{ et } X_n \xrightarrow{L^1} X \quad \text{ssi} \quad X_n \xrightarrow{\mathbb{P}} X \text{ et } (X_n)_{n \geq 1} \text{ est u.i.}$$

Plus généralement, pour tout $p \geq 1$, toute suite de v.a. $(X_n)_{n \geq 1}$ dans L^p et toute variable aléatoire X ,

$$X \in L^p \text{ et } X_n \xrightarrow{L^p} X \quad \text{ssi} \quad X_n \xrightarrow{\mathbb{P}} X \text{ et } (|X_n|^p)_{n \geq 1} \text{ est u.i.}$$

Améliore le théorème de convergence dominée puisque $(X_n)_{n \geq 1}$ est u.i. lorsque $\sup_n |X_n| \leq X$ avec $X \in L^1$. Ce théorème est dû à G. Vitali tandis que le théorème de convergence dominée est dû à H. Lebesgue.

Démonstration. Supposons que $X \in L^1$ et $X_n \rightarrow X$ dans L^1 . Alors comme déjà vu, $X_n \rightarrow X$ en probabilité par l'inégalité de Markov, et $(X_n)_{n \geq 1} \cup \{X\}$ est u.i. (exemples de familles u.i. plus haut).

Démontrons la réciproque. Supposons que $X_n \rightarrow X$ en probabilité et que $(X_n)_{n \geq 1}$ est u.i. Alors on sait que $(X_n)_{n \in \mathbb{N}}$ est bornée dans L^1 . Soit $(X_{n_k})_{k \in \mathbb{N}}$ une sous-suite qui converge p.s. qui peut s'obtenir par exemple au moyen du lemme de Borel – Cantelli. Alors $X \in L^1$ car par le lemme de Fatou :

$$\mathbb{E}(|X|) = \mathbb{E}(\liminf_{k \rightarrow \infty} |X_{n_k}|) \leq \liminf_{k \rightarrow \infty} \mathbb{E}(|X_{n_k}|) \leq \sup_{n \in \mathbb{N}} \mathbb{E}(|X_n|) < \infty.$$

Ensuite, par le critère epsilon-delta, $(X_n - X)_{n \geq 1}$ est u.i. car $(X_n)_{n \geq 1}$ est u.i. et $X \in L^1$. Alors en utilisant la convergence en probabilité et le critère epsilon-delta, on obtient que pour tout $\varepsilon > 0$ et n assez grand,

$$\mathbb{E}(|X_n - X|) = \mathbb{E}(|X_n - X| \mathbb{1}_{|X_n - X| \geq \varepsilon}) + \mathbb{E}(|X_n - X| \mathbb{1}_{|X_n - X| < \varepsilon}) \leq 2\varepsilon.$$

Démontrons enfin la version L^p . Si $X \in L^p$ et $X_n \rightarrow X$ dans L^p , alors $X_n \rightarrow X$ en probabilité (inégalité de Markov), et $(|X_n|^p)_{n \in \mathbb{N}}$ est u.i. (exemples de familles u.i. plus haut). Réciproquement, supposons que $X_n \rightarrow X$ en probabilité et $(|X_n|^p)_{n \geq 1}$ est u.i. Alors $X \in L^p$ (procéder comme pour $p = 1$). De plus $|X_n - X|^p \rightarrow 0$ en probabilité, et $(|X_n - X|^p)_{n \in \mathbb{N}}$ est u.i. d'où $|X_n - X|^p \rightarrow 0$ dans L^1 c'est-à-dire $X_n \rightarrow X$ dans L^p . ■

Remarque 1.9.5. Variables aléatoires ou lois.

Pour une famille $(X_i)_{i \in I}$ de v.a. dans L^1 , être u.i. est une propriété de la famille de leurs lois $(\mu_i)_{i \in I}$, $X_i \sim \mu_i$. Les (i), (ii), et (iii) de la définition de u.i. se réécrivent comme suit en terme de lois :

(i) (définition) $\lim_{R \rightarrow +\infty} \int_{|x| \geq R} x d\mu_i(x) = 0.$

(ii) (epsilon-delta) $\sup_{i \in I} \int |x| d\mu_i(x) < \infty$ et $\forall \varepsilon > 0, \exists \delta > 0, \forall A \in \mathcal{A}, \mathbb{P}(A) \leq \delta \Rightarrow \sup_{i \in I} \int_A |x| d\mu_i(x) \leq \varepsilon.$

(iii) (de la Vallée Poussin) il existe $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ telle que $\lim_{x \rightarrow +\infty} \varphi(x)/x = +\infty$ et $\sup_{i \in I} \int \varphi(|x|) d\mu_i(x) < \infty.$

Remarque 1.9.6. Intégrabilité uniforme comme relative compacité.

En analyse fonctionnelle, l'intégrabilité uniforme est reliée à la relative compacité pour une topologie faible sur L^1 . Plus précisément, le théorème de Dunford – Pettis, dont une preuve se trouve par exemple dans [17], affirme que pour toute famille $(X_i)_{i \in I} \subset L^1$, les propriétés suivantes sont équivalentes :

- $(X_i)_{i \in I}$ est uniformément intégrable.
- $(X_i)_{i \in I}$ est relativement compacte pour la topologie faible $\sigma(L^1, L^\infty)$.
- $(X_i)_{i \in I}$ est relativement séquentiellement compacte pour la topologie faible $\sigma(L^1, L^\infty)$.

Théorème 1.9.7. Tension d'une famille de mesures de probabilité.

Soit $(\mu_i)_{i \in I}$ une famille de mesures de probabilités sur un espace topologique E muni de sa tribu borélienne. Les deux propriétés suivantes sont équivalentes, et on dit alors que $(\mu_i)_{i \in I}$ est tendue :

(i) pour tout $\varepsilon > 0$, il existe un compact $K_\varepsilon \subset E$ tel que $\sup_{i \in I} \mu_i(K_\varepsilon^c) \leq \varepsilon.$

(ii) il existe $\varphi : E \rightarrow [0, \infty]$ mesurable à ensembles de sous-niveaux compacts telle que $\sup_{i \in I} \int \varphi d\mu_i < \infty.$

En considérant une variable aléatoire $X_i \sim \mu_i$ pour tout $i \in I$, les deux propriétés s'écrivent

$$\sup_{i \in I} \mathbb{P}(X_i \notin K_\varepsilon) \leq \varepsilon \quad \text{et} \quad \sup_{i \in I} \mathbb{E}(\varphi(X_i)) < \infty.$$

Lorsque $E = \mathbb{R}$ et φ est surlinéaire, on reconnaît le critère d'u.i. L'u.i. est un concept de tension fort à base de moment d'ordre > 1 . La tension est un concept important en théorie des probabilités. La tension est à la convergence en loi ce que l'u.i. est à la convergence L^1 , en quelque sorte.

Démonstration. Les ensembles de sous-niveau de φ sont $\{x \in E : \varphi(x) \leq r\}$, $r \geq 0$. Si E n'est pas compact alors il ne peut pas être un ensemble de sous-niveau de φ et donc φ n'est pas bornée, tandis que si E est compact alors la seconde propriété a immédiatement lieu avec φ constante.

Preuve de (ii) \Rightarrow (i). Par l'inégalité de Markov, pour tout $r > 0$,

$$\sup_{i \in I} \mu_i(\{x \in E : \varphi(x) \leq r\}^c) \leq \frac{1}{r} \sup_{i \in I} \int \varphi d\mu_i = \frac{C}{r},$$

qui conduit à prendre $r = r_\varepsilon = C/\varepsilon$ et $K_\varepsilon = \{x \in E : \varphi(x) \leq r_\varepsilon\}$, pour tout $\varepsilon > 0$.

Preuve de (i) \Rightarrow (ii). Soit $(\varepsilon_n)_{n \geq 1}$ une suite de réels positifs telle que $\sum_n \varepsilon_n < \infty$. Soit K_{ε_n} le compact fourni par l'hypothèse. Considérons le compact $K'_n := \cup_{k=1}^n K_{\varepsilon_k}$. Par construction, la suite $(K'_n)_{n \geq 1}$ est croissante pour l'inclusion, et pour tout $i \in I$, $\mu_i(K_n^c) \leq \mu_i(K_{\varepsilon_n}^c) \leq \varepsilon_n$. Maintenant, pour tout $x \in F = \cup_n K'_n$, il existe n_x tel que $x \in K'_m$ pour tout $m \geq n_x$, et donc $\sum_n \mathbb{1}_{K_n^c}(x) \leq n_x - 1 < \infty$. Par conséquent, si on définit

$$\varphi := \sum_n \mathbb{1}_{K_n^c}$$

alors $\varphi < \infty$ sur F tandis que $\varphi = \infty$ sur F^c , et φ a des ensembles de sous-niveau compacts car $\{x \in E : \varphi(x) \leq n-1\} = K'_n$ pour tout $n \geq 1$. D'un autre côté, par définition de K'_n ,

$$\sup_{i \in I} \int \varphi d\mu_i \leq \sum_n \varepsilon_n < \infty.$$

■

Remarque 1.9.8. Tension et relative compacité.

En analyse fonctionnelle et théorie des probabilités, le théorème de tension de Prokhorov affirme que pour toute famille de mesures de probabilité les propriétés suivantes sont équivalentes :

- la famille est tendue.
- la famille est relativement compacte pour la topologie de la convergence étroite.

En pratique, il est courant d'utiliser ce théorème comme suit : si une suite de mesures de probabilités est tendue, grâce à l'existence d'une fonction φ comme ci-dessus, alors il ne reste plus qu'à établir l'unicité de la valeur d'adhérence de la suite pour montrer qu'elle converge. Il s'agit de l'analogue du théorème de Bolzano – Weierstrass pour les mesures de probabilités.

Pour aller plus loin 1.9.9.

1. Théorème de tension de Prokhorov [7, Th. 25.10].
2. Inégalité maximale de Kolmogorov [7, Th. 22.4].
3. TLC par méthode de remplacement de Lindeberg (couplage) [13].
4. TLC par méthode de Stein (intégration par parties) [27].
5. Simulation de variables aléatoires [5].
6. Formule de Lévy – Khintchine, lois infiniment divisibles, lois stables [7].
7. Théorème d'extension de Carathéodory – Kolmogorov.
8. Problème des moments et caractérisation de la loi par les moments, théorème de Stieltjes.
9. Loi du zéro-un de Kolmogorov, loi du zéro-un de Hewitt – Savage.

Chapitre 2

Espérance conditionnelle

2.1 Introduction

Si X et Y sont deux v.a. discrètes définies sur un même espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$, on peut naturellement définir l'espérance conditionnelle de X sachant $\{Y = y\}$ comme la moyenne de la loi conditionnelle :

$$\mathbb{E}(X | Y = y) = \sum_x x \mathbb{P}(X = x | Y = y)$$

qui fait sens si X est positive ou intégrable et $\mathbb{P}(Y = y) > 0$. L'interprétation de la quantité déterministe $\mathbb{E}(X | Y = y)$ est une moyennisation de l'aléa résiduel dans X conditionnellement à $\{Y = y\}$. En notant $h(y) := \mathbb{E}(X | Y = y)$, on a

$$\mathbb{E}(X \mathbb{1}_{Y=y}) = \mathbb{E}(X | Y = y) \mathbb{P}(Y = y) = \mathbb{E}(h(Y) \mathbb{1}_{Y=y}), \quad \text{pour tout } y.$$

Comme Y est discrète, la tribu $\sigma(Y)$ est engendré par les atomes $\{Y = y\}$, d'où par linéarité

$$\mathbb{E}(X \mathbb{1}_B) = \mathbb{E}(h(Y) \mathbb{1}_B), \quad \text{pour tout } B \in \sigma(Y).$$

Ceci exprime le fait que la variable aléatoire $\mathbb{E}(X | Y) := h(Y)$ est une approximation $\sigma(Y)$ -mesurable de X , au sens faible. C'est ce qu'on appelle l'espérance conditionnelle de X sachant Y ou sachant $\sigma(Y)$. Comment la définir au-delà des variables discrètes? Quelles sont ses propriétés? C'est l'objet du reste de ce chapitre.

Rappelons que pour une v.a. $X \geq 0$, si $\mathbb{E}(X) = 0$ alors $X = 0$ p.s.

Rappelons également que les éléments de L^1 sont des classes d'équivalence de v.a. pour l'égalité p.s.

Lemme 2.1.1. Égalité au sens faible implique égalité presque sûre.

Si X et Y sont des v.a. positives ou intégrables, définies sur un même espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$, égales au sens faible : $\mathbb{E}(XZ) = \mathbb{E}(YZ)$ pour toute v.a. $Z = \mathbb{1}_A$, $A \in \mathcal{A}$, alors $X = Y$ p.s.

Démonstration. Le choix $A = \{X \geq Y\} \in \mathcal{A}$, donne $0 = \mathbb{E}((X - Y)Z) = \mathbb{E}((X - Y)^+)$, d'où $(X - Y)^+ = 0$ p.s. d'où $X \leq Y$ p.s., tandis que le choix $A = \{X \leq Y\}$ donne $(X - Y)^- = 0$ p.s. d'où $X \geq Y$ p.s. ■

L'espérance conditionnelle émerge en se restreignant aux Z mesurables pour une sous-tribu \mathcal{B} .

Théodéf 2.1.2. Espérance conditionnelle.

Soit $X : \Omega \rightarrow \mathbb{R}$ une v.a. positive (respectivement intégrable) sur $(\Omega, \mathcal{A}, \mathbb{P})$, et soit \mathcal{B} une sous-tribu de \mathcal{A} . Alors il existe une unique v.a. Y positive (respectivement intégrable) vérifiant les deux propriétés suivantes :

- (mesurabilité) Y est \mathcal{B} -mesurable.
- (approximation) $\mathbb{E}(XZ) = \mathbb{E}(YZ)$ pour toute v.a. $Z = \mathbb{1}_B$, $B \in \mathcal{B}$

Cette v.a. Y , notée $\mathbb{E}(X | \mathcal{B})$, est appelée espérance conditionnelle de X sachant \mathcal{B} .

Ce théorème exprime le fait que $\mathbb{E}(X | \mathcal{B})$ est la meilleure approximation, au sens faible, de X par une v.a. \mathcal{B} -mesurable, au p.s. près. En particulier, si X est \mathcal{B} -mesurable, alors $\mathbb{E}(X | \mathcal{B}) = X$ p.s.

Lorsque $Z = 1$, on obtient $\mathbb{E}(\mathbb{E}(X | \mathcal{B})) = \mathbb{E}(X)$. En particulier $\mathbb{E}(\mathbb{E}(\mathbb{1}_A | \mathcal{B})) = \mathbb{P}(A)$ pour tout $A \in \mathcal{A}$. Cela rend naturel de poser $\mathbb{P}(A | \mathcal{B}) := \mathbb{E}(\mathbb{1}_A | \mathcal{B})$, qui est une v.a. vérifiant $\mathbb{E}(\mathbb{P}(A | \mathcal{B})) = \mathbb{P}(A)$.

Le lemme 2.1.1 fournit immédiatement l'unicité p.s. de l'espérance conditionnelle. Pour l'existence, c'est-à-dire la construction, nous abordons d'abord le cas où \mathcal{B} est discrète, déjà subtil bien qu'élémentaire, puis le cas où X est de carré intégrable, agréablement lié à une notion de projection orthogonale, puis le cas général.

2.2 Construction lorsque la tribu est discrète

On se place dans le cas où \mathcal{B} est discrète au sens où elle est engendrée par un système complet d'événements $(B_n)_{n \in \mathbb{N}} : B_n \in \mathcal{A}$ pour tout n , $B_n \cap B_m = \emptyset$ si $n \neq m$, et $\sum_n \mathbb{P}(B_n) = 1$. L'ensemble $B_\infty := \Omega \setminus \bigcup_n B_n$ vérifie $\mathbb{P}(B_\infty) = 0$, et $(B_n)_{n \in \mathbb{N} \cup \{\infty\}}$ est une partition dénombrable de Ω . En particulier on a

$$\mathcal{B} = \sigma((B_n)_{n \in \mathbb{N}}) = \{\bigcup_{i \in I} B_i : I \subset \mathbb{N} \cup \{\infty\}\}.$$

Si Y est une variable aléatoire discrète prenant les valeurs $(y_n)_{n \in \mathbb{N}}$ alors les $B_n := \{Y = y_n\}$ forment un système complet d'événements de la tribu $\sigma(Y)$ engendrée par Y , qui est donc discrète. Le lemme suivant affirme la réciproque : les variables aléatoires mesurables par rapport à une tribu discrète sont discrètes.

Lemme 2.2.1. Variables aléatoires mesurables pour une tribu discrète.

Si \mathcal{B} est une sous-tribu de \mathcal{A} discrète engendrée par un système complet d'événements $(B_n)_{n \in \mathbb{N}}$, et si Y est une variable aléatoire \mathcal{B} -mesurable alors elle est constante sur les B_n c'est-à-dire que $Y = \sum_n y_n \mathbb{1}_{B_n}$ p.s.

Pour une variable de Bernoulli, le système complet est de la forme $\{B, B^c\}$ et $\mathcal{B} = \{\emptyset, B, B^c, \Omega\}$.

Démonstration. Supposons par l'absurde qu'il existe $n \in \mathbb{N}$ et $\omega, \omega' \in B_n$ tels que $Y(\omega) \neq Y(\omega')$. Soit C et C' deux boréliens disjoints de \mathbb{R} , par exemple des intervalles, contenant respectivement $Y(\omega)$ et $Y(\omega')$. Comme Y est \mathcal{B} -mesurable, on a $Y^{-1}(C) \in \mathcal{B}$ et $Y^{-1}(C') \in \mathcal{B}$, donc $Y^{-1}(C) = \bigcup_{m \in I} B_m$ et $Y^{-1}(C') = \bigcup_{m \in I'} B_m$, avec $I, I' \subset \mathbb{N} \cup \{\infty\}$. Nécessairement $n \in I \cap I'$ car $\omega, \omega' \in B_n$, or $\emptyset \neq B_n \subset Y^{-1}(C) \cap Y^{-1}(C') = Y^{-1}(C \cap C') = Y^{-1}(\emptyset) = \emptyset$. ■

Continuons avec notre sous-tribu discrète \mathcal{B} générée par $(B_n)_{n \in \mathbb{N}}$. Pour tout $A \in \mathcal{A}$, on a

$$\mathbb{P}(A) = \sum_{n: \mathbb{P}(B_n) > 0} \mathbb{P}(A | B_n) \mathbb{P}(B_n).$$

Il est alors naturel de considérer la variable aléatoire \mathcal{B} -mesurable, de moyenne $\mathbb{P}(A)$, suivante :

$$\mathbb{P}(A | \mathcal{B}) := \sum_{n: \mathbb{P}(B_n) > 0} \mathbb{P}(A | B_n) \mathbb{1}_{B_n} = \sum_{n: \mathbb{P}(B_n) > 0} \mathbb{E}\left(\mathbb{1}_A \frac{\mathbb{1}_{B_n}}{\mathbb{P}(B_n)}\right) \mathbb{1}_{B_n}.$$

Plus généralement, pour une v.a. X positive ou intégrable, on définit la v.a. \mathcal{B} -mesurable

$$\mathbb{E}(X | \mathcal{B}) := \sum_{n: \mathbb{P}(B_n) > 0} \left(\frac{1}{\mathbb{P}(B_n)} \int_{B_n} X d\mathbb{P}\right) \mathbb{1}_{B_n} = \sum_{n: \mathbb{P}(B_n) > 0} \mathbb{E}\left(X \frac{\mathbb{1}_{B_n}}{\mathbb{P}(B_n)}\right) \mathbb{1}_{B_n},$$

de sorte que $\mathbb{E}(\mathbb{1}_A | \mathcal{B}) = \mathbb{P}(A | \mathcal{B})$. Cette variable aléatoire $\mathbb{E}(X | \mathcal{B})$ s'obtient, sur chaque atome B_n , en remplaçant X par sa moyenne sur B_n , ce qui correspond bien à l'idée intuitive de meilleure approximation \mathcal{B} -mesurable.

Avec cette définition, la v.a. $\mathbb{E}(X | \mathcal{B})$ est constante sur chaque atome B_n de \mathcal{B} , elle est donc \mathcal{B} -mesurable d'après le lemme 2.2.1. Pour vérifier la propriété d'approximation de l'espérance conditionnelle, il suffit par linéarité de le faire pour une v.a. $Z = \mathbb{1}_{B_n}$ avec $\mathbb{P}(B_n) > 0$, ce qui donne, grâce au fait que les B_n sont deux à deux disjoints :

$$\mathbb{E}(\mathbb{E}(X | \mathcal{B}) Z) = \int_{B_n} X d\mathbb{P} = \mathbb{E}(X Z).$$

Notons que si X est \mathcal{B} -mesurable, alors elle est constante sur chaque B_n , et on a bien $\mathbb{E}(X | \mathcal{B}) = X$.

2.3 Construction pour les variables de carré intégrable

Soit $X \in H := L^2(\Omega, \mathcal{A}, \mathbb{P})$, et \mathcal{B} une sous-tribu quelconque de \mathcal{A} . L'espace de Hilbert $F := L^2(\Omega, \mathcal{B}, \mathbb{P})$ est un sous-espace vectoriel fermé (car complet) de l'espace de Hilbert H . En particulier, la projection orthogonale $Y := p_F(X)$ de X sur F existe et est unique, caractérisée par les deux propriétés suivantes :

- $Y \in F$
- $\langle X, Y \rangle = \langle X, Z \rangle$, c'est-à-dire $\mathbb{E}(XY) = \mathbb{E}(XZ)$, pour tout $Z \in F$.

Alternativement, on peut appliquer le théorème de représentation de Riesz à la forme linéaire $Z \in F \mapsto \langle X, Z \rangle$.

Comme F contient les v.a. \mathcal{B} -mesurables bornées, nous venons de construire $\mathbb{E}(X | \mathcal{B})$, géométriquement! La meilleure approximation au sens faible est réalisée dans ce cadre hilbertien par une projection orthogonale. Plus précisément, la meilleure approximation est ici au sens des moindres carrés. En résumé, si $X \in L^2$ alors

$$\mathbb{E}(X | \mathcal{B}) := \arg \min_{Y \in L^2(\Omega, \mathcal{B}, \mathbb{P})} \mathbb{E}((X - Y)^2).$$

On a bien que $\mathbb{E}(X | \mathcal{B}) = X$ si X est \mathcal{B} -mesurable, car $p_F(X) = X$ si $X \in F$.

D'autre part, le conditionnement conserve la positivité : si $X \geq 0$ alors $\mathbb{E}(X | \mathcal{B}) = Y \geq 0$ p.s. En effet, avec $Z = \mathbb{1}_B$ où $B = \{Y \leq 0\}$, il vient $0 \leq \mathbb{E}(X \mathbb{1}_B) = \mathbb{E}(Y \mathbb{1}_B) \leq 0$, d'où $\mathbb{E}(Y \mathbb{1}_B) = 0$, d'où $-Y \mathbb{1}_B = 0$ p.s.

2.4 Construction pour les variables positives

Soit X une v.a. positive sur $(\Omega, \mathcal{A}, \mathbb{P})$, prenant ses valeurs dans $[0, \infty]$, et \mathcal{B} une sous-tribu de \mathcal{A} .

Pour tout $n \geq 0$, on introduit la v.a. bornée $X_n := \min(X, n) = X \wedge n$, qui est en particulier de carré intégrable. On sait alors définir $Y_n := \mathbb{E}(X_n | \mathcal{B})$ par projection comme précédemment. À présent, grâce à la linéarité de la projection, au fait que $X_{n+1} - X_n \geq 0$, et à la conservation de la positivité par conditionnement,

$$Y_{n+1} - Y_n = p_{L^2(\Omega, \mathcal{B}, \mathbb{P})}(X_{n+1} - X_n) = \mathbb{E}(X_{n+1} - X_n | \mathcal{B}) \geq 0.$$

Donc $(Y_n)_{n \in \mathbb{N}}$ est positive et croissante p.s. Par convergence monotone sur $(\Omega, \mathcal{B}, \mathbb{P})$, la v.a. $Y := \lim_{n \rightarrow \infty} Y_n$ existe. Elle est \mathcal{B} mesurable et positive. Pour s'assurer que Y est bien l'espérance conditionnelle recherchée, $\mathbb{E}(X | \mathcal{B})$, il reste à vérifier la propriété d'approximation. Or en combinant le théorème de convergence monotone et la propriété d'approximation pour $Y_n = \mathbb{E}(X_n | \mathcal{B})$, on obtient, pour tout $B \in \mathcal{B}$,

$$\mathbb{E}(Y \mathbb{1}_B) = \mathbb{E}(\lim_{n \rightarrow \infty} (Y_n \mathbb{1}_B)) = \lim_{n \rightarrow \infty} \mathbb{E}(Y_n \mathbb{1}_B) = \lim_{n \rightarrow \infty} \mathbb{E}(X_n \mathbb{1}_B) = \mathbb{E}(\lim_{n \rightarrow \infty} (X_n \mathbb{1}_B)) = \mathbb{E}(X \mathbb{1}_B).$$

De plus, si $X \geq 0$ est intégrable, alors, en prenant $B = \Omega$, il vient que $\mathbb{E}(Y) = \mathbb{E}(X) < \infty$ donc Y est intégrable.

En résumé, si X prend ses valeurs dans $[0, \infty]$ alors

$$\mathbb{E}(X | \mathcal{B}) := \lim_{n \rightarrow \infty} \mathbb{E}(X \wedge n | \mathcal{B}).$$

2.5 Construction pour les variables intégrables

Soit $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ et \mathcal{B} sous-tribu de \mathcal{A} . On se ramène au cas positif intégrable via la décomposition habituelle $X = X^+ - X^-$ où $X^\pm := \max(\pm X, 0) \in L^1(\Omega, \mathcal{A}, \mathbb{P})$, en posant

$$\mathbb{E}(X | \mathcal{B}) := \mathbb{E}(X^+ | \mathcal{B}) - \mathbb{E}(X^- | \mathcal{B}),$$

qui hérite bien des propriétés de l'espérance conditionnelle (mesurabilité, intégrabilité, approximation).

Remarque 2.5.1. Presque sûrement sans complications.

L'espérance conditionnelle est définie presque sûrement. En général, cela ne pose pas de problème, et peut même être passé sous silence. Mais il faut faire preuve de prudence, car il arrive que cela soit source de complications, comme par exemple pour l'inégalité de Jensen pour l'espérance conditionnelle ci-dessous.

2.6 Propriétés de l'espérance conditionnelle

La construction de l'espérance conditionnelle est semblable à la construction de l'espérance. Elle en vérifie les mêmes propriétés (linéarité, normalisation, positivité, théorèmes limites, inégalité de Jensen, etc.), ainsi que des propriétés spécifiques liées à l'approximation (ou projection dans le cadre L^2). En réalité, l'espérance n'est qu'une espérance conditionnelle spéciale, par rapport à la tribu triviale engendrée par les v.a. constantes!

Ci-dessous, et sauf mention du contraire, les v.a. sont sur $(\Omega, \mathcal{A}, \mathbb{P})$ et \mathcal{B}, \mathcal{G} sont des sous-tribus de \mathcal{A} .

En pratique, comme l'espérance conditionnelle est unique p.s., pour s'assurer qu'on a bien affaire à elle, il suffit de vérifier les propriétés qui la caractérisent : mesurabilité et approximation, ainsi que positivité ou intégrabilité selon le cas. Aussi, ci-dessous, rien n'est dit sur la preuve d'un item lorsque qu'elle découle d'une vérification directe et sans surprise des propriétés caractéristiques de l'espérance conditionnelle.

La tribu engendrée par les constantes est $\{\emptyset, \Omega\}$, les constantes sont mesurables par rapport à toute tribu. En revanche la tribu engendrée par une v.a. égale p.s. à une constante peut être plus grosse que $\{\emptyset, \Omega\}$.

1. *Égalité.* Si X_1 et X_2 sont positives (ou intégrables) et $X_1 = X_2$ p.s. alors $\mathbb{E}(X_1 | \mathcal{B}) = \mathbb{E}(X_2 | \mathcal{B})$ p.s.
Un point clé est que pour tout $B \in \mathcal{B}$, comme $X_1 \mathbb{1}_B = X_2 \mathbb{1}_B$ p.s., il vient que $\mathbb{E}(X_1 \mathbb{1}_B) = \mathbb{E}(X_2 \mathbb{1}_B)$.
2. *Normalisation.* $\mathbb{E}(1 | \mathcal{B}) = 1$ p.s., et plus généralement $\mathbb{E}(c | \mathcal{B}) = c$ p.s. pour toute constante c .
3. *Positivité.* Si $X \geq 0$ p.s. alors $\mathbb{E}(X | \mathcal{B}) \geq 0$ p.s. Il en découle par linéarité que si X_1 et X_2 sont intégrables et vérifient $X_1 \leq X_2$ p.s. alors $\mathbb{E}(X_1 | \mathcal{B}) \leq \mathbb{E}(X_2 | \mathcal{B})$ p.s.
4. *Linéarité.* L'application $X \in L^1(\Omega, \mathcal{A}, \mathbb{P}) \mapsto \mathbb{E}(X | \mathcal{B}) \in L^1(\Omega, \mathcal{B}, \mathbb{P})$ est linéaire : si X_1 et X_2 sont intégrables et λ_1, λ_2 sont des constantes, alors

$$\mathbb{E}(\lambda_1 X_1 + \lambda_2 X_2 | \mathcal{B}) = \lambda_1 \mathbb{E}(X_1 | \mathcal{B}) + \lambda_2 \mathbb{E}(X_2 | \mathcal{B}) \quad \text{p.s.}$$

De plus cette propriété subsiste au-delà de l'intégrabilité, lorsque $X_1, X_2, \lambda_1, \lambda_2$ sont toutes positives.

5. *Convergence monotone.* Si $0 \leq X_n \nearrow X$ p.s. alors $\mathbb{E}(X_n | \mathcal{B}) \nearrow \mathbb{E}(X | \mathcal{B})$ p.s.
En effet, pour tout $B \in \mathcal{B}$, $0 \leq X_n \mathbb{1}_B \nearrow X \mathbb{1}_B$ p.s., et la propriété d'approximation pour $\mathbb{E}(X | \mathcal{B})$ est alors héritée de celle de $\mathbb{E}(X_n | \mathcal{B})$ par convergence monotone pour l'espérance.
6. *Lemme de Fatou.* Exercice!
7. *Convergence dominée.* Si $X_n \in L^1 \rightarrow X$ p.s., $\sup_n |X_n| \leq Y$ avec $Y \in L^1$, alors $\mathbb{E}(X | \mathcal{B}) = \lim_{n \rightarrow \infty} \mathbb{E}(X_n | \mathcal{B})$ p.s.
En effet, tout d'abord, le fait que $X \in L^1$ n'a rien à voir avec le conditionnement et découle du lemme de Fatou car $\mathbb{E}(|X|) = \mathbb{E}(\liminf_{n \rightarrow \infty} |X_n|) \leq \liminf_{n \rightarrow \infty} \mathbb{E}(|X_n|) \leq \mathbb{E}(Y) < \infty$. D'autre part, pour tout $B \in \mathcal{B}$, $X_n \mathbb{1}_B \rightarrow X \mathbb{1}_B$ p.s. et $|X_n \mathbb{1}_B| \leq Y$, et la propriété d'approximation pour $\mathbb{E}(X | \mathcal{B})$ est alors héritée de celle de $\mathbb{E}(X_n | \mathcal{B})$ par convergence dominée pour l'espérance.
8. *Inégalité de Jensen.* Si φ est une fonction réelle convexe et si X et $\varphi(X)$ sont intégrables alors

$$\varphi(\mathbb{E}(X | \mathcal{B})) \leq \mathbb{E}(\varphi(X) | \mathcal{B}) \quad \text{p.s.}$$

En particulier : $|\mathbb{E}(X | \mathcal{B})| \leq \mathbb{E}(|X| | \mathcal{B})$ p.s. et $\mathbb{E}(X | \mathcal{B})^2 \leq \mathbb{E}(X^2 | \mathcal{B})$ p.s.

La preuve est semblable à celle de l'inégalité de Jensen pour l'espérance, à une subtilité p.s. près. Plus précisément, l'égalité est atteinte si φ est affine, tandis que dans le cas général, la convexité de φ fait qu'elle est enveloppe d'une famille au plus dénombrable de fonctions affines : il existe $D \subset \mathbb{R}^2$ au plus dénombrable tel que $\varphi(x) = \sup_{(a,b) \in D} (ax + b)$ pour tout $x \in \mathbb{R}$. On a alors $\varphi(X) \geq aX + b$ pour tout $(a, b) \in D$, et donc, pour tout $B \in \mathcal{B}$, grâce aux propriétés d'approximation de $\mathbb{E}(X | \mathcal{B})$ et $\mathbb{E}(\varphi(X) | \mathcal{B})$,

$$\mathbb{E}(\mathbb{E}(\varphi(X) | \mathcal{B}) \mathbb{1}_B) = \mathbb{E}(\varphi(X) \mathbb{1}_B) \geq \mathbb{E}((aX + b) \mathbb{1}_B) = \mathbb{E}((a\mathbb{E}(X | \mathcal{B}) + b) \mathbb{1}_B).$$

Par conséquent $\mathbb{E}(\varphi(X) | \mathcal{B}) \geq a\mathbb{E}(X | \mathcal{B}) + b$ p.s. et ce p.s. dépend de (a, b) . Ensuite, comme $(a, b) \in D$ et D est au plus dénombrable, on obtient $\mathbb{E}(\varphi(X) | \mathcal{B}) \geq \sup_{(a,b) \in D} (a\mathbb{E}(X | \mathcal{B}) + b) = \varphi(\mathbb{E}(X | \mathcal{B}))$ p.s.

9. *Inégalité de Markov pour la probabilité conditionnelle.* Si $X \in L^1$ est positive alors pour tout $x > 0$,

$$\mathbb{P}(X \geq x | \mathcal{F}) = \mathbb{E}(\mathbb{1}_{\{X \geq x\}} | \mathcal{F}) \leq \frac{\mathbb{E}(X | \mathcal{F})}{x}.$$

10. *Inégalité de Cauchy-Schwarz et de Hölder pour l'espérance conditionnelle.* Exercice!

3 → 4

Voici enfin une liste de propriétés plus spécifiques au caractère conditionnel de l'espérance conditionnelle.

1. *Projection.* Si X est \mathcal{B} -mesurable positive ou intégrable, alors $\mathbb{E}(X | \mathcal{B}) = X$ p.s.
C'est le cas notamment lorsque X est constante et lorsque $X = \mathbb{1}_B$, $B \in \mathcal{B}$.
2. *Approximation.* Si X est positive (respectivement intégrable), alors $\mathbb{E}(XZ) = \mathbb{E}(\mathbb{E}(X | \mathcal{B})Z)$ pour toute v.a. Z \mathcal{B} -mesurable positive (respectivement bornée). S'obtient par approximation par des fonctions étagées, linéarité de l'espérance, et convergence monotone (respectivement dominée).
3. *Moyenne.* Si X est intégrable alors $\mathbb{E}(\mathbb{E}(X | \mathcal{B})) = \mathbb{E}(X)$.
Découle de la propriété d'approximation avec $B = \Omega \in \mathcal{B}$ car alors $\mathbb{1}_B = 1$.
4. *Espérance classique.* Si $\mathcal{B} = \{\emptyset, \Omega\}$ et si X est intégrable alors $\mathbb{E}(X | \mathcal{B}) = \mathbb{E}(X)$ p.s.
On retrouve l'idée que la meilleure approximation de X par une constante est son espérance.
5. *Composition ou emboîtement.* Si X est positive ou intégrable et si $\mathcal{B} \subset \mathcal{G} \subset \mathcal{A}$, alors

$$\mathbb{E}(\mathbb{E}(X | \mathcal{G}) | \mathcal{B}) = \mathbb{E}(\mathbb{E}(X | \mathcal{B}) | \mathcal{G}) = \mathbb{E}(X | \mathcal{B}) \quad \text{p.s.}$$

Propriété naturelle du point de vue de la meilleure approximation, ou de la projection (cas L^2).
On retrouve l'item précédent en considérant le cas particulier $\mathcal{G} = \{\emptyset, \Omega\}$.

6. *Indépendance.* Si X est intégrable et indépendante de \mathcal{B} alors $\mathbb{E}(X | \mathcal{B}) = \mathbb{E}(X)$ p.s.
En effet, pour tout $B \in \mathcal{B}$, l'indépendance donne $\mathbb{E}(X \mathbb{1}_B) = \mathbb{E}(X) \mathbb{E}(\mathbb{1}_B) = \mathbb{E}(\mathbb{E}(X) \mathbb{1}_B)$.
L'item *Moyenne* précédent est un cas particulier car toute v.a. est indépendante de $\{\emptyset, \Omega\}$.
7. *Extraction du mesurable.* Si Z est \mathcal{B} -mesurable et X, XZ intégrables ou bien X, Z positives, alors

$$\mathbb{E}(XZ | \mathcal{B}) = Z \mathbb{E}(X | \mathcal{B}) \quad \text{p.s.}$$

(lorsque $X = 1$ on retrouve comme cas particulier l'item *Projection*).

En effet, la propriété d'approximation se vérifie tout d'abord directement pour $Z = \mathbb{1}_B$, $B \in \mathcal{B}$, puis par linéarité pour les Z étagées, puis par convergence monotone pour les Z positives, après s'être ramené à X positive via $X = X^+ - X^-$, puis par $Z^+ - Z^-$ pour le cas général.

8. *Aléa résiduel après moyennisation de l'aléa indépendant.* Si X et Y sont deux v.a. à valeurs dans des espaces mesurables E et F respectivement, avec X indépendante de \mathcal{B} et Y mesurable par rapport à \mathcal{B} , et si $f : E \times F \rightarrow \mathbb{R}$ est une fonction mesurable positive (resp. telle que $f(X, Y)$ est intégrable), alors

$$\mathbb{E}(f(X, Y) | \mathcal{B}) = g(Y) \quad \text{p.s. où } g(y) := \mathbb{E}(f(X, y)).$$

En effet, pour toute v.a. Z \mathcal{B} -mesurable positive (resp. bornée), grâce au théorème de Fubini – Tonelli,

$$\begin{aligned} \mathbb{E}(f(X, Y) Z) &= \int f(x, y) z d\mathbb{P}_{X, Y, Z}(x, y, z) \\ &= \int f(x, y) z d(\mathbb{P}_X \otimes \mathbb{P}_{Y, Z})(x, y, z) \\ &= \int \left(\int f(x, y) d\mathbb{P}_X(x) \right) z d\mathbb{P}_{Y, Z}(y, z) \\ &= \mathbb{E}(g(Y) Z). \end{aligned}$$

Cette propriété donne une nouvelle interprétation intuitive de l'espérance conditionnelle, à base d'information plutôt que d'analyse (approximation) ou de géométrie (projection) : la variable $f(X, Y)$ est constituée de deux sources d'aléa, un aléa mesurable Y et un aléa indépendant X , et son espérance conditionnelle est donnée par l'aléa résiduel après moyennisation de l'aléa indépendant!

9. *Variance.* Soit X de carré intégrable. Notons $H := L^2(\Omega, \mathcal{A}, \mathbb{P})$, $F := L^2(\Omega, \mathcal{B}, \mathbb{P})$, et $F_0 := L^2(\Omega, \{\emptyset, \Omega\}, \mathbb{P})$ l'ensemble des v.a. constantes¹. Par construction, $\mathbb{E}(X | \mathcal{B}) = p_F(X)$ est le projeté orthogonal de X sur F , et

$$\begin{aligned} \text{dist}(X, F)^2 &:= \min_{Y \in F} \mathbb{E}((X - Y)^2) \\ &= \mathbb{E}((X - \mathbb{E}(X | \mathcal{B}))^2) \\ &= \mathbb{E}(X^2) - 2\mathbb{E}(X\mathbb{E}(X | \mathcal{B})) + \mathbb{E}(\mathbb{E}(X | \mathcal{B})^2) \\ &= \mathbb{E}(\text{Var}(X | \mathcal{B})) \quad \text{où } \text{Var}(X | \mathcal{B}) := \mathbb{E}(X^2 | \mathcal{B}) - \mathbb{E}(X | \mathcal{B})^2. \end{aligned}$$

Dans le cas particulier extrême de F_0 , on a $p_{F_0}(X) = \mathbb{E}(X)$ et la formule variationnelle devient

$$\text{dist}(X, F_0)^2 = \min_{c \in \mathbb{R}} \mathbb{E}((X - c)^2) = \mathbb{E}((X - \mathbb{E}(X))^2) = \text{Var}(X).$$

Comme $F_0 \subset F$, le théorème de Pythagore donne la décomposition de la variance suivante :

$$\text{Var}(X) = \text{dist}(X, F_0)^2 = \text{dist}(X, F)^2 + \text{dist}(p_F(X), F_0)^2 = \mathbb{E}(\text{Var}(X | \mathcal{B})) + \text{Var}(\mathbb{E}(X | \mathcal{B})),$$

qui se lit espérance de la variance conditionnelle plus variance de l'espérance conditionnelle.

2.7 Espérance conditionnelle par rapport à une variable aléatoire

Si X est une v.a. positive ou intégrable, et si Y est une v.a. alors on définit la variable aléatoire

$$\mathbb{E}(X | Y) := \mathbb{E}(X | \sigma(Y))$$

où $\sigma(Y)$ sa tribu engendrée par Y . Or le lemme de Doob affirme que comme $\mathbb{E}(X | Y)$ est mesurable par rapport à la sous-tribu $\sigma(Y)$, il existe h borélienne telle que $\mathbb{E}(X | Y) = h(Y)$. Cela conduit à la notation

$$\mathbb{E}(X | Y = y) := h(y).$$

Ce n'est qu'une notation. On note également $\mathbb{P}(A | Y = y) := h(y)$ lorsque $X = \mathbb{1}_A$, $A \in \mathcal{A}$.

Si Y est discrète et prend ses valeurs dans $(y_n)_{n \in \mathbb{N}}$, alors, en notant $B_n := \{Y = y_n\}$, on sait que

$$\mathbb{E}(X | Y) = \sum_n \left(\frac{1}{\mathbb{P}(B_n)} \int_{B_n} X d\mathbb{P} \right) \mathbb{1}_{B_n} =: h(Y).$$

Dans ce cas, comme Y prend par définition la valeur y_n sur B_n ,

$$h(y) = \sum_n \left(\frac{1}{\mathbb{P}(Y = y_n)} \int_{\{Y = y_n\}} X d\mathbb{P} \right) \mathbb{1}_{y=y_n}.$$

1. Pour les v.a. constantes p.s. il faut faire grossir la tribu $\{\emptyset, \Omega\}$ avec tous les éléments de \mathcal{A} de probabilité nulle.

En particulier, lorsque $X = \mathbb{1}_A$, $A \in \mathcal{A}$, cela donne

$$h(y) = \sum_n \mathbb{P}(A | Y = y_n) \mathbb{1}_{y=y_n}, \quad \text{de sorte que} \quad h(y_n) = \mathbb{P}(A | Y = y_n).$$

Mais cette formule ne fait plus sens lorsque Y n'est pas discrète, car $\mathbb{P}(Y = y)$ peut alors s'annuler partout.

Quittons l'univers des variables aléatoires discrètes, et considérons à présent le cas où le couple (X, Y) admet une densité f par rapport à la mesure de Lebesgue sur \mathbb{R}^2 , et X est intégrable. Vérifions alors que

$$h(y) = \frac{\int x f(x, y) dx}{\int f(x, y) dx}.$$

Le dénominateur est la densité marginale, de Y . Il suffit pour cela de vérifier que

$$\mathbb{E}(X | Y) = h(Y) = \frac{\int x f(x, Y) dx}{\int f(x, Y) dx}.$$

En effet, la v.a. $h(Y)$ est bien $\sigma(Y)$ -mesurable, et pour tout $B \in \sigma(Y)$, il existe un borélien C tel que $B = Y^{-1}(C)$, et donc, grâce au lemme de transfert, au théorème de Fubini – Tonelli, et à la définition de h ,

$$\begin{aligned} \mathbb{E}(h(Y) \mathbb{1}_B) &= \mathbb{E}(h(Y) \mathbb{1}_C(Y)) = \iint h(y) \mathbb{1}_C(y) f(x, y) dx dy \\ &= \int h(y) \mathbb{1}_C(y) \left(\int f(x, y) dx \right) dy \\ &= \int \left(\int x f(x, y) dx \right) \mathbb{1}_C(y) dy \\ &= \iint x \mathbb{1}_C(y) f(x, y) dx dy \\ &= \mathbb{E}(X \mathbb{1}_C(Y)) = \mathbb{E}(X \mathbb{1}_B). \end{aligned}$$

2.8 Noyau conditionnel et loi conditionnelle

Poursuivons dans le même esprit. Si $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ est borélienne bornée, un raisonnement similaire donne

$$\mathbb{E}(\varphi(X) | Y) = h_\varphi(Y) \quad \text{où} \quad h_\varphi(y) := \frac{\int \varphi(x) f(x, y) dx}{\int f(x, y) dx} =: \mathbb{E}(\varphi(X) | Y = y).$$

C'est une explicitation du lemme de Doob. On a également

$$\mathbb{E}(\varphi(X) | Y = y) = h_\varphi(y) = \int \varphi(x) K_y(dx) \quad \text{où} \quad K_y(dx) := \frac{f(x, y)}{\int f(x, y) dx} dx.$$

On dit que K est un noyau conditionnel, d'où la notation K , on aurait pu utiliser la lettre ν également². Pour presque tout y , K_y est une mesure de probabilité sur \mathbb{R} . En particulier, si $\varphi = \mathbb{1}_C$ avec C borélien, alors

$$\mathbb{P}(X \in C | Y = y) = \mathbb{E}(\mathbb{1}_C(X) | Y = y) = \int_C K_y(dx) = K_y(C).$$

C'est une sorte de lemme du transfert pour la probabilité conditionnelle.

On dit que $K_y(\cdot)$ est la loi conditionnelle de X sachant $Y = y$, une sorte de mesure image de $\mathbb{P}(X \in \cdot | Y = y)$.

En notant, abusivement, $f_{(X, Y)}(x, y) := f(x, y)$ la densité du couple (X, Y) , $f_{(X)}(x) := \int f(x, y) dy$ et $f_{(Y)}(y) := \int f(x, y) dx$ les densités de ses lois marginales, et $f_{X|Y=y}(x)$ la densité de $K_y(dx)$, il vient que

$$f_{(X, Y)}(x, y) = f_{X|Y=y}(x) f_{(Y)}(y).$$

De plus X et Y sont indépendantes ssi $f_{X|Y=y} = f_{(X)}$ sur le support de $f_{(Y)}$.

2. K comme *kernel* et ν comme *noyau*.

Cela se généralise immédiatement au cas où (X, Y) est un vecteur aléatoire de $\mathbb{R}^d = \mathbb{R}^p \times \mathbb{R}^q$. Dans ce cas l'espérance conditionnelle $\mathbb{E}(X | Y)$ désigne le vecteur aléatoire de \mathbb{R}^p donné par $(\mathbb{E}(X_i | Y))_{1 \leq i \leq p}$.

Soit $(X, Y) : \Omega \rightarrow E \times F$ un couple de variables aléatoires défini sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$, où (E, \mathcal{E}) et (F, \mathcal{F}) sont des espaces mesurables. On appelle *noyau conditionnel* toute application

$$K : F \times \mathcal{E} \rightarrow [0, 1] \quad \text{telle que} \quad \begin{cases} y \mapsto K(y, C) \text{ est mesurable, pour tout } C \in \mathcal{E} \\ C \mapsto K(y, C) \text{ est une mesure de probabilité sur } (E, \mathcal{E}), \text{ pour tout } y \in F \end{cases}$$

Dans le cas où $E = F$, on parle de *noyau de transition*, utile pour les chaînes de Markov du chapitre 5.

On utilise également les notations naturelles $K_Y(C) := K(y, C)$ et $K_Y(\varphi) := \int \varphi(x) K_Y(dx)$.

Si φ est mesurable, positive ou bornée, alors $y \mapsto K_Y(\varphi)$ est mesurable.

Théodéf 2.8.1. Loi conditionnelle au-delà du cas à densité.

Soit $(X, Y) : \Omega \rightarrow E \times F$ un couple de variable aléatoires défini sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$, où (E, \mathcal{E}) et (F, \mathcal{F}) sont des espaces mesurables boréliens polonais^a. Alors il existe un unique *noyau conditionnel* $K : F \times \mathcal{E} \rightarrow [0, 1]$ tel que pour tout $y \in F$ et toute fonction $\varphi : E \rightarrow \mathbb{R}$ mesurable, positive ou bornée,

$$\mathbb{E}(\varphi(X) | Y) = K_Y(\varphi) \quad \text{p.s.,} \quad \text{autrement dit} \quad \mathbb{E}(\varphi(X) | Y = y) = \int_E \varphi(x) K_Y(dx).$$

La loi conditionnelle de X sachant $Y = y$ est définie par $\text{Loi}(X | Y = y) = K_Y(\cdot)$.

De plus, lorsque $E = \mathbb{R}^n$, sa moyenne est $\mathbb{E}(X | Y = y)$.

a. Un espace topologique est polonais lorsqu'il est séparable et qu'il est métrisé par une distance pour lequel il est complet.

Ce théorème est un résultat de théorie de la mesure, abordé par exemple dans [17, III.70–74]³. Dans ces notes, nous ne faisons intervenir que des instances plus élémentaires et explicites, soit discrètes, soit à densité.

La donnée du noyau conditionnel K et de la loi de Y suffit à déterminer la loi du couple (X, Y) , et en particulier celle de X qui est sa première loi marginale. En effet, il suffit de tester le couple (X, Y) sur des fonctions test produit mesurables, bornées ou positives, de la forme $(x, y) \mapsto \varphi(x)\psi(y)$, ce qui donne

$$\mathbb{E}(\varphi(X)\psi(Y)) = \mathbb{E}(\mathbb{E}(\varphi(X)\psi(Y) | Y)) = \mathbb{E}(\mathbb{E}(\varphi(X) | Y)\psi(Y)) = \mathbb{E}(K_Y(\varphi)\psi(Y)).$$

Aussi X et Y sont *indépendantes* ssi $K_Y(\varphi) = \mathbb{E}(\varphi(X) | Y)$ est constante et égale p.s. à $\mathbb{E}(\varphi(Y))$ (qui est son espérance!) pour tout φ , c'est-à-dire que $\text{Loi}(X | Y = y)$ est p.s. *constante* en y et égale à la loi de X .

La formulation suivante du théorème 2.8.1 porte le nom de *désintégration de la mesure* :

$$\int \varphi(x)\psi(y)\mathbb{P}_{(X,Y)}(dx, dy) = \int \left(\int \varphi(x)K_Y(dx) \right) \psi(y)\mathbb{P}_{(Y)}(dy).$$

Remarque 2.8.2. Algorithme de simulation.

Du point de vue algorithmique ou *Monte Carlo*, pour *simuler une réalisation* du couple (X, Y) , il est possible de simuler d'abord une réalisation de Y , notée y , puis de simuler la loi conditionnelle $\text{Loi}(X | Y = y)$.

Remarque 2.8.3. Conditionnement par rapport à un événement de probabilité nulle.

Lorsque $\mathbb{P}(Y = y) = 0$, on peut chercher à approcher $\mathbb{E}(\varphi(X) | Y = y)$ par $\lim_{n \rightarrow \infty} \mathbb{E}(\varphi(X) | Y \in C_n)$ où $\mathbb{P}(Y \in C_n) > 0$ et $\cap_n C_n = \{y\}$. Mais cette limite n'est pas universelle, elle dépend de la suite $(C_n)_{n \in \mathbb{N}}$ choisie. C'est le cas notamment dans le *paradoxe de Borel–Kolmogorov* sur $\text{Loi}(X | X \in C)$ avec X uniforme sur la sphère, et C grand cercle, limite de deux grossissements différents : latitude ou longitude. « *The concept of a conditional probability with regard to an isolated hypothesis whose probability equals 0 is inadmissible.* » – Andrey Kolmogorov. Plus généralement, il est naturel de considérer la notion de mesure trace sur un ensemble de mesure nulle, en utilisant un grossissement motivé par un problème, et d'en déduire une loi uniforme. La notion de *mesure de Hausdorff* donne aussi une réponse dans un contexte métrique. Cela permet de définir la loi uniforme sur des sous-variétés compactes (surfaces) de \mathbb{R}^n . Pour les variétés intrinsèques plutôt qu'extrinsèques, munies d'une métrique riemannienne, on dispose d'une mesure de volume,

3. « *The theorem of desintegration of measure has a bad reputation, and probabilists often try to avoid the use of conditional distributions... But it is really simple and easy to prove.* » – Claude Dellacherie et Paul-André Meyer [17, III.70].

jouant le rôle de mesure de Lebesgue ou de loi uniforme qd la variété est compacte et la mesure normalisée.

Pour aller plus loin 2.8.4.

- Notion d'indépendance conditionnelle.
- Construction du noyau conditionnel (désintégration de la mesure).
- Théorème de de Finetti pour les suites échangeables et inégalité de Diaconis – Freedman.
- Paradoxe de Borel – Kolmogorov, mesure de Hausdorff, mesure de surface, mesure de cône.

4 → 5

2.9 Motivation pour la suite : processus de Galton – Watson, processus auto-régressif

Le processus de Bienaymé – Galton – Watson⁴ décrit l'évolution d'une population asexuée, à générations séparées, dans laquelle chaque individu fait des enfants de manière i.i.d. par rapport aux autres. L'évolution s'écrit

$$Z_{n+1} = \sum_{k=1}^{Z_n} X_{n+1,k}$$

où Z_n est le nombre d'individus à la génération $n \in \mathbb{N}$, et où $(X_{n,k})_{n \geq 1, k \geq 1}$ sont des v.a. i.i.d. de même loi μ sur \mathbb{N} appelée loi de reproduction. La v.a. $X_{n+1,k}$ modélise le nombre aléatoire d'enfants de l'individu k de la génération n , c'est-à-dire sa contribution à la génération $n+1$. La suite $(Z_n)_{n \in \mathbb{N}}$ donne l'évolution de la taille d'un arbre aléatoire, et on dit que $(Z_n)_{n \in \mathbb{N}}$ est un processus de branchement. Il en existe moult variantes et généralisations. Supposons à présent que μ a une moyenne m et $Z_0 = 1$. L'équation de récurrence aléatoire précédente donne

$$\mathbb{E}(Z_{n+1} | Z_n = z) = zm,$$

d'où

$$\mathbb{E}(Z_{n+1}) = \mathbb{E}(\mathbb{E}(Z_{n+1} | Z_n)) = m\mathbb{E}(Z_n) = m^n \mathbb{E}(Z_0) = m^n \begin{cases} \xrightarrow{n \rightarrow \infty} \infty & \text{si } m > 1 \text{ (cas sur-critique)} \\ = 1 & \text{si } m = 1 \text{ (cas critique)} \\ \xrightarrow{n \rightarrow \infty} 0 & \text{si } m < 0 \text{ (cas sous-critique)} \end{cases}.$$

Au-delà du comportement en moyenne, l'étude de la trajectoire $n \mapsto Z_n$, de son explosion ou extinction, de son maximum, du temps d'extinction, etc., peut être menée en exploitant le concept de martingale et celui de chaîne de Markov. En particulier et à titre d'exemple, la suite $(\frac{Z_n}{m^n})_{n \in \mathbb{N}}$ vérifie, pour sa filtration canonique $(\mathcal{F}_n)_{n \in \mathbb{N}}$,

$$\mathbb{E}\left(\frac{Z_{n+1}}{m^{n+1}} \mid \mathcal{F}_n\right) = \frac{Z_n}{m^n},$$

propriété de projection typique de la structure martingale, tandis que la récurrence aléatoire

$$Z_{n+1} = f(Z_n, \varepsilon_{n+1}) \quad \text{où } \varepsilon_{n+1} := (X_{n+1,1}, X_{n+1,2}, \dots) \text{ est indépendante de } Z_0,$$

vérifiée par la suite $(Z_n)_{n \in \mathbb{N}}$, est typique de la structure chaîne de Markov.

Les structures martingale et chaîne de Markov sont également présentes dans le processus auto-régressif gaussien $(Y_n)_{n \in \mathbb{N}}$ vérifiant $Y_{n+1} = \theta Y_n + \varepsilon_{n+1}$ où $\theta \in \mathbb{R}$ et où cette fois-ci $(\varepsilon_n)_{n \in \mathbb{N}}$ est une suite de v.a. i.i.d. de loi $\mathcal{N}(0, I_d)$. Cela inclut la marche aléatoire à incréments gaussiens, dont la limite d'échelle est le mouvement brownien. Cela nous conduit également, dans un premier temps, à étudier les vecteurs gaussiens, qui font l'objet du chapitre 3.

4. Pour ce modèle, et bien d'autres, on pourra consulter par exemple [11] pour les mathématiques et [2] pour l'histoire.

Chapitre 3

Vecteurs aléatoires gaussiens



FIGURE 3.1 – Bibliothèque de l’Institut Mittag Leffler à Djursholm, près de Stockholm, Suède. L’univers des vecteurs gaussiens a notamment été exploré par Harald Cramér (1893 – 1985), mathématicien et actuaire suédois, professeur à l’université de Stockholm, qui s’est intéressé à la théorie des nombres, aux probabilités, et à la statistique.

Les vecteurs gaussiens, ou plutôt leurs lois appelées lois gaussiennes, apparaissent naturellement par le phénomène TLC. En particulier, elles sont incontournables pour la modélisation des erreurs et du bruit. Le mouvement Brownien, qui émerge comme limite d’échelle de marches aléatoires à pas de carré intégrable par le phénomène TLC, est un vecteur gaussien de dimension infinie, et sa loi, appelée mesure de Wiener, est une loi gaussienne sur l’espace des fonctions continues. Les lois gaussiennes sont également des objets riches et structurés, profondément reliés à l’analyse harmonique, à l’algèbre linéaire, et à la géométrie de la sphère de grande dimension. Elles permettent bien souvent des calculs explicites, y compris de nature combinatoire.

Si $X = (X_1, \dots, X_n)^\top$ est un vecteur colonne aléatoire de \mathbb{R}^n à composantes de carré intégrable, alors son vecteur moyenne et sa matrice de covariance sont donnés par

$$m = \mathbb{E}(X) = (\mathbb{E}(X_1), \dots, \mathbb{E}(X_n))^\top \quad \text{et} \quad K = (\mathbb{E}((X_j - m_j)(X_k - m_k)))_{1 \leq j, k \leq n} = \mathbb{E}((X - m)(X - m)^\top) = \mathbb{E}(XX^\top) - mm^\top.$$

- La matrice K est symétrique et semi-définie positive. C’est la moyenne de la matrice symétrique semi-définie positive aléatoire $(X - m)(X - m)^\top$ (qui est de rang 1, tandis que K peut très bien être de rang plein).
- On dit que X est isotrope si $K = \sigma^2 I_n$ pour $\sigma > 0$.
- Si les composantes X_1, \dots, X_n sont indépendantes alors K est diagonale.
- Si $a \in \mathbb{R}^d$ et $A \in \mathcal{M}_{d,n}(\mathbb{R})$ alors $a + AX$ a pour moyenne $a + Am$ et covariance AKA^\top .
En effet, si X est centré alors $\mathbb{E}(AX \otimes AX) = \mathbb{E}(AX(AX)^\top) = A\mathbb{E}(XX^\top)A^\top = AKA^\top$.
En particulier, en prenant $d = 1$, si $x \in \mathbb{R}^n$ alors $\langle X, x \rangle$ a pour moyenne $\langle m, x \rangle$ et covariance $\langle Kx, x \rangle$.
- En diagonalisant ou par Choleski, toute matrice symétrique à spectre ≥ 0 est une matrice de covariance.
- $X \in L^2$ ssi la fonction caractéristique $\varphi_X(\cdot) := \mathbb{E}(e^{i\langle X, \cdot \rangle})$ est dérivable deux fois en 0, et dans ce cas

$$m = \mathbb{E}(X) = -i\nabla\varphi_X(0) \quad \text{et} \quad K + mm^\top = \mathbb{E}(XX^\top) = -\text{Hess}\varphi_X(0).$$

3.1 Définition et simulation

Théodéf 3.1.1. Vecteurs gaussiens et lois gaussiennes.

Soit $X = (X_1, \dots, X_n)^\top$ un vecteur colonne aléatoire. Les propriétés suivantes sont équivalentes :

- (i) la variable aléatoire réelle $\langle t, X \rangle = t_1 X_1 + \dots + t_n X_n$ est gaussienne, pour tout $t \in \mathbb{R}^n$.
- (ii) Les cumulants d'ordre > 2 sont tous nuls : $\varphi_X(t) := \mathbb{E}(e^{i\langle t, X \rangle}) = e^{i\langle t, m \rangle - \frac{1}{2}\langle Kt, t \rangle}$, pour tout $t \in \mathbb{R}^n$.
- (iii) $X \stackrel{\text{loi}}{=} m + AZ$, où $m \in \mathbb{R}^n$, $A \in \mathcal{M}_n(\mathbb{R})$, $Z = (Z_1, \dots, Z_n)$, Z_1, \dots, Z_n i.i.d. $\mathcal{N}(0, 1)$.

On dit alors que X est un vecteur gaussien. Sa loi ne dépend que de m et K : loi gaussienne $\mathcal{N}(m, K)$.

- On dit également loi normale, d'où la notation \mathcal{N} . Gaussienne standard : $\mathcal{N}(0, I_n) = \mathcal{N}(0, 1)^{\otimes n}$.
- Le (iii) dit que les lois gaussiennes sont les déformations affines de la loi gaussienne standard $\mathcal{N}(0, I_n)$.
- Toute matrice symétrique à spectre positif est la matrice de covariance d'un vecteur aléatoire gaussien.
- Simulation de $\mathcal{N}(m, K)$ par réduction à $\mathcal{N}(0, 1)$ via $K = AA^\top$ (diagonalisation, ou mieux : Choleski).

Démonstration. Rappelons que $\varphi_{\mathcal{N}(m, \sigma^2)}(s) = e^{ims - \frac{1}{2}\sigma^2 s^2}$ et pour s'en souvenir, se ramener à $m = 0$ par translation, puis vecteur propre de Fourier lorsque $\sigma^2 = 1$, et principe d'incertitude pour σ^2 au numérateur!

(i) \Rightarrow (ii). Les X_1, \dots, X_n sont gaussiennes (prendre $t = e_i$) donc de carré intégrable, donc X possède un vecteur moyenne m et une matrice de covariance K . Si $t \in \mathbb{R}^n$ alors $\langle t, X \rangle = t_1 X_1 + \dots + t_n X_n \sim \mathcal{N}(\langle t, m \rangle, \langle Kt, t \rangle)$ donc $\varphi_X(t) = \varphi_{\langle t, X \rangle}(1) = \exp(i\langle t, m \rangle - \frac{1}{2}\langle Kt, t \rangle)$. Donc les cumulants de $\langle t, X \rangle$ d'ordre > 2 sont tous nuls.

(ii) \Rightarrow (i). Pour tout $t \in \mathbb{R}^n$, comme $\varphi_X(t) = \varphi_{\langle t, X \rangle}(1)$, on en déduit que $\langle t, X \rangle \sim \mathcal{N}(\langle t, m \rangle, \langle Kt, t \rangle)$.

(iii) \Rightarrow (i) Les composantes de X sont des combinaisons linéaires de v.a. gaussiennes indépendantes.

(ii) \Rightarrow (iii) Soit A telle que $AA^\top = K$ (factorisation de Choleski ou diagonalisation). Alors pour tout $t \in \mathbb{R}^n$,

$$\varphi_{m+AZ}(t) = e^{i\langle m, t \rangle} \mathbb{E}(e^{i\langle A^\top t, Z \rangle}) = e^{i\langle m, t \rangle} \varphi_Z(A^\top t) = e^{i\langle m, t \rangle} e^{-\frac{1}{2}\|A^\top t\|^2} = e^{i\langle m, t \rangle} e^{-\frac{1}{2}\langle AA^\top t, t \rangle} = \varphi_X(t).$$

■

3.2 Indépendance des composantes et densité

Théorème 3.2.1. Stabilité par transformation affine et indépendance des composantes.

- (i) Si $X \sim \mathcal{N}(m, K)$ alors pour tout $m' \in \mathbb{R}^d$ et tout $A \in \mathcal{M}_{d,n}(\mathbb{R})$, on a $m' + AX \sim \mathcal{N}(m' + Am, AK A^\top)$.
- (ii) Si $X \sim \mathcal{N}(m, K)$ alors les composantes X_1, \dots, X_n de X sont indépendantes ssi K est diagonale.
- (iii) Si (X, Y) est un vecteur gaussien centré, alors X et Y indépendants ssi $X_i \perp Y_j$ dans L^2 pour tous i, j .

— Contrexemple : si $Z \sim \mathcal{N}(0, 1)$ et $\varepsilon \sim \frac{1}{2}(\delta_{-1} + \delta_1)$ alors $X = (Z, \varepsilon Z)$ n'est pas gaussien car $X_1 + X_2 = Z + \varepsilon Z = (1 + \varepsilon)Z$ vérifie $\mathbb{P}(X_1 + X_2 = 0) = 1/2$. La covariance de X est diagonale car $\mathbb{E}(X_1 X_2) = \mathbb{E}(\varepsilon)\mathbb{E}(Z^2) = 0$ mais X_1 et X_2 ne sont pas indépendantes car $\text{Loi}(X_1 | X_2 = x) = \frac{1}{2}(\delta_{-x} + \delta_x)$.

— Le (ii) indique également que pour les composantes d'un vecteur gaussien, indépendance mutuelle et indépendance deux à deux sont équivalentes.

Démonstration.

(i) Pour tout $t \in \mathbb{R}^n$, $\langle t, m' + AX \rangle = \langle t, m' \rangle + \langle A^\top t, X \rangle$ qui est gaussienne car X est gaussien.

(ii) Si $K = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$ alors pour tout $t \in \mathbb{R}^n$, $\varphi_X(t) = \prod_{i=1}^n \exp(it_i m_i - \frac{1}{2}\sigma_i^2 t_i^2) = \prod_{i=1}^n \varphi_{X_i}(t_i)$ donc X_1, \dots, X_n sont indépendantes. La réciproque n'a rien de gaussien (covariance nulle si indépendance).

(iii) Découle immédiatement du (ii).

■

Théorème 3.2.2. Densité gaussienne.

La loi gaussienne $\mathcal{N}(m, K)$ sur \mathbb{R}^n a une densité ssi K est inversible, donnée par

$$x \in \mathbb{R}^n \mapsto \frac{\exp\left(-\frac{1}{2}\langle K^{-1}(x - m), x - m \rangle\right)}{\sqrt{(2\pi)^n \det(K)}},$$

et dans le cas contraire $\mathcal{N}(m, K)$ est portée par le sous-espace affine strict $m + \text{Im}(K)$ de \mathbb{R}^n .

— La loi gaussienne standard $\mathcal{N}(0, I_n) = \mathcal{N}(0, 1)^{\otimes n}$ a pour densité $x \in \mathbb{R}^n \mapsto (2\pi)^{-\frac{n}{2}} e^{-\frac{|x|^2}{2}}$.

Démonstration. Soit $A \in \mathcal{M}_{n,n}(\mathbb{R})$ telle que $K = AA^\top$, de sorte que $X \stackrel{d}{=} m + AZ$, où $Z \sim \mathcal{N}(0, I_n) = \mathcal{N}(0, 1)^{\otimes n}$. Si K est inversible, alors A est inversible, et comme Z a pour densité $\prod_{i=1}^n (2\pi)^{-\frac{1}{2}} \exp(-\frac{1}{2}x_i^2) = (2\pi)^{-\frac{n}{2}} \exp(-\frac{1}{2}|x|^2)$, il vient par changement de variable affine que $m + AZ$ a pour densité la formule de l'énoncé. Si K n'est pas inversible, alors A n'est pas inversible et $m + AZ$ prend ses valeurs dans $m + A\mathbb{R}^n$. Or $\text{Im}(A) = \text{Im}(K)$. ■

3.3 Cas isotrope : théorème de Cochran, théorème de Maxwell, équation de la chaleur

— Un vecteur aléatoire X de \mathbb{R}^n est isotrope qd sa loi est invariante par rotation : $RX \stackrel{\text{loi}}{=} X$ pour tout $R \in \mathcal{O}_n(\mathbb{R})$. La loi gaussienne $\mathcal{N}(m, K)$ est isotrope ssi $m = 0$ et $K = \sigma^2 I_n$ pour $\sigma \geq 0$.

Les lois gaussiennes isotropes sont des lois produit : $\mathcal{N}(0, \sigma^2 I_n) = \mathcal{N}(0, \sigma^2)^{\otimes n}$. Densité $x \in \mathbb{R}^n \mapsto \frac{1}{\sqrt{2\pi\sigma^2}^n} e^{-\frac{|x|^2}{2\sigma^2}}$.

— Cas extrêmes : $\mathcal{N}(0, \sigma^2 I_n) \rightarrow \delta_0$ si $\sigma \rightarrow 0$, et $\mathcal{N}(0, \sigma^2 I_n) \rightarrow \text{Lebesgue}$ si $\sigma \rightarrow \infty$. Plus précisément, soit $Z \sim \mathcal{N}(0, I_n)$. Alors $\sigma Z \sim \mathcal{N}(0, \sigma^2 I_n)$, et pour toute $f : \mathbb{R}^n \rightarrow \mathbb{R}$ continue et bornée, par convergence dominée,

$$\mathbb{E}(f(\sigma Z)) \xrightarrow{\sigma \rightarrow 0} f(0) = \int f d\delta_0.$$

D'autre part, pour tout borélien $I \subset \mathbb{R}^n$ borné de mesure de Lebesgue $|I| > 0$,

$$\mathbb{P}(\sigma Z \in I) \xrightarrow{\sigma \rightarrow \infty} 0 \quad \text{et} \quad \text{Loi}(\sigma Z \mid \sigma Z \in I) \xrightarrow{\text{loi}} \text{Uniforme}(I).$$

La masse se disperse et s'uniformise. En effet, pour tout borélien $J \subset I$, par convergence dominée,

$$\mathbb{P}(\sigma Z \in I) = (2\pi\sigma^2)^{-\frac{n}{2}} \int_I e^{-\frac{|x|^2}{2\sigma^2}} dx \xrightarrow{\sigma \rightarrow \infty} 0 \quad \text{et} \quad \mathbb{P}(\sigma Z \in J \mid \sigma Z \in I) = \frac{\int_J e^{-\frac{1}{2\sigma^2}|x|^2} dx}{\int_I e^{-\frac{1}{2\sigma^2}|x|^2} dx} \xrightarrow{\sigma \rightarrow \infty} \frac{|J|}{|I|}.$$

Le noyau de la chaleur $\mathcal{N}(0, 2tI_n)$ interpole entre δ_0 ($t = 0$) et Lebesgue ($t = \infty$).

— Si \bar{X} suit la loi uniforme sur la sphère $\{x \in \mathbb{R}^n : |x| = \sqrt{n}\}$, alors

$$p_{\mathbb{R}^k}(X) := (X_1, \dots, X_k) \xrightarrow[n \rightarrow \infty]{\text{loi}} \mathcal{N}(0, I_k).$$

En effet, $X \stackrel{\text{loi}}{=} \sqrt{n} \frac{Z}{|Z|}$ où $Z \sim \mathcal{N}(0, I_n) = \mathcal{N}(0, 1)^{\otimes n}$, et par la loi des grands nombres,

$$\sqrt{n} \frac{(Z_1, \dots, Z_k)}{\sqrt{Z_1^2 + \dots + Z_n^2}} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} (Z_1, \dots, Z_k) \sim \mathcal{N}(0, I_k).$$

Toute gaussienne standard est donc proche d'une loi uniforme sur une sphère de grande dimension¹.

Théorème 3.3.1. Théorème de Cochran sur les gaussiennes isotropes.

Si $Z = (Z_1, \dots, Z_n) \sim \mathcal{N}(m, \sigma^2 I_n)$, et si $\mathbb{R}^n = E_1 \oplus \dots \oplus E_r$ avec $E_j \perp E_k$ si $j \neq k$, alors les vecteurs aléatoires $\text{proj}_{E_k}(Z) = P_k Z$, $1 \leq k \leq r$, sont gaussiens indépendants avec $P_k Z \sim \mathcal{N}(P_k m, \sigma^2 P_k)$ de \mathbb{R}^n .

Démonstration. On se ramène par translation à $m = 0$. Comme les s.e.v. E_i sont \perp et la loi de Z est invariante par rotation, on se ramène au cas où $E_k = \text{vect}(e_i : i \in I_k)$, $1 \leq k \leq r$, pour une partition $I_1 \cup \dots \cup I_r = \{1, \dots, n\}$. Le résultat découle alors du théorème 3.2.1 : la matrice P_k de projection orthogonale sur E_k est diagonale, et $(P_k)_{i,i} = \mathbb{1}_{i \in I_k}$. ■

1. Cela est lié au fait que la loi uniforme se concentre en grande dimension autour des grands cercles, phénomène lui-même lié au fait que les ensembles optimaux pour le problème isopérimétrique sont les calottes sphériques. Ce résultat dû à Paul Lévy a été étendu aux variétés riemanniennes à courbure de Ricci positive par Misha Gromov, puis aux semi-groupes de Markov par Dominique Bakry et Michel Ledoux.

Corollaire 3.3.2. Échantillons gaussiens et studentisation.

Si les X_1, \dots, X_n sont des v.a. i.i.d. $\mathcal{N}(m, \sigma^2)$ alors la moyenne empirique et la variance empirique

$$m_n := \frac{X_1 + \dots + X_n}{n} \quad \text{et} \quad \sigma_n^2 := \frac{(X_1 - m_n)^2 + \dots + (X_n - m_n)^2}{n-1}$$

sont indépendantes, $m_n \sim \mathcal{N}(m, \frac{\sigma^2}{n})$, $\frac{n-1}{\sigma^2} \sigma_n^2 \sim \chi^2(n-1)$, d'où $\sqrt{n} \frac{m_n - m}{\sigma_n} \sim \frac{\mathcal{N}(0,1)}{\sqrt{\frac{\chi^2(n-1)}{n-1}}} =: \text{Student}(n-1)$.

- On qualifie d'empirique ce qui est issu de l'expérience, ici l'échantillonnage d'un phénomène aléatoire.
- Le théorème de Cochran et son corollaire permettent la construction de tests d'hypothèses statistiques.

Démonstration. $\mathbb{R}^n = E_1 \oplus E_2$, $E_1 := \mathbb{R}(1, \dots, 1)$, $E_2 = E_1^\perp$, d'où $\text{proj}_{E_1}(X) = m_n$, $\text{proj}_{E_2}(X) = X - m_n$. ■

Théorème 3.3.3. Caractérisation géométrique de Maxwell des gaussiennes isotropes.

Une loi μ sur \mathbb{R}^n , $n \geq 2$, est produit et invariante par rotation ssi $\mu = \mathcal{N}(0, \sigma^2 I_n)$ pour un $\sigma \geq 0$.

- Quand $\sigma = 0$, on dispose de la convention très naturelle $\mathcal{N}(0, I_n) = \delta_0 = \delta_0^{\otimes n}$.
- Quand $n = 1$, toute loi est produit, et l'invariance par rotation devient la symétrie, or toute loi centrée symétrique n'est pas gaussienne, le théorème n'a donc pas lieu pour $n = 1$. La loi de Rademacher $\frac{1}{2}\delta_{-\sigma} + \frac{1}{2}\delta_\sigma$ est centrée et symétrique, ses deux premiers moments, 0 et σ^2 , sont identiques à ceux de $\mathcal{N}(0, \sigma^2)$.
- Maxwell a obtenu ce résultat dans son exploration de la distribution des vitesses en théorie cinétique des gaz, domaine scientifique dans lequel on a coutume de parler de maxwellienne plutôt que de gaussienne! L'argument de symétrie de Maxwell est contestable, et a été critiqué notamment par Paul Lévy [47].

Démonstration. Pour tout $\sigma > 0$, $\mathcal{N}(0, \sigma^2 I_n) = \mathcal{N}(0, \sigma^2)^{\otimes n}$, de densité $e^{-\frac{|x|^2}{2\sigma^2} - n \log \sqrt{2\pi\sigma^2}}$ est produit et invariante par rotation. Réciproquement, soit μ une loi sur \mathbb{R}^n , $n \geq 2$, produit et invariante par rotation. Quitte à remplacer μ par $\mu * \mathcal{N}(0, \varepsilon I_n)$, $\varepsilon > 0$, c'est-à-dire $X + \varepsilon Z$ en terme de v.a. on peut supposer sans perte de généralité² que μ a une densité lisse $f : \mathbb{R}^n \rightarrow (0, \infty)$. L'invariance par rotation donne $\log(f(x)) = g(|x|^2)$ et donc

$$\partial_i \log(f(x)) = 2g'(|x|^2)x_i.$$

D'autre part la nature produit donne $\log(f(x)) = h_1(x_1) + \dots + h_n(x_n)$ et donc

$$\partial_i \log(f(x)) = h'_i(x_i).$$

Donc $\partial_i \log(f(x))$, qui dépend de $|x|$ via $g'(|x|)$, ne dépend que de x_i . Or $n \geq 2$ par hypothèse, donc g' est constante. Ensuite, comme g' est constante, il existe $a, b \in \mathbb{R}$ tels que $g(u) = au + b$ pour tout u , et donc $f(x) = e^{a|x|^2 + b}$. Comme f est une densité, $a < 0$ et $e^b = (\pi/a)^{-n/2}$, et $\mu = \mathcal{N}(0, \sigma^2 I_n)$ avec $\sigma^2 = -1/(2a)$. ■

Le théorème suivant est au carrefour de l'analyse harmonique, des probabilités, de l'analyse des équations aux dérivées partielles, de l'analyse fonctionnelle, de la physique statistique, et de la théorie de l'information. Le laplacien possède des analogues dans des contextes très variés : graphes, réseaux, variétés, formes modulaires, ...

Théorème 3.3.4. Noyau, équation, et semi-groupe de la chaleur, entropie, information.

Soit μ une mesure de probabilité sur \mathbb{R}^n , et pour tout $t \geq 0$, soit $\mu_t := \mu * \mathcal{N}(0, tI_n)$. Alors :

- $\mu_t \rightarrow \mu$ étroitement quand $t \rightarrow 0$.
- Pour tout $t > 0$, la mesure de probabilité μ_t a pour densité $f_t := \mu * p_t$ où p_t est la densité de $\mathcal{N}(0, tI_n)$.
- Pour tout $t \in \mathbb{R}_+$, $\mu \mapsto \mu_t$ est un opérateur linéaire sur les mesures de probabilités sur \mathbb{R}^n , et la famille $\{(\cdot)_t : t \in \mathbb{R}_+\}$ est un semi-groupe abélien : $(\cdot)_0 = \text{id}$, et $(\cdot)_s \circ (\cdot)_t = (\cdot)_t \circ (\cdot)_s = (\cdot)_{s+t}$ pour tous $s, t \in \mathbb{R}_+$.
- La fonction $(t, x) \in (0, \infty) \times \mathbb{R}^n \mapsto f(t, x) = f_t(x)$ est solution de l'équation de la chaleur :

$$\partial_t f_t = \frac{1}{2} \Delta_x f_t \quad \text{où} \quad \Delta_x = \partial_{x_1}^2 + \dots + \partial_{x_n}^2.$$

2. Plus précisément, comme $X_\varepsilon := X + \varepsilon Z$ et Z sont L^2 car gaussiens, et centrés, le vecteur X l'est aussi. Si K est la matrice de covariance de X , alors la matrice de covariance de X_ε est $K_\varepsilon := K + \varepsilon^2 I_n$, et donc pour $t \in \mathbb{R}^n$, par indépendance, $\varphi_X(t) = \varphi_{X_\varepsilon}(t) / \varphi_Z(\varepsilon t) = \exp(-\frac{1}{2} \langle (K_\varepsilon - \varepsilon^2 I_n)t, t \rangle) = \varphi_{\mathcal{N}(0, K)}(t)$. Cela ne nécessite pas $\varepsilon \rightarrow 0$. Alternativement, quand $\varepsilon \rightarrow 0$, d'une part $\varphi_{X_\varepsilon}(t) = \varphi_X(t) \varphi_Z(\varepsilon t) \rightarrow \varphi_X(t)$ et d'autre part les deux premiers moments de X_ε convergent vers ceux de X , et comme X_ε est gaussien, on obtient que X l'est aussi.

(v) Si μ a une densité $f \in \mathcal{C}^3(\mathbb{R}^n, \mathbb{R})$ avec f''' bornée alors pour tout $x \in \mathbb{R}^3$,

$$f_t(x) = f(x) + t(\frac{1}{2}\Delta f)(x) + o_{t \rightarrow 0}(t) \quad \text{autrement dit} \quad \lim_{t \rightarrow 0} \frac{f_t(x) - f(x)}{t} = (\frac{1}{2}\Delta f)(x).$$

(vi) Identité de de Bruijn sur entropie de Boltzmann – Shannon et information de Fisher : pour tout $t > 0$,

$$\partial_t \left(- \int_{\mathbb{R}^n} f_t \log f_t dx \right) = \frac{1}{2} \int_{\mathbb{R}^n} \frac{|\nabla_x f_t|^2}{f_t} dx \geq 0.$$

- On dit que $p_t(x) = (2\pi t)^{-\frac{n}{2}} e^{-\frac{|x|^2}{2t}}$ est le noyau de la chaleur et $\{(\cdot)_t : t \in \mathbb{R}_+\}$ le semi-groupe de la chaleur.
- Si $(B_t)_{t \in \mathbb{R}_+}$ est un mouvement brownien issu de $B_0 \sim \mu$ (chapitre 7), alors $B_t \sim \mu_t$ pour tout $t \geq 0$.
- L'identité de de Bruijn est un théorème-H à la Boltzmann pour l'équation de la chaleur.
- Sur un espace non-linéaire comme une variété, on ne dispose pas de gaussiennes, mais avec une structure riemannienne, on dispose d'un laplacien, et le semi-groupe de la chaleur associé peut alors servir à la fois à régulariser (temps petit) et à interpoler (entre masses de Dirac et mesure de volume).
- En quelque sorte $(\cdot)_t = e^{t\frac{1}{2}\Delta}$ et on dit que $\frac{1}{2}\Delta$ est le générateur infinitésimal du semi-groupe de la chaleur.
- Comme $f_t > 0$ dès que $t > 0$, même si $\mu = \delta_0$, la vitesse de propagation est infinie, ce qui est peu réaliste pour la diffusion de la chaleur, mais des versions à vitesse maximale existent (dites relativistes, non-linéaires). Les modèles mathématiques sont des caricatures, plus ou moins utiles à travers les échelles d'approximation.

Démonstration.

(i) Pour toute $f : \mathbb{R}^n \rightarrow \mathbb{R}$ continue et bornée, et avec $X \sim \mu$ et $Z \sim \mathcal{N}(0, I_n)$ indépendantes,

$$\int f d\mu_t = \mathbb{E}(f(X + \sqrt{t}Z)) \xrightarrow{t \rightarrow 0} \mathbb{E}(f(X)) = \int f d\mu \quad \text{par convergence dominée.}$$

- (ii) Avec les notations précédentes, si $t > 0$ alors $X + \sqrt{t}Z$ a pour densité $\mu * p_t := \int p_t(\cdot - y) d\mu(y)$.
- (iii) En effet par associativité $(\mu_s)_t = \mu * p_s * p_t = \mu * p_{s+t} = \mu_{s+t}$ car $\mathcal{N}(0, sI_n) * \mathcal{N}(0, tI_n) = \mathcal{N}(0, (s+t)I_n)$.
- (iv) Dérivation sous le signe somme et le fait que $\partial_t p_t = \frac{1}{2}\Delta_x p_t$ pour $t > 0$.
- (v) Formule de Taylor $f(y) = f(x) + \sum_{i=1}^n \partial_{x_i} f(x)(y-x)_i + \frac{1}{2} \sum_{1 \leq i, j \leq n} \partial_{x_i x_j}^2 f(x)(y-x)_i (y-x)_j + O(\|x-y\|^3)$, et formule du noyau de la chaleur $f_t(x) = \int_{\mathbb{R}^n} f(y) p_t(x-y) dy = \mathbb{E}(f(x + \sqrt{t}Z))$, avec $\mathbb{E}(Z_i) = 0$ et $\mathbb{E}(Z_i Z_j) = t\mathbb{1}_{i=j}$.
- (vi) Par dérivation sous le signe somme, (iii), et intégration par parties,

$$\partial_t \int_{\mathbb{R}^n} f_t \log f_t dx = \frac{1}{2} \int_{\mathbb{R}^n} (1 + \log f_t) \Delta_x f_t dx = -\frac{1}{2} \int_{\mathbb{R}^n} \frac{|\nabla_x f_t|^2}{f_t} dx.$$

Les termes de bord disparaissent car f_t et des dérivées sont nulles à l'infini (à vérifier). ■

3.4 Espérance conditionnelle et loi conditionnelle

Le théorème suivant affirme que, pour un vecteur gaussien, la loi conditionnelle d'un bloc de composantes par rapport à un autre bloc de composantes disjoint est toujours gaussienne, de moyenne et de covariance explicites. Cela joue un rôle important dans l'étude des modèles linéaires gaussiens en statistique.

Théorème 3.4.1. Espérance conditionnelle et loi conditionnelle.

Soit $(X, Y) : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}^d$ un vecteur gaussien de $\mathbb{R}^d = \mathbb{R}^p \times \mathbb{R}^q$, de matrice de covariance

$$\Gamma = \begin{pmatrix} \text{Cov}(X, X) & \text{Cov}(X, Y) \\ \text{Cov}(Y, X) & \text{Cov}(Y, Y) \end{pmatrix} = \begin{pmatrix} K & C \\ C^\top & L \end{pmatrix}$$

inversible (implique que K et L le sont également). Alors :

- (i) $\mathbb{E}(Y | X)$ est un vecteur gaussien, fonction affine de X donnée par $\mathbb{E}(Y | X) = \mathbb{E}(Y) + C^\top K^{-1}(X - \mathbb{E}(X))$.
- (ii) Loi $Y | X = x = \mathcal{N}(\mathbb{E}(Y) + C^\top K^{-1}(x - \mathbb{E}(X)), L - C^\top K^{-1}C)$.
- (iii) $Y = \mathbb{E}(Y | X) + \varepsilon$ où $\varepsilon \sim \mathcal{N}(0, L - C^\top K^{-1}C)$ et ε est indépendant de X .

— Cela reste valable quand Γ n'est pas inversible en utilisant une pseudo-inverse ou inverse généralisée de K .

Démonstration. On se ramène par translation au cas où le vecteur gaussien (X, Y) est centré. Nous pouvons alors bénéficier de la caractérisation de l'indépendance par l'orthogonalité dans L^2 (théorème 3.2.1).

- (i) Soit F_0 le s.e.v. de $F := L^2(\Omega, \sigma(X), \mathbb{P})$ engendré par la famille de v.a.r. gaussiennes $\{a + b^\top X : a \in \mathbb{R}, b \in \mathbb{R}^p\}$. Pour tout $1 \leq j \leq q$, $Y_j \in L^2$, donc $\mathbb{E}(Y_j | X) = p_F(Y_j)$. La v.a. $\Delta_j := p_F(Y_j) - p_{F_0}(Y_j) = Y_j - p_F(Y_j) - (Y_j - p_{F_0}(Y_j))$ est $\perp X_i$ dans L^2 pour tout i , donc indépendante de X car (Δ_j, X) est gaussien centré (théorème 3.2.1). Or Δ_j est aussi $\sigma(X)$ -mesurable, elle est donc constante, donc nulle car centrée. Donc $\mathbb{E}(Y_j | X) = p_{F_0}(Y_j) = b_j^\top X$ pour un $b_j \in \mathbb{R}^q$. Enfin l'orthogonalité de $Y_j - b_j^\top X$ et X_i dans L^2 donne $C = Kb$ d'où $b^\top = C^\top K^{-1}$.
- (ii) Par construction, la moyenne de $\text{Loi}(Y | X = x)$ est $\mathbb{E}(Y | X = x)$. Comme Γ et K sont inversibles, d'après le théorème 3.2.2, les vecteurs aléatoires (X, Y) et X centrés ont pour densités

$$z \mapsto ((2\pi)^d \det(\Gamma))^{-\frac{1}{2}} e^{-\frac{1}{2} \langle \Gamma^{-1} z, z \rangle} \quad \text{et} \quad x \mapsto ((2\pi)^p \det(K))^{-\frac{1}{2}} e^{-\frac{1}{2} \langle K^{-1} x, x \rangle},$$

où $z := (x, y)^\top$. Le rapport des deux donne la densité de la loi conditionnelle $\text{Loi}(Y | X = x)$:

$$y \mapsto \sqrt{\frac{\det(K)}{(2\pi)^q \det(\Gamma)}} e^{-\frac{1}{2} (\langle \Gamma^{-1} z, z \rangle - \langle K^{-1} x, x \rangle)}.$$

Or $\Gamma^{-1} = \begin{pmatrix} K^{-1} + K^{-1} C S^{-1} C^\top K^{-1} & -K^{-1} C S^{-1} \\ -S^{-1} C^\top K^{-1} & S^{-1} \end{pmatrix}$ où $S := L - C^\top K^{-1} C$, de plus $\det(\Gamma) = \det(K) \det(S)$.

Après simplification, la densité de $\text{Loi}(Y | X = x)$ s'écrit (S apparaît comme une variance conditionnelle)

$$y \mapsto ((2\pi)^q \det(S))^{-\frac{1}{2}} e^{-\frac{1}{2} \langle S^{-1} (y - m), y - m \rangle}, \quad \text{où} \quad m := C^\top K^{-1} x,$$

qui est bien celle de $\mathcal{N}(m, S)$. Alternativement, on peut utiliser la fonction caractéristique conditionnelle.

- (iii) Découle de (ii). La preuve de (i) donne aussi $Y = \mathbb{E}(Y | X) + \varepsilon$ où ε est gaussien centré indépendant de X . ■

6 → 7

Ceci n'est pas abordé lors du cours oral.

Théodéf 3.4.2: Indépendance conditionnelle

Pour un vecteur aléatoire $Z = (X, Y)$ de $\mathbb{R}^d = \mathbb{R}^p \times \mathbb{R}^q$, on dit que les composantes de Y sont *conditionnellement indépendantes* sachant X lorsque $\text{Loi}(Y | X = x)$ est produit pour tout x .

Si Z est un vecteur gaussien de matrice de covariance inversible Γ , alors les composantes de Y sont conditionnellement indépendantes sachant X ssi la sous-matrice $(\Gamma^{-1})_{p < k \leq d, p < k \leq d}$ est diagonale.

Ainsi, pour un vecteur gaussien, les zéros dans la matrice de covariance indiquent l'indépendance tandis que les zéros dans l'inverse de la matrice de covariance indiquent l'indépendance conditionnelle.

Démonstration. On adopte la décomposition en blocs de Γ du théorème 3.4.1. Comme dans la preuve du théorème 3.4.1, comme Γ est inversible, les matrices K et $S := L - C^\top K^{-1} C$ le sont également, et pour tout x , $\text{Loi}(Y | X = x)$ est gaussienne de matrice de covariance S . Le théorème 3.2.1 donne alors que les composantes de Y sont conditionnellement indépendantes sachant X ssi S est diagonale, c'est-à-dire ssi S^{-1} est diagonale. Enfin la formule d'inversion par blocs de Γ donne $(S^{-1})_{k,l} = (\Gamma^{-1})_{k,l}$ pour tout $p < k, l \leq d$. ■

3. Formule d'inversion par blocs. La matrice carrée S est appelée complément de Schur de K dans la matrice par blocs Γ . La matrice Γ est (symétrique positive) inversible ssi K et son complément de Schur S dans Γ le sont. Cette géométrie des matrices inversibles et du cône des matrices symétriques positives est importante, bien au-delà des vecteurs gaussiens. Voir par exemple [58] pour un point de vue panoramique.

Exemple 3.4.3: Processus auto-régressif gaussien

Le processus auto-régressif gaussien AR(1) est la suite $(X_n)_{n \in \mathbb{N}}$ définie par

$$X_{n+1} = \theta X_n + \varepsilon_{n+1}$$

où θ est un réel fixé, et où $(\varepsilon_n)_{n \in \mathbb{N}}$ sont des v.a. i.i.d. de loi $\mathcal{N}(0, 1)$. On observe que $(\varepsilon_0, \dots, \varepsilon_{n+1})^\top = T(X_0, \dots, X_n)^\top$ où T est la matrice $(n+1) \times (n+1)$ triangulaire inférieure bidiagonale

$$T_{i,i} = 1 \quad \text{et} \quad T_{i,i-1} = -\theta.$$

Il en découle que (X_0, \dots, X_n) est un vecteur gaussien centré de matrice de covariance $(T^\top T)^{-1}$. Son inverse est donc la matrice tridiagonale $T^\top T$. Par conséquent, d'après le théorème 3.4.2, pour tout $1 \leq i < j \leq d$, si $|i - j| > 1$ alors X_i et X_j sont conditionnellement indépendantes sachant les autres composantes de (X_0, \dots, X_n) . On le vérifie par exemple sur la loi de $(X_0, X_2) = (\varepsilon_0, \theta X_1 + \varepsilon_2)$ sachant $X_1 = \theta \varepsilon_0 + \varepsilon_1 = x_1$, égale à la loi de $(\theta^{-1}(x_1 - \varepsilon_1), \theta x_1 + \varepsilon_2)$, qui est bien produit. Les processus auto-régressifs, gaussiens ou pas, sont emblématiques de la modélisation des séries chronologiques ou séries temporelles, et constituent des exemples naturels de martingales et de chaînes de Markov.

Pour aller plus loin 3.4.4. Vecteurs gaussiens.

- Caractérisation des gaussiennes comme maximum d'entropie de Boltzmann – Shannon [12].
- Combinatoire des moments des lois gaussiennes et formule de Wick [5].
- Intégration par parties gaussienne et équation de Stein.
- Maximum de v.a. i.i.d. gaussiennes, $\sqrt{\log(n)}$, loi de Gumbel, $\sqrt{n \log(n)}$.
- Polynômes orthogonaux d'Hermite, décomposition en chaos, calcul de Malliavin.
- Processus auto-régressifs ARMA, filtrage, théorie du signal, pour les séries temporelles.
- Matrices aléatoires de Wishart et factorisation de Bartlett, matrices aléatoires GOE/GUE [12].
- Inégalité de Sobolev logarithmique gaussienne [12].
- Chaos multiplicatif gaussien, champ libre gaussien.

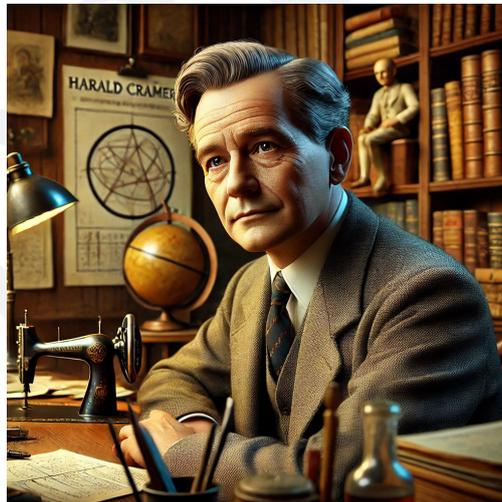


FIGURE 3.2 – Portrait imaginaire de Harald Cramér généré par ChatGPT. La génération d'image par apprentissage automatique profond fait appel à des processus de diffusion gaussiens, qui sont des vecteurs gaussiens de grande dimension, de la même famille que le mouvement brownien. Les vecteurs ou processus gaussiens, incontournables en raison du phénomène TLC, sont largement utilisés dans tous les champs de la modélisation stochastique.

Chapitre 4

Martingales à temps discret



FIGURE 4.1 – Table de jeu avec roulette, dans un casino de Las Vegas, Nevada, dans les années 1950. Initialement développées comme une généralisation des marches aléatoires pour la modélisation de jeux de hasard, les martingales constituent aujourd’hui un outil puissant, polyvalent, et élégant pour l’étude des processus stochastiques.

« Le temps, cher ami, le temps amène l’occasion; l’occasion c’est la martingale de l’homme : plus on a engagé, plus on gagne quand on sait attendre. » Alexandre Dumas (1802 – 1870), Les trois mousquetaires (1844), chapitre 42.

4.1 Processus, filtrations, martingales

Dans ce chapitre, on se place toujours et sauf mention du contraire sur un espace de probabilité. $(\Omega, \mathcal{F}, \mathbb{P})$.

Une suite croissante $(\mathcal{F}_n)_{n \in \mathbb{N}}$ de sous-tribus de \mathcal{F} est appelée filtration : $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \dots \subset \mathcal{F}$, tandis que le quadruplet $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \in \mathbb{N}}, \mathbb{P})$ est un espace de probabilité filtré. Cette notion provient des processus.

Un processus stochastique à temps discret est une suite $X = (X_n)_{n \in \mathbb{N}}$ de v.a. modélisant une expérience au fil du temps. L’information ou champ des possibles accumulé au temps n est donné par la sous-tribu

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

Cela signifie que \mathcal{F}_n^X est constituée par tous les événements que nous pouvons écrire avec X_0, \dots, X_n .

La suite $(\mathcal{F}_n^X)_{n \in \mathbb{N}}$ est une filtration appelée filtration canonique du processus.

Un processus $(X)_{n \in \mathbb{N}}$ est ...

- adapté à une filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$ lorsque X_n est mesurable par rapport à \mathcal{F}_n , pour tout n et la filtration canonique du processus est la plus petite tribu rendant le processus adapté.
- prévisible lorsque X_n est mesurable par rapport à \mathcal{F}_{n-1} pour tout $n \geq 1$, et $X_0 = 0$.
- intégrable lorsque $\mathbb{E}(|X_n|) < \infty$ c’est-à-dire $X_n \in L^1$ pour tout n .
- de carré intégrable lorsque $\mathbb{E}(X_n^2) < \infty$ c’est-à-dire $X_n \in L^2$ pour tout n .
- borné dans L^p lorsque $\sup_{n \in \mathbb{N}} \mathbb{E}(|X_n|^p) < +\infty$.
- borné lorsqu’il existe une constante réelle $C > 0$ telle que $|X_n| \leq C$ p.s. pour tout n .
- fini lorsque $|X_n| < +\infty$ p.s. pour tout n .
- positif lorsque $X_n \geq 0$ p.s. pour tout n .

Un processus $X = (X_n)_{n \in \mathbb{N}}$, adapté à une filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$ et intégrable, est une

- *martingale* lorsque $\mathbb{E}(X_{n+1} | \mathcal{F}_n) = X_n$ p.s. pour tout n .
- *sous-martingale* lorsque $\mathbb{E}(X_{n+1} | \mathcal{F}_n) \geq X_n$ p.s. pour tout n .
- *surmartingale* lorsque $\mathbb{E}(X_{n+1} | \mathcal{F}_n) \leq X_n$ p.s. pour tout n .

Être une martingale est équivalent à être à la fois une sous-martingale et une surmartingale.

Les propriétés de l'espérance conditionnelle indiquent qu'il est possible de remplacer $n + 1$ par tout $m \geq n + 1$.

Dans la suite de ce chapitre, on omet bien souvent le p.s. pour alléger les notations. Méfiance, toutefois.

Les martingales sont des suites aléatoires constantes en moyenne :

$$\mathbb{E}(M_0) = \dots = \mathbb{E}(M_n) = \dots.$$

Une information sur M_0 est donc conservée¹ au fil du temps. De même, l'espérance est croissante le long des sous-martingales, suites croissantes en moyenne, et décroissante le long des surmartingales, suites décroissantes en moyenne. Une question naturelle, élucidée dans ce chapitre, est le comportement asymptotique des trajectoires du processus, au-delà de la moyenne, notamment la convergence presque sûre et en moyenne de $(M_n)_{n \in \mathbb{N}}$.

Remarque 4.1.1. Comment ne plus se tromper entre sous-martingale et surmartingale.

- Astuce mnémotechnique visuelle de l'auteur de ces notes : « Pour une suite convergente monotone, sous la limite = croître, sur la limite = décroître ». Mais il est possible de faire autrement, comme par exemple Nathanaël Enriquez : « Quand je demande aux étudiants de me dire si un truc qui monte est plutôt "sous" ou "sur", ils me disent tous "sur". Il faut donc retenir que c'est l'inverse de l'intuition! ».
- À propos de la dénomination, voici un extrait d'une interview de J. L. Doob par J. L. Snell [59] :
Doob : « [...] The martingale definition led at once to the idea of sub and super martingales, and it was clear that these were the appropriate names but, as I remarked in my 1984 book (Classical Potential Theory and Its Probabilistic Counterpart, Springer, 1984), the name supermartingale was spoiled for me by the fact that every evening the exploits of "Superman" were played on the radio by one of my children. If I had been doing my work at the university rather than at home I am sure I would not have used the ridiculous names semi- and lower semimartingales for sub- and supermartingales in my 1953 book. Perhaps I should have noted that one reason for the success of that book is the prestigious sounding title, a translation of a name in a German Khintchine paper. »
- Ce que sous-entend Doob est que si B est un mouvement brownien alors $f(B)$ est une sous-martingale si f est sous-harmonique, et une surmartingale si f est surharmonique, et il en va de même pour la marche aléatoire simple pour le laplacien discret. Reste à comprendre comment distinguer (sous-/sur)harmonique. En fait, les fonctions surharmoniques sont littéralement sur les fonctions harmoniques, au sens suivant : si f est surharmonique et si u est harmonique avec $f = u$ sur $\partial\Omega$ alors $f \geq u$. Avant de devenir probabiliste, Doob était analyste harmonique. D'autre part, certains préfèrent les opérateurs positifs, et l'intégration par parties $\int (-\Delta f) f dx = \int |\nabla f|^2 dx \geq 0$ fait que $-\Delta$ est un opérateur positif, ce qui rend naturel de dire que f est surharmonique si $-\Delta f \geq 0$.
- Extrait du livre de 1984 de Doob [21, p. 808] : « Before martingales had been formally christened, Lévy (1935,1937), Bernstein (1937), and other mathematicians had analyzed some of their properties in special contexts; usually the martingales in question arose as partial sums $\sum_{j=0}^n y_j$ of a sequence y of random variables under the condition $\mathbb{E}(y_j | y_0, \dots, y_{j-1}) = 0$ so that the sums arose as generalizations of sums of independent random variables with zero means. Ville (1939) defined a martingale very nearly as a positive martingale is now defined but tied it to a sequence of independent random variables under analysis. His fundamental tool, a fact he proved, was that almost every sample sequence of a positive martingale is bounded. Doob (1940) discussed martingales and proved the basic convergence properties under the name "family of random variables with the property E." ("E" was chosen not as the initial letter of "expectation" but as the first letter in the alphabet following "D".) Under the respective names "semimartingale" and "lower semimartingale," submartingales and supermartingales were introduced in (Snell 1952) and (Doob 1953). This obviously inappropriate nomenclature was chosen under the malign influence of the noise level of radio's SUPERman program, a favorite supper-time program of Doob's son during the writing of (Doob 1953)". For further work in martingale theory see (Neveu 1972) (discrete parameter context) and (Dellacherie and Meyer 1980) (continuous parameter context). ».

Les martingales sont des généralisations des sommes de variables aléatoires indépendantes et centrées ou

1. C'est une loi de conservation, où *loi* a le même sens que dans *loi de la nature* et *loi des grands nombres*. Alternativement, cette loi de conservation peut être vue comme un invariant de la suite. Identifier les invariants d'un modèle est toujours utile pour son étude.

marches aléatoires, avec l'idée que chaque accroissement est conditionnellement centré sachant les précédents :

$$M_n \in L^1(\mathcal{F}_n) \quad \text{et} \quad \mathbb{E}(M_{n+1} - M_n | \mathcal{F}_n) = 0 \quad \text{pour tout } n.$$

Pour les martingales de carré intégrable, cela devient

$$M_{n+1} - M_n \perp L^2(\mathcal{F}_n) \quad \text{pour tout } n.$$

Dans toute la suite, la tendance est de noter les martingales M , et autrement les autres processus.

4.2 Exemples de martingales

Les propriétés de l'espérance conditionnelle font que les martingales sont polymorphes et assez répandues. Dans les exemples suivants, la filtration est à deviner, en général il s'agit d'une filtration canonique.

— *Somme.* $S_0 = 0$, $S_n = X_1 + \dots + X_n$, X_1, X_2, \dots indépendantes de moyenne m ,

$$M_n = S_n - nm.$$

C'est une marche aléatoire à pas indépendants centrés $X_1 - m, \dots, X_n - m$.

Dans le même esprit, S_n est une surmartingale si $m \leq 0$ et une sous-martingale si $m \geq 0$.

— *Produit.* $P_0 = 1$, $P_n = X_1 \cdots X_n$, X_1, X_2, \dots indépendantes, positives, de moyenne m ,

$$M_n = \frac{P_n}{m^n}.$$

Le logarithme $\log(M_n)$ fait retrouver le cas précédent, on parle également de martingale exponentielle.

— *Somme carrée.* $S_0 = 0$, $S_n = X_1 + \dots + X_n$, X_1, X_2, \dots indépendantes, de moyenne m et de variance σ^2 ,

$$M_n = (S_n - nm)^2 - n\sigma^2.$$

— *Exponentielle de Wald.* $S_0 = 0$, $S_n = X_1 + \dots + X_n$, X_1, X_2, \dots i.i.d. Rademacher $p\delta_1 + q\delta_{-1}$,

$$M_n = \left(\frac{q}{p}\right)^{S_n}.$$

Plus généralement, $S_0 = x$, $S_n = x + X_1 + \dots + X_n$, X_1, X_2, \dots i.i.d. $\varphi(\alpha) = \log(\mathbb{E}(e^{-\alpha X_1}))$, $\alpha > 0$,

$$M_n = \exp(-\alpha S_n - n\varphi(\alpha)).$$

Cette martingale et les deux précédentes interviennent dans l'étude de la ruine du joueur (exemple 4.4.6). Les martingales somme $S_n - nm$ et somme carrée $(S_n - nm)^2 - n\sigma^2$ sont les deux premières dérivées en $\alpha = 0$ de la martingale exponentielle $\exp(-\alpha S_n - n\varphi(\alpha))$. Plus généralement, l'ensemble des martingales pour une même filtration est stable par combinaison linéaire, c'est un espace vectoriel. D'autre part, les martingales précédentes sont toutes construites de la même façon : une fonction (linéaire ou pas) d'une marche aléatoire, corrigée additivement ou multiplicativement pour avoir une espérance constante.

— *Martingale fermée ou de Doob.* $X \in L^1$, $(\mathcal{F})_{n \geq \mathbb{N}}$ filtration,

$$M_n = \mathbb{E}(X | \mathcal{F}_n).$$

Dans le même esprit, si un processus $(X_n)_{n \in \mathbb{N}}$ est intégrable, adapté à une filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$, et décroissant (respectivement croissant), alors c'est une surmartingale (respectivement sous-martingale).

— *Modèle de Wright-Fisher (génétique des populations).* Une population est formée de N individus, de type A ou B . Soit X_n le nombre d'individus de type A à la génération n . La génération $n+1$ est fabriquée en générant de manière i.i.d. les types des N individus en utilisant la fréquence du type A dans la génération n . Cela s'écrit $X_{n+1} = \sum_{k=1}^N \mathbb{1}_{\{U_{n+1,k} \leq X_n/N\}}$ où $(U_{n+1,k})_{n \in \mathbb{N}, 1 \leq k \leq N}$ sont i.i.d. de loi uniforme sur $[0, 1]$, indépendantes de X_0 . Cela donne $\text{Loi}(X_{n+1} | X_0, \dots, X_n) = \text{Bin}(N, X_n/N)$, avec $0 \leq X_0 \leq N$, et $(X_n)_{n \in \mathbb{N}}$ est une martingale.

— *Urnes de Pólya (mécanisme de renforcement).* Une urne contient, au temps 0, $A_0 = 1$ boule blanche et $B_0 = 1$ boule noire. On tire uniformément une boule, puis on la remet dans l'urne avec une nouvelle boule de cette couleur, ce qui produit A_1 et B_1 , etc. Si $(U_n)_{n \in \mathbb{N}}$ i.i.d. uniformes sur $[0, 1]$, alors

$$\begin{pmatrix} A_0 \\ B_0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} A_{n+1} \\ B_{n+1} \end{pmatrix} = \begin{pmatrix} A_n + \mathbb{1}_{U_n \leq M_n} \\ B_n + \mathbb{1}_{U_n > M_n} \end{pmatrix}, \quad \text{et} \quad \left(M_n := \frac{A_n}{A_n + B_n} = \frac{A_n}{n+2}\right)_{n \in \mathbb{N}} \text{ est une martingale.}$$

- *Processus de Galton–Watson (dynamique des populations).* À la génération n , la population comporte Z_n individus. Sachant Z_n , ces Z_n individus font des enfants de manière i.i.d. pour donner la génération $n + 1$, d'où la récurrence aléatoire $Z_{n+1} = \sum_{k=1}^{Z_n} X_{n+1,k}$, avec $X_{n+1,k}$ i.i.d. de même loi de reproduction μ sur \mathbb{N} de moyenne m . Soit $\rho \in [0, 1]$ solution de $\mathbb{E}(\rho^{Z_1} | Z_0 = 1) = \rho \leq 1$. Alors on dispose de deux martingales :

$$\rho^{Z_n} \quad \text{et} \quad \frac{Z_n}{m^n}.$$

- *Chaînes de Markov.* Si $(X_n)_{n \in \mathbb{N}}$ est une chaîne de Markov (chapitre 5) d'espace d'état E au plus dénombrable, et de générateur L , et si $h : E \rightarrow \mathbb{R}$ est harmonique : $Lh = 0$, alors $(h(X_n) - h(X_0))_{n \in \mathbb{N}}$ est une martingale.

Lemme 4.2.1. Déformations convexes.

Soit $(X_n)_{n \in \mathbb{N}}$ un processus adapté, $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ convexe telle que $\varphi(X_n) \in L^1$ pour tout n . Alors :

- (i) Si $(X_n)_{n \in \mathbb{N}}$ est une martingale alors $(\varphi(X_n))_{n \in \mathbb{N}}$ est une sous-martingale.
- (ii) Si $(X_n)_{n \in \mathbb{N}}$ est une sous-martingale et φ est croissante alors $(\varphi(X_n))_{n \in \mathbb{N}}$ est une sous-martingale.

En particulier :

- (a) si $(X_n)_{n \in \mathbb{N}}$ est une martingale alors $(|X_n|)_{n \in \mathbb{N}}$ est une sous-martingale.
- (b) si $(X_n)_{n \in \mathbb{N}}$ est une martingale de carré intégrable alors $(X_n^2)_{n \in \mathbb{N}}$ est une sous-martingale.
- (c) si $(X_n)_{n \in \mathbb{N}}$ est une sous-martingale alors $((X_n)^+)_{n \in \mathbb{N}}$ est une sous-martingale.

Comme $(X_n)_{n \in \mathbb{N}}$ est une sous-martingale ssi $(-X_n)_{n \in \mathbb{N}}$ est une surmartingale, et réciproquement, bien des résultats ne sont énoncés que pour l'une ou l'autre (attention alors aux autres hypothèses : positivité, etc.). En mathématiques, la tendance est à préférer la positivité à la négativité, la convexité à la concavité, le sous-harmonique au surharmonique, la croissance à la décroissance, les sous-martingales aux surmartingales, sauf quand ces dernières sont positives! Par ailleurs, on omet souvent de préciser la filtration quand elle peut être devinée par le contexte.

Démonstration. Découle de l'inégalité de Jensen pour l'espérance conditionnelle. ■

Si $(X_n)_{n \in \mathbb{N}}$ et $(H_n)_{n \in \mathbb{N}}$ sont deux processus, le processus $H \cdot X = ((H \cdot X)_n)_{n \in \mathbb{N}}$ est défini par

$$(H \cdot X)_0 = 0 \quad \text{et} \quad (H \cdot X)_n = \sum_{k=1}^n H_k (X_k - X_{k-1}) \quad \text{si } n \geq 1.$$

La v.a. H_0 n'intervient pas. Si $H \equiv 1$ alors $H \cdot X = X - X_0$. C'est une intégrale stochastique discrète : $(H \cdot X)_n = \int_0^n H dX$. Si X est une marche aléatoire, alors $H \cdot X$ est une version à pas pondérés de X .

Le lemme suivant est fondamental, et intervient pour établir les théorèmes d'arrêt et de convergence.

Lemme 4.2.2. Transformation de martingale ou intégrale stochastique.

On se place sur un espace de probabilité filtré (cette hypothèse est toujours implicite par la suite).

- (i) Si H est adapté (par exemple prévisible) et X est adapté alors $H \cdot X$ est adapté.
- (ii) Si H est prévisible borné et X est une martingale alors $H \cdot X$ est une martingale.
- (iii) Si H est prévisible borné positif et X est une surmartingale alors $H \cdot X$ est une surmartingale.
- (iv) Si H est prévisible borné positif et X est une sous-martingale alors $H \cdot X$ est une sous-martingale.

Dans l'intégrale discrète $(H \cdot X)_n$, le processus H est évalué à droite de chaque intervalle de temps discret $\{k-1, k\}$ lié à $X_k - X_{k-1}$, ce qui conduit à supposer H prévisible pour que $H \cdot X$ soit une martingale. On aurait pu alternativement évaluer H à gauche de chaque intervalle, ce qui aurait conduit à le supposer adapté.

Démonstration. La propriété (i) est immédiate. Pour les autres, le processus $H \cdot X$ est adapté et intégrable car H est prévisible borné et X est adapté intégrable. De plus, comme H est prévisible, pour tout $n \in \mathbb{N}$,

$$\mathbb{E}((H \cdot X)_{n+1} | \mathcal{F}_n) = \sum_{k=1}^{n+1} H_k \mathbb{E}(X_k - X_{k-1} | \mathcal{F}_n),$$

et de plus, comme X est intégrable et adapté,

$$\mathbb{E}(X_k - X_{k-1} | \mathcal{F}_n) = \begin{cases} X_k - X_{k-1} & \text{si } 1 \leq k \leq n \\ \mathbb{E}(\mathbb{E}(X_{n+1} | \mathcal{F}_n) - X_n | \mathcal{F}_n) & \text{si } k = n + 1 \end{cases},$$

ce qui permet d'exploiter la propriété de martingale sur X , et celle de sous/surmartingale si H est positif. ■

Lemme 4.2.3. Représentation de martingales ou universalité de l'intégrale stochastique discrète.

Si $(X_n)_{n \in \mathbb{N}}$ est une marche aléatoire à incréments i.i.d. de loi $\frac{1}{2}(\delta_1 + \delta_{-1})$, issue de 0, alors pour toute martingale $(M_n)_{n \in \mathbb{N}}$ pour la filtration canonique de X , il existe un processus prévisible $(H_n)_{n \in \mathbb{N}}$ t.q. $M - M_0 = H \cdot X$.

De manière équivalente $M_n - M_{n-1} = H_n(X_n - X_{n-1})$ avec $X_n - X_{n-1}$ prenant ses valeurs dans $\{\pm 1\}$. L'universalité porte sur M , pas sur X . Lorsque X est plus général, cette forme particulière n'est plus suffisante pour décrire les martingales X mesurables. Par exemple pour la martingale carrée $M_n = (Z_1 + \dots + Z_n)^2 - n$ avec $(Z_n)_{n \geq 1}$ i.i.d. centrées de variance unité, il vient $M_n - M_{n-1} = 2Z_n(Z_1 + \dots + Z_{n-1}) + Z_n^2 - 1$, qui n'est plus de la forme espérée quand $Z_n^2 \neq 1$.

Il se trouve que le lemme 4.2.3 possède un analogue à temps et espace continu pour l'intégrale stochastique d'Itô par rapport au mouvement brownien, ce qui n'est pas surprenant (phénomène TLC).

Démonstration. Pour tout $n \geq 1$, M_n est mesurable pour $\mathcal{F}_n = \sigma(X_0, X_1, \dots, X_n) = \sigma(Z_1, \dots, Z_n)$ où $Z_n = X_n - X_{n-1}$. Donc il existe $f_n : \mathbb{R}^n \rightarrow \mathbb{R}$ borélienne telle que $M_n = f_n(Z_1, \dots, Z_n)$, d'où

$$M_n = F_n^- \frac{1 - Z_n}{2} + F_n^+ \frac{1 + Z_n}{2} = \frac{F_n^+ + F_n^-}{2} + \frac{F_n^+ - F_n^-}{2} Z_n \quad \text{où} \quad F_n^\pm := f_n(Z_1, \dots, Z_{n-1}, \pm 1).$$

Or $\mathbb{E}(M_n | \mathcal{F}_{n-1}) = M_{n-1}$, et Z_n est indépendante de \mathcal{F}_{n-1} et centrée, d'où $M_{n-1} = \frac{F_n^+ + F_n^-}{2}$, ce qui donne

$$M_n - M_{n-1} = H_n Z_n \quad \text{où} \quad H_n := \frac{F_n^+ - F_n^-}{2}.$$

Observons enfin que comme $X_0 = 0$ et M_0 est $\sigma(X_0)$ -mesurable, il vient que M_0 est constante. ■

4.3 Martingales comme modélisation de stratégies

La notion mathématique de martingale² a été développée il y a un peu plus d'un siècle pour modéliser certains jeux de hasard, comme la célèbre roulette des casinos. Cette approche conduit en particulier à démontrer l'inexistence de stratégies gagnantes. Au fil du temps, les martingales sont devenues un outil mathématique puissant de la théorie des probabilités, bien au-delà de la modélisation des jeux de hasard. Alternativement, il est plaisant de penser qu'il y a des jeux de hasard dans bien des modèles aléatoires. Aussi, il est recommandé de rechercher les martingales cachées dans chaque modèle aléatoire afin de pouvoir bénéficier de l'arsenal de résultats pour ce type de suites aléatoires (arrêt, convergence, voir suite du chapitre).

Au casino, un joueur, à l'instant $n \geq 1$, mise une quantité d'argent H_n , qu'il décide librement au vu des parties précédentes. S'il gagne, il empoche $r > 1$ fois sa mise, soit rH_n , et sinon il perd sa mise H_n . Son gain au temps n est modélisable par $H_n Z_n$, où Z_n est une v.a. prenant les valeurs $\{-1, r\}$. La fortune du joueur à l'instant n est

$$\sum_{k=1}^n H_k Z_k = (H \cdot X)_n \quad \text{où} \quad X_n := Z_1 + \dots + Z_n.$$

La tribu $\mathcal{F}_n = \sigma(Z_1, \dots, Z_n, H_1, \dots, H_n) = \sigma(X_1, \dots, X_n, H_1, \dots, H_n)$ modélise l'information disponible au joueur après la n^e partie, de sorte que H_n est \mathcal{F}_{n-1} -mesurable tandis que X_n est \mathcal{F}_n -mesurable. Les processus de la forme $H \cdot X$ avec H prévisible et X marche aléatoire à pas indépendants et centrés modélisent une manière de jouer basée sur une stratégie H , et c'est ce système de jeu que Jean Ville qualifie de martingale vers 1939. Ces processus sont à l'origine du concept mathématique de martingale introduit par Joseph Leo Doob vers 1940, et non l'inverse! Autrement dit ces processus $H \cdot X$ sont des martingales originelles et archétypales.

Le casino fait en sorte que $(Z_n)_{n \geq 1}$ soit une suite de v.a. i.i.d. centrées³, ce qui fait de X une marche aléatoire à pas i.i.d. centrés. Le jeu est équitable : en misant toujours la même somme, le gain est nul en moyenne à chaque instant. Pour notre joueur, le lemme 4.2.2 est une mauvaise nouvelle, la conservation de la propriété de martingale indique que le jeu reste équitable pour tout H , la moyenne de la fortune reste constante. Il n'existe donc pas de stratégie gagnante, de choix de H qui fasse en sorte que $\mathbb{E}(M_n) > \mathbb{E}(M_0) = 0$ où $M = H \cdot X$, même pour n grand. Les martingales, qui sont une déception pour le joueur de casino, s'avèrent utiles pour l'étude de modèles aléatoires variés, précisément grâce à leur loi de conservation. Se ruiner au casino, mais gagner au jeu de la quête de certitude!

Le joueur peut décider de quitter le jeu au bout d'un temps aléatoire T , par exemple dès que sa fortune atteint une valeur donnée, grande fortune ou ruine, ou plus généralement lorsqu'un critère est vérifié par l'information disponible accumulée : cela se modélise par la contrainte $\{T = n\} \in \mathcal{F}_n$ pour tout n . Supposons que T soit fini p.s.

2. Le concept mathématique de martingale est présent, sans être nommé de la sorte, dès les années 1930, dans l'œuvre de Paul Lévy. Le terme *martingale* apparaît vers 1939 dans l'œuvre de Jean Ville comme synonyme de « système de jeu », puis dans celle de Joseph Leo Doob des années 1940 et 1950 pour nommer le concept mathématique que nous connaissons aujourd'hui.

3. La v.a. Z_n suit une loi de Bernoulli centrée à support asymétrique $\{-1, r\}$: $p := \mathbb{P}(Z_n = r) = 1 - \mathbb{P}(Z_n = -1) = 1/(r+1)$.

Alors la fortune du joueur à sa sortie du jeu est M_T . Le théorème 4.4.1 d'arrêt de Doob ci-après affirme que si T est borné p.s. alors $\mathbb{E}(M_T) = \mathbb{E}(M_0) = 0$. Par conséquent le joueur ne peut pas garantir un gain en moyenne en quittant le jeu au bout d'un temps borné! Cette modélisation ne tient pas compte du fait qu'en réalité, la fortune du joueur ne peut pas être arbitrairement négative, il y a un seuil. Le cas $r = 1$ et $H \equiv 1$ est étudié dans l'exemple 4.4.6 (ruine du joueur). Au-delà, il existe un théorème fameux de Dubbins et Savage sur les stratégies optimales [23, 48].

Devoir maison !

7 → 8

4.4 Temps d'arrêt et théorème d'arrêt de Doob

Une variable aléatoire $T : \Omega \rightarrow \mathbb{N} \cup \{\infty\}$ est un temps d'arrêt pour une filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$ lorsque

$$\{T = n\} \in \mathcal{F}_n \quad \text{pour tout } n \in \mathbb{N},$$

ou de manière équivalente, lorsque $\{T \leq n\} = \cup_{k=0}^n \{T = k\} \in \mathcal{F}_n$ pour tout $n \in \mathbb{N}$. L'exemple typique est le temps d'atteinte $T_C := \inf\{n \in \mathbb{N} : X_n \in C\}$ d'un borélien $C \subset \mathbb{R}$ par un processus adapté $(X_n)_{n \in \mathbb{N}}$:

$$\{T_C = n\} = \{X_0 \notin C, \dots, X_{n-1} \notin C, X_n \in C\} = \{X_0 \in C\}^c \cap \dots \cap \{X_{n-1} \in C\}^c \cap \{X_n \in C\} \in \mathcal{F}_n.$$

Notons que $\inf \emptyset = +\infty$. L'événement $\{T = +\infty\}$ signifie que X n'atteint jamais C . De manière générale,

$$\{T = +\infty\} = \left(\cup_{n=0}^{\infty} \{T = n\} \right)^c \in \sigma \left(\cup_{n=0}^{\infty} \mathcal{F}_n \right).$$

Un temps d'arrêt modélise un temps de prise de décision, et cela ne peut se faire qu'avec l'information disponible (accumulée) au moment de la décision, d'où la définition. En particulier, pour le temps d'atteinte, l'événement $\{T_C = n\}$ ne fait intervenir la connaissance de la trajectoire de X que jusqu'au temps n . Contrexemple : pour tout N fixé, la v.a. $S_C := \sup\{n \leq N : X_n \in C\}$ n'est pas un temps d'arrêt, car

$$\{S_C = n\} = \{X_n \in C\} \cap \{X_{n+1} \in C\}^c \cap \dots \cap \{X_N \in C\}^c \notin \mathcal{F}_n \text{ en général.}$$

Il est possible de s'arrêter au premier feu rouge, mais pas au dernier car cette notion dépend du futur, inconnu.

Si $X = (X_n)_{n \in \mathbb{N}}$ est adapté et T est un temps d'arrêt alors le processus arrêté $X^T := (X_{n \wedge T})_{n \in \mathbb{N}}$ est adapté. En effet, cela découle du lemme 4.2.2 et de la décomposition

$$X_{n \wedge T} = X_0 + \sum_{k=1}^n \mathbb{1}_{\{T \geq k\}} (X_k - X_{k-1}) = X_0 + (H \cdot X)_n \quad \text{où } H_n := \mathbb{1}_{\{T \geq n\}} = 1 - \mathbb{1}_{\{T < n\}},$$

car H est prévisible donc adapté, car T est un temps d'arrêt : $\{T < n\} = \{T \leq n-1\} \in \mathcal{F}_{n-1} \subset \mathcal{F}_n$. Alternativement, on peut découper selon la valeur de T pour se ramener au cas déterministe : pour tout borélien $C \subset \mathbb{R}$ et tout $n \in \mathbb{N}$,

$$\{X_{n \wedge T} \in C\} = \cup_{k \in \mathbb{N} \cup \{\infty\}} (\{X_{n \wedge T} \in C\} \cap \{T = k\}) = \cup_{k=0}^n (\{X_k \in C\} \cap \{T = k\}) \cup (\{X_n \in C\} \cap \{T \leq n\}^c) \in \mathcal{F}_n.$$

L'arrêt est une technique de troncature ou seuillage qui conserve l'adaptation! La trajectoire du processus arrêté X^T suit celle de X puis reste constante (s'arrête) à partir du temps (aléatoire) T . Cela conduit au théorème suivant :

Dessin !

Théorème 4.4.1. Théorème d'arrêt de Doob.

Soit $(M_n)_{n \in \mathbb{N}}$ une martingale et T un temps d'arrêt (sous-entendu pour la même filtration). Alors :

- (i) Le processus arrêté $M^T = (M_{n \wedge T})_{n \in \mathbb{N}}$ est une martingale, en particulier

$$M_{n \wedge T} \in L^1 \quad \text{et} \quad \mathbb{E}(M_{n \wedge T}) = \mathbb{E}(M_0) \quad \text{pour tout } n.$$

- (ii) Si T vérifie l'une des trois conditions suivantes :

- (a) T est borné (p.s. par une constante).
- (b) T est fini (p.s.) et $(M_{n \wedge T})_{n \in \mathbb{N}}$ est u.i. (par exemple dominée par une v.a. intégrable).
- (c) T est intégrable et M est à oscillations bornées : $\sup_{n \in \mathbb{N}} \mathbb{E}(|M_{n+1} - M_n| | \mathcal{F}_n)$ bornée p.s..

alors

$$M_T \in L^1 \quad \text{et} \quad \mathbb{E}(M_T) = \mathbb{E}(M_0).$$

De même en remplaçant martingale par sur ou sous-martingale, et $=$ par \leq et \geq respectivement.

Démonstration.

- (i) Pour tout $n \in \mathbb{N}$, $M_{n \wedge T} = M_0 + (H \cdot M)_n$ où $H_n := \mathbb{1}_{\{T \geq n\}} = 1 - \mathbb{1}_{\{T < n\}}$ (somme télescopique).
 Or le processus $(H_n)_{n \geq 1}$ est borné et prévisible. Le résultat découle alors du lemme 4.2.2.
 On peut aussi réduire au déterminisme par découpage : $M_{(n+1) \wedge T} = \sum_{k=0}^n M_k \mathbb{1}_{\{T=k\}} + M_{n+1} \mathbb{1}_{\{T \leq n\}}^c$.
- (ii) (a) Soit N tel que $T \leq N$ p.s. Alors par (i), $M_T = M_{N \wedge T} = M_N^T \in L^1$ et $\mathbb{E}(M_T) = \mathbb{E}(M_{N \wedge T}) = \mathbb{E}(M_0)$.
 (b) Comme $T < \infty$ p.s. la v.a. M_T fait sens p.s. et comme $M_{n \wedge T} \rightarrow M_T$ p.s. et $(M_{n \wedge T})_{n \in \mathbb{N}}$ est u.i., le théorème de convergence dominée généralisé u.i. donne $M_T \in L^1$ et $\mathbb{E}(M_{n \wedge T}) \rightarrow \mathbb{E}(M_T)$.
 (c) Comme $\mathbb{E}(T) < \infty$, il vient que $T < \infty$ p.s. ce qui donne un sens à M_T p.s. et nous savons que $M_{n \wedge T} \rightarrow M_T$ p.s. Montrons que $(M_{n \wedge T})_{n \in \mathbb{N}}$ est dominée par une v.a. intégrable, ce qui donnera immédiatement le résultat par le théorème de convergence dominée. Nous avons

$$M_{n \wedge T} = M_0 + \sum_{k=1}^{n \wedge T} (M_k - M_{k-1}) = M_0 + \sum_{k=1}^{\infty} (M_k - M_{k-1}) \mathbb{1}_{\{n \wedge T \geq k\}}.$$

Or le théorème de convergence monotone⁴ et les hypothèses faites sur T et M donnent

$$\mathbb{E} \left| \sum_{k=1}^{\infty} (M_k - M_{k-1}) \mathbb{1}_{\{n \wedge T \geq k\}} \right| \leq \sum_{k=1}^{\infty} \mathbb{E}(\mathbb{1}_{\{T \geq k\}} \mathbb{E}(|M_k - M_{k-1}| | \mathcal{F}_{k-1})) \leq C \mathbb{E}(T) < \infty.$$

Donc $|M_{n \wedge T}| \leq |M_0| + \sum_{k=1}^{\infty} \mathbb{1}_{\{T \geq k\}} |M_k - M_{k-1}| \in L^1$. ■

Remarque 4.4.2. Théorème d'arrêt pour les sous-martingales et les surmartingales.

Dans la dernière propriété, le temps 0 peut être remplacé par un temps d'arrêt S , avec la contrainte $S \leq T$ dans le cas sous-martingale ou surmartingale. Pour le voir, il est possible d'utiliser le fait que si $T \leq N$ p.s. alors $X_T - X_S = (H \cdot X)_N$ où H est donné par $H_n := \mathbb{1}_{\{S < n \leq T\}} = \mathbb{1}_{\{S \leq n-1\}} - \mathbb{1}_{\{T \leq n-1\}} \geq 0$.

Exemple 4.4.3. Application du théorème d'arrêt au temps de sortie d'un intervalle.

Soit $(M_n)_{n \in \mathbb{N}}$ une martingale à valeurs dans \mathbb{Z} et à accroissements aux plus proches voisins (assure une continuité discrète des trajectoires) : $M_{n+1} - M_n \in \{-1, 0, 1\}$ pour tout n . Soient $a < b$ dans \mathbb{Z} , et $T = \inf\{n \in \mathbb{N} : M_n \in \{a, b\}\}$. Supposons que M_0 prend ses valeurs dans $[a, b]$. Le processus arrêté $(M_{n \wedge T})_{n \in \mathbb{N}}$ est alors borné par construction, en particulier u.i. Supposons de plus que T est fini p.s. Alors le théorème 4.4.1 d'arrêt de Doob donne $x = \mathbb{E}(M_0) = \mathbb{E}(M_T) = a \mathbb{P}(M_T = a) + b \mathbb{P}(M_T = b)$, tandis que $\mathbb{P}(M_T = a) + \mathbb{P}(M_T = b) = 1$, d'où

$$\mathbb{P}(M_T = a) = \frac{b-x}{b-a} \quad \text{et} \quad \mathbb{P}(M_T = b) = \frac{x-a}{b-a}.$$

Considérons à présent le cas particulier de la marche aléatoire $M_n = M_0 + Z_1 + \dots + Z_n$, où $(Z_n)_{n \geq 1}$ est une suite de v.a. i.i.d. centrées de variance unité, et M_0 est \mathcal{F}_0 -mesurable. Alors $\frac{M_n - M_0}{\sqrt{n}} \rightarrow \mathcal{N}(0, 1)$ en loi (TLC), d'où $\mathbb{P}(T = \infty) \leq \mathbb{P}(T > n) \leq \mathbb{P}(M_n \in [a, b]) \rightarrow 0$, d'où $T < \infty$ p.s. Comme M est à oscillations bornées, il aurait suffi d'établir que T est intégrable. En fait T est majoré par une v.a. géométrique non-dégénérée^a, il admet donc un moment exponentiel et en particulier tous ses moments sont finis, en particulier T est intégrable. D'autre part, en utilisant une martingale exponentielle, il est possible d'obtenir la transformée de Laplace de T . Voir aussi l'exemple 4.4.6 (ruine du joueur).

^a En effet, $\{T < \infty\} \supset \cup_{n \in \mathbb{N}} A_n \supset \overline{\lim}_n A_n$ où $A_n := \{\sup_{0 \leq k \leq \ell-1} |M_{n\ell+k} - M_{n\ell}| > \ell\}$, du moment que $\ell > b - a$. Comme les incréments de M sont i.i.d., les A_n sont indépendants et de même probabilité, et d'autre part, comme les accroissements ne sont pas identiquement nuls, $p := \mathbb{P}(A_n) > 0$. Le lemme de Borel-Cantelli donne alors $T < \infty$ p.s. En fait, la suite $(\mathbb{1}_{A_n})_{n \in \mathbb{N}}$ est un jeu de pile ou face, et $T \leq T'$ où $T' := \inf\{n \in \mathbb{N} : \mathbb{1}_{A_n} = 1\} \sim \text{Geom}_{\mathbb{N}}(p)$ est le temps de premier succès du jeu de pile ou face.

Exemple 4.4.4. Contrexemple.

Si M est une martingale, et $x \neq \mathbb{E}(M_0)$, alors le temps d'arrêt d'atteinte $T = \inf\{n \in \mathbb{N} : M_n = x\}$ ne peut pas vérifier les hypothèses du théorème 4.4.1 car cela donnerait $\mathbb{E}(M_0) = \mathbb{E}(M_T) = x$. Dans le cas de la marche aléatoire simple symétrique sur \mathbb{Z} , M est à oscillations bornées, par conséquent, d'après le théorème d'arrêt de Doob, T n'est pas intégrable. De plus, T n'est pas fini p.s. ou bien $(M_{n \wedge T})_{n \in \mathbb{N}}$ n'est pas u.i.

4. Ou le théorème de Fubini-Tonelli, pour commuter \mathbb{E} et $\sum_{n=0}^{\infty}$.

Exemple 4.4.5. Première occurrence de motif.

Considérons un jeu de pile ou face modélisé par une suite $X = (X_n)_{n \geq 1}$ de v.a. i.i.d. de loi de Bernoulli $\mathbb{P}(X_1 = 1) = p$, $\mathbb{P}(X_1 = 0) = q = 1 - p$, $0 < p \leq 1$, et le temps de première occurrence du motif 111 dans X :

$$T = \inf\{n \geq 3 : X_n = X_{n-1} = X_{n-2} = 1\}.$$

Définissons la suite $(U_n)_{n \in \mathbb{N}}$ par

$$U_0 := 0 \quad \text{et} \quad U_{n+1} = (U_n + 1) \frac{X_{n+1}}{p},$$

de sorte que $U_T \equiv u_3 := \frac{1}{p}(1 + \frac{1}{p}(1 + \frac{1}{p}))$, quantité déterministe, égale à U_3 sur $(X_0, X_1, X_2) = (1, 1, 1)$, car sinon on retombe à zéro. D'autre part, $(M_n := U_n - n)_{n \in \mathbb{N}}$ est une martingale pour la filtration naturelle de X , vérifiant $M_0 = 0$. L'élimination des recouvrements partiels conduit à $T \leq 3T'$ où $T' = \inf\{n \in \mathbb{N} : (X_{3n}, X_{3n+1}, X_{3n+2}) = (1, 1, 1)\}$, et T' suit la loi géométrique de paramètre $p^3 > 0$, donc intégrable et en particulier finie p.s. Il en va donc de même de T . En outre, le processus arrêté $M^T = (M_{n \wedge T})_{n \in \mathbb{N}}$ est majoré par $u_3 + T \leq u_3 + 3T' \in L^1$. Donc M^T est u.i., et le théorème 4.4.1 d'arrêt de Doob donne alors $\mathbb{E}(U_T) = \mathbb{E}(T)$, d'où

$$\mathbb{E}(T) = \mathbb{E}(U_T) = U_T = u_3 = \frac{1}{p} \left(1 + \frac{1}{p} \left(1 + \frac{1}{p} \right) \right) = \frac{1}{p} + \frac{1}{p^2} + \frac{1}{p^3} \left(\leq \frac{3}{p^3} = 3\mathbb{E}(T') \right).$$

Dans cet exemple, la martingale est purement instrumentale, et fournit à la fin une formule pour $\mathbb{E}(T)$.

8 → 9

Vu aussi en TD !

Exemple 4.4.6. Ruine du joueur de pile ou face.

Un joueur de pile ou face possède au temps $n = 0$ une fortune initiale de a €. Son but est d'atteindre la richesse de b €, avant d'être ruiné. On suppose que $a, b \in \mathbb{N}$, et $b > a > 0$. Au temps $n \geq 1$, il joue à pile ou face, gagne 1 € avec probabilité $p \in (0, 1)$, et perd 1 € avec probabilité $q = 1 - p$, et ces parties successives sont i.i.d. Son gain à l'issue de la n^{e} partie est $S_n = Z_1 + \dots + Z_n$, où $(Z_n)_{n \geq 1}$ est une suite de v.a. i.i.d. de loi de Rademacher $p\delta_1 + q\delta_{-1}$, portée par $\{-1, 1\}$, de moyenne $m := p - q = 2p - 1$ et de variance $\sigma^2 := 4pq$. Posons $S_0 := 0$. Le temps d'arrêt d'intérêt est

$$T = \inf\{n \geq 1 : S_n \in \{-a, b\}\}.$$

Le jeu se termine par la ruine si $S_T = -a$ et par la fortune si $S_T = b$. Le processus $(S_n - nm)_{n \in \mathbb{N}}$ est une marche aléatoire centrée et en procédant comme dans l'exemple 4.4.3, on obtient que T est majoré par une v.a. géométrique, il est donc en particulier de carré intégrable. Considérons les martingales

$$A := (S_n - nm)_{n \in \mathbb{N}} \quad \text{et} \quad B := ((S_n - nm)^2 - n\sigma^2)_{n \in \mathbb{N}}.$$

Les martingales arrêtées A^T et B^T ne sont pas forcément bornées, mais sont u.i. car dominées par une v.a. intégrable. En effet, comme S^T est bornée par construction et que $T \in L^2$ (donc $T \in L^1$), on a

$$|A_n^T| \leq |S_n^T| + |m|T \in L^1 \quad \text{et} \quad |B_n^T| \leq 2|S_n^T|^2 + 2m^2T^2 + \sigma^2T \in L^1.$$

Le théorème 4.4.1 d'arrêt de Doob donne alors

$$\mathbb{E}(S_T) = m\mathbb{E}(T) \quad \text{et} \quad \mathbb{E}((S_T - mT)^2) = \sigma^2\mathbb{E}(T).$$

Considérons le cas équitable $p = q = 1/2$. Alors $m = 0$ et $\sigma^2 = 1$, ce qui donne

$$0 = \mathbb{E}(S_T) = -a\mathbb{P}(S_T = -a) + b\mathbb{P}(S_T = b) \quad \text{et} \quad \mathbb{E}(T) = \mathbb{E}(S_T^2) = a^2\mathbb{P}(S_T = -a) + b^2\mathbb{E}(S_T = b).$$

Comme $\mathbb{P}(S_T = -a) + \mathbb{P}(S_T = b) = 1$, il en découle que dans le cas équitable, la probabilité que le jeu se termine par la fortune et la durée moyenne du jeu sont données par

$$\mathbb{P}(S_T = b) = \frac{a}{a+b} \quad \text{et} \quad \mathbb{E}(T) = ab.$$

Considérons le cas inéquitable $p \neq 1/2$, c'est-à-dire $q/p \neq 1$. On dispose de la martingale

$$C := \left(\left(\frac{q}{p} \right)^{S_n} \right)_{n \in \mathbb{N}}.$$

La martingale arrêtée C^T est bornée car S^T est bornée, et le théorème 4.4.1 d'arrêt de Doob donne

$$1 = \mathbb{E}(C_0) = \mathbb{E}(C_T) = \left(\left(\frac{q}{p} \right)^b - \left(\frac{q}{p} \right)^{-a} \right) \mathbb{P}(S_T = b) + \left(\frac{q}{p} \right)^{-a},$$

d'où

$$\mathbb{P}(S_T = b) = \frac{p^b q^a - p^{a+b}}{q^{a+b} - p^{a+b}} \quad \text{et} \quad \mathbb{E}(T) = \frac{\mathbb{E}(S_T)}{m} = \frac{1}{p-q} \left((a+b) \frac{p^b q^a - p^{a+b}}{q^{a+b} - p^{a+b}} - a \right).$$

Il est enfin possible d'obtenir la transformée de Laplace de T en utilisant la martingale

$$\left(e^{\alpha S_n - n \log L(\alpha)} \right)_{n \in \mathbb{N}} = \left(\frac{e^{\alpha S_n}}{L(\alpha)^n} \right)_{n \in \mathbb{N}} \quad \text{où} \quad L(\alpha) := \mathbb{E}(e^{\alpha Z_1}) = pe^\alpha + qe^{-\alpha}, \alpha > 0.$$

Théorème 4.4.7. Caractérisation des martingales ou réciproque du théorème d'arrêt.

Pour tout processus $(X_n)_{n \in \mathbb{N}}$, les deux propriétés suivantes sont équivalentes :

- (i) X est adapté et intégrable et $\mathbb{E}(X_T) = \mathbb{E}(X_0)$ pour tout temps d'arrêt T borné.
- (ii) X est une martingale.

Démonstration. (ii) \Rightarrow (i). Cela provient de la définition de martingale et du théorème 4.4.1 d'arrêt de Doob.

(i) \Rightarrow (ii). Pour tout n et tout $A \in \mathcal{F}_n$, (i) appliquée aux temps d'arrêt bornés $T = n\mathbb{1}_A + (n+1)\mathbb{1}_{A^c}$ et $T = n+1$, donne $\mathbb{E}(X_0) = \mathbb{E}(X_n \mathbb{1}_A + X_{n+1} \mathbb{1}_{A^c}) = \mathbb{E}(X_{n+1})$, d'où $\mathbb{E}(X_{n+1} \mathbb{1}_A) = \mathbb{E}(X_n \mathbb{1}_A)$, d'où $\mathbb{E}(X_{n+1} | \mathcal{F}_n) = X_n$. ■

4.5 Arrêt optimal et enveloppe de Snell

Le problème du contrôle consiste à déterminer les valeurs des paramètres d'un système dynamique pour qu'il adopte un comportement donné : comment régler à chaque instant les moteurs d'une fusée pour qu'elle décolle de la terre et se pose sur la lune! Lorsque les ingrédients sont aléatoires on parle de contrôle stochastique. Cela permet de modéliser la prise de décision⁵ : faut-il approvisionner le stock? louer l'appartement qu'on vient de visiter? vendre ou acheter un actif? arrêter ou démarrer la centrale nucléaire? terminer le processus de recrutement? arrêter de suivre ce cours? Ces exemples sont tous des cas particuliers du problème de l'arrêt optimal.

Le problème de l'arrêt optimal, emblématique du contrôle stochastique, peut être formalisé comme suit : pour un processus adapté et intégrable $(X_n)_{n \in \mathbb{N}}$ et un temps fini N appelé horizon, on s'intéresse à

$$S_N := \sup_{T \in \mathcal{T}_N} \mathbb{E}(X_T)$$

où \mathcal{T}_N est l'ensemble des temps d'arrêts $T \leq N$. Un temps d'arrêt $T_* \in \mathcal{T}_N$ est optimal lorsque $\mathbb{E}(X_{T_*}) = S_N$. Le but ultime est de proposer un algorithme pour le calcul approché de T_* et de $S_N = \mathbb{E}(X_{T_*})$. Comme l'horizon N est fini, on peut se contenter de ne connaître le processus X que jusqu'au temps $N : X_0, \dots, X_N$.

Le problème est élémentaire si X est une martingale car dans ce cas le théorème 4.4.1 d'arrêt de Doob donne $\mathbb{E}(X_T) = \mathbb{E}(X_0)$ pour tout $T \in \mathcal{T}_N$, et tout temps d'arrêt dans \mathcal{T}_N est optimal, en particulier $T = 0$.

Lorsque X n'est pas une martingale, le théorème 4.5.1 ci-dessous construit la meilleure majoration de X par une surmartingale, et donc, en l'utilisant pour $-X$, la meilleure minoration de X par une sous-martingale. Ces deux processus enveloppants coïncident au temps terminal N , et en tout temps $0 \leq n \leq N$ lorsque X est une martingale.

Théodéf 4.5.1. Enveloppe de Snell.

Soit $(X_n)_{0 \leq n \leq N}$ un processus à horizon fini N , adapté à une filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$, et intégrable. Alors :

- (i) La récurrence rétrograde suivante fait sens et définit une surmartingale $(Y_n)_{0 \leq n \leq N}$ majorant X :

$$Y_N := X_N \quad \text{et} \quad Y_n := \max(X_n, \mathbb{E}(Y_{n+1} | \mathcal{F}_n)) \quad \text{pour } 0 \leq n \leq N-1.$$

- (ii) Le processus Y est la plus petite surmartingale majorant X .
- (iii) La v.a. $T_* := \min\{0 \leq n \leq N : Y_n = X_n\}$ est un temps d'arrêt et $Y^{T_*} = (Y_{n \wedge T_*})_{0 \leq n \leq N}$ est une martingale.
- (iv) Le temps d'arrêt T_* est optimal et $\mathbb{E}(Y_0) = \sup_{T \in \mathcal{T}_N} \mathbb{E}(X_T) = \mathbb{E}(X_{T_*}) = \mathbb{E}(Y_{T_*})$.

5. C'est l'occasion de signaler que l'université Paris-Dauphine - PSL compte un Département de formation d'Informatique et de Mathématiques de la Décision (MIDO) ainsi qu'un Centre de Recherche en Mathématiques de la Décision (CEREMADE). Le master offre notamment des cours de contrôle des chaînes de Markov (M1) et de contrôle des équations différentielles stochastiques (M2).

Le processus Y est appelé enveloppe de Snell de X , c'est une enveloppe par le haut, majorante.

J. L. Snell était un élève de J. L. Doob. Il a développé les applications des martingales notamment en finance mathématique. Il a part ailleurs écrit un livre [22] avec P. G. Doyle sur les marches aléatoires et les réseaux électriques.

Démonstration.

(i) Le processus Y est bien défini car $Y_N = X_N \in L^1$, et $Y_n \in L^1 \Rightarrow Y_{n-1} \in L^1$ car

$$\mathbb{E}(|Y_{n-1}|) \leq \mathbb{E}(|X_{n-1}|) + \mathbb{E}(|\mathbb{E}(Y_n | \mathcal{F}_{n-1})|) \leq \mathbb{E}(|X_{n-1}|) + \mathbb{E}(|Y_n|) < +\infty.$$

De plus, par construction, Y_n est \mathcal{F}_n -mesurable, et $Y_n \geq \mathbb{E}(Y_{n+1} | \mathcal{F}_n)$.

(ii) Soit $(Z_n)_{0 \leq n \leq N}$ une surmartingale majorante : $Z \geq X$. Par définition, $Z_N \geq X_N = Y_N$. Montrons que $Z_n \geq Y_n \Rightarrow Z_{n-1} \geq Y_{n-1}$. Comme Z est une surmartingale et $Z_n \geq Y_n$, il vient que

$$Z_{n-1} \geq \mathbb{E}(Z_n | \mathcal{F}_{n-1}) \geq \mathbb{E}(Y_n | \mathcal{F}_{n-1}),$$

or $Z_{n-1} \geq X_{n-1}$, d'où $Z_{n-1} \geq \max(X_{n-1}, \mathbb{E}(Y_n | \mathcal{F}_{n-1})) = Y_{n-1}$.

(iii) Comme $T_* = \min\{0 \leq n \leq N : Y_n - X_n = 0\}$, et $Y - X$ est adapté, c'est un temps d'arrêt.

Sur $\{T_* \geq n+1\}$, il vient que $Y_n > X_n$ donc $Y_n = \mathbb{E}(Y_{n+1} | \mathcal{F}_n)$, d'où

$$Y_{n+1}^{T_*} - Y_n^{T_*} = (Y_{n+1} - Y_n) \mathbb{1}_{\{T_* \geq n+1\}} = (Y_{n+1} - \mathbb{E}(Y_{n+1} | \mathcal{F}_n)) \mathbb{1}_{\{T_* \geq n+1\}}.$$

Or $\{T_* \geq n+1\} = \{T_* \leq n\}^c \in \mathcal{F}_n$, d'où la propriété de martingale pour Y^{T_*} .

(iv) La combinaison du fait que $X_{T_*} = Y_{T_*}$ et du théorème 4.4.1 d'arrêt de Doob pour la martingale Y^{T_*} donne

$$\mathbb{E}(X_{T_*}) = \mathbb{E}(Y_{T_*}) = \mathbb{E}(Y_0).$$

De plus, comme Y est une surmartingale, par le théorème 4.4.1 d'arrêt de Doob, $\mathbb{E}(Y_T) \leq \mathbb{E}(Y_0)$. Or $Y \geq X$ donne en particulier $\mathbb{E}(Y_T) \geq \mathbb{E}(X_T)$. Il en découle que $\mathbb{E}(X_T) \leq \mathbb{E}(Y_0)$, avec égalité si $T = T_*$. ■

Exemple 4.5.2. Problème du recrutement.

Examinons le cas d'un problème d'arrêt optimal célèbre et relativement accessible à l'étude mathématique, connu sous le nom de *problème des secrétaires* ou *problème du mariage*. Il s'agit d'examiner successivement des candidatures pour un recrutement, et d'arrêter le processus de recrutement en un temps d'arrêt qui maximise la probabilité de recruter la meilleure candidature. S'il y a N candidatures au total, on modélise la qualité de la i^e candidature par une v.a. Z_i , et on suppose que les Z_1, \dots, Z_N sont i.i.d. et de loi continue pour éviter les ex aequo. Il existe alors p.s. une unique permutation σ de $\{1, \dots, N\}$ telle que $Z_{\sigma(1)} > \dots > Z_{\sigma(N)}$. Cette permutation aléatoire σ suit la loi uniforme sur le groupe symétrique. Le temps aléatoire

$$\inf\{1 \leq n \leq N : \sigma(n) = 1\} = \sigma^{-1}(1)$$

n'est pas utilisable car ce n'est pas un temps d'arrêt pour la filtration $(\mathcal{F}_n)_{n \geq 1}$, où $\mathcal{F}_n := \sigma(Z_1, \dots, Z_n)$. Il dépend de toutes les candidatures. Pour pallier ce problème, considérons, pour tout $0 \leq n \leq N$, la v.a.

$$X_n := \mathbb{P}(\sigma(n) = 1 | \mathcal{F}_n) = \mathbb{E}(\mathbb{1}_{\{\sigma(n)=1\}} | \mathcal{F}_n).$$

Il s'agit de la probabilité que la n^e candidature soit la meilleure des N sachant l'information disponible au temps n , modélisée par \mathcal{F}_n . Pour tout temps d'arrêt $T \leq N$, on a

$$\begin{aligned} \mathbb{E}(X_T) &= \sum_{n=1}^N \mathbb{E}(\mathbb{1}_{\{T=n\}} X_n) = \sum_{n=1}^N \mathbb{E}(\mathbb{1}_{\{T=n\}} \mathbb{E}(\mathbb{1}_{\{\sigma(n)=1\}} | \mathcal{F}_n)) \\ &= \sum_{n=1}^N \mathbb{E}(\mathbb{E}(\mathbb{1}_{\{T=n\}} \mathbb{1}_{\{\sigma(n)=1\}} | \mathcal{F}_n)) \\ &= \sum_{n=1}^N \mathbb{E}(\mathbb{E}(\mathbb{1}_{\{T=n\}} \mathbb{1}_{\{\sigma(T)=1\}} | \mathcal{F}_n)) \\ &= \mathbb{P}(\sigma(T) = 1). \end{aligned}$$

Ainsi, trouver la meilleure stratégie en optimisant T revient à considérer l'enveloppe de Snell de X .

Pour certains modèles aléatoires, il est possible d'expliciter T_* .

Ceci n'est pas abordé lors du cours oral.

Exemple 4.5.3: Problème du recrutement : calcul effectif

Poursuivons l'étude entamée dans l'exemple 4.5.2. On dispose de la formule explicite

$$X_n = \frac{n}{N} \mathbb{1}_{\{R_n=1\}},$$

où R_n est le rang de la n^e candidature parmi les candidatures $1, \dots, n$ examinées au temps n . Ce rang est plus petit que ou égal au rang $\sigma(n)$ par rapport à toutes les candidatures $1, \dots, N$. Notons que $R_1 = 1$ et $R_N = \sigma(N)$. Intuitivement n/N est la probabilité pour que la meilleure candidature figure parmi les n premières candidatures examinées, tandis que $\{R_n = 1\}$ exprime le fait qu'elle est la meilleure au temps n . Pour établir la formule, comme $\{\sigma(n) = 1\} = \{R_n = 1, R_{n+1} \neq 1, \dots, R_N \neq 1\}$, et comme les v.a. R_1, \dots, R_N sont indépendantes avec R_n \mathcal{F}_n -mesurable et de loi uniforme sur $\{1, \dots, n\}$ (cette propriété nécessite une démonstration, omise ici), il vient

$$X_n = \mathbb{E}(\mathbb{1}_{\{\sigma(n)=1\}} | \mathcal{F}_n) = \mathbb{1}_{\{R_n=1\}} \mathbb{P}(R_{n+1} \neq 1, \dots, R_N \neq 1 | \mathcal{F}_n) = \mathbb{1}_{\{R_n=1\}} \frac{n}{n+1} \cdots \frac{N-1}{N} = \frac{n}{N} \mathbb{1}_{\{R_n=1\}}.$$

Calculons à présent l'enveloppe de Snell Y de X , dont la définition rétrograde s'écrit

$$Y_N = \mathbb{1}_{\{R_N=1\}} \quad \text{et} \quad Y_n = \max\left(\frac{n}{N} \mathbb{1}_{\{R_n=1\}}, \mathbb{E}(Y_{n+1} | \mathcal{F}_n)\right).$$

À ce stade, l'indépendance des R_1, \dots, R_N et une récurrence rétrograde donnent que Y_n est une fonction de R_n , d'où $\mathbb{E}(Y_{n+1} | \mathcal{F}_n) = \mathbb{E}(Y_{n+1})$. Soit à présent $f : \{0, \dots, N\} \rightarrow \mathbb{R}$ définie par

$$f(n) := \mathbb{E}(Y_{n+1}).$$

Cette fonction f est décroissante (Y est une surmartingale!). Le temps d'arrêt optimal T_* est donc

$$T_* = N \wedge \inf\left\{1 \leq n \leq N : \frac{n}{N} \mathbb{1}_{\{R_n=1\}} \geq f(n)\right\}.$$

Comme f est décroissante, il existe donc un entier déterministe m dépendant de f tel que

$$T_* = T_m \quad \text{où} \quad T_m := N \wedge \inf\{m \leq n \leq N : R_n = 1\}.$$

Pour trouver m , il suffit de calculer $g(m) := \mathbb{E}(X_{T_m})$ et d'optimiser en m . Or

$$\begin{aligned} g(m) &= \mathbb{P}(\sigma(T_m) = 1) = \sum_{n=m}^N \mathbb{E}(\mathbb{1}_{\{T_m=n\} \cap \{\sigma(n)=1\}}) \\ &= \sum_{n=m}^N \mathbb{P}(R_m \neq 1, \dots, R_{n-1} \neq 1, R_n = 1, R_{n+1} \neq 1, \dots, R_N \neq 1) \\ &= \sum_{n=m}^N \frac{1}{n} \prod_{\substack{i=m \\ i \neq n}}^N \frac{i-1}{i} = \frac{m-1}{N} \sum_{n=m}^N \frac{1}{n-1}. \end{aligned}$$

À présent, pour tout $x \in (0, 1)$, une comparaison série-intégrale donne

$$g(\lfloor xN \rfloor) = \frac{\lfloor xN \rfloor}{N} \sum_{n=\lfloor xN \rfloor}^N \frac{1}{n-1} \xrightarrow{N \rightarrow \infty} -x \log(x).$$

Ainsi, quand N est grand, la valeur maximale $\approx 1/e$ est atteinte en $m \approx N/e$. Cela signifie au bout du compte qu'une stratégie optimale approchée consiste à rejeter les $\lfloor N/e \rfloor \approx 0.37N$ premières candidatures, puis à poursuivre et s'arrêter à la première candidature qui fait mieux que toutes celles déjà examinées! En pratique toutefois, l'hypothèse faite sur la qualité des candidatures est discutable. ...

4.6 Tribu d'arrêt et théorème d'arrêt généralisé

Lemme 4.6.1. Propriétés des temps d'arrêt et tribu d'arrêt.

- (i) Si S et T sont des temps d'arrêt, alors $S \wedge T$ et $S \vee T$ le sont.
- (ii) Si $(T_k)_{k \in \mathbb{N}}$ est une suite de temps d'arrêt, alors $\inf_k T_k$, $\sup_k T_k$, $\underline{\lim}_k T_k$, et $\overline{\lim}_k T_k$ le sont.
- (iii) Si T est un temps d'arrêt, alors la famille d'événements suivante est une tribu :

$$\mathcal{F}_T := \{A \in \mathcal{F} : \forall n, A \cap \{T = n\} \in \mathcal{F}_n\}.$$

De plus T est \mathcal{F}_T -mesurable. Enfin si T est constant et égal à n , alors $\mathcal{F}_T = \mathcal{F}_n$, d'où la notation.

- (iv) Si S et T sont des temps d'arrêt et $S \leq T$ alors $\mathcal{F}_S \subset \mathcal{F}_T$, en particulier $\mathcal{F}_{n \wedge T} \subset \mathcal{F}_T$ pour tout $n \in \mathbb{N}$.

La tribu \mathcal{F}_T est appelée tribu d'arrêt de T ou tribu des événements antérieurs à T . Elle regroupe tous les événements qu'on peut écrire avec les processus adaptés arrêtés en T , c'est-à-dire avec leurs trajectoires jusqu'au temps aléatoire T , c'est-à-dire avec $(X_n)_{0 \leq n \leq T}$, c'est-à-dire le passé du processus par rapport au temps aléatoire T .

Démonstration.

- (i) Découle de $\{S \wedge T \leq n\} = \{S \leq n\} \cup \{T \leq n\}$, et $\{S \vee T \leq n\} = \{S \leq n\} \cap \{T \leq n\}$.
- (ii) Découle des formules $\{\inf_k T_k \leq n\} = \cup_k \{T_k \leq n\}$ et $\{\liminf_k T_k \leq n\} = \cup_m \cap_{k \geq m} \{T_k \leq n\}$, et, symétriquement, des formules $\{\sup_k T_k \leq n\} = \cap_k \{T_k \leq n\}$ et $\{\limsup_k T_k \leq n\} = \cap_m \cup_{k \geq m} \{T_k \leq n\}$.
- (iii) Se vérifie directement à partir des définitions.
- (iv) Si $A \in \mathcal{F}_S$, alors pour tout n , $A \cap \{T = n\} = \cup_{k=0}^n (A \cap \{S = k\}) \cap \{T = n\} \in \mathcal{F}_n$.

■

Théorème 4.6.2. Extension du théorème 4.4.1 d'arrêt de Doob pour les tribus d'arrêt.

Soit M une martingale et S et T deux temps d'arrêt tels que $S \leq T$.

Si S et T vérifient l'une des hypothèses du (ii) du théorème 4.4.1 d'arrêt de Doob, alors $\mathbb{E}(M_T | \mathcal{F}_S) = M_S$.

Il en découle que M est une martingale ssi pour toute suite croissante $(T_n)_{n \in \mathbb{N}}$ de temps d'arrêts bornés, le processus $(M_{T_n})_{n \in \mathbb{N}}$ est une martingale pour la filtration $(\mathcal{F}_{T_n})_{n \in \mathbb{N}}$. En effet, la nécessité de la condition provient du théorème, tandis que sa suffisance s'obtient en prenant $T_n = n$ car alors $\mathcal{F}_{T_n} = \mathcal{F}_n$.

Ce théorème d'arrêt pour les tribus d'arrêt est lui aussi disponible pour les sous-martingales et les surmartingales, en remplaçant $=$ par \leq et \geq respectivement, avec essentiellement la même preuve.

Démonstration. Considérons le cas où T est borné p.s. : $T \leq N$ p.s. Alors $M_T = M_{N \wedge T}$, et par le théorème 4.4.1, $M_S \in L^1$, $M_T \in L^1$, et la propriété de martingale de M^T donne $\mathbb{E}(M_T | \mathcal{F}_n) = \mathbb{E}(M_{N \wedge T} | \mathcal{F}_n) = M_{n \wedge T}$ pour tout $n \leq N$. Ensuite, pour tout $A \in \mathcal{F}_S$, le découpage selon les valeurs de S ($\leq T \leq N$) donne

$$\mathbb{E}(\mathbb{1}_A M_T) = \sum_{n=0}^N \mathbb{E}(\mathbb{1}_{A \cap \{S=n\}} \mathbb{E}(M_T | \mathcal{F}_n)) = \sum_{n=0}^N \mathbb{E}(\mathbb{1}_{A \cap \{S=n\}} M_{n \wedge T}) = \sum_{n=0}^N \mathbb{E}(\mathbb{1}_{A \cap \{S=n\}} M_{S \wedge T}) = \sum_{n=0}^N \mathbb{E}(\mathbb{1}_{A \cap \{S=n\}} M_S) = \mathbb{E}(\mathbb{1}_A M_S).$$

Considérons le cas où S et T sont finis p.s. et M^S et M^T sont u.i. Pour tout N , les temps d'arrêt bornés p.s. $S \wedge N$ et $T \wedge N$ vérifient $S \wedge N \leq T \wedge N$, et l'étude du cas borné précédent donne alors $\mathbb{E}(\mathbb{1}_A M_{T \wedge N}) = \mathbb{E}(\mathbb{1}_A M_{S \wedge N})$ pour tout $A \in \mathcal{F}_{S \wedge N}$. Comme $\mathcal{F}_{S \wedge N}$ est croissante en N , le théorème de convergence dominée généralisée u.i. donne $\mathbb{E}(\mathbb{1}_A M_T) = \mathbb{E}(\mathbb{1}_A M_S)$ pour tout $A \in \mathcal{F}_{S \wedge N}$, et donc pour tout $A \in \mathcal{F}_S = \sigma(\cup_N \mathcal{F}_{S \wedge N})$.

Considérons enfin le cas où S et T sont intégrables et M est à oscillations bornées. L'approche du cas précédent s'adapte, par convergence dominée, comme dans la preuve du théorème 4.4.1 d'arrêt de Doob. ■

4.7 Décomposition de Doob et inégalité maximale de Doob

Théodéf 4.7.1. Décomposition de Doob, compensateur, processus croissant.

- (i) Si $(X_n)_{n \in \mathbb{N}}$ est adapté et intégrable, alors les processus $(M_n)_{n \in \mathbb{N}}$ et $(A_n)_{n \in \mathbb{N}}$ définis par $M_0 = A_0 = 0$ et

$$M_n := X_n - X_0 - A_n \quad \text{et} \quad A_n := \sum_{k=1}^n \mathbb{E}(X_k - X_{k-1} | \mathcal{F}_{k-1}), \quad n \geq 1,$$

sont respectivement une martingale et un processus prévisible intégrable. De plus cette décomposition $X - X_0 = M + A$ en somme de martingale et de processus prévisible issus de 0 est unique.

(ii) Si X est une sous-martingale, alors A est croissant et positif, et donc, dans $[0, \infty]$,

$$0 \leq A_n \nearrow_{n \rightarrow \infty} A_\infty \text{ p.s.}$$

Le processus A est le compensateur de X : $X - A$ est une martingale obtenue en compensant X par A .

(iii) Si M est une martingale de carré intégrable, alors le compensateur de la sous-martingale $X = M^2$, appelé processus croissant de M , noté $\langle M \rangle$ et appelé crochet (oblique) de M , vérifie, pour tout $n \geq 1$,

$$\langle M \rangle_n := \sum_{k=1}^n \mathbb{E}((M_k - M_{k-1})^2 | \mathcal{F}_{k-1}) = \sum_{k=1}^n \mathbb{E}(M_k^2 - M_{k-1}^2 | \mathcal{F}_{k-1}) \quad \text{et} \quad \mathbb{E}(\langle M \rangle_\infty) = \sup_{n \in \mathbb{N}} \mathbb{E}(M_n^2) - \mathbb{E}(M_0^2).$$

En particulier M est bornée dans L^2 ssi $\langle M \rangle_\infty \in L^1$.

La propriété (i) exprime le fait qu'il y a une manière naturelle de corriger additivement un processus adapté intégrable pour en faire une martingale, en lui retranchant la somme de ses accroissements conditionnés, que ce terme correctif additif constitue un processus prévisible, et que la décomposition obtenue de la sorte est unique.

Si par exemple $M_n = Z_1 + \dots + Z_n$ est une marche aléatoire à pas i.i.d. centrés et de variance σ^2 , alors on sait que $M_n^2 - \sigma^2 n$ est une martingale, ce qui fait que le processus croissant de M est donné par $\langle M \rangle_n = \sigma^2 n$.

Si $(M_n)_{n \in \mathbb{N}}$ et $(M'_n)_{n \in \mathbb{N}}$ sont des martingales de carré intégrable pour la même filtration, alors $(M_n M'_n)_{n \in \mathbb{N}}$ est adapté et intégrable, et son compensateur est noté $\langle M, M' \rangle_n = \frac{1}{4}(\langle M + M' \rangle_n - \langle M - M' \rangle_n)$ (identité de polarisation de forme quadratique). Dans la même veine, pour une filtration fixée, l'ensemble des martingales bornées dans L^2 est un espace de Hilbert pour le produit scalaire $\langle M, M' \rangle := \langle M, M' \rangle_\infty := \frac{1}{4}(\langle M + M' \rangle_\infty - \langle M - M' \rangle_\infty)$.

Démonstration. La décomposition proposée vérifie bien les propriétés annoncées. Pour l'unicité, si $X = X_0 + M + A$ est une décomposition avec M martingale et A prévisible et $M_0 = A_0 = 0$, alors $A = X - X_0 - M$ est intégrable, et

$$\mathbb{E}(X_n - X_{n-1} | \mathcal{F}_{n-1}) = A_n - A_{n-1} + \mathbb{E}(M_n - M_{n-1} | \mathcal{F}_{n-1}) = A_n - A_{n-1},$$

et comme $A_0 = 0$, on obtient, pour tout $n \geq 1$, $A_n = \sum_{k=1}^n \mathbb{E}(X_k - X_{k-1} | \mathcal{F}_{k-1})$, ce qui fixe aussi $M_n = X_n - X_0 - A_n$.

Pour finir, si X est une sous-martingale, alors pour tout $n \geq 1$, on a

$$A_n - A_{n-1} = \mathbb{E}(A_n - A_{n-1} | \mathcal{F}_{n-1}) = \mathbb{E}(X_n - X_{n-1} | \mathcal{F}_{n-1}) - \mathbb{E}(M_n - M_{n-1} | \mathcal{F}_{n-1}) = \mathbb{E}(X_n - X_{n-1} | \mathcal{F}_{n-1}) \geq 0 \quad \text{p.s.}$$

Pour le (iii), la seconde formule pour $\langle M \rangle_n$ provient de la propriété de martingale de M :

$$\mathbb{E}((M_k - M_{k-1})^2 | \mathcal{F}_{k-1}) = \mathbb{E}(M_k^2 | \mathcal{F}_{k-1}) - 2M_{k-1}\mathbb{E}(M_k | \mathcal{F}_{k-1}) + M_{k-1}^2 = \mathbb{E}(M_k^2 - M_{k-1}^2 | \mathcal{F}_{k-1}).$$

D'autre part, comme $\langle M \rangle_0 = 0$ et $\langle M \rangle_n \nearrow \langle M \rangle_\infty$ dans $[0, \infty]$, en combinant propriété de martingale pour $M^2 - \langle M \rangle$ et convergence monotone pour $\langle M \rangle$, il vient que $\mathbb{E}(M_n^2) = \mathbb{E}(\langle M \rangle_n) + \mathbb{E}(M_0^2) \nearrow \mathbb{E}(\langle M \rangle_\infty) + \mathbb{E}(M_0^2)$ dans $[0, \infty]$. ■

Si $X_k \in L^p$ pour tout $0 \leq k \leq n$, alors $\max_{0 \leq k \leq n} |X_k| \leq |X_0| + \dots + |X_n| \in L^p$, et ceci ne nécessite rien. Peut-on rendre cela quantitatif? Qu'advient-il en horizon infini $n \rightarrow \infty$? Voici une réponse pour les sous-martingales positives.

Théorème 4.7.2. Inégalité maximale de Doob.

Soit $(X_n)_{n \in \mathbb{N}}$ une sous-martingale positive, par exemple une fonction convexe positive d'une martingale.

(i) Pour tout $n \in \mathbb{N}$ et tout réel $r > 0$,

$$\mathbb{P}\left(\max_{0 \leq k \leq n} X_k \geq r\right) \leq \frac{\mathbb{E}(X_n \mathbb{1}_{\{\max_{1 \leq k \leq n} X_k \geq r\}})}{r} \leq \frac{\mathbb{E}(X_n)}{r}.$$

(ii) Pour tout $n \in \mathbb{N}$ et tout réel $p > 1$,

$$\max_{0 \leq k \leq n} \mathbb{E}(X_k^p) \leq \mathbb{E}\left(\max_{0 \leq k \leq n} X_k^p\right) \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}(X_n^p),$$

en particulier, par convergence monotone,

$$\sup_{n \in \mathbb{N}} \mathbb{E}(X_n^p) \leq \mathbb{E}\left(\sup_{n \in \mathbb{N}} X_n^p\right) \leq \left(\frac{p}{p-1}\right)^p \sup_{n \in \mathbb{N}} \mathbb{E}(X_n^p).$$

En particulier, pour toute sous-martingale positive ou martingale $(X_n)_{n \in \mathbb{N}}$ et tout réel $p > 1$:
— X est bornée dans L^p ssi X est dominée dans L^p au sens où $\sup_{n \in \mathbb{N}} |X_n| \in L^p$.

La dernière propriété relie, pour les martingales, deux critères d'uniforme intégrabilité (exemple 1.9.2).

Observons que $q = p/(p-1)$ est l'indice conjugué de Hölder de p . Si $p = 2$ alors $(p/(p-1))^p = 4$.

Il est possible d'établir que la constante $p/(p-1)$ est optimale, et qu'il n'y a pas d'inégalité maximale pour $p = 1$.

L'inégalité maximale de Doob exprime le fait que le moment (ou queue de distribution) du maximum sur un intervalle de temps est contrôlable par le moment du temps terminal. Il s'agit d'une généralisation au cadre martingale de l'inégalité maximale de Kolmogorov qui concerne les marches aléatoires (sommées de v.a. i.i.d.).

Démonstration.

(i) Si $T := \inf\{n \in \mathbb{N} : X_n \geq r\}$, alors $\{T = k\} = \{X_0 < r, \dots, X_{k-1} < r, X_k \geq r\} \in \mathcal{F}_k$ pour tous $0 \leq k \leq n$ et $r > 0$, d'où

$$r \mathbb{1}_{\{T=k\}} \leq X_k \mathbb{1}_{\{T=k\}} \leq \mathbb{E}(X_n | \mathcal{F}_k) \mathbb{1}_{\{T=k\}} = \mathbb{E}(X_n \mathbb{1}_{\{T=k\}} | \mathcal{F}_k),$$

d'où, en prenant l'espérance, $r \mathbb{P}(T = k) \leq \mathbb{E}(X_n \mathbb{1}_{\{T=k\}})$, puis en sommant sur $0 \leq k \leq n$,

$$r \mathbb{P}(T \leq n) \leq \mathbb{E}(X_n \mathbb{1}_{\{T \leq n\}}).$$

Il ne reste plus qu'à observer que $\{\max_{0 \leq k \leq n} X_k \geq r\} = \{T \leq n\}$ pour obtenir l'inégalité attendue.

(ii) L'inégalité de gauche est immédiate par monotonie. Établissons l'inégalité de droite. De (i), pour tout $r > 0$,

$$r \mathbb{P}(X_n^* \geq r) \leq \mathbb{E}(X_n \mathbb{1}_{\{X_n^* \geq r\}}) \quad \text{où} \quad X_n^* := \max_{0 \leq k \leq n} X_k.$$

Exploitions l'indicateur dans le membre de droite pour obtenir une estimée auto-référente. L'idée est d'utiliser la liaison entre moments et queue de distribution⁶. Plus précisément, on écrit, pour tout réel $p > 1$,

$$\int_0^\infty r \mathbb{P}(X_n^* \geq r) p r^{p-2} dr \leq \int_0^\infty \mathbb{E}(X_n \mathbb{1}_{\{X_n^* \geq r\}}) p r^{p-2} dr,$$

qui se réécrit au moyen du théorème de Fubini–Tonelli en

$$\mathbb{E} \int_0^\infty r \mathbb{1}_{\{X_n^* \geq r\}} p r^{p-2} dr \leq \mathbb{E} \int_0^\infty X_n \mathbb{1}_{\{X_n^* \geq r\}} p r^{p-2} dr,$$

c'est-à-dire (ici on utilise $p > 1$ qui donne $p-1 > 0$)

$$\mathbb{E} \int_0^{X_n^*} p r^{p-1} dr \leq \frac{p}{p-1} \mathbb{E} \left(X_n \int_0^{X_n^*} (p-1) r^{p-2} dr \right),$$

en d'autres termes,

$$\mathbb{E}((X_n^*)^p) \leq \frac{p}{p-1} \mathbb{E}(X_n (X_n^*)^{p-1}).$$

L'inégalité de Hölder appliquée au membre de droite donne

$$\mathbb{E}(X_n (X_n^*)^{p-1}) \leq \mathbb{E}(X_n^p)^{1/p} \mathbb{E}((X_n^*)^p)^{1-1/p}, \quad \text{d'où enfin} \quad \mathbb{E}((X_n^*)^p) \leq \left(\frac{p}{p-1} \right)^p \mathbb{E}(X_n^p).$$

■

4.8 Convergence par la voie de carré intégrable (vallée de Doob)

10 → 11

Théorème 4.8.1. Martingales bornées dans L^2 .

Si $(M_n)_{n \in \mathbb{N}}$ est une martingale bornée dans L^2 : $\sup_{n \in \mathbb{N}} \mathbb{E}(M_n^2) < \infty$, alors il existe $M_\infty \in L^2$ t.q.

$$\lim_{n \rightarrow \infty} M_n = M_\infty \quad \text{p.s. et dans } L^2.$$

En particulier $\lim_{n \rightarrow \infty} M_n = M_\infty$ dans L^1 et $\mathbb{E}(M_0) = \mathbb{E}(M_\infty)$. De plus $\mathbb{E}(M_0^2) = \mathbb{E}(M_\infty^2) - \mathbb{E}(\langle M \rangle_\infty)$.

Rappelons que si $X_n \in L^p \rightarrow X \in L^p$ dans L^p alors $\mathbb{E}(|X_n|^p) \rightarrow \mathbb{E}(|X|^p)$ car la norme L^p est continue.

6. Le théorème de Fubini–Tonelli donne $\mathbb{E}(U(|X|)) - U(0) = \mathbb{E} \int_0^{|X|} U'(r) dr = \int_0^\infty U'(r) \mathbb{P}(|X| \geq r) dr$.

Démonstration. Pour tous $n \geq 1$ et $1 \leq k < n$, on a, en notant $\Delta M_k := M_k - M_{k-1}$,

$$\mathbb{E}(\Delta M_k \Delta M_n) = \mathbb{E}(\mathbb{E}(\Delta M_k \Delta M_n | \mathcal{F}_{n-1})) = \mathbb{E}(\Delta M_k \mathbb{E}(\Delta M_n | \mathcal{F}_{n-1})) = 0.$$

Cette orthogonalité des accroissements successifs donne, pour tous $n, m \in \mathbb{N}$,

$$\mathbb{E}((M_{n+m} - M_n)^2) = \sum_{k=n+1}^{n+m} \mathbb{E}((\Delta M_k)^2).$$

En particulier, comme $\sup_{n \in \mathbb{N}} \mathbb{E}((M_n - M_0)^2) < \infty$ car $\sup_{n \in \mathbb{N}} \mathbb{E}(M_n^2) < \infty$ par hypothèse, on obtient

$$\sum_{k \in \mathbb{N}} \mathbb{E}((\Delta M_k)^2) < \infty.$$

De plus $(M_n)_{n \in \mathbb{N}}$ est de Cauchy dans L^2 , donc elle converge dans L^2 vers une limite notée $M_\infty \in L^2$. Il reste à établir la convergence p.s. Il suffit de montrer que $(M_n)_{n \in \mathbb{N}}$ est de Cauchy p.s. Pour cela, définissons

$$X_n := \sup_{i, j \geq n} |M_i - M_j|.$$

Maintenant il suffit d'établir que $\lim_{n \rightarrow \infty} X_n = 0$ p.s. En fait, $0 \leq X_n \searrow X_\infty$ p.s. quand $n \rightarrow \infty$, où $X_\infty \geq 0$. Par conséquent, il suffit de montrer que $\mathbb{E}(X_\infty^2) = 0$ où le carré est choisi pour faciliter le calcul par la suite. Par le lemme de Fatou, il suffit d'établir que $\lim_{n \rightarrow \infty} \mathbb{E}(X_n^2) = 0$. Or

$$X_n \leq \sup_{i \geq n} |M_i - M_n| + \sup_{j \geq n} |M_n - M_j| = 2Y_n \quad \text{où} \quad Y_n := \sup_{k \geq n} |M_k - M_n|.$$

Il suffit d'établir que $\mathbb{E}(Y_n^2) \rightarrow 0$. L'inégalité maximale de Doob (th. 4.7.2) pour la martingale $(M_{n+k} - M_n)_{k \geq 0}$ donne

$$\mathbb{E}(Y_n^2) \leq 4 \sup_{k \geq n} \mathbb{E}((M_k - M_n)^2) = 4 \sum_{k=n+1}^{\infty} \mathbb{E}((\Delta M_k)^2),$$

reste d'une série convergente! Enfin les limites L^2 et p.s. coïncident par unicité de la limite en probabilité. ■

Ceci n'est pas abordé lors du cours oral.

Remarque 4.8.2: Formules de Wald

Si M est une martingale de carré intégrable et T un temps d'arrêt fini p.s. tel que $M_T \in L^2$ alors

$$\mathbb{E}(M_T) = \mathbb{E}(M_0) \quad \text{et} \quad \mathbb{E}(M_T^2) = \mathbb{E}(M_0^2) + \mathbb{E}(\langle M \rangle_T).$$

En effet, par le théorème 4.4.1 d'arrêt de Doob, $(M_{n \wedge T})_{n \in \mathbb{N}}$ est une martingale, $\mathbb{E}(M_{n \wedge T}) = \mathbb{E}(M_0)$, et la décomposition de Doob de $M_{n \wedge T}^2$ donne $\mathbb{E}(M_{n \wedge T}^2) = \mathbb{E}(M_0^2) + \mathbb{E}(\langle M \rangle_{n \wedge T})$. Or par convergence monotone, $\mathbb{E}(\langle M \rangle_{n \wedge T}) \rightarrow \mathbb{E}(\langle M \rangle_T)$, tandis qu'en procédant comme dans la preuve du théorème 4.8.1, on a que $(M_{n \wedge T})_{n \in \mathbb{N}}$ est de Cauchy dans L^2 , donc $M_{n \wedge T} \rightarrow M_T$ dans L^2 et dans L^1 , d'où les formules. Les martingales M et $M^2 - \langle M \rangle$ et le temps d'arrêt T ne vérifient pas forcément les hypothèses du théorème 4.4.1 d'arrêt de Doob, mais les conclusions attendues ont lieu d'après ce qui précède. Lorsque $M_n = S_n - nm$ avec $S_n = Z_1 + \dots + Z_n$ et $(Z_n)_{n \geq 1}$ i.i.d. de moyenne m et de variance σ^2 , on a $\langle M \rangle_n = \sigma^2 n$. Cela peut être utilisé pour la ruine du joueur de pile ou face de l'exemple 4.4.6.

Théorème 4.8.3. Surmartingales positives.

Soit $(X_n)_{n \in \mathbb{N}}$ une surmartingale positive. Alors :

- (i) Il existe $X_\infty \geq 0$ telle que $X_\infty \in L^1$ et $\lim_{n \rightarrow \infty} X_n = X_\infty$ p.s.
- (ii) Pour tout n , $X_n \geq \mathbb{E}(X_\infty | \mathcal{F}_n)$, donc $\mathbb{E}(X_n) \geq \mathbb{E}(X_\infty)$.
- (iii) Il y a de plus convergence dans L^1 ssi $(X_n)_{n \in \mathbb{N}}$ est u.i.

En somme, un certain type de suites aléatoires positives et décroissantes en moyenne sont convergentes.

Démonstration. La convergence dans L^1 n'a rien à voir avec les martingales et concerne le lien général entre les convergences p.s. et L^1 des suites de v.a. et la notion d'u.i. Quant à la minoration de X_n , elle découle directement de la propriété de surmartingale, du lemme de Fatou conditionnel, et de la convergence presque sûre :

$$X_n \geq \liminf_{m \rightarrow \infty} \mathbb{E}(X_m | \mathcal{F}_n) = \mathbb{E}(\liminf_{m \rightarrow \infty} X_m | \mathcal{F}_n) = \mathbb{E}(X_\infty | \mathcal{F}_n).$$

Cela montre également au passage que $X_\infty \in L^1$ car $X_\infty \geq 0$ et $\mathbb{E}(X_\infty) \leq \mathbb{E}(X_n) < \infty$.

Établissons la convergence presque sûre. L'idée est de se ramener à une sous-martingale bornée par déformation, puis à une martingale bornée dans L^2 par la décomposition de Doob. Plus précisément, comme e^{-x} est convexe et décroissante, le processus $(Y_n := e^{-X_n})_{n \in \mathbb{N}}$ est une sous-martingale, de décomposition de Doob (théorème 4.7.1)

$$Y_n = Y_0 + M_n + A_n, \quad n \geq 1, \quad M_0 = A_0 = 0,$$

où M est une martingale et A prévisible. Comme Y est une sous-martingale, A est croissant positif : $0 \leq A_n \nearrow A_\infty$ p.s. dans $[0, \infty]$. Il nous suffit d'établir que M est bornée dans L^2 et d'exploiter le théorème 4.8.1. Comme $X_n \geq 0$, la v.a. Y_n est à valeurs dans $[0, 1]$, et par convergence monotone, $\mathbb{E}(A_\infty) = \lim_{n \rightarrow \infty} \mathbb{E}(A_n) = \lim_{n \rightarrow \infty} \mathbb{E}(Y_n - Y_0) < \infty$ car $Y_n - Y_0$ est dans $[-1, 1]$. En particulier A et donc $M = Y - Y_0 - A$ est bornée dans L^1 . Mais à ce stade, il n'est pas évident que M soit de carré intégrable. Cependant M est majorée par une constante, car A est positif et Y est à valeurs dans un intervalle borné. La propriété d'orthogonalité des accroissements de martingales, déjà exploitée dans la preuve du théorème 4.8.1 pour les martingales de carré intégrable, reste valable pour les martingales positives ou minorées par une constante, ou négatives ou majorées par une constante. Pour M cela donne, pour tous n, m , dans $[0, \infty]$,

$$\mathbb{E}((M_{n+m} - M_n)^2) = \sum_{k=n+1}^{n+m} \mathbb{E}((\Delta M_k)^2) \quad \text{où} \quad \Delta M_k := M_k - M_{k-1}.$$

Écrivons $Y_n^2 = Y_0^2 + \sum_{k=1}^n (Y_k^2 - Y_{k-1}^2)$. Comme $Y_k = Y_{k-1} + \Delta M_k + \Delta A_k$, on obtient

$$Y_n^2 = Y_0^2 + \sum_{k=1}^n [(\Delta M_k)^2 + (\Delta A_k)^2 + 2Y_{k-1}\Delta M_k + 2Y_{k-1}\Delta A_k + 2\Delta M_k\Delta A_k].$$

Maintenant $Y_0^2 + \sum_k (\Delta A_k)^2 \geq 0$ et $2 \sum_k Y_{k-1} \Delta A_k \geq 0$ car $Y \geq 0$ et $\Delta A \geq 0$. Donc

$$\sum_{k=1}^n (\Delta M_k)^2 + 2 \sum_{k=1}^n (Y_{k-1} + \Delta A_k) \Delta M_k \leq Y_n^2 \leq 1.$$

À ce stade, on observe que $\mathbb{E}((Y_{k-1} + \Delta A_k) \Delta M_k) = \mathbb{E}((Y_{k-1} + \Delta A_k) \mathbb{E}(\Delta M_k | \mathcal{F}_{k-1})) = 0$, d'où

$$\mathbb{E}(M_n^2) = \mathbb{E}((M_n - M_0)^2) = \sum_{k=1}^n \mathbb{E}((\Delta M_k)^2) \leq 1.$$

■

Théorème 4.8.4. Sous-martingales bornées dans L^1 .

Soit $(X_n)_{n \in \mathbb{N}}$ une sous-martingale bornée dans L^1 . Alors :

- (i) Il existe $X_\infty \in L^1$ telle que $\lim_{n \rightarrow \infty} X_n = X_\infty$ p.s.
- (ii) Il y a de plus convergence dans L^1 ssi $(X_n)_{n \in \mathbb{N}}$ est u.i.
- (iii) La bornitude dans L^1 est équivalente à l'hypothèse d'apparence plus faible $\sup_{n \in \mathbb{N}} \mathbb{E}(X_n^+) < \infty$.

Démonstration. La propriété (ii) n'a rien à voir avec les martingales et concerne le lien général entre les convergences p.s. et L^1 des suites de v.a. Le fait que $X_\infty \in L^1$ découle tout simplement du lemme de Fatou :

$$\mathbb{E}(|X_\infty|) = \mathbb{E}(\liminf_n |X_n|) \leq \liminf_n \mathbb{E}(|X_n|) \leq \sup_n \mathbb{E}(|X_n|) < \infty.$$

Concernant la reformulation de l'hypothèse de bornitude, comme $\mathbb{E}(X_n^+) - \mathbb{E}(X_n^-) = \mathbb{E}(X_n) \geq \mathbb{E}(X_0)$, il vient

$$\mathbb{E}(X_n^-) \leq \sup_{n \in \mathbb{N}} \mathbb{E}(X_n^+) - \mathbb{E}(X_0).$$

Démontrons à présent la convergence p.s. Soit $C := \sup_n \mathbb{E}(|X_n|) < \infty$. Il suffit d'établir que

$$X = Y - Z$$

où $(Y_n)_{n \in \mathbb{N}}$ et $(Z_n)_{n \in \mathbb{N}}$ sont des surmartingales positives et d'utiliser le théorème 4.8.3.

Comme $x^+ = \max(x, 0)$ est convexe et croissante, le processus $(X_n^+)_{n \in \mathbb{N}}$ est une sous-martingale. Soit

$$X_n^+ = X_0^+ + M_n + A_n$$

sa décomposition de Doob (théorème 4.7.1), où $0 \leq A_n \nearrow A_\infty$ p.s. dans $[0, \infty]$. Comme $\mathbb{E}(A_n) = \mathbb{E}(X_n^+) - \mathbb{E}(X_0^+) \leq C$, il vient que $\mathbb{E}(A_\infty) \leq C$ par convergence monotone. Introduisons à présent le processus $(Y_n)_{n \in \mathbb{N}}$ défini par

$$Y_n := X_0^+ + M_n + \mathbb{E}(A_\infty | \mathcal{F}_n).$$

Le processus $(Y_n)_{n \in \mathbb{N}}$ est une martingale en particulier une surmartingale positive car $Y_n \geq X_0^+ + M_n + A_n = X_n^+ \geq 0$. Enfin $Z_n = Y_n - X_n$ définit une surmartingale (somme d'une martingale et d'une surmartingale), positive car

$$Z_n \geq X_n^+ - X_n = X_n^- \geq 0.$$

■

Remarque 4.8.5. Martingales positives.

Une martingale positive est toujours bornée dans L^1 . En tant que sous-martingale bornée dans L^1 (théorème 4.8.4) ou surmartingale positive (théorème 4.8.3), elle converge p.s. vers une v.a. intégrable. C'est le cas notamment pour le modèle des urnes de Pólya et celui de Wright – Fisher.

Remarque 4.8.6. Équivalence de deux théorèmes de convergence.

La preuve du théorème 4.8.4 (sous-martingales bornées dans L^1) ci-dessus est une réduction au théorème 4.8.3 (surmartingales positives) faisant usage de la décomposition de Doob (théorème 4.7.1). Or il est à l'inverse possible de déduire très simplement le théorème 4.8.3 du théorème 4.8.4. En effet, si $(X_n)_{n \in \mathbb{N}}$ est une surmartingale positive, alors $(-X_n)_{n \in \mathbb{N}}$ est une sous-martingale vérifiant $\sup_{n \in \mathbb{N}} \mathbb{E}((-X_n)^+) = 0 < \infty$. Cette observation suggère de rechercher une preuve alternative directe du théorème 4.8.4. Une telle preuve est proposée dans la section 4.9, au moyen de l'inégalité des montées et des descentes de Doob.

Le théorème suivant généralise le théorème 4.8.1 de convergence des martingales bornées dans L^2 .

Théorème 4.8.7. Martingales bornées L^p , $p > 1$.

Soit $(M_n)_{n \in \mathbb{N}}$ une martingale ou une sous-martingale positive.

Si $(M_n)_{n \in \mathbb{N}}$ est bornée dans L^p : $\sup_{n \in \mathbb{N}} \mathbb{E}(|M_n|^p) < \infty$, pour un réel $p > 1$, alors :

- (i) Il existe $M_\infty \in L^p$ tel que $\lim_{n \rightarrow \infty} M_n = M_\infty$ p.s. et dans L^p .
- (ii) $M^* := \sup_{n \in \mathbb{N}} |M_n| \in L^p$ et $\mathbb{E}(M_*^p) \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}(|M_\infty|^p)$.

Une telle martingale est une martingale fermée (théorème 4.10.1).

Démonstration. Comme $(M_n)_{n \in \mathbb{N}}$ est une sous-martingale bornée dans L^1 , $M_n \rightarrow M_\infty$ p.s. où $M_\infty \in L^1$ (théorème 4.8.4). De plus le lemme de Fatou donne $\mathbb{E}(|M_\infty|^p) \leq \liminf_{n \rightarrow \infty} \mathbb{E}(|M_n|^p) \leq \sup_{n \in \mathbb{N}} \mathbb{E}(|M_n|^p) < \infty$, donc $M_\infty \in L^p$.

Pour tout $1 \leq r < p$, comme $(|M_n|^r)_{n \in \mathbb{N}}$ est bornée dans $L^{p/r}$, $p/r > 1$; elle est u.i. par le critère de de la Vallée Poussin (théorème 1.9.1), et le théorème 1.9.4 de convergence dominée u.i. L^r indique que $M_\infty \in L^r$ et $M_n \rightarrow M_\infty$ dans L^r . Cela établit la convergence dans L^r , $1 \leq r < p$, du (i) sans utiliser la propriété de martingale de $(M_n)_{n \in \mathbb{N}}$.

Reste à obtenir la convergence dans L^p , et (ii). L'inégalité maximale de Doob du théorème 4.7.2 pour la sous-martingale positive $(|M_n|)_{n \in \mathbb{N}}$ donne

$$\mathbb{E}\left(\sup_{n \in \mathbb{N}} |M_n|^p\right) \leq \left(\frac{p}{p-1}\right)^p \sup_{n \in \mathbb{N}} \mathbb{E}(|M_n|^p) < \infty,$$

d'où $\sup_{n \in \mathbb{N}} |M_n| \in L^p$, autrement dit $\sup_{n \in \mathbb{N}} |M_n|^p \in L^1$, d'où, par convergence dominée,

$$\mathbb{E}(|M_\infty|^p) = \lim_{n \rightarrow \infty} \mathbb{E}(|M_n|^p) = \sup_{n \in \mathbb{N}} \mathbb{E}(|M_n|^p) < \infty,$$

où la dernière égalité vient de la croissance de $(\mathbb{E}(|M_n|^p))_{n \in \mathbb{N}}$, car $(|M_n|^p)_{n \in \mathbb{N}}$ est aussi une sous-martingale car $x \geq 0 \mapsto x^p$ est convexe croissante. Ceci redémontre que $M_\infty \in L^p$, donne $\mathbb{E}(|M_\infty|^p) = \sup_{n \in \mathbb{N}} \mathbb{E}(|M_n|^p)$, et donne également (ii). Enfin, la convergence dans L^p dans (i) s'obtient par convergence dominée en observant que

$$|M_n - M_\infty|^p \leq 2^{p-1} (\sup_{n \in \mathbb{N}} |M_n|^p + |M_\infty|^p) \in L^1.$$

■

Exemple 4.8.8. Contrexemple de martingales non u.i. : Galton – Watson critique.

Considérons le processus de branchement de Galton – Watson, de la forme

$$M_{n+1} = X_{n+1,1} + \dots + X_{n+1,M_n}, \quad M_0 := 1,$$

où $(X_{j,k})_{j,k \geq 1}$ sont i.i.d. de loi μ sur \mathbb{N} telle que $\mu(0) > 0$. Plaçons-nous dans le cas critique où μ est de moyenne 1. Comme $(M_n)_{n \geq 1}$ est une martingale positive, elle converge p.s. vers $M_\infty \in L^1$ (théorème 4.8.3). C'est aussi une chaîne de Markov d'espace d'état \mathbb{N} . L'état 0 est absorbant et tous les autres états peuvent mener à 0 et sont donc transitoires. Il en découle que p.s. soit $(M_n)_{n \in \mathbb{N}}$ converge vers 0 soit vers ∞ , et comme $M_\infty \in L^1$, il vient que $M_\infty = 0$ p.s. La convergence de $(M_n)_{n \in \mathbb{N}}$ ne peut pas avoir lieu dans L^1 car cela donnerait la contradiction suivante : $1 = \mathbb{E}(M_n) \rightarrow \mathbb{E}(M_\infty) = 0$. La martingale $(M_n)_{n \in \mathbb{N}}$ n'est donc pas u.i. Étant positive, elle est évidemment bornée dans L^1 , mais dans aucun L^p pour $p > 1$ car non u.i. Comme elle n'est pas u.i., elle n'est pas dominée dans L^1 : $\sup_{n \in \mathbb{N}} |M_n| \notin L^1$, le maximum de sa trajectoire a une queue lourde. Cela explique le caractère plutôt déroutant des simulations naïves des trajectoires de ce processus.

4.9 Convergence par la voie des montées et descentes (arête de Doob)

Dans la section précédente, le théorème 4.8.3 de convergence de surmartingales positives puis le théorème 4.8.4 de convergence de sous-martingales bornées dans L^1 , qui sont équivalents, sont démontrés successivement, via la décomposition de Doob (théorème 4.7.1), par réduction au théorème 4.8.1 de convergence des martingales bornées dans L^2 , lui-même démontré via l'inégalité maximale de Doob (théorème 4.7.2). Dans cette section, on adopte l'approche d'origine de Doob, basée sur les montées et les descentes, plus inattendue, qui établit directement le théorème 4.8.4 de convergence des sous-martingales bornées dans L^1 . Avec le recul, cette approche par montées et descentes reste très spécifique aux martingales, tandis que celle par réduction au cas hilbertien est plus universelle (c'est par exemple celle adoptée pour la construction de l'espérance conditionnelle).

Soit $x = (x_n)_{n \in \mathbb{N}}$ une suite réelle déterministe. Pour tous réels $a < b$, soit $(S_k(x))_{k \geq 1}$ et $(T_k(x))_{k \geq 1}$ les suites à valeurs dans $\mathbb{N} \cup \{+\infty\}$ définies par récurrence comme suit ($\inf \emptyset = +\infty$) :

$$\begin{cases} S_{k+1}(x) & := \inf\{n \geq T_k(x) : x_n \leq a\} \\ T_{k+1}(x) & := \inf\{n \geq S_{k+1}(x) : x_n \geq b\}, \end{cases} \quad \text{avec } T_0(x) := 0.$$

En particulier $T_k(x) \leq S_{k+1}(x) \leq T_{k+1}(x)$ pour tout $k \geq 0$. Posons à présent

$$N_n([a, b], x) := \sum_{k=1}^{\infty} \mathbb{1}_{\{T_k(x) \leq n\}} \quad \text{et} \quad N_\infty([a, b], x) := \sum_{k=1}^{\infty} \mathbb{1}_{\{T_k(x) < \infty\}}.$$

La quantité $N_n([a, b], x)$ est le nombre de montées pour l'intervalle $[a, b]$ effectuées par la suite x_0, \dots, x_n . La quantité $N_\infty([a, b], x)$ est le nombre de montées pour l'intervalle $[a, b]$ effectuées par la suite $(x_n)_{n \in \mathbb{N}}$.

Lemme 4.9.1. Caractérisation de la convergence par les montées.

Une suite $(x_n)_{n \in \mathbb{N}}$ converge dans $\overline{\mathbb{R}} = [-\infty, +\infty]$ ssi $N_\infty([a, b], x) < \infty$ pour tous rationnels $a < b$.

Démonstration. Supposons que $x_n \rightarrow x \in \overline{\mathbb{R}}$. Si $x \in \mathbb{R}$, alors $(x_n)_{n \in \mathbb{N}}$ est à partir d'un certain rang dans un intervalle arbitrairement petit contenant x , et si son diamètre est plus petit que celui de $[a, b]$, alors le nombre de montées pour $[a, b]$ est fini. Si $x = +\infty$ (respectivement $x = -\infty$) alors, à partir d'un certain rang, $(x_n)_{n \in \mathbb{N}}$ reste au-dessus (respectivement au-dessous) de $[a, b]$, et le nombre de montées pour $[a, b]$ est fini.

Réciproquement, supposons que $(x_n)_{n \in \mathbb{N}}$ ne converge pas dans $\overline{\mathbb{R}}$. Alors en considérant des rationnels $a < b$ tels que $\underline{\lim}_n x_n < a < b < \overline{\lim}_n x_n$, on obtient une infinité de termes $< a$ et une infinité de termes $> b$, et en particulier une infinité de montées à travers $[a, b]$. ■

Nous allons exploiter ce lemme d'analyse pour les trajectoires de sous-martingales.

Si $(X_n)_{n \in \mathbb{N}}$ est adapté alors $S_k(X)$ et $T_k(X)$ sont des temps d'arrêts, et donc $N_n([a, b], X)$ est \mathcal{F}_n -mesurable :

$$\{T_k(X) \leq n\} = \cup_{0 \leq m_1 < n_1 < \dots < m_k < n_k \leq n} \{X_{m_1} \leq a, X_{n_1} \geq b, \dots, X_{m_k} \leq a, X_{n_k} \geq b\} \in \mathcal{F}_n \quad (\text{idem pour } S_k).$$

Lemme 4.9.2. Inégalité sur les montées de Doob.

Si $(X_n)_{n \in \mathbb{N}}$ est une sous-martingale alors pour tous réels $a < b$ et pour tout $n \in \mathbb{N}$,

$$(b - a)\mathbb{E}(N_n([a, b], X)) \leq \mathbb{E}((X_n - a)^+ - (X_0 - a)^+).$$

Démonstration. Comme $x \mapsto (x - a)^+$ est convexe et positive, le processus $(Y_n := (X_n - a)^+)_{n \in \mathbb{N}}$ est une sous-martingale. Notons S_k, T_k , et N_n sans X ni $[a, b]$ pour alléger. Pour faire apparaître $(b - a)$, l'idée est de considérer les traversées de $[a, b]$, c'est-à-dire les montées qui suivent une descente. Pour cela, on introduit, pour tout $n \in \mathbb{N}$,

$$Y_n - Y_0 = \sum_{k=1}^n (Y_k - Y_{k-1}) = \sum_{k=1}^n (H_k + K_k)(Y_k - Y_{k-1}) =: ((K + H) \cdot Y)_n$$

où

$$H_n := \sum_{k=1}^{\infty} \mathbb{1}_{\{S_k < n \leq T_k\}} = \begin{cases} 1 & \text{si } n \text{ est entre une descente et la montée suivante} \\ 0 & \text{sinon} \end{cases} \quad \text{et } K_n := 1 - H_n.$$

Comme $(H + K) \cdot Y = H \cdot Y + K \cdot Y$, nous pouvons évaluer $(H \cdot Y)_n$ et $(K \cdot Y)_n$ séparément. Or

$$(H \cdot Y)_n = \sum_{k=1}^n H_k(Y_k - Y_{k-1}) = \sum_{k=1}^{N_n} (Y_{T_k} - Y_{S_k}) + \mathbb{1}_{\{S_{N_n+1} < n\}}(Y_n - Y_{S_{N_n+1}}).$$

Comme $Y_{S_{N_n+1}} = 0$ sur $\{S_{N_n+1} < \infty\}$ et comme $Y_n \geq 0$, il vient $(H \cdot Y)_n \geq \sum_{k=1}^{N_n} (Y_{T_k} - Y_{S_k})$, d'où

$$\mathbb{E}((H \cdot Y)_n) \geq (b - a)\mathbb{E}(N_n).$$

D'autre part $(H_n)_{n \geq 1}$ est prévisible car $\{S_k < n \leq T_k\} = \{S_k \leq n - 1\} \cap \{T_k \leq n - 1\}^c \in \mathcal{F}_{n-1}$ car S_k et T_k sont des temps d'arrêt, et comme il prend ses valeurs dans $[0, 1]$, le processus $K = 1 - H$ est également positif, et prévisible. Il en découle, par le lemme 4.2.2, que le processus $K \cdot Y$ est une sous-martingale, et en particulier

$$\mathbb{E}((K \cdot Y)_n) \geq \mathbb{E}((K \cdot Y)_0) = 0.$$

En combinant les deux minoration en moyenne obtenues, cela donne finalement

$$(b - a)\mathbb{E}(N_n) \leq \mathbb{E}((H \cdot Y)_n) \leq \mathbb{E}((H \cdot Y)_n + (K \cdot Y)_n) = \mathbb{E}(Y_n - Y_0).$$

■

Preuve du théorème 4.8.4 de convergence des martingales bornées dans L^1 via montées et descentes. Pour tous rationnels $a < b$, l'inégalité sur les montées de Doob (lemme 4.9.2) donne, pour tout $n \geq 1$,

$$(b - a)\mathbb{E}(N_n([a, b], X)) \leq \mathbb{E}((X_n - a)^+) \leq |a| + \sup_{k \in \mathbb{N}} \mathbb{E}(X_k^+),$$

d'où, par convergence monotone,

$$(b - a)\mathbb{E}(N_\infty([a, b], X)) < \infty,$$

qui donne $N_\infty([a, b], X) < \infty$ p.s. Soit $\Omega_{a,b} \in \mathcal{A}$ l'ensemble p.s. associé. Comme $\cap_{a,b \in \mathbb{Q}: a < b} \Omega_{a,b}$ est encore p.s. il vient que p.s. pour tous rationnels $a < b$, $N_\infty([a, b], X) < \infty$. Le lemme d'analyse 4.9.1 donne alors la convergence p.s. de X . Enfin le fait que la limite p.s. est intégrable, et donc en particulier finie p.s., provient du lemme de Fatou. ■

4.10 Martingales fermées ou uniformément intégrables

Théorème 4.10.1. Martingales bornées dans L^1 .

Si $(M_n)_{n \in \mathbb{N}}$ est une martingale alors les conditions suivantes sont équivalentes :

- (i) $(M_n)_{n \in \mathbb{N}}$ converge p.s. et dans L^1 vers $M_\infty \in L^1$.
- (ii) $(M_n)_{n \in \mathbb{N}}$ est une martingale fermée : il existe $X \in L^1$ t.q. $M_n = \mathbb{E}(X | \mathcal{F}_n)$ pour tout $n \in \mathbb{N}$.
- (iii) $(M_n)_{n \in \mathbb{N}}$ est uniformément intégrable.

De plus, dans ce cas, $M_\infty = \mathbb{E}(X | \mathcal{F}_\infty)$ où $\mathcal{F}_\infty := \sigma(\cup_{n=0}^{\infty} \mathcal{F}_n)$.

C'est le cas en particulier des martingales bornées dans L^p , $p > 1$, du théorème 4.8.7.

Démonstration. (i)⇒(ii). Pour tous $m \geq n$, $M_n = \mathbb{E}(M_m | \mathcal{F}_n)$, d'où $M_n = \mathbb{E}(M_\infty | \mathcal{F}_n)$ quand $m \rightarrow \infty$, par convergence dominée conditionnelle ou par contraction $\mathbb{E}(|M_n - \mathbb{E}(M_\infty | \mathcal{F}_n)|) = \mathbb{E}(|\mathbb{E}(M_m - M_\infty | \mathcal{F}_n)|) \leq \mathbb{E}(|M_m - M_\infty|)$.

(ii)⇒(i). La suite $(M_n)_{n \in \mathbb{N}}$ est une sous-martingale bornée dans L^1 , donc par le théorème 4.8.4, il existe $X \in L^1$ telle que $M_n \rightarrow X$ p.s. La convergence dans L^1 découle enfin du fait qu'une martingale fermée est toujours u.i.

(i)⇒(iii). Découle du fait que la convergence dans L^1 implique l'uniforme intégrabilité.

(iii)⇒(i). Comme $(M_n)_{n \in \mathbb{N}}$ est u.i. elle est bornée dans L^1 , d'où (i) par le théorème 4.8.4.

Il ne reste plus qu'à établir que si (i)-(ii)-(iii) sont réalisées, alors $M_\infty = \mathbb{E}(X | \mathcal{F}_\infty)$. Comme M_n est \mathcal{F}_∞ -mesurable pour tout n , il en est de même pour M_∞ . De plus, pour tout $B \in \mathcal{F}_n$, $\mathbb{E}(X \mathbb{1}_B) = \mathbb{E}(M_n \mathbb{1}_B) = \mathbb{E}(M_\infty \mathbb{1}_B)$. Ceci implique que $\mathbb{E}(X \mathbb{1}_B) = \mathbb{E}(M_\infty \mathbb{1}_B)$ sur $\cup_{n=0}^\infty \mathcal{F}_n$, donc sur $\sigma(\cup_{n=0}^\infty \mathcal{F}_n) = \mathcal{F}_\infty$ par le théorème des classes monotones. ■

Ceci n'est pas abordé lors du cours oral.

Si $(M_n)_{n \in \mathbb{N}}$ est une martingale u.i. alors M_T fait sens pour tout temps d'arrêt T , fini ou pas, en posant $M_T := M_\infty$ sur $\{T = \infty\}$. Cela conduit naturellement à une version étendue du théorème d'arrêt de Doob :

Théorème 4.10.2: Théorème d'arrêt pour les martingales u.i.

Si $(M_n)_{n \in \mathbb{N}}$ est une martingale u.i. alors pour tout temps d'arrêt T , fini ou pas,

$$M_T = \mathbb{E}(M_\infty | \mathcal{F}_T), \quad \text{en particulier} \quad \mathbb{E}(M_T) = \mathbb{E}(M_\infty) = \mathbb{E}(M_n) \quad \text{pour tout } n.$$

De plus, si S est un autre temps d'arrêt tel que $S \leq T$ alors

$$M_S = \mathbb{E}(M_T | \mathcal{F}_S), \quad \text{et en particulier} \quad \mathbb{E}(M_S) = \mathbb{E}(M_T).$$

Si $(T_n)_{n \in \mathbb{N}}$ est une suite croissante de temps d'arrêts, alors $(M_{T_n})_{n \in \mathbb{N}}$ est une $(\mathcal{F}_{T_n})_{n \in \mathbb{N}}$ -martingale.

Démonstration. Le découpage selon T donne, en notant $\mathcal{F}_\infty := \sigma(M_\infty)$, que $M_T \in L^1$:

$$\mathbb{E}(|M_T|) = \sum_{n \in \mathbb{N} \cup \{\infty\}} \mathbb{E}(\mathbb{1}_{\{T=n\}} |\mathbb{E}(M_\infty | \mathcal{F}_n)|) \leq \sum_{n \in \mathbb{N} \cup \{\infty\}} \mathbb{E}(\mathbb{1}_{\{T=n\}} \mathbb{E}(|M_\infty| | \mathcal{F}_n)) = \mathbb{E}(|M_\infty|) < \infty,$$

où la commutation $\mathbb{E} \sum_n = \sum_n \mathbb{E}$ provient du théorème de Fubini – Tonelli. De même $M_T = \mathbb{E}(M_\infty | \mathcal{F}_T)$ car pour tout $A \in \mathcal{F}_T$, grâce au fait que $A \cap \{T = n\} \in \mathcal{F}_n$, et $M_n = \mathbb{E}(M_\infty | \mathcal{F}_n)$ (théorème 4.10.1),

$$\mathbb{E}(\mathbb{1}_A M_T) = \sum_{n \in \mathbb{N} \cup \{\infty\}} \mathbb{E}(\mathbb{1}_{A \cap \{T=n\}} M_n) = \sum_{n \in \mathbb{N} \cup \{\infty\}} \mathbb{E}(\mathbb{1}_{A \cap \{T=n\}} \mathbb{E}(M_\infty | \mathcal{F}_n)) = \mathbb{E}(\mathbb{1}_A M_\infty).$$

Enfin $S \leq T$ donne $\mathcal{F}_S \subset \mathcal{F}_T$, d'où $M_S = \mathbb{E}(M_\infty | \mathcal{F}_S) = \mathbb{E}(\mathbb{E}(M_\infty | \mathcal{F}_T) | \mathcal{F}_S) = \mathbb{E}(M_T | \mathcal{F}_S)$. ■

4.11 Martingales rétrogrades ou renversées et application à la loi des grands nombres

Dans cette section, on se place sur $(\Omega, \mathcal{F}, \mathbb{P})$ muni d'une suite décroissante $(\mathcal{F}_n)_{n \in \mathbb{N}}$ de sous-tribus de \mathcal{F} . Un processus $(X_n)_{n \in \mathbb{N}}$ est une martingale renversée lorsque pour tout n ,

$$X_n \in L^1(\mathcal{F}_n) \quad \text{et} \quad \mathbb{E}(X_n | \mathcal{F}_{n+1}) = X_{n+1}.$$

Exemple 4.11.1. Moyenne empirique de v.a. i.i.d. intégrables.

Soit $(Z_n)_{n \geq 1}$ une suite de v.a. i.i.d. intégrables et, pour tout $n \geq 1$,

$$X_n = \frac{S_n}{n} \quad \text{où} \quad S_n := Z_1 + \dots + Z_n \quad \text{et} \quad \mathcal{F}_n := \sigma(X_n, X_{n+1}, \dots) = \sigma(S_n, S_{n+1}, \dots) = \sigma(S_n, Z_{n+1}, Z_{n+2}, \dots).$$

Par définition, $X_n \in L^1(\mathcal{F}_n)$ pour tout n . De plus, pour tout $n \geq 1$,

$$S_n = \mathbb{E}(S_n | \mathcal{F}_n) = \sum_{k=1}^n \mathbb{E}(Z_k | \mathcal{F}_n) \stackrel{*}{=} \sum_{k=1}^n \mathbb{E}(Z_k | S_n) \stackrel{**}{=} n \mathbb{E}(Z_1 | S_n) = n \mathbb{E}(X_1 | S_n) = n \mathbb{E}(X_1 | \mathcal{F}_n).$$

Donc $X_n = \frac{S_n}{n} = \mathbb{E}(X_1 | \mathcal{F}_n)$ pour tout n . Mais alors $\mathcal{F}_n \supset \mathcal{F}_{n+1}$ donne, pour tout $n \geq 1$,

$$\mathbb{E}(X_n | \mathcal{F}_{n+1}) = \mathbb{E}(\mathbb{E}(X_1 | \mathcal{F}_n) | \mathcal{F}_{n+1}) = \mathbb{E}(X_1 | \mathcal{F}_{n+1}) = X_{n+1}.$$

Quitte à poser $X_0 = X_1$, $(X_n)_{n \in \mathbb{N}}$ est donc une martingale renversée.

Pour * : si \mathcal{G} et \mathcal{G}' sont des tribus indépendantes, et U positive ou intégrable, indépendante de \mathcal{G}' , alors

$$\mathbb{E}(U | \sigma(\mathcal{G} \cup \mathcal{G}')) = \mathbb{E}(U | \mathcal{G}),$$

utilisé avec $\mathcal{G} = \sigma(S_n)$ et $\mathcal{G}' = \sigma(Z_{n+1}, Z_{n+2}, \dots)$ de sorte que $\mathcal{F}_n = \sigma(S_n, Z_{n+1}, Z_{n+2}, \dots) = \sigma(\mathcal{G}, \mathcal{G}')$.

Pour ** : $\mathbb{E}(Z_k | S_n) = h_k(S_n)$ et $h_k(s) = \mathbb{E}(Z_k | S_n = s)$ ne dépend pas de k , les lois conditionnelles sont identiques car la loi de (Z_1, \dots, Z_n) est échangeable, c'est-à-dire invariante par permutation des indices.

Théorème 4.11.2. Convergence des martingales renversées.

Si $(X_n)_{n \in \mathbb{N}}$ est une martingale renversée alors il existe $X_\infty \in L^1$ t.q. $X_n \rightarrow X$ p.s. et dans L^1 .

L'usage d'un temps négatif, avec $(\mathcal{F}_n)_{n \leq 0}$ croissante, permet la formule habituelle $\mathbb{E}(X_{n+1} | \mathcal{F}_n) = X_n$ pour $(X_n)_{n \leq 0}$. Avec ces notations renversantes, le théorème 4.11.2 s'écrit $X_n \rightarrow X_{-\infty} \in L^1$ p.s. et dans L^1 , quand $n \rightarrow -\infty$.

Démonstration. Pour tout entier K , définissons

$$Y_n^K = \begin{cases} X_{K-n} & \text{si } 0 \leq n \leq K \\ X_0 & \text{si } n > K \end{cases} \quad \text{et} \quad \mathcal{F}_n^K = \begin{cases} \mathcal{F}_{K-n} & \text{si } 0 \leq n \leq K \\ \mathcal{F}_0 & \text{si } n > K \end{cases}.$$

Alors $(Y_n^K)_{n \in \mathbb{N}}$ est une martingale pour la filtration $(\mathcal{F}_n^K)_{n \in \mathbb{N}}$. L'inégalité sur les montées de Doob du lemme 4.9.2 pour la sous-martingale Y^K appliquée au temps K et à un intervalle quelconque $[a, b]$ donne

$$(b-a)N_K([a, b], Y^K) \leq \mathbb{E}(Y_K^K - a)^+ = \mathbb{E}(X_0 - a)^+.$$

Or $N_K([a, b], Y^K) \nearrow N_\infty([a, b], X)$ quand $K \rightarrow \infty$, d'où, par convergence monotone,

$$(b-a)\mathbb{E}(N_\infty([a, b], X)) \leq \mathbb{E}((X_0 - a)^+) < \infty,$$

d'où $X_n \rightarrow X$ p.s. dans $\overline{\mathbb{R}}$ (lemme 4.9.1). La propriété de martingale renversée implique la bornitude dans L^1 :

$$\mathbb{E}(|X_{n+1}|) \leq \mathbb{E}(\mathbb{E}(|X_n| | \mathcal{F}_{n+1})) = \mathbb{E}(|X_n|) \leq \dots \leq \mathbb{E}(|X_0|) < \infty.$$

Il en découle par le lemme de Fatou que la limite p.s. est dans L^1 . D'autre part, la propriété de martingale renversée donne également $X_n = \mathbb{E}(X_0 | \mathcal{F}_n)$, donc X est u.i. car fermée et la convergence a donc aussi lieu dans L^1 . La décroissance des tribus des martingales renversées simplifie l'étude de la convergence! ■

Comme corollaire, la loi des grands nombres classique pour les sommes de variables indépendantes et identiquement distribuées, avec une preuve assez courte et distincte de la preuve classique par troncature.

Corollaire 4.11.3. Loi des grands nombres pour les v.a. i.i.d..

Si $(Z_n)_{n \geq 1}$ est une suite de v.a. i.i.d. intégrables alors $\frac{Z_1 + \dots + Z_n}{n} \rightarrow \mathbb{E}(Z_1)$ p.s. et dans L^1 .

Démonstration. Le théorème 4.11.2 pour la martingale renversée de l'exemple 4.11.1 donne $X_\infty \in L^1$ t.q. $\frac{S_n}{n} \rightarrow X_\infty$, p.s. et dans L^1 . En particulier, $\mathbb{E}(Z_1) = \mathbb{E}(\frac{S_n}{n}) \rightarrow \mathbb{E}(X_\infty)$. De plus, par la loi du zéro-un, la v.a. $\frac{S_n}{n}$ est mesurable pour la tribu terminale de $(Z_n)_{n \geq 1}$, qui se réduit à $\{\emptyset, \Omega\}$, donc X_∞ est constante p.s. ■

Le théorème 4.11.2 possède le raffinement suivant : $X_\infty = \mathbb{E}(X_n | \mathcal{F}_\infty)$ pour tout n , où $\mathcal{F}_\infty = \bigcap_n \mathcal{F}_n$. D'autre part, si X est une martingale renversée, alors $X_n = \mathbb{E}(X_0 | \mathcal{F}_n)$, d'où $\mathbb{E}(|X_n|) \leq \mathbb{E}(|X_0|)$, d'où sa bornitude dans L^1 . Le théorème 4.11.2 peut être généralisé aux surmartingales et aux sous-martingales renversées bornées dans L^1 .

Au-delà de la loi des grands nombres pour les suites i.i.d., les martingales renversées permettent d'étudier les suites échangeables, d'établir la loi du zéro-un de Hewitt-Savage, et le théorème de représentation de de Finetti.

L'expérience montre que l'inégalité maximale de Doob et les martingales renversées ont peu d'usage en dehors de ceux mis en avant ici. En revanche les théorèmes de convergence de martingales (bornées dans L^2 , ou L^p , $p > 1$), de surmartingales positives, et de sous-martingales bornées dans L^1 sont très utilisés dans les applications.

C'est le moment d'un rappel sur les sources de convergence p.s. La plupart des preuves de convergence p.s. reposent au bout du compte sur l'une des deux propriétés élémentaires suivantes (parfois via le lemme de Borel-Cantelli) : si X est à valeurs dans $[0, \infty]$, alors $\mathbb{E}(X) < \infty$ implique $X < \infty$ p.s. et $\mathbb{E}(X) = 0$ implique $X = 0$ p.s.

4.12 Théorème limite central pour les martingales de carré intégrable

Théorème 4.12.1. Théorème limite central pour les martingales de carré intégrable.

Soit $(M_n)_{n \in \mathbb{N}}$ une martingale de carré intégrable et $0 \leq a_n \nearrow +\infty$ une suite déterministe telle que :

- (i) $\frac{\langle M \rangle_n}{a_n} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \sigma^2$ pour une constante $\sigma^2 > 0$, autrement dit $\langle M \rangle_n \underset{n \rightarrow \infty}{\sim} \sigma^2 a_n$ en probabilité.
- (ii) $\frac{1}{a_n} \sum_{k=1}^n \mathbb{E}(|M_k - M_{k-1}|^2 \mathbb{1}_{\{|M_k - M_{k-1}| \geq \varepsilon \sqrt{a_n}\}} | \mathcal{F}_{k-1}) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0$ pour tout $\varepsilon > 0$ (condition de Lindeberg).

Alors

$$\frac{M_n}{\sqrt{a_n}} \xrightarrow[n \rightarrow \infty]{\text{loi}} \mathcal{N}(0, \sigma^2) \quad \text{autrement dit} \quad \frac{M_n}{\sqrt{\langle M \rangle_n}} \xrightarrow[n \rightarrow \infty]{\text{loi}} \mathcal{N}(0, 1).$$

— Le lien entre les deux conclusions découle du lemme de Slutsky.

Il s'agit d'un TLC autonormalisé par la variance conditionnelle $\langle M \rangle_n = \sum_{k=1}^n \mathbb{E}((M_k - M_{k-1})^2 | \mathcal{F}_{k-1})$. Comme $\langle M \rangle_n \rightarrow +\infty$ en probabilité, cela donne en particulier une LGN faible :

$$\frac{M_n}{\langle M \rangle_n} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0.$$

Cette convergence peut être renforcée en une convergence presque sûre (LGN forte).

— La condition de Lindeberg (ii) exprime le fait qu'il est possible de procéder à une troncature pour se ramener à une martingale à accroissements bornés⁷. Cette condition est automatiquement vérifiée lorsque M est bornée car $a_n \rightarrow +\infty$, ce qui annule les indicateurs pour n assez grand. Une condition suffisante est donnée par la condition de Liapounov, qui est plus simple :

- (ii') il existe un réel $p > 1$ tel que $\frac{1}{a_n^p} \sum_{k=1}^n \mathbb{E}(|M_k - M_{k-1}|^{2p} | \mathcal{F}_{k-1}) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0$.

En effet, en notant $\Delta_k := M_k - M_{k-1}$, l'inégalité de Markov⁸ $\mathbb{1}_{\{|\Delta_k| \geq \varepsilon \sqrt{a_n}\}} \leq \frac{|\Delta_k|}{\varepsilon \sqrt{a_n}}$ donne

$$\mathbb{E}(|\Delta_k|^2 \mathbb{1}_{\{|\Delta_k| \geq \varepsilon \sqrt{a_n}\}} | \mathcal{F}_{k-1}) = \mathbb{E}(|\Delta_k|^{2p-1} \mathbb{1}_{\{|\Delta_k| \geq \varepsilon \sqrt{a_n}\}} | \mathcal{F}_{k-1}) \leq \frac{1}{(\varepsilon \sqrt{a_n})^{2(p-1)}} \mathbb{E}(|\Delta_k|^{2p} | \mathcal{F}_{k-1}).$$

— Le théorème s'applique en particulier à la martingale $M_n = Z_1 + \dots + Z_n$ où $(Z_n)_{n \geq 1}$ est une suite de v.a. indépendantes, centrées, de carré intégrable, mais pas forcément de même loi. C'est dans ce contexte non i.i.d. qu'on a développé historiquement les conditions de Lindeberg et de Liapounov. Dans ce cas cela donne $\langle M \rangle_n = \sigma_1^2 + \dots + \sigma_n^2$ où $\sigma_i^2 := \mathbb{E}(Z_i^2)$, de sorte que si $\sigma_i^2 = 1/i$ alors $\langle M \rangle_n \underset{n \rightarrow \infty}{\sim} \log(n)$. Ce TLC peut donc donner accès à des fluctuations à vitesses différentes de \sqrt{n} ! Il est très utilisé dans les applications.

Démonstration. Pour simplifier, et éviter notamment un lourd argument de troncature, nous nous contentons de la preuve dans le cas restreint où $a_n = n$ et où (ii) est remplacé par

(ii'') M est conditionnellement bornée dans L^3 : pour tout $n \geq 1$, $\mathbb{E}(|M_n - M_{n-1}|^3 | \mathcal{F}_{n-1}) \leq 1$.

ce qui implique la condition de Liapounov (ii') avec $p = 3/2$, donc de Lindeberg (ii).

Pour tout $n \geq 1$, soit $\Delta \langle M \rangle_n := \langle M \rangle_n - \langle M \rangle_{n-1} = \mathbb{E}((M_n - M_{n-1})^2 | \mathcal{F}_{n-1}) = \mathbb{E}((\Delta M)_n^2 | \mathcal{F}_{n-1})$.

Observons dès à présent que l'inégalité de Jensen pour l'espérance conditionnelle et (ii'') donnent

$$0 \leq \Delta \langle M \rangle_n = \mathbb{E}((\Delta M)_n^2 | \mathcal{F}_{n-1}) \leq \mathbb{E}(|\Delta M_n|^3 | \mathcal{F}_{n-1})^{2/3} \leq 1.$$

Le résultat à démontrer est équivalent à $\mathbb{E}(e^{i \frac{t}{\sqrt{n}} M_n + \frac{\sigma^2 t^2}{2n}}) \rightarrow 1$ pour tout $t \in \mathbb{R}$. Or (i) implique que $e^{\frac{t^2}{2n} \langle M \rangle_n} \rightarrow e^{\frac{\sigma^2 t^2}{2}}$ en probabilité et donc dans L^1 car la suite est u.i. car bornée car $0 \leq \langle M \rangle_n \leq n$, d'où

$$|e^{i \frac{t}{\sqrt{n}} M_n + \frac{\sigma^2 t^2}{2n}} - e^{i \frac{t}{\sqrt{n}} M_n + \frac{t^2}{2n} \langle M \rangle_n}| = |e^{\frac{\sigma^2 t^2}{2n} \langle M \rangle_n} - e^{\frac{t^2}{2n} \langle M \rangle_n}| \rightarrow 0 \text{ dans } L^1.$$

Il suffit donc d'établir que $\mathbb{E}(e^{i \frac{t}{\sqrt{n}} M_n + \frac{t^2}{2n} \langle M \rangle_n}) \rightarrow 1$ pour tout $t \in \mathbb{R}$. Quitte à remplacer M par la martingale $M - M_0$, qui a le même comportement asymptotique, on peut supposer que $M_0 = 0$. Soit $u \in [-1, 1]$. Posons

$$L_n := e^{iu M_n + \frac{u^2}{2} \langle M \rangle_n} = L_{n-1} Z_n \quad \text{où} \quad Z_n := e^{iu \Delta M_n + \frac{u^2}{2} \Delta \langle M \rangle_n}.$$

7. La troncature est une technique classique pour établir les résultats asymptotiques comme le théorème limite central.

8. Alternativement, il est possible d'utiliser l'inégalité de Hölder puis de Markov conditionnelles, mais c'est plus long et moins élégant.

Montrons que $\mathbb{E}(L_n) \rightarrow 1$ quand $n \rightarrow \infty$, avec une estimée quantitative dépendant de u . Nous avons

$$L_0 = 1 \quad \text{et} \quad \mathbb{E}(L_n | \mathcal{F}_{n-1}) = L_{n-1} \mathbb{E}(Z_n | \mathcal{F}_{n-1}), \quad n \geq 1.$$

Cela permet d'utiliser le lemme 4.12.2 avec $x = \Delta M_n$ et $y = \Delta \langle M \rangle_n \in [0, 1]$, ce qui donne, grâce au fait que $\mathbb{E}(\Delta M_n | \mathcal{F}_{n-1}) = 0$ et $\mathbb{E}(\Delta \langle M \rangle_n - (\Delta M_n)^2 | \mathcal{F}_{n-1}) = 0$ (car M et $M^2 - \langle M \rangle$ sont des martingales!),

$$\mathbb{E}(Z_n | \mathcal{F}_{n-1}) = \mathbb{E}\left(1 + iu\Delta M_n + \frac{u^2}{2}(\Delta \langle M \rangle_n - (\Delta M_n)^2) | \mathcal{F}_{n-1}\right) + r_n(u) = 1 + r_n(u)$$

où $r_n(u) := \mathbb{E}(R(u, \Delta M_n, \Delta \langle M \rangle_n) | \mathcal{F}_{n-1})$. De plus (ii'') donne $\mathbb{E}((1 + |\Delta M_n|)^3 | \mathcal{F}_{n-1}) \leq \mathbb{E}(4 + 4|\Delta M_n|^3 | \mathcal{F}_{n-1}) \leq 8$, d'où, par le lemme 4.12.2, $|r_n(u)| \leq \frac{32}{3}|u|^3 e^{\frac{u^2}{2}}$. Ceci donne la majoration récursive suivante :

$$|\mathbb{E}(L_n) - 1| = |\mathbb{E}(L_{n-1} \mathbb{E}(Z_n | \mathcal{F}_{n-1}) - 1)| = |\mathbb{E}(L_{n-1} - 1 + (\mathbb{E}(Z_n | \mathcal{F}_{n-1}) - 1)L_{n-1})| \leq |\mathbb{E}(L_{n-1}) - 1| + \mathbb{E}(|L_{n-1}|) \frac{32}{3}|u|^3 e^{\frac{u^2}{2}}.$$

Or comme $\langle M \rangle_n \leq n$, il vient que $|L_{n-1}| \leq e^{\frac{(n-1)u^2}{2}}$, d'où finalement, pour tout réel $u \in [-1, 1]$ et tout entier $n \geq 0$,

$$|\mathbb{E}(L_n) - 1| \leq \frac{32}{3}n|u|^3 e^{\frac{nu^2}{2}}.$$

Ceci donne bien $\lim_{n \rightarrow \infty} \mathbb{E}(e^{i\frac{t}{\sqrt{n}}M_n + \frac{t^2}{2n}\langle M \rangle_n}) = 1$, pour tout $t \in \mathbb{R}$, en posant $u = \frac{t}{\sqrt{n}}$ car alors $|u| \leq 1$ pour $n \geq t^2$. ■

Lemme 4.12.2. Approximation uniforme à l'usage des fonctions caractéristiques.

Pour tout $x \in \mathbb{R}$, tout $y \in [0, 1]$, et tout $u \in [-1, 1]$,

$$\Phi(u) := e^{iux + \frac{u^2}{2}y} = 1 + iux + \frac{u^2}{2}(y - x^2) + R(u, x, y), \quad \text{avec} \quad |R(u, x, y)| \leq \frac{4}{3}(1 + |x|)^3|u|^3 e^{\frac{u^2}{2}}.$$

Démonstration. Une formule de Taylor avec reste intégral donne

$$\Phi(u) = \Phi(0) + \Phi'(0)u + \Phi''(0)\frac{u^2}{2} + \frac{1}{2}\int_0^u \Phi'''(s)(u-s)^2 ds = 1 + iux + \frac{u^2}{2}(y - x^2) + R(u, x, y).$$

Or comme $\Phi'''(s) = (ix + sy)(3y + (ix + sy)^2)\Phi(s)$, et comme $|s| \leq |u| \leq 1$ et $y \in [0, 1]$, il vient que $|\Phi'''(s)| \leq 4(1 + |x|)^3 e^{\frac{s^2}{2}}$, d'où la majoration annoncée du reste : $|R(u, x, y)| \leq \frac{4}{3}(1 + |x|)^3|u|^3 e^{\frac{u^2}{2}}$. ■

Pour aller plus loin 4.12.3.

- Échangeabilité et martingales renversées, loi du zéro-un de Hewitt–Savage [43] th. de de Finetti [7].
- Plus sur (la LGN et) le TLC pour les martingales [34, Cor. 3.1] et [16, Sec. 2.8.2].
- Plus d'applications des martingales en modélisation [5, 11, 8].

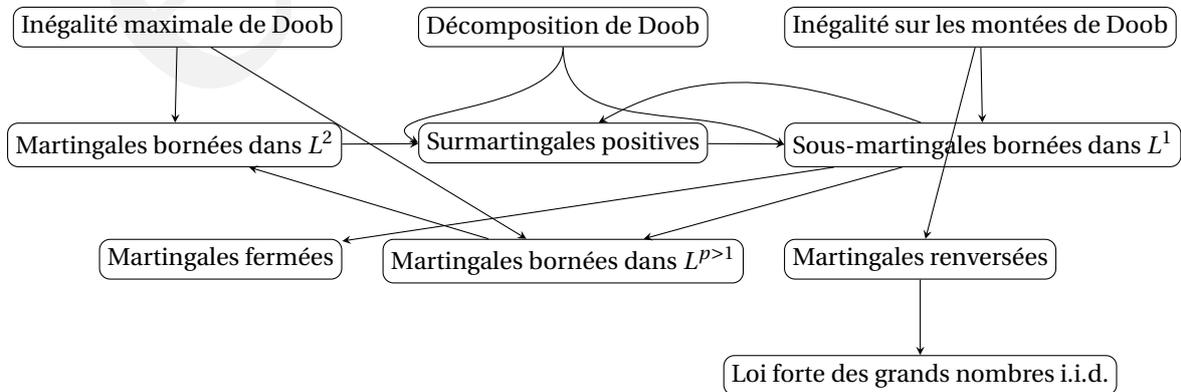


FIGURE 4.2 – Liens entre théorèmes de convergence de martingales.

Chapitre 5

Chaînes de Markov à temps et espace discrets



FIGURE 5.1 – Explosion de Ivy Mike, première bombe à hydrogène, le 1^{er} novembre 1952. Les algorithmes Markov Chain Monte Carlo (MCMC) sont nés des travaux autour de cette bombe H, à Los Alamos, notamment un article paru en 1953 et intitulé *Equation of State Calculations by Fast Computing Machines*, de Nicholas Metropolis, Arianna W. Rosenbluth, Marshall Rosenbluth, Augusta H. Teller et Edward Teller, étendu en 1970 par Wilfred Keith Hastings. Il semble que Metropolis, en charge du MANIAC I, n'a fait que fournir la ressource informatique. Metropolis semble être co-inventeur du terme Monte Carlo, dans un article antérieur écrit avec Stanisław Ulam. D'autre part, Stanisław Ulam, co-concepteur de la bombe H avec Edward Teller, a également développé l'étude des processus de branchement de type Galton – Watson, pour la modélisation des neutrons dans les réactions en chaîne. De nos jours, les algorithmes MCMC font partie des plus utilisés au monde, dans tous les champs applicatifs.

Ce chapitre est consacré aux chaînes de Markov homogènes en temps, à temps et espaces discrets. Ce sont les suites récurrentes aléatoires $(X_n)_{n \in \mathbb{N}}$ de la forme $X_{n+1} = g(X_n, \varepsilon_{n+1})$ où $(\varepsilon_n)_{n \in \mathbb{N}}$ est une suite de v.a. i.i.d. Tout comme les martingales, les chaînes de Markov sont basées sur une structure de dépendance, qui leur confère des propriétés remarquables, et un rôle important en modélisation. Le concept de chaînes de Markov a été introduit par Andrei Andreyevich Markov vers 1906, à l'âge de cinquante ans, pour explorer le phénomène loi des grands nombres au-delà du cadre i.i.d. Cela s'inscrivait dans un contexte philosophique de contestation de la loi des grands nombres, concept probabiliste scandaleux qui mettait en contradiction libre arbitre individuel (indépendance) et déterminisme du comportement collectif (convergence de la moyenne empirique) [40]. Contrairement aux martingales, qui sont devenues des outils figés, les chaînes de Markov portent encore aujourd'hui des problèmes ouverts.

Dans toute la suite, un ensemble au plus dénombrable E est toujours muni de la tribu de toutes ses parties, qui rend continue toute fonction $f : E \rightarrow \mathbb{R}$. Une mesure de Borel μ sur E vérifie $\mu(A) = \sum_{x \in A} \mu(x)$ pour tout $A \subset E$, où $\mu(x) = \mu(\{x\})$. Une fonction $f : E \rightarrow \mathbb{R}$ est μ -intégrable lorsque la somme au plus dénombrable $\sum_{x \in E} |f(x)| \mu(x)$ converge, et dans ce cas, $\int_E f(x) d\mu(x) = \sum_{x \in E} f(x) \mu(x)$. On dit qu'une mesure μ sur E charge l'état $x \in E$ lorsque $\mu(x) > 0$. Les mesures sont toujours supposées non identiquement nulles : elles chargent donc au moins un état.

Il est possible d'aborder ce thème sous un angle algorithmique, sans théorie de la mesure, comme dans [29].

5.1 Suites récurrentes aléatoires et noyaux de transition

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a. à valeurs dans un ensemble au plus dénombrable E . La loi de la suite est une mesure de probabilité sur $E^{\mathbb{N}}$ muni de la tribu engendrée par les cylindres. Le théorème de Carathéodory – Kolmogorov indique qu'elle est caractérisée par ses marges de dimension finie, c'est-à-dire par les lois des vecteurs aléatoires (X_0, \dots, X_m) pour tout $m \in \mathbb{N}$. Or E est au plus dénombrable, et donc la loi de (X_0, \dots, X_m) est caractérisée par la donnée de $\mathbb{P}(X_m = x_m, \dots, X_0 = x_0)$ pour tous x_0, \dots, x_m dans E . Par conditionnements successifs, on peut écrire

$$\mathbb{P}(X_m = x_m, \dots, X_0 = x_0) = \prod_{n=0}^{m-1} \mathbb{P}(X_{n+1} = x_{n+1} \mid X_n = x_n, \dots, X_0 = x_0) \mathbb{P}(X_0 = x_0).$$

Si un événement de la forme $\{X_n = x_n, \dots, X_0 = x_0\}$ est négligeable, il en va de même de $\{X_m = x_m, \dots, X_0 = x_0\}$. Aussi, dans toute la suite, ces cas particuliers sont systématiquement omis par souci de concision.

L'indice n de la suite $(X_n)_{n \in \mathbb{N}}$ est interprété comme un temps. La variable X_n représente la position spatiale à l'instant n , la tribu $\sigma(X_0, \dots, X_{n-1})$ représente son *passé* tandis que la tribu $\sigma(X_{n+1}, X_{n+2}, \dots)$ représente son *futur*.

Définition 5.1.1. Chaîne de Markov.

On dit qu'une suite $(X_n)_{n \in \mathbb{N}}$ de v.a. à valeurs dans un ensemble au plus dénombrable E est une *chaîne de Markov d'espace d'état E* lorsque pour tous $n \in \mathbb{N}$ et $x_0, \dots, x_{n+1} \in E$,

$$\mathbb{P}(X_{n+1} = x_{n+1} \mid X_n = x_n, \dots, X_0 = x_0) = \mathbb{P}(X_{n+1} = x_{n+1} \mid X_n = x_n).$$

Cela s'écrit également

$$\text{Loi}(X_{n+1} \mid X_n, \dots, X_0) = \text{Loi}(X_{n+1} \mid X_n) \quad \text{pour tout } n \in \mathbb{N}.$$

On dit que la chaîne est *homogène* (en temps) lorsque pour tout $n \in \mathbb{N}$ et tous $x, y \in E$,

$$\mathbb{P}(X_{n+1} = y \mid X_n = x) = \mathbb{P}(X_1 = y \mid X_0 = x).$$

Cela s'écrit également

$$\text{Loi}(X_{n+1} \mid X_n = x) = \text{Loi}(X_1 \mid X_0 = x) \quad \text{pour tous } n \in \mathbb{N} \text{ et } x \in E.$$

Les chaînes de Markov modélisent des processus stochastiques sans mémoire. Elles sont incontournables, présentes pratiquement partout : informatique, physique, statistique, finance, ingénierie, biologie, sciences humaines et sociales, etc. Sur le plan mathématique, leur apparente simplicité conceptuelle est trompeuse. Ce domaine porte, encore aujourd'hui, des questions ouvertes ardues et une recherche internationale bien vivante et de haut niveau.

Exemple 5.1.2. Suites récurrentes aléatoires.

Soit X_0 une variable aléatoire à valeurs dans un ensemble au plus dénombrable E .

Soit $(\varepsilon_{n+1})_{n \in \mathbb{N}}$ une suite de v.a. indépendantes à valeurs dans un espace mesurable F , indépendante de X_0 .

Pour tout $n \in \mathbb{N}$, soit $g_n : E \times F \rightarrow E$ mesurable.

La suite récurrente aléatoire $(X_n)_{n \in \mathbb{N}}$ définie par $X_{n+1} = g_n(X_n, \varepsilon_{n+1})$ est une chaîne de Markov sur E .

Elle est homogène en temps si $\text{Loi}(\varepsilon_{n+1})$ et g_n ne dépendent pas du temps n : $(\varepsilon_{n+1})_{n \in \mathbb{N}}$ est i.i.d. et $g_n = g$.

En effet, comme (X_0, \dots, X_n) est $\sigma(X_0, \varepsilon_1, \dots, \varepsilon_n)$ -mesurable donc indépendant de ε_{n+1} ,

$$\mathbb{P}(X_{n+1} = x_{n+1} \mid X_n = x_n, \dots, X_0 = x_0) = \mathbb{P}(g(x_n, \varepsilon_{n+1}) = x_{n+1}).$$

Exemple de la marche aléatoire sur \mathbb{Z} : $g(x, \varepsilon) = x + \varepsilon$ d'où $X_{n+1} = X_0 + \varepsilon_1 + \dots + \varepsilon_{n+1}$.

Exemple de la marche aléatoire à droite sur le groupe symétrique \mathbb{S}_n : $g(x, \varepsilon) = x\varepsilon$ d'où $X_{n+1} = X_0\varepsilon_1 \dots \varepsilon_{n+1}$.

Réciproquement, toute chaîne de Markov est une suite récurrente aléatoire de ce type (théorème 5.5.1).

Pour un entier $r \geq 1$, considérons à présent une suite récurrente aléatoire $(X_n)_{n \in \mathbb{N}}$ d'ordre r dans E , donnée par $X_{n+1} = f(X_n, \dots, X_{n-r}, \varepsilon_{n+1})$, où $(\varepsilon_{n+1})_{n \geq r}$ est une suite de v.a. i.i.d. indépendantes de X_0, \dots, X_{r-1} . On dit qu'il s'agit d'une chaîne de Markov d'ordre r , c'est-à-dire que $((X_{n+r-1}, \dots, X_n))_{n \in \mathbb{N}}$ est une CM sur E^r .

Noyaux de transition

Soit (X_n) une chaîne de Markov d'espace d'état E . Son *noyau de transition* $\mathbf{P} : E \times E \rightarrow [0, 1]$ est défini par

$$\mathbf{P}(x, y) = \mathbb{P}(X_1 = y \mid X_0 = x), \quad x, y \in E.$$

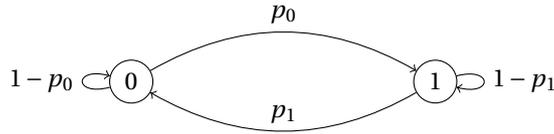


FIGURE 5.2 – Graphe des transitions de la chaîne de Markov à deux états.

Par homogénéité il vient que $\mathbb{P}(X_{n+1} = y \mid X_n = x) = \mathbf{P}(x, y)$ pour tout $n \in \mathbb{N}$ et tous $x, y \in E$. Si $\nu = \text{Loi}(X_0)$, la définition 5.1.1 donne pour tout $n \in \mathbb{N}$ et tous x_0, \dots, x_n dans E ,

$$\mathbb{P}(X_n = x_n, \dots, X_0 = x_0) = \nu(x_0) \prod_{k=0}^{n-1} \mathbf{P}(x_k, x_{k+1}).$$

Ainsi, la loi de la suite $(X_n)_{n \in \mathbb{N}}$ est entièrement caractérisée par la *loi initiale* $\nu = \text{Loi}(X_0)$ et le *noyau de transition* \mathbf{P} . La notation $\text{CM}(E, \nu, \mathbf{P})$ signifie « chaîne de Markov d'espace d'état E , de loi initiale ν , et de noyau de transition \mathbf{P} ». Chaque couple (ν, \mathbf{P}) définit une unique loi sur $E^{\mathbb{N}}$. Dans la suite, on omet parfois ν et/ou \mathbf{P} dans cette notation, pour alléger. Si $(X_n)_{n \in \mathbb{N}}$ et $(Y_n)_{n \in \mathbb{N}}$ ont même loi sur $E^{\mathbb{N}}$ et X est une $\text{CM}(E, \nu, \mathbf{P})$ alors l'est aussi Y aussi.

Exemple 5.1.3. Suites récurrentes aléatoires.

Le noyau de transition de la chaîne de Markov $(X_n)_{n \in \mathbb{N}}$ de l'exemple 5.1.2 est $\mathbf{P}(x, y) = \mathbb{P}(g(x, \varepsilon_1) = y)$.

La notion de noyau peut être isolée du concept de chaîne de Markov. Une application $\mathbf{P} : E \times E \rightarrow \mathbb{R}$ est un noyau de transition sur E ssi pour tous x , $\mathbf{P}(x, \cdot)$, vue comme $\sum_{y \in E} \mathbf{P}(x, y) \delta_y$, est une loi sur E :

$$0 \leq \mathbf{P}(x, y) \leq 1 \quad \text{et} \quad \sum_{z \in E} \mathbf{P}(x, z) = 1.$$

Quitte à numéroter les éléments de E , un noyau de transition \mathbf{P} peut être vu comme une matrice carrée, infinie si E est infini, appelée matrice de transition, dont chaque ligne est une loi sur E .

$$E = \{x_0, x_1, x_2, \dots\} \quad \text{et} \quad \mathbf{P} = \begin{pmatrix} \mathbf{P}(x_0, x_0) & \mathbf{P}(x_0, x_1) & \mathbf{P}(x_0, x_2) & \cdots \\ \mathbf{P}(x_1, x_0) & \mathbf{P}(x_1, x_1) & \mathbf{P}(x_1, x_2) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Soient p_0 et p_1 dans $[0, 1]$. Le noyau de transition \mathbf{P} à deux états est le noyau sur $E = \{0, 1\}$ qui s'écrit

$$\mathbf{P}(0, 1) = 1 - \mathbf{P}(0, 0) = p_0 \quad \text{et} \quad \mathbf{P}(1, 0) = 1 - \mathbf{P}(1, 1) = p_1.$$

Vue comme une matrice 2×2 , le noyau \mathbf{P} s'écrit

$$\mathbf{P} = \begin{pmatrix} \mathbf{P}(0, 0) & \mathbf{P}(0, 1) \\ \mathbf{P}(1, 0) & \mathbf{P}(1, 1) \end{pmatrix} = \begin{pmatrix} 1 - p_0 & p_0 \\ p_1 & 1 - p_1 \end{pmatrix}.$$

Une chaîne de Markov avec ce noyau est une chaîne à deux états. C'est une marche aléatoire sur $\mathbb{Z}/2\mathbb{Z}$ si $p_0 = p_1$. La matrice de transition de la marche aléatoire sur \mathbb{Z} est tridiagonale à diagonales constantes (Toeplitz).

Il est utile de représenter un noyau de transition avec un graphe des transitions, comme dans la figure 5.2.

Structure de l'ensemble des noyaux

Si \mathbf{P} et \mathbf{Q} sont deux noyaux de transition sur E , et si α et β sont deux réels positifs tels que $\alpha + \beta = 1$, alors $\alpha \mathbf{P}(x, y) + \beta \mathbf{Q}(x, y)$ définit un noyau de transition sur E noté $\alpha \mathbf{P} + \beta \mathbf{Q}$. D'autre part, $\sum_{z \in E} \mathbf{P}(x, z) \mathbf{Q}(z, y)$ définit également un noyau de transition sur E noté $\mathbf{P}\mathbf{Q}$. Ainsi, l'ensemble des noyaux de transition sur E est convexe, et constitue un semi-groupe non-abélien pour le produit, d'élément neutre \mathbf{I} défini par $\mathbf{I}(x, y) = \mathbb{1}_{x=y}$. D'autre part, les matrices de permutation de E forment un groupe, inclus dans ce semi-groupe, dont l'action à droite, à gauche, ou par conjugaison sur le semi-groupe correspondent à l'invariance par permutation des lignes et/ou colonnes.

Noyau itéré et écriture matricielle

Soit $(X_n)_{n \in \mathbb{N}}$ une $\text{CM}(E, \nu, \mathbf{P})$. L'application $\mathbf{P}^n : E \times E \rightarrow [0, 1]$ définie par

$$\mathbf{P}^n(x, y) := \mathbb{P}(X_n = y \mid X_0 = x)$$

est un noyau de transition sur E . En particulier $\mathbf{P}^1 = \mathbf{P}$, et $\mathbf{P}^0 = \mathbf{I}$. Pour tous $n \geq 1$ et x_0, x_n dans E ,

$$\mathbf{P}^n(x_0, x_n) = \mathbf{P}(X_n = x_n | X_0 = x_0) = \sum_{x_1, \dots, x_{n-1} \in E} \mathbb{P}(X_n = x_n, \dots, X_1 = x_1 | X_0 = x_0).$$

La suite x_0, \dots, x_n constitue un chemin dans E . Par récurrence sur n , il vient

$$\mathbf{P}^n(x_0, x_n) = \sum_{x_1, \dots, x_{n-1} \in E} \mathbf{P}(x_0, x_1) \cdots \mathbf{P}(x_{n-1}, x_n) = \sum_{x \in E} \mathbf{P}^{n-1}(x_0, x) \mathbf{P}(x, x_n).$$

Ces formules sont celles du produit matriciel : $\mathbf{P}^n = \mathbf{P}\mathbf{P}^{n-1} = \mathbf{P}^{n-1}\mathbf{P}$.

Pour le processus de Bernoulli de l'exemple 5.2.2, le noyau itéré est $\mathbf{P}^n(x, \cdot) = \text{Loi}(B_n | B_0 = x) = \delta_x * \text{Bin}(n, p)$.

En assimilant toute loi ν sur E à un *vecteur ligne*, et toute fonction $f : E \rightarrow \mathbb{R}$ à un *vecteur colonne*, la moyenne de f pour ν s'écrit, lorsque cela a un sens, $\nu f = \sum_{x \in E} f(x) \nu(x)$ (série si E est infini). En assimilant tout noyau de transition à une matrice carrée infinie, si $(X_n)_{n \in \mathbb{N}}$ est une CM(E, ν, \mathbf{P}), alors $\text{Loi}(X_n) = \nu \mathbf{P}^n$ et $\mathbb{E}(f(X_n)) = \nu \mathbf{P}^n f$, car

$$\mathbb{E}(f(X_n)) = \sum_{y \in E} \mathbb{P}(X_n = y) f(y) = \sum_{x, y \in E} \nu(x) \mathbf{P}^n(x, y) f(y) = \nu \mathbf{P}^n f.$$

La famille $\{\mathbf{P}^n : n \in \mathbb{N}\}$ est un semi-groupe (de Markov) d'opérateurs linéaires agissant sur les mesures de probabilités sur E . En notant $\nu_n := \text{Loi}(X_n) = \nu \mathbf{P}^n$, la suite $(\nu_n)_{n \in \mathbb{N}}$ se calcule par récurrence linéaire

$$\nu_0 = \nu \quad \text{et} \quad \nu_{n+1} = \nu_n \mathbf{P}.$$

Les trajectoires $(X_n)_{n \in \mathbb{N}}$ ne sont ni déterministes ni linéaires, mais l'évolution des lois instantanées est déterministe et linéaire. Ainsi, la structure markovienne des phénomènes d'évolution aléatoire sans mémoire sur E fait apparaître une structure linéaire, non pas sur E qui n'est pas forcément structuré, mais sur l'ensemble des lois sur E . Faire émerger une équation d'évolution déterministe de loi de probabilité se retrouve dans l'équation de Boltzmann (non-linéaire) en théorie cinétique des gaz, avant Markov, et chez Euler, avant Boltzmann, en mécanique des fluides, en concevant la probabilité comme un fluide incompressible (lois de conservation : positivité, et intégrale).

Notation conditionnelle

Si $(X_n)_{n \in \mathbb{N}}$ est une chaîne de Markov sur E , alors pour tout $x \in E$, tout $A \subset E$, et toute v.a. Z , on note

$$\mathbb{P}^x(A) = \mathbb{P}(A | X_0 = x) \quad \text{et} \quad \mathbb{E}^x(Z) = \mathbb{E}(Z | X_0 = x).$$

5.2 Exemples de chaînes de Markov

Exemple 5.2.1. Marches aléatoires sur les groupes et sur les graphes.

La marche aléatoire^a simple (MAS) sur \mathbb{Z}^d , $d \geq 1$, est la suite récurrente aléatoire $(X_n)_{n \in \mathbb{N}}$ de \mathbb{Z}^d définie par

$$X_{n+1} = X_n + \varepsilon_{n+1} = X_0 + \varepsilon_1 + \cdots + \varepsilon_n$$

où $(\varepsilon_{n+1})_{n \in \mathbb{N}}$ est une suite de v.a. i.i.d. de loi uniforme sur $\{\pm e_1, \dots, \pm e_d\}$ où $\{e_1, \dots, e_d\}$ est la base canonique de \mathbb{R}^d , et où X_0 est une v.a. sur \mathbb{Z}^d indépendante de $(\varepsilon_n)_{n \geq 1}$. C'est une CM($\mathbb{Z}^d, \text{Loi}(X_0), \mathbf{P}$) de noyau

$$\mathbf{P}(x, y) = \begin{cases} \frac{1}{2d} & \text{si } x \text{ et } y \text{ sont voisins dans } \mathbb{Z}^d, \text{ c'est-à-dire } |x - y|_1 = 1 \\ 0 & \text{sinon} \end{cases}.$$

Cela modélise une diffusion, comme le mouvement brownien, qui en constitue la limite d'échelle. Il est possible de définir une MA sur le cercle discret $\mathbb{Z}/r\mathbb{Z}$ en considérant l'addition modulo r dans \mathbb{Z} . Plus généralement, il est possible de définir une marche aléatoire sur un groupe et même sur un semi-groupe, commutatif ou non, à partir d'accroissements i.i.d. de loi uniforme sur une partie du groupe, typiquement génératrice du groupe, comme par exemple les transpositions sur le groupe symétrique. Une MA sur le groupe symétrique peut modéliser par exemple le battage d'un paquet de cartes à jouer. Dans une autre direction, il est possible de définir une marche aléatoire sur les sommets d'un graphe, en faisant en sorte que $\mathbf{P}(x, \cdot)$ soit la loi uniforme sur les voisins de x dans le graphe. Les deux concepts peuvent se rejoindre via le graphe de Cayley d'un groupe finiment engendré. Une MA sur un graphe peut modéliser par exemple le robot d'archivage du moteur de recherche de Google. Il est également possible de considérer la MA sur

le graphe dual, ce qui revient à marcher sur les arêtes plutôt que les sommets. Il est enfin possible d'utiliser des accroissements non-uniformes. Les MA sont des objets fondamentaux et ubiquitaires des probabilités.

a. Le terme est attribué à George Pólya, à qui on doit la classification des états de la marche sur \mathbb{Z}^d en toute dimension d vers 1920.

Exemple 5.2.2. Processus de Bernoulli du jeu de pile ou face.

Les résultats successifs dans un jeu de pile ou face sont modélisés par une suite $(\varepsilon_{n+1})_{n \in \mathbb{N}}$ de v.a. i.i.d. de loi de Bernoulli $p\delta_1 + (1-p)\delta_0$, $p \in [0, 1]$. Le nombre de gains en n parties est $B_n = \varepsilon_1 + \dots + \varepsilon_n$. En posant $B_0 := 0$, la suite $(B_n)_{n \in \mathbb{N}}$ est une CM($\mathbb{N}, \delta_0, \mathbf{P}$) appelée processus de Bernoulli, de noyau de transition

$$\mathbf{P}(x, y) = \begin{cases} p & \text{si } y = x + 1 \\ 1 - p & \text{si } y = x \\ 0 & \text{sinon} \end{cases}.$$

C'est une marche aléatoire sur le semi-groupe commutatif $(\mathbb{N}, +)$.

Exemple 5.2.3. Autres exemples.

Dans les modèles aléatoires, les chaînes de Markov apparaissent souvent explicitement sous forme de suites récurrentes aléatoires. Elles ne se confondent pas forcément avec les éventuelles martingales présentes dans le modèle. Voici quelques exemples, en plus des marches aléatoires mentionnées précédemment :

- *Marche aléatoire sur l'hypercube discret.* Considérons la suite $(X_n)_{n \in \mathbb{N}}$ sur $E = \{0, 1\}^d = (\mathbb{Z}/2\mathbb{Z})^d$, construite comme suit : sachant que $X_n = x$, pour fabriquer X_{n+1} , on choisit uniformément une coordonnée parmi $\{1, \dots, d\}$, puis on la transforme en 1 si elle était égale à 0 et réciproquement. C'est une chaîne de Markov, et une marche aléatoire sur un groupe produit. C'est le modèle le plus simple exhibant des effets dimensionnels comme le phénomène de convergence abrupte à l'équilibre.
- *Urne d'Ehrenfest.* On dispose de $d > 1$ molécules ou particules réparties dans deux récipients ou urnes adjacentes A et B connectés par un petit trou. On note X_n le nombre de particules dans A au temps n . Sachant que $X_n = x$, pour fabriquer X_{n+1} , on choisit uniformément l'une des d particules, et on la change de récipient. La suite $(X_n)_{n \in \mathbb{N}}$ est une chaîne de Markov sur $E = \{0, 1, \dots, d\}$. Ehrenfest, ancien élève de Boltzmann, a introduit ce modèle pour proposer un dépassement du paradoxe entre théorème de récurrence de Poincaré et irréversibilité en théorie cinétique des gaz de Boltzmann.
- *Modèle de Wright–Fisher.* La martingale $(X_n)_{n \in \mathbb{N}}$ considérée dans la section 4.2 est une suite récurrente aléatoire donc une chaîne de Markov sur $\{0/N, 1/N, \dots, N/N\}$.
- *Urnes de Pólya.* La suite $(A_n, B_n)_{n \in \mathbb{N}}$ considérée dans la section 4.2 est une suite récurrente aléatoire et donc une chaîne de Markov sur \mathbb{N}^2 , qui diffère de la martingale considérée dans cet exemple.
- *Processus de Galton–Watson.* La suite $(Z_n)_{n \in \mathbb{N}}$ considérée dans la section 4.2 est une suite récurrente aléatoire, donc une chaîne de Markov sur \mathbb{N} , distincte des martingales considérées quand $m \neq 1$.
- *Recherche de motif.* La suite $(U_n)_{n \in \mathbb{N}}$ de l'exemple 4.4.5 est une suite récurrente aléatoire, donc une chaîne de Markov, qui n'est pas la martingale utilisée dans cet exemple.

5.3 Propriété de Markov forte

Propriété de Markov faible : si $(X_n)_{n \in \mathbb{N}}$ est une CM(E, ν, \mathbf{P}) alors pour tous $m, n \geq 0$, et tous $A \in E^n$ et $C \in E^m$,

$$\begin{aligned} \mathbb{P}((X_{m+n}, \dots, X_m) \in C, (X_m, \dots, X_0) \in A \mid X_m = x) &= \mathbb{P}((X_{m+n}, \dots, X_n) \in C \mid X_m = x) \mathbb{P}((X_n, \dots, X_0) \in A \mid X_m = x) \\ &= \mathbb{P}((X_n, \dots, X_0) \in C \mid X_0 = x) \mathbb{P}((X_m, \dots, X_0) \in A \mid X_m = x). \end{aligned}$$

Autrement dit, pour tout $m \in \mathbb{N}$, et conditionnellement à $\{X_m = x\}$, la suite $(X_{m+n})_{n \in \mathbb{N}}$ est une CM(E, δ_x, \mathbf{P}) indépendante de $\sigma(X_0, \dots, X_m)$. Il est peut-être éclairant et amusant d'imaginer une trajectoire aléatoire, punaisée en x aux temps m , et qui s'agite des deux côtés, de manière indépendante. Cela s'étend à des temps d'arrêt :

Théorème 5.3.1. Propriété de Markov forte.

Soit $(X_n)_{n \in \mathbb{N}}$ une CM(E), et $T : \Omega \rightarrow \mathbb{N} \cup \{\infty\}$ un temps d'arrêt pour la filtration naturelle $(\mathcal{F}_n)_{n \in \mathbb{N}}$. Alors pour tout $x \in E$, sachant $\{T < \infty \text{ et } X_T = x\}$, $(X_{T+n})_{n \in \mathbb{N}}$ est une CM(E, δ_x, \mathbf{P}), indépendante de \mathcal{F}_T .

Autrement dit, pour tout $A \in \mathcal{F}_T$, tout événement cylindrique $C \subset E^{\mathbb{N}}$, et tout $x \in E$,

$$\mathbb{P}(\{(X_{T+n})_{n \in \mathbb{N}} \in C\} \cap A \mid T < \infty, X_T = x) = \mathbb{P}^x(\{(X_n)_{n \in \mathbb{N}} \in C\} \mid T < \infty, X_T = x).$$

Autrement dit, pour tout $A \in \mathcal{F}_T$, toute fonction mesurable positive ou bornée $f : E^{\mathbb{N}} \rightarrow \mathbb{R}$, et tout $x \in E$,

$$\mathbb{E}(f((X_{T+n})_{n \in \mathbb{N}}) \mathbb{1}_A \mid T < \infty, X_T = x) = \mathbb{E}^x(f((X_n)_{n \in \mathbb{N}})) \mathbb{P}(A \mid T < \infty, X_T = x).$$

La propriété de Markov faible correspond au cas où T est constant et égal à m .

Démonstration. Pour tous $m \in \mathbb{N}$, $A \in \mathcal{F}_T$, et $x \in E$, $A \cap \{T = m\} \cap \{X_T = x\} \in \mathcal{F}_m$, et par propriété de Markov faible,

$$\mathbb{P}(\{X_{T+n} = x_n, \dots, X_T = x_0\} \cap A \cap \{T = m\} \cap \{X_T = x\}) = \mathbb{P}(X_n = x_n, \dots, X_0 = x_0 \mid X_0 = x_0) \mathbb{P}(A \cap \{T = m\} \cap \{X_T = x\})$$

pour tous $x_1, \dots, x_n \in E$ et avec $x_0 := x$. En sommant sur $m \in \mathbb{N}$, il vient

$$\mathbb{P}(\{X_{T+n} = x_n, \dots, X_T = x_0\} \cap A \cap \{T < \infty\} \cap \{X_T = x\}) = \mathbb{P}(X_n = x_n, \dots, X_0 = x_0 \mid X_0 = x_0) \mathbb{P}(A \cap \{T < \infty\} \cap \{X_T = x\}).$$

La propriété désirée s'obtient alors en divisant par $\mathbb{P}(T < \infty \text{ et } X_T = x)$. Pour la loi, prendre $A = \Omega$. ■

Temps de premier passage, temps d'atteinte, nombre de visites, potentiel

Si $(X_n)_{n \in \mathbb{N}}$ est une CM(E) et $F \subset E$, alors le temps de premier passage en F et le temps d'atteinte de F sont

$$T_F^* := \inf\{n \geq 1 : X_n \in F\} \quad \text{et} \quad T_F := \inf\{n \geq 0 : X_n \in F\}.$$

Ces v.a. prennent leurs valeurs dans $\{1, 2, \dots\} \cup \{\infty\}$ et $\{0, 1, 2, \dots\} \cup \{\infty\}$, et $T_F^* \geq T_F$. Ce sont des temps d'arrêt pour la filtration naturelle de X , égaux sur $\{X_0 \notin F\}$. Sur $\{X_0 \in F\}$, T_F^* est aussi le temps de retour en F .

Le nombre de visites en F est comptabilisé par les v.a. suivantes, qui prennent leurs valeurs dans $\mathbb{N} \cup \{\infty\}$:

$$N_F^* := \sum_{n=1}^{\infty} \mathbb{1}_{\{X_n \in F\}} \quad \text{et} \quad N_F := \sum_{n=0}^{\infty} \mathbb{1}_{\{X_n \in F\}} = \mathbb{1}_{\{X_0 \in F\}} + N_F^*.$$

En particulier $\{T_F^* = \infty\} = \{N_F^* = 0\}$ et $\{T_F = \infty\} = \{N_F = 0\}$. Notons également que $N_F = \infty$ ssi $N_F^* = \infty$.

Lorsque $F = \{x\}$, on note T_x^* , T_x , N_x^* , et N_x .

Les nombres moyens de visites en y partant de x sont donnés par les quantités suivantes dans $\mathbb{N} \cup \{\infty\}$:

$$\mathbf{G}^*(x, y) := \mathbb{E}^x(N_y^*) = \sum_{n=1}^{\infty} \mathbf{P}^n(x, y) \quad \text{et} \quad \mathbf{G}(x, y) := \mathbb{E}^x(N_y) = \sum_{n=0}^{\infty} \mathbf{P}^n(x, y) = (\mathbf{I} + \mathbf{G}^*)(x, y).$$

Les fonctions \mathbf{G}^* et \mathbf{G} jouent un rôle important et similaire, en particulier $\mathbf{G}(x, y) = \infty$ ssi $\mathbf{G}^*(x, y) = \infty$.

Pour tous $x, y \in E$, la quantité suivante joue un rôle clé dans l'étude des trajectoires : $\rho(x, y) := \mathbb{P}^x(T_y^* < \infty)$.

Théorème 5.3.2. Temps de retour et potentiel.

- (i) Pour tous $x, y \in E$, $\mathbb{P}^x(N_y = \infty) = \mathbb{P}^y(N_y = \infty) \rho(x, y)$.
En particulier, si $\rho(x, x) < 1$, alors $\mathbb{P}^x(N_x = \infty) = 0$.
- (ii) Pour tout $y \in E$, si $\mathbf{G}^*(y, y) < \infty$, alors $\mathbf{G}^*(x, y) < \infty$ pour tout $x \in E$, et $\rho(x, y) = \frac{\mathbf{G}^*(x, y)}{\mathbf{G}^*(y, y)} = \frac{\mathbf{G}^*(x, y)}{1 + \mathbf{G}^*(y, y)}$.
En particulier, $\rho(x, x) = 1$ ssi $\mathbf{G}^*(x, x) = \infty$.
- (iii) Pour tous $x, y \in E$, la transformée de Laplace de T_y^* est donnée, pour tout $\lambda > 0$, par

$$\mathbb{E}^x(e^{-\lambda T_y^*}) = \frac{\mathbf{G}_\lambda^*(x, y)}{1 + \mathbf{G}_\lambda^*(y, y)} \quad \text{où} \quad \mathbf{G}_\lambda^*(x, y) := \sum_{n=1}^{\infty} e^{-\lambda n} \mathbf{P}^n(x, y).$$

Démonstration.

- (i) La propriété de Markov forte pour T_y^* et le fait que $X_{T_y^*} = y$ donnent

$$\begin{aligned} \mathbb{P}^x(N_y = \infty) &= \mathbb{P}^x(M_y = \infty) = \mathbb{E}^x(\mathbb{1}_{\{T_y^* < \infty\}} \mathbb{1}_{\sum_{n=0}^{\infty} \mathbb{1}_{\{X_{T_y^*+n} = y\}} = \infty}) = \mathbb{E}^y(\mathbb{1}_{\{\sum_{n=0}^{\infty} \mathbb{1}_{\{X_n = y\}} = \infty\}}) \mathbb{P}^x(T_y^* < \infty, X_{T_y^*} = y) \\ &= \mathbb{P}^y(N_y = \infty) \mathbb{P}^x(T_y^* < \infty). \end{aligned}$$

(ii) De même pour (i), nous avons, dans $[0, \infty]$,

$$\mathbf{G}^*(x, y) = \mathbb{E}^x \left(\mathbb{1}_{\{T_y^* < \infty\}} \sum_{n=0}^{\infty} \mathbb{1}_{\{X_{T_y^*+n}=y\}} \right) = \mathbb{E}^x \left(\sum_{n=0}^{\infty} \mathbb{1}_{\{X_n=y\}} \right) \mathbb{P}^x(T_y^* < \infty, X_{T_y^*} = y) = \mathbf{G}(y, y) \mathbb{P}^x(T_y^* < \infty).$$

(iii) La série $\mathbf{G}_\lambda^*(x, y)$ converge car $0 \leq \mathbf{P}^n(x, y) \leq 1$. Par la propriété de Markov forte donne (iii) via

$$\mathbf{G}_\lambda^*(x, y) = \mathbb{E}^x \left(\mathbb{1}_{\{T_y^* < \infty\}} \sum_{n=0}^{\infty} e^{-\lambda(T_y^*+n)} \mathbb{1}_{\{X_{T_y^*+n}=y\}} \right) = \mathbb{E}^y \left(\sum_{n=0}^{\infty} e^{-\lambda n} \mathbb{1}_{\{X_n=y\}} \right) \mathbb{E}^x(\mathbb{1}_{\{T_y^* < \infty\}} e^{-\lambda T_y^*}) = (1 + \mathbf{G}_\lambda^*(y, y)) \mathbb{E}^x(e^{-\lambda T_y^*}).$$

Note : (iii) donne (ii) car par convergence monotone, $\lim_{\lambda \rightarrow 0} \mathbf{G}_\lambda^*(x, y) = \mathbf{G}^*(x, y)$ et $\lim_{\lambda \rightarrow 0} \mathbb{E}^x(e^{-\lambda T_y^*}) = \mathbb{P}^x(T_y^* < \infty)$. ■

5.4 Classification des états, récurrence et transience

Soit $(X_n)_{n \in \mathbb{N}}$ une CM(E, ν, \mathbf{P}). Un état $x \in E$ est dit...

— récurent lorsque $\rho(x, x) = 1$, ou de manière équivalente, lorsque $\mathbf{G}(x, x) = \infty$.

— transitoire lorsque $\rho(x, x) < 1$, ou de manière équivalente, lorsque $\mathbf{G}(x, x) < \infty$.

Si x est récurrent, alors la propriété de Markov donne p.s. une infinité de temps de retour successifs en x , qui délimitent des excursions indépendantes et de même loi enracinées en x . Ce découpage de la trajectoire indique que si la probabilité de passer par y est non-nulle lors d'une excursion, alors le nombre total de passages en y le long de la trajectoire est infini p.s. (jeu de pile ou face). Autrement dit, si $\rho(x, x) = 1$ et $\rho(x, y) > 0$, alors $\mathbb{P}^x(N_y = \infty) = 1$.

Si x est transitoire, alors le nombre de retours ou d'excursions est fini p.s. et de loi géométrique. Autrement dit, si $\rho(x, x) < 1$, alors $\text{Loi}(N_x | X_0 = x) = \text{Geo}_{\mathbb{N}^*}(1 - \rho(x, x))$, de moyenne $1/(1 - \rho(x, x)) = \mathbf{G}(x, x)$ (pile ou face).

La table 5.1 regroupe quelques propriétés remarquables de ce type, pour certaines issues du théorème 5.3.2.

État x récurrent	État x transitoire
$\rho(x, x) = 1, \mathbf{G}(x, x) = \infty$	$\rho(x, x) < 1, \mathbf{G}(x, x) < \infty$
$\rho(x, y) > 0 \Rightarrow \mathbb{P}^x(N_y = \infty) = 1$	$\mathbb{P}^x(N_x = \infty) = 0$
$\rho(y, x) = \mathbb{P}^y(N_x = \infty)$	$\text{Loi}(N_x X_0 = x) = \text{Geo}_{\mathbb{N}^*}(1 - \rho(x, x))$
$\rho(y, x) > 0 \Rightarrow \mathbf{G}(y, x) = \infty$	$\mathbf{G}^*(y, x) = \rho(y, x)/(1 - \rho(x, x)) < \infty$

TABLE 5.1 – Propriétés des états récurrents et transitoires, valables pour tout $y \in E$.

On note E_R l'ensemble des états récurrents et $E_T = E \setminus E_R$ celui des états transitoires.

On dit que $F \subset E$ est clos lorsque $\mathbb{P}^x(T_{F^c}^* < \infty) = 0$ pour tout $x \in F$. La chaîne ne peut pas en sortir.

Un état x est absorbant lorsque $\{x\}$ est clos, ou de manière équivalente lorsque $\mathbf{P}(x, x) = 1$.

Exemple 5.4.1. Processus de Bernoulli.

Considérons le processus de Bernoulli de l'exemple 5.2.2. Si $p > 0$, alors tous les états sont transitoires car le processus ne fait que croître, $\mathbb{P}^x(T_x^* = \infty) = 1$. Si $p = 0$, tous les états sont absorbants (chaîne constante).

Exemple 5.4.2. Marche aléatoire sur \mathbb{Z} .

Considérons la MA sur \mathbb{Z} de l'exemple 5.2.1. Comme $\mathbf{P}^n(x, x) = \mathbb{P}(U_1 + \dots + U_n = 0) = \mathbf{P}^n(0, 0)$, il vient que

$$\mathbf{G}(x, x) = \mathbf{G}(0, 0) = \sum_{n=0}^{\infty} \mathbf{P}^n(0, 0).$$

Les états sont donc soit tous récurrents soit tous transitoires. Or $(U_1 + 1 + \dots + U_n + 1)/2 \sim \text{Bin}(n, 1/2)$, d'où $\mathbf{P}^{2k+1}(0, 0) = 0$ et $\mathbf{P}^{2k}(0, 0) = 2^{-2k} (2k)! / (k!)^2$, d'où $\mathbf{G}(0, 0) = \infty$ (Stirling), et tous les états sont récurrents.

On dit que x mène ou conduit à y , et on note $x \rightarrow y$, lorsque x et y sont reliés par un chemin qui suit les flèches dans le graphe des transitions de \mathbf{P} , c'est-à-dire qu'il existe $n > 0$ tel que $\mathbf{P}^n(x, y) > 0$. On dit que x et y communiquent, et on note $x \leftrightarrow y$, lorsque $x = y$ ou bien lorsqu'à la fois $x \rightarrow y$ et $y \rightarrow x$. Avec cette définition, on a toujours $x \leftrightarrow x$, même lorsque $\mathbf{P}(x, x) = 0$. Si $x \neq y$, alors $x \rightarrow y$ ssi $\mathbf{G}(x, y) > 0$. Si $x \rightarrow y$ et $y \not\rightarrow x$ alors x est transitoire.

Théodéf 5.4.3. Décomposition de l'espace d'état en classes irréductibles.

Pour tout noyau de transition \mathbf{P} sur E :

- (i) La relation binaire \leftrightarrow est une relation d'équivalence sur E .
Ces classes d'équivalence, qui partitionnent E , sont appelées classes irréductibles.
- (ii) Dans chaque classe, les états sont de même nature : tous récurrents ou tous transitoires.
- (iii) Les classes constituées d'états récurrents sont closes et appelées classes de récurrence.
Les classes constituées d'état transitoires sont appelées classes transitoires.

- Les ensembles clos, comme par exemple les classes de récurrence, sont absorbants.
Le cas des états absorbants est singulier puisque leur classe de récurrence est réduite à eux-mêmes.
- Presque sûrement, une chaîne partant d'un état récurrent repasse une infinité de fois par son état initial, ne s'échappe jamais de sa classe de récurrence, et visite une infinité de fois tous les états de cette classe.
- Presque sûrement, une chaîne partant d'un état transitoire ne peut repasser qu'un nombre fini de fois par son état initial, et peut être capturée par une classe de récurrence ou visiter d'autres états transitoires (un nombre fini de fois pour chacun). En particulier, si E est fini, alors il existe forcément un état récurrent.
- L'ensemble des états transitoires peut être infini, et la chaîne peut ne jamais être capturée par une classe de récurrence. C'est toujours le cas en l'absence d'état récurrent, comme pour le processus de Bernoulli avec $p > 0$, pour lequel il existe une infinité dénombrable de classes transitoires avec des passages à sens unique.
- Contrairement aux classes de récurrence, les classes transitoires peuvent ne pas être closes. Des passages (à sens unique) peuvent exister entre elles, ainsi que vers les classes de récurrence. Sur le graphe des transitions, les flèches qui pénètrent dans une classe close proviennent toujours d'états transitoires.

On dit que \mathbf{P} est irréductible lorsque E est constitué d'une unique classe, c'est-à-dire lorsque tous les états communiquent : pour tous x et y dans E , il existe $n \geq 0$ tel que $\mathbf{P}^n(x, y) > 0$. On dit que \mathbf{P} est récurrent lorsque $E_T = \emptyset$, et transitoire lorsque $E_R = \emptyset$. Ces termes sont également utilisés, au féminin, pour toute CM de noyau \mathbf{P} .

$$\begin{aligned} \text{noyau irréductible} & \begin{cases} (1) \text{ une seule classe de récurrence (noyau récurrent)} \\ (2) \text{ une seule classe transitoire (noyau transitoire)} \end{cases} \\ \text{noyau réductible} & \begin{cases} (3) \text{ plusieurs classes de récurrence et aucun état transitoire} \\ (4) \text{ au moins une classe de récurrence et des états transitoires} \\ (5) \text{ plusieurs classes transitoires} \end{cases} \end{aligned}$$

L'étude des classes de récurrence des cas (3-4) se ramène au cas (1).

L'étude des classes transitoires closes des cas (4-5) se ramène au cas (2).

Exemple 5.4.4.

Considérons le noyau de transition \mathbf{P} sur $E = \{a, b, c, d\}$ donné par

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

L'état a est absorbant, donc récurrent. L'état b est transitoire car $b \rightarrow c$ mais aucun état ne mène à b . Les états c, d sont récurrents car le temps d'atteinte de c partant de d et de d partant de c sont finis p.s. (lois géom.). Les classes irréductibles de \mathbf{P} sont $\{a\}$, $\{c, d\}$, $\{b\}$. Les deux premières sont closes (et récurrentes).

Remarque 5.4.5. Cas des suites de v.a. i.i.d.

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a. i.i.d. de loi μ sur E , telle que $\mu(x) > 0$ pour tout $x \in E$. La suite X est une CM (E, μ, \mathbf{P}) où $\mathbf{P}(x, y) := \mu(y)$ pour tous x et y . Ce noyau est irréductible car $\mathbf{P}(x, y) > 0$ pour tous x et y . De plus $T_x^* \sim \text{Geo}_{\mathbb{N}}(\mu(x))$, $\mathbb{P}^x(T_x^* < \infty) = 1$, et $\mathbb{E}^x(T_x^*) = 1/\mu(x)$ (jeu de pile ou face). Ainsi, \mathbf{P} est irréductible récurrent. De plus $\mathbf{P}^n(x, y) = \mu(y)$, et $\mathbf{G}(x, x) = \sum_{n=0}^{\infty} \mathbf{P}^n(x, x) = \sum_{n=0}^{\infty} \mu(x) = \infty$. Ainsi, dans le cas i.i.d., la loi initiale se confond avec le noyau. En général, le comportement d'une CM dépend à la fois de la loi initiale et de la classification des états associée au noyau. Cependant, pour certains noyaux, le comportement asymptotique des chaînes associées ne dépend que du noyau et pas de la loi initiale (suite du chapitre!).

5.5 Simulation des trajectoires : suites récurrentes aléatoires, générateur, puissances

Approche récursive par construction de suite récurrente aléatoire

La propriété de Markov faible indique que pour simuler une trajectoire $(x_n)_{n \in \mathbb{N}}$ d'une CM $(X_n)_{n \in \mathbb{N}}$, il suffit de procéder récursivement : simuler la loi initiale $\nu = \text{Loi}(X_0)$, ce qui donne x_0 , puis la loi $\mathbf{P}(x_0, \cdot)$, ce qui donne x_1 , etc.

Théorème 5.5.1. Suites récurrentes aléatoires.

Pour tout noyau de transition \mathbf{P} sur E , et toute loi ν sur E , il existe des fonctions

$$f : [0, 1] \rightarrow E \quad \text{et} \quad g : E \times [0, 1] \rightarrow E$$

telles que la suite récurrente aléatoire $(X_n)_{n \in \mathbb{N}}$ définie par

$$X_0 = f(U_0) \quad \text{et} \quad X_{n+1} = g(X_n, U_{n+1})$$

est une CM (E, ν, \mathbf{P}) , pour toute suite $(U_n)_{n \in \mathbb{N}}$ de v.a. i.i.d. de loi uniforme sur $[0, 1]$.

Démonstration. Après avoir identifié E à \mathbb{N} , on pose, pour tous $n \in \mathbb{N}$ et $u \in [0, 1]$,

$$f(u) := \min\{m \in \mathbb{N} : \nu(0) + \dots + \nu(m) \geq u\} \quad \text{et} \quad g(n, u) := \min\{m \in \mathbb{N} : \mathbf{P}(n, 0) + \dots + \mathbf{P}(n, m) \geq u\}.$$

Ce sont les fonctions de répartition inverses des lois ν et $\mathbf{P}(n, \cdot)$. ■

Pour simuler récursivement une trajectoire $(x_n)_{n \in \mathbb{N}}$ d'une CM (E, ν, \mathbf{P}) , on simule une réalisation $x_0 = f(u_0)$ de la loi initiale ν , puis le reste de la trajectoire par la formule de récurrence $x_{n+1} = g(x_n, u_{n+1})$. Cela fait envisager les CM comme des suites récurrentes aléatoires. Le théorème 5.5.1 réduit également la construction effective des CM (E, ν, \mathbf{P}) à celle des suites i.i.d. de loi uniforme sur $[0, 1]$, elle-même réductible à un jeu de pile ou face (bits).

Approche par sauts et générateur (découpage événementiel de la trajectoire)

Soit $(X_n)_{n \in \mathbb{N}}$ une CM (E, ν, \mathbf{P}) et $(S_n)_{n \in \mathbb{N}}$ ses temps de saut définis par récurrence par

$$S_0 := 0 \quad \text{et} \quad S_{n+1} := \inf\{k > S_n : X_k \neq X_{S_n}\}.$$

La suite $(S_n)_{n \in \mathbb{N}}$ est à valeurs dans $\mathbb{N} \cup \{\infty\}$, strictement croissante tant qu'elle est finie, et éventuellement absorbée par ∞ si la trajectoire se termine par un état absorbant. La suite $(X_n)_{n \in \mathbb{N}}$ est constante et égale à X_{S_n} sur l'intervalle de temps aléatoire $\{m \in \mathbb{N} : S_n \leq m < S_{n+1}\}$. Par la propriété de Markov forte, pour tous m, x, y ,

$$\begin{aligned} \mathbb{P}(S_{n+1} - S_n = m, X_{S_{n+1}} = y \mid S_n < \infty, X_{S_n} = x) &= \mathbb{P}(X_{S_n+m} = y, X_{S_n+m-1} = x, \dots, X_{S_n} = x \mid S_n < \infty, X_{S_n} = x) \\ &= \begin{cases} \mathbf{P}(x, x)^{m-1} \mathbf{P}(x, y) & \text{si } m > 0 \text{ et } y \neq x \\ 0 & \text{sinon} \end{cases}. \end{aligned}$$

Il en découle que pour tout état x et tout temps n ,

- si $\mathbf{P}(x, x) = 1$, alors $\mathbb{P}(S_{n+1} = \infty \mid S_n < \infty, X_{S_n} = x) = 1$ et on dit que x est absorbant.
- si $\mathbf{P}(x, x) < 1$, alors $\mathbb{P}(S_{n+1} < \infty \mid S_n < \infty, X_{S_n} = x) = 1$, $S_{n+1} - S_n$ et $X_{S_{n+1}}$ sont indépendantes conditionnellement à $\{S_n < \infty, X_{S_n} = x\}$, et de plus, pour tout $m > 0$ et tous $y \neq x$, le jeu de pile ou face donne

$$\begin{aligned} \mathbb{P}(S_{n+1} - S_n = m \mid S_n < \infty, X_{S_n} = x) &= \mathbf{P}(x, x)^{m-1} (1 - \mathbf{P}(x, x)) \\ \mathbb{P}(X_{S_{n+1}} = y \mid S_n < \infty, X_{S_n} = x) &= \frac{\mathbf{P}(x, y)}{1 - \mathbf{P}(x, x)}. \end{aligned}$$

Cela donne un nouvel algorithme de simulation des trajectoires, basé sur la simulation des deux lois

$$\text{Geo}_{\mathbb{N}^*}(-\mathbf{L}(x, x)) \quad \text{et} \quad \sum_{y \neq x} \frac{-\mathbf{L}(x, y)}{\mathbf{L}(x, x)} \delta_y \quad \text{où} \quad \mathbf{L} := \mathbf{P} - \mathbf{I}.$$

Elles correspondent aux temps et positions des sauts. L'application $\mathbf{L} : E \times E \rightarrow \mathbb{R}$, appelée générateur, vérifie

$$\mathbf{L}(x, x) = \mathbf{P}(x, x) - 1, \quad \mathbf{L}(x, y) = \mathbf{P}(x, y) \text{ si } x \neq y, \quad \text{et} \quad \sum_y \mathbf{L}(x, y) = 0.$$

Ce nouvel algorithme par sauts est d'autant plus préférable que les coefficients diagonaux de \mathbf{P} sont proches de 1. La terminologie vient du fait que \mathbf{L} est le générateur du semi-groupe $(\mathbf{P}^n)_{n \in \mathbb{N}}$ au sens où $\partial_n^+ \mathbf{P}^n = \mathbf{P}^{n+1} - \mathbf{P}^n = \mathbf{L} \mathbf{P}^n$.

Remarque 5.5.2. Potentiel et inverse du générateur.

Soit \mathbf{P} un noyau de transition de potentiel $\mathbf{G} = \sum_{n=0}^{\infty} \mathbf{P}^n$ et de générateur $\mathbf{L} = \mathbf{P} - \mathbf{I}$. On a $\mathbf{G} = \lim_{n \rightarrow \infty} \mathbf{G}_n$ où $\mathbf{G}_n := \sum_{k=0}^n \mathbf{P}^k$, et $\mathbf{L}\mathbf{G}_n = \mathbf{G}_n\mathbf{L} = \mathbf{P}^{n+1} - \mathbf{I}$. Ainsi, si \mathbf{P} est transitoire, alors \mathbf{G} est un inverse de $-\mathbf{L}$. Pour la marche aléatoire simple, \mathbf{L} est un laplacien (discret), dont l'étude constitue la théorie du potentiel (discrète). Par analogie, la fonction $\mathbf{G} : E \times E \rightarrow \mathbb{N} \cup \{\infty\}$ est appelée fonction de Green ou potentiel de l'opérateur \mathbf{L} .

Approche matricielle

Dans certains cas, comme par exemple pour le jeu de pile ou face, le noyau itéré \mathbf{P}^n est explicite et simple. Pour simuler X_n sur l'événement $\{X_0 = x_0\}$ sans passer par la trajectoire x_0, \dots, x_n , il suffit d'utiliser la loi $\mathbf{P}^n(x_0, \cdot)$ sur E . Lorsque E est fini, le coût peut être réduit en calculant par récurrence la loi $\text{Loi}(X_n) = \nu \mathbf{P}^n$ par multiplication du vecteur ligne $\nu \mathbf{P}^k$ par la matrice \mathbf{P} , pour $1 \leq k \leq n$. Cela nécessite de l'ordre de $n|E|^2$ opérations élémentaires.

5.6 Atteinte, absorption, harmonicité, martingales

Lorsque que $F \subset E$ est clos, le temps d'atteinte $T_F = \inf\{n \geq 0 : X_n \in F\}$ de F devient un temps d'absorption puisque la chaîne ne peut pas s'échapper de F . C'est le cas lorsque F est une classe de récurrence. Le vecteur des probabilités d'atteinte $(a_F(x))_{x \in E}$ et le vecteur des temps moyens d'atteinte $(m_F(x))_{x \in E}$ sont définis par

$$a_F(x) := \mathbb{P}^x(T_F < \infty) \quad \text{et} \quad m_F(x) := \mathbb{E}^x(T_F).$$

Les valeurs prises sont respectivement dans $[0, 1]$ et $[0, \infty]$. On a $m_F(x) = \mathbb{E}^x(T_F^*)$ si $x \notin F$. On a toujours $a_F = 0$ et $m_F = \infty$ sur tout ensemble clos G disjoint de F . C'est le cas avec $G = E_R \setminus F$ lorsque F est une classe de récurrence.

Théorème 5.6.1. Atteinte et absorption.

Si $(X_n)_{n \in \mathbb{N}}$ est une CM(E, ν, \mathbf{P}), et $F \subset E$, alors les propriétés suivantes ont lieu.

(i) Le vecteur $(a_F(x))_{x \in E}$ est solution positive ou nulle minimale ^a du système linéaire

$$a_F(x) = \begin{cases} 1 & \text{si } x \in F, \\ \sum_{y \in E} \mathbf{P}(x, y) a_F(y) & \text{si } x \notin F. \end{cases}$$

Si F est clos, cela s'écrit $\mathbf{P}a_F = a_F$ où a_F est un vecteur colonne, c'est-à-dire $\mathbf{L}a_F = 0$ (harmonicité).

(ii) Le vecteur $(m_F(x))_{x \in E}$ est solution positive ou nulle minimale du système linéaire (dans $[0, \infty]$)

$$m_F(x) = \begin{cases} 0 & \text{si } x \in F, \\ 1 + \sum_{y \in E} \mathbf{P}(x, y) m_F(y) & \text{si } x \notin F. \end{cases}$$

Si F est clos, cela s'écrit $\mathbb{1}_{F^c} + \mathbf{P}m_F = m_F$, c'est-à-dire $\mathbf{L}m_F = -\mathbb{1}_{F^c}$.

^a. Signifie que si a est également une solution positive ou nulle, alors $a_F(x) \leq a(x)$ pour tout $x \in E$.

Seule la somme sur F^c intervient dans l'équation de m_F car $m_F = 0$ sur F .

Démonstration. L'écriture pour F clos est immédiate. Les propriétés (i) et (ii) s'établissent de la même manière. Si $x \in F$ alors $a_F = 1$. Soit $x \notin F$. Nous avons $a_F(x) = \sum_{y \in E} \mathbb{P}^x(T_F < \infty \mid X_1 = y) \mathbf{P}(x, y)$. Or $T_F \geq 1$ sur $\{X_0 = x\}$ car $x \notin F$, et donc par la propriété de Markov faible $\mathbb{P}^x(T_F < \infty \mid X_1 = y) = \mathbb{P}^y(T_F < \infty) = a_F(y)$ pour tout y . Ainsi, a_F est bien solution. Soit a une autre solution positive. Alors $a = a_F = 1$ sur F . D'autre part, si $x \notin F$, alors en opérant des substitutions successives dans la formule $a(x) = \sum_{y \in F} \mathbf{P}(x, y) + \sum_{y \notin F} \mathbf{P}(x, y) a(y)$, on obtient, pour tout n ,

$$a(x) = \mathbb{P}^x(X_1 \in F) + \dots + \mathbb{P}^x(X_1 \notin F, \dots, X_{n-1} \notin F, X_n \in F) + \varepsilon_n = \mathbb{P}^x(T_F \leq n) + \varepsilon_n,$$

où $\varepsilon_n = \sum_{y_1, \dots, y_n \notin F} \mathbf{P}(x, y_1) \mathbf{P}(y_1, y_2) \dots \mathbf{P}(y_{n-1}, y_n) a(y_n) \geq 0$, d'où $a(x) \geq \mathbb{P}^x(T_F \leq n) \nearrow a_F(x)$ quand $n \rightarrow \infty$. ■

Exemple 5.6.2. Marche aléatoire sur \mathbb{N} et ruine du joueur.

Soit $0 < p < 1$. Considérons un joueur qui, à chaque étape, peut gagner 1 Euro avec probabilité p ou perdre 1 Euro avec probabilité $1 - p$. Le joueur commence le jeu avec une fortune initiale X_0 , et cesse de jouer dès que sa fortune est nulle. La suite (X_n) est une chaîne de Markov sur \mathbb{N} de noyau de transition \mathbf{P} vérifiant

$\mathbf{P}(0, 0) = 1$, $\mathbf{P}(x, x+1) = p$ et $\mathbf{P}(x, x-1) = 1-p$ pour tout $x > 0$. Une telle chaîne constitue une marche aléatoire sur \mathbb{N} , tuée ou absorbée en 0. L'état 0 est absorbant donc $\{0\}$ est une classe de récurrence, et tous les états non nuls communiquent, et mènent à 0, donc \mathbb{N}^* est une classe transitoire. Mais cela ne permet pas d'affirmer que la chaîne est absorbée par l'origine p.s. Par le théorème 5.6.1 pour l'ensemble clos $F = \{0\}$, le vecteur des probabilités d'absorption $a = a_F$ est solution positive minimale de l'équation donnée par

$$a(x) = pa(x+1) + (1-p)a(x-1), \text{ si } x > 0, \text{ et } a(0) = 1.$$

Supposons maintenant que le jeu est équitable : $p = 1/2$. Les solutions de la récurrence linéaire sont alors de la forme $a(x) = \alpha + \beta x$. Les conditions $a(0) = 1$ et $0 \leq a \leq 1$ entraînent $\alpha = 1$ et $\beta = 0$. Ainsi, $a(x) = 1$ pour tout x , ce qui signifie que quelle que soit sa fortune initiale x , le joueur finira ruiné presque sûrement! Ce modèle tolère des fortunes arbitrairement grandes au cours du jeu, ce qui n'est pas réaliste. Cette étude markovienne de la ruine est à comparer à celle basée sur les martingales de l'exemple 4.4.6.

Théorème 5.6.3. Martingale de chaîne de Markov.

Soit $(X_n)_{n \in \mathbb{N}}$ une CM(E, ν, \mathbf{P}), $\mathbf{L} = \mathbf{P} - \mathbf{I}$ son générateur, et $f : E \rightarrow \mathbb{R}$ intégrable pour toutes les lois $\nu \mathbf{P}^n$. Alors la suite $(M_n)_{n \in \mathbb{N}}$ ci-dessous est une martingale pour la filtration naturelle de X :

$$M_0 := 0 \text{ et } M_n := f(X_n) - f(X_0) - \sum_{k=0}^{n-1} (\mathbf{L}f)(X_k), \quad n \geq 1.$$

Autrement dit, la décomposition de Doob de $(f(X_n))_{n \in \mathbb{N}}$ s'écrit $f(X_n) = f(X_0) + M_n + \sum_{k=0}^{n-1} (\mathbf{L}f)(X_k)$.

Si f est harmonique : $\mathbf{L}f = 0$, c'est-à-dire que $\mathbf{P}f = f$, alors $(f(X_n))_{n \in \mathbb{N}}$ est une martingale.

Si f est sous-harmonique : $\mathbf{L}f \geq 0$, c'est-à-dire $\mathbf{P}f \geq f$, alors $(f(X_n))_{n \in \mathbb{N}}$ est une sous-martingale.

Si f est surharmonique : $\mathbf{L}f \leq 0$, c'est-à-dire $\mathbf{P}f \leq f$, alors $(f(X_n))_{n \in \mathbb{N}}$ est une surmartingale.

La terminologie vient du fait que le générateur d'une marche aléatoire simple sur \mathbb{Z}^d est un laplacien discret. Les termes (sous-|sur)martingale viennent de ce lien avec les fonctions (sous-|sur)harmoniques.

Démonstration. Le processus est adapté. Par hypothèse $\mathbb{E}(f(X_n)) = \nu \mathbf{P}^n f$ et $\mathbb{E}((\mathbf{L}f)(X_k)) = \nu(\mathbf{P}^{k+1} - \mathbf{P}^k)f$ sont des quantités bien définies et finies. Pour tout $n \geq 1$, en notant $Z_k := f(X_k) - \mathbb{E}(f(X_k) | X_{k-1}) = f(X_k) - \mathbb{E}(f(X_k) | \mathcal{F}_{k-1})$,

$$M_n = \sum_{k=0}^{n-1} [f(X_{k+1}) - f(X_k) - (\mathbf{P}f)(X_k) + f(X_k)] = \sum_{k=1}^n [f(X_k) - (\mathbf{P}f)(X_{k-1})] = \sum_{k=1}^n Z_k, \text{ et } \mathbb{E}(Z_k | \mathcal{F}_{k-1}) = 0.$$

Par exemple, le vecteur a_F des probabilités d'atteinte de $F \subset E$ clos est harmonique et borné (théorème 5.6.1).

Si $(X_n)_{n \in \mathbb{N}}$ est une CM(E, \mathbf{P}) et $E \subset \mathbb{R}$, si X_0 est intégrable et si $\mathbf{P}(x, \cdot)$ a pour moyenne x pour tout $x \in E$, alors l'application identité $x \mapsto x$ est harmonique, et donc $(X_n)_{n \in \mathbb{N}}$ est une martingale.

Théorème 5.6.4. Représentation probabiliste de Kakutani de la solution du problème de Dirichlet.

Soit $(X_n)_{n \in \mathbb{N}}$ une CM(E, ν, \mathbf{P}) et $\mathbf{L} = \mathbf{P} - \mathbf{I}$ son générateur.

Soit $D \subset E$ et $\bar{D} := D \cup \partial D$, où $\partial D := \{x \notin D : \exists y \in D, \mathbf{P}(x, y) > 0\}$ est son bord extérieur pour \mathbf{P} .

Supposons que D et ∂D sont finis et non-vides, et que $T_{\partial D} := \inf\{n \in \mathbb{N} : X_n \in \partial D\} < \infty$ p.s. sur $\{X_0 \in \bar{D}\}$.

Alors pour tout $g : \partial D \rightarrow \mathbb{R}$, la fonction $f : \bar{D} \rightarrow \mathbb{R}$ définie par

$$f(x) = \mathbb{E}^x(g(X_{T_{\partial D}})) \text{ est solution de } \begin{cases} \mathbf{L}f = 0 & \text{sur } D \\ f = g & \text{sur } \partial D \end{cases}.$$

De plus cette fonction de Kakutani est l'unique solution de ce problème de Dirichlet discret.

- La quantité $\mathbf{L}(f)(x)$ est bien définie pour $f : \bar{D} \rightarrow \mathbb{R}$ et $x \in D$ car par définition de ∂D , elle ne nécessite pas de connaître les valeurs de f en dehors de \bar{D} . Autrement dit, elle ne dépend pas du prolongement à tout E .
- La formule de Kakutani indique que la solution hérite de g le fait d'être nulle/positive/négative.
- Ce théorème exprime sous forme probabiliste la solution d'un système d'équations linéaires.
- Ce théorème permet d'utiliser la méthode de Monte Carlo pour l'approximation de la solution.
- Ce théorème fournit une fonction harmonique sur D à condition au bord prescrite.

- Comme \bar{D} est fini, toute solution est bornée, car définie sur \bar{D} .
- Dans le cas de la marche aléatoire simple sur \mathbb{Z}^d , l'ensemble ∂D coïncide avec le bord dans \mathbb{Z}^d pour la norme ℓ^1 . De plus, si D est fini, alors ∂D et donc \bar{D} aussi, tandis que $T_{\partial D} < \infty$ p.s. par l'argument habituel de découpage de la trajectoire en morceaux de longueur fixe supérieure au diamètre de D et réduction au premier succès dans un jeu de pile ou face ou au lemme de Borel–Cantelli.
- La fonction a_F du théorème 5.6.1 pour $F \subset E$ clos résout un problème de Dirichlet sur $D = F^c$ avec $g \equiv 1$.
- Historiquement, le problème de Dirichlet et la solution de Kakutani ont été introduits pour le laplacien continu $\Delta = \sum_{i=1}^d \partial_i^2$ sur \mathbb{R}^d , et la solution de Kakutani fait alors intervenir le mouvement brownien, dont le laplacien est le générateur. Le théorème 5.6.4 couvre le cas de la marche aléatoire simple sur \mathbb{Z}^d , dont le générateur est le laplacien discret $(\Delta f)(x) = \frac{1}{2d} \sum_{y \sim x} f(y) - f(x)$. Le phénomène TLC fait le lien.
- Le bord ∂D fait intervenir la géométrie du graphe des transitions, qui peut être très variée, incluant graphe complet, arbres, courbure positive, etc. La géométrie des graphes est beaucoup plus diverse que celle des variétés différentielles de la géométrie différentielle, qui sont localement des espaces euclidiens. La marche aléatoire simple sur \mathbb{Z}^d constitue une version graphique ou discrète du mouvement brownien sur \mathbb{R}^d .

Démonstration. Montrons que la fonction f donnée par la formule de Kakutani est solution. Pour tout $x \in \partial D$, $f(x) = g(x)$, tandis que pour tout $x \in D$, $(\Delta f)(x) = 0$ car sur $\{X_0 = x\}$, $T_{\partial D} \geq 1$, et la propriété de Markov faible donne

$$f(x) = \sum_{y \in E} \mathbb{E}^x(g(X_{T_{\partial D}}) \mathbb{1}_{\{X_1=y\}}) = \sum_{y \in E} \mathbf{P}(x, y) \mathbb{E}^y(g(X_{T_{\partial D}})) = \mathbf{P}(f)(x).$$

Démontrons l'unicité avec une martingale. Soit $\tilde{f} : \bar{D} \rightarrow \mathbb{R}$ une solution du problème de Dirichlet. Prolongeons \tilde{f} par 0 en dehors de \bar{D} . Ce prolongement est borné, et le théorème 5.6.3 indique que $(\tilde{f}(X_n) - \sum_{k=0}^{n-1} (\Delta \tilde{f})(X_k))_{n \in \mathbb{N}}$ est une martingale. Supposons à présent que $X_0 = x \in \bar{D}$. Pour tout $0 \leq k < T_{\partial D}$, $X_k \in D$, donc $(\Delta \tilde{f})(X_k) = 0$ car \tilde{f} est harmonique sur D , ainsi $\sum_{k=0}^{n \wedge T_{\partial D} - 1} (\Delta \tilde{f})(X_k) = 0$ pour tout $n \in \mathbb{N}$. Donc par le théorème 4.4.1 d'arrêt de Doob, $(\tilde{f}(X_{n \wedge T_{\partial D}}))_{n \in \mathbb{N}}$ est une martingale, bornée, et $\tilde{f}(x) = \mathbb{E}(\tilde{f}(X_0)) = \mathbb{E}(\tilde{f}(X_{T_{\partial D}})) = \mathbb{E}(g(X_{T_{\partial D}}))$ car $\tilde{f} = g$ sur ∂D .

Il est possible de démontrer l'unicité avec le principe du maximum. Par linéarité, il suffit d'établir que la fonction identiquement nulle est l'unique solution lorsque $g = 0$. Soit donc \tilde{f} solution avec $g = 0$. Comme \bar{D} est fini, \tilde{f} atteint son maximum sur \bar{D} . Il suffit d'établir que \tilde{f} ne peut atteindre son maximum que sur ∂D . Car dans ce cas, par symétrie, il en sera de même pour le minimum, or $\tilde{f} = 0$ sur ∂D , d'où $\tilde{f} = 0$ sur \bar{D} . Supposons donc qu'il existe $x \in D$ tel que $\tilde{f}(x) = M := \max_{y \in \bar{D}} \tilde{f}(y)$. Comme \tilde{f} est solution, il vient, en itérant, pour tout $n \geq 1$,

$$M = \tilde{f}(x) = \sum_{y \in \bar{D}} \mathbf{P}(x, y) \tilde{f}(y) = \sum_{y \in D} \mathbf{P}(x, y) \tilde{f}(y) = \dots = \sum_{y \in D} \mathbf{P}^n(x, y) \tilde{f}(y) \leq M \mathbf{P}^n(x, D).$$

Or comme $T < \infty$ p.s. il vient que $\mathbf{P}^n(x, \partial D) > 0$ pour n assez grand, donc $\mathbf{P}^n(x, D) < 1$, d'où la contradiction. ■

Remarque 5.6.5. H-transformée de Doob.

Dans le même registre que le problème de Dirichlet, soit \mathbf{P} un noyau de transition sur E , $h : E \rightarrow \mathbb{R}$ harmonique positive, et $E_h = \{h > 0\} = \{h \neq 0\}$. Alors $\mathbf{P}^h(x, y) := \mathbf{P}(x, y)h(y)/h(x)$ définit un noyau de transition sur E_h . Cette déformation du processus (h-transformée de Doob) a pour effet d'empêcher la sortie de E_h .

Remarque 5.6.6. Martingales et chaînes de Markov.

Les martingales à temps discret du chapitre 4 prennent leurs valeurs dans \mathbb{R} . Les chaînes de Markov à temps et espace discrets de ce chapitre prennent leur valeurs dans un ensemble E qui n'a pas de structure en général. Si $(\varepsilon_{n+1})_{n \in \mathbb{N}}$ est une suite de v.a. réelles i.i.d. centrées indépendantes de X_0 , alors la suite récurrente aléatoire $(X_n)_{n \in \mathbb{N}}$ définie par $X_{n+1} = \varepsilon_{n+1} X_0 + X_n$ est une martingale mais pas une chaîne de Markov. Une chaîne de Markov est une suite récurrente aléatoire d'ordre 1, c'est-à-dire un processus autoregressif d'ordre 1 possiblement non-linéaire, pas forcément à valeurs réelles ou vectorielles. Au contraire, la propriété de martingale tolère des récursions d'ordre arbitraire, mais impose une structure réelle ou vectorielle.

Remarque 5.6.7. Fonction de chaîne de Markov et critère de Dynkin.

Si $X = (X_n)_{n \in \mathbb{N}}$ est une CM(E, \mathbf{P}) et $f : E \rightarrow F$ est injective alors $f(X) = (f(X_n))_{n \in \mathbb{N}}$ est une CM(F), car si $X_{n+1} = g(X_n, \varepsilon_{n+1})$ alors $f(X_{n+1}) = f(g(f^{-1}(f(X_n)), \varepsilon_{n+1}))$. En revanche, ce n'est plus toujours vrai si f n'est pas injective. Plus généralement, on dispose par exemple d'une condition suffisante appelée critère de Dyn-

kin^a : si pour tout x la loi $\mathbf{P}(x, \cdot)$ ne dépend de x que via $f(x)$, alors $f(X)$ est une CM. En effet, cela revient à dire que $g(x, \varepsilon)$ ne dépend de x que via $a = f(x)$, et donc, en notant $\tilde{g}(a, \varepsilon)$ cette valeur commune, il vient

$$f(X_{n+1}) = f(g(X_n, \varepsilon_{n+1})) = \sum_{a \in F} \sum_{x \in f^{-1}(a)} f(g(x, \varepsilon_{n+1})) \mathbb{1}_{\{X_n=x\}} = \sum_{a \in F} f(\tilde{g}(a, \varepsilon_{n+1})) \mathbb{1}_{\{f(X_n)=a\}} = f(\tilde{g}(f(X_n), \varepsilon_{n+1})).$$

Le noyau \mathbf{Q} de $f(X)$ est alors donné pour tous $a, b \in F$ par

$$\mathbf{Q}(a, b) = \mathbb{P}(f(\tilde{g}(a, \varepsilon_1)) = b) = \sum_{\substack{y \in f^{-1}(b) \\ x \in f^{-1}(a)}} \mathbb{P}(g(x, \varepsilon_1) = y) = \sum_{x \in f^{-1}(a)} \mathbf{P}(x, f^{-1}(b)) = |f^{-1}(a)| \tilde{\mathbf{P}}(a, f^{-1}(b)).$$

a. Proposé par Eugene Dynkin vers 1965.

5.7 Chaînes irréductibles, mesures invariantes, récurrence positive, théorème ergodique

L'étude des classes closes, transitoire ou récurrentes, revient à étudier les chaînes irréductibles. En revanche, l'étude des classes transitoires non closes ne s'y ramène pas, et doit être menée spécifiquement.

Mesures et lois invariantes

Une mesure $\mu \neq 0$ sur E est invariante pour le noyau de transition \mathbf{P} lorsque pour tout y ,

$$\mu(y) = \sum_{x \in E} \mu(x) \mathbf{P}(x, y).$$

Cela s'écrit $\mu = \mu \mathbf{P}$, ou encore $\mu \mathbf{L} = 0$, c'est une notion duale¹ de celle de fonction harmonique positive. En particulier, μ est aussi invariante pour \mathbf{P}^n pour tout n . Pour une mesure de probabilité ou loi μ , son invariance pour \mathbf{P} revient à dire que pour si $(X_n)_{n \in \mathbb{N}}$ est une CM(E, \mathbf{P}), alors

$$X_0 \sim \mu \Rightarrow X_n \sim \mu \text{ pour tout temps } n.$$

Il s'agit d'un équilibre en loi invariant par la dynamique, et l'équation $\mu = \mu \mathbf{P}$ est une équation de point fixe.

Soit μ une mesure invariante pour \mathbf{P} . Soit x un état repéré sur le graphe des transitions de \mathbf{P} . Multiplions le poids de chaque flèche arrivant en x par le poids $\mu(y)$ où y est l'état de départ de la flèche. La somme totale de ces produits doit être égale à $\mu(x)$ (équation bilan). Il en découle que si, sur le graphe des transitions, un état x ne reçoit aucune flèche mais envoie au moins une flèche à l'extérieur, alors il ne peut pas exister de mesure invariante, car cela violerait le bilan de masse. Un tel état constitue une classe transitoire singleton (état fuyant en quelque sorte).

Exemple 5.7.1. Mesures et lois invariantes.

- *Cas i.i.d.* Si $(X_n)_{n \in \mathbb{N}}$ est une suite i.i.d. de loi μ sur E , alors X est une CM sur E et μ est invariante.
- *Chaîne à deux états.* Une mesure μ est invariante pour la chaîne à deux états de matrice

$$\begin{pmatrix} 1 - p_0 & p_0 \\ p_1 & 1 - p_1 \end{pmatrix}$$

ssi $p_0 \mu(0) = p_1 \mu(1)$. En particulier, lorsque $p_0 + p_1 > 0$, la mesure μ définie par $\mu(0) = p_1 / (p_0 + p_1)$ et $\mu(1) = p_0 / (p_0 + p_1)$ est l'unique loi invariante. Si $p_0 = p_1 = 0$, toutes les mesures sont invariantes.

- *Processus de Bernoulli.* Pour le processus de Bernoulli de l'exemple 5.2.2, une mesure μ sur \mathbb{N} invariante ssi $\mu(0) = (1 - p)\mu(0)$ et $\mu(x) = (1 - p)\mu(x) + p\mu(x - 1)$ pour tout $x > 0$. Ainsi, si $0 < p < 1$, alors il n'y a pas de mesure invariante (la classe transitoire $\{0\}$ perd de la masse), si $p = 0$, alors toutes les mesures sont invariantes, tandis que si $p = 1$, alors toute mesure invariante μ est constante.
- *Multiplicité des mesures invariantes pour un noyau irréductible.* Considérons la marche aléatoire sur \mathbb{Z} de l'exemple 5.2.1 de loi d'incrément $p\delta_1 + (1 - p)\delta_0$, $0 < p < 1$. Une mesure μ est invariante ssi $\mu(x) = \alpha + \beta(p/(1 - p))^x$ pour tout $x \in \mathbb{Z}$, où $\alpha, \beta \in \mathbb{R}$. Il n'y a donc pas unicité à normalisation près.
- *Absence de mesure invariante pour un noyau irréductible.* Le processus de Derman est la chaîne de Markov irréductible sur $E = \mathbb{N}$ de noyau $\mathbf{P}(x, x + 1) = p_x$ et $\mathbf{P}(x, 0) = 1 - p_x$, $x \geq 1$, et $\mathbf{P}(0, 1) = 1$, où $0 < p_x < 1$ et $\prod_{x=1}^{\infty} p_x > 0$. Si μ est une mesure invariante, alors $\mu(x + 1) = \mu(x) p_x$, d'où $\mu(x + 1) = \prod_{y=1}^x p_y$, tandis que $\mu(0) = \sum_{x=1}^{\infty} \mu(x) (1 - p_x) = \sum_{x=1}^{\infty} (\mu(x) - \mu(x + 1)) = \mu(1) - \lim_{x \rightarrow \infty} \mu(x) = \mu(1) - \prod_{x=1}^{\infty} p_x$. Or $\mu(1) = \mu(0)$ car $\mathbf{P}(0, 1) = 1$ est la seule transition menant à 1, d'où une contradiction car $\prod_{x=1}^{\infty} p_x \neq 0$.

1. Lorsque E est fini, vecteur propre de \mathbf{P} à gauche plutôt qu'à droite, associé à la valeur propre 1, et à coordonnées positives.

Structure de l'ensemble des mesures invariantes

La linéarité de l'équation $\mu\mathbf{P} = \mu$ entraîne que l'ensemble des mesures invariantes est un cône convexe : si μ_1 et μ_2 sont invariantes et si α_1 et α_2 sont deux réels positifs avec $\alpha_1 + \alpha_2 > 0$, alors la mesure $\alpha_1\mu_1 + \alpha_2\mu_2$ est également invariante. L'ensemble des lois invariantes est également convexe (prendre $\alpha_1 + \alpha_2 = 1$). Si μ est une mesure invariante, alors tout multiple de μ est également invariante. En particulier, si μ est invariante et vérifie $\mu(E) < \infty$ (toujours vrai si E est fini), alors $\mu(E)^{-1}\mu$ est une loi invariante.

Récurrence positive et récurrence nulle

Si $(X_n)_{n \in \mathbb{N}}$ est une CM(E), pour tous $x, y \in E$, le nombre moyen de passages en y avant retour en x est

$$\mu_x(y) := \mathbb{E}^x \left(\sum_{n=0}^{T_x^* - 1} \mathbb{1}_{\{X_n=y\}} \right) \in [0, \infty].$$

Notons que $\mu_x(x) = 1$, et $\mu_x(y) = \delta_x(y)$ si x est absorbant. De plus, $\mu_x(y) > 0$ si et seulement si $x \rightarrow y$.

Lemme 5.7.2. Nombre moyen de passages avant retour.

Si \mathbf{P} est un noyau de transition irréductible sur E , alors pour tout $x \in E$:

- (i) $\mu_x \geq \mu_x \mathbf{P}$ c'est-à-dire que $\mu_x(y) \geq \sum_{z \in E} \mu_x(z) \mathbf{P}(z, y)$ pour tout $y \in E$.
- (ii) μ_x est une mesure qui charge tous les états, c'est-à-dire que $0 < \mu_x(y) < \infty$ pour tout $y \in E$.
- (iii) μ_x a pour masse totale $\mathbb{E}^x(T_x^*) \in [0, \infty]$, c'est-à-dire que $\mu_x(E) = \mathbb{E}^x(T_x^*)$.
- (iv) si \mathbf{P} est de plus récurent alors il y a égalité dans (i) et μ_x est une mesure invariante.

Le (i), qui ne nécessite pas l'irréductibilité, s'écrit aussi $\mu_x \mathbf{L} \leq 0$, notion duale de la surharmonicité.

Démonstration. Pour (i), nous avons $\mu_x(y) \geq \mathbb{E}^x \left(\sum_{n=1}^{T_x^*} \mathbb{1}_{\{X_n=y\}} \right)$ avec égalité si \mathbf{P} est récurrent car dans ce cas $T_x^* < \infty$ et $X_{T_x^*} = x$. Ce décalage du temps permet d'exploiter une transition en fin de trajectoire, plus précisément

$$\mathbb{E}^x \left(\sum_{n=1}^{T_x^*} \mathbb{1}_{\{X_n=y\}} \right) = \mathbb{E}^x \left(\sum_{n=1}^{\infty} \mathbb{1}_{\{X_n=y, T_x^* \geq n\}} \right) = \sum_{n=1}^{\infty} \mathbb{P}^x(X_n = y, T_x^* \geq n),$$

ce qui permet alors un découpage selon X_{n-1} et l'usage de la propriété de Markov faible, ce qui donne (i) :

$$\begin{aligned} \sum_{n=1}^{\infty} \mathbb{P}^x(X_n = y, T_x^* \geq n) &= \sum_{n=1}^{\infty} \sum_{z \in E} \mathbb{P}^x(X_n = y; X_{n-1} = z; T_x^* \geq n) \\ &= \sum_{z \in E} \sum_{n=1}^{\infty} \mathbb{P}(X_n = y \mid X_{n-1} = z, T_x^* \geq n) \mathbb{P}^x(X_{n-1} = z, T_x^* \geq n) \\ &\stackrel{*}{=} \sum_{z \in E} \sum_{n=1}^{\infty} \mathbb{P}(X_n = y \mid X_{n-1} = z) \mathbb{P}^x(X_{n-1} = z, T_x^* > n-1) \\ &= \sum_{z \in E} \mathbf{P}(z, y) \sum_{n=0}^{\infty} \mathbb{P}^x(X_n = z, T_x^* > n) \\ &= \sum_{z \in E} \mathbf{P}(z, y) \mu_x(z). \end{aligned}$$

Pour (*), $\{T_x^* \geq n\} = \{X_1 \neq x, \dots, X_{n-1} \neq x\} \in \mathcal{F}_{n-1}$, et d'autre part, si $z \neq x$, alors $\{X_{n-1} = z, X_{n-1} \neq x\} = \{X_{n-1} = z\}$, tandis que si $z = x$ alors $\mathbb{P}^x(X_{n-1} = z, T_x^* \geq n) = 0$, d'où l'usage de la propriété de Markov faible.

Pour (ii), par (i), $\mu_x(x) \mathbf{P}^{m_1}(x, y) \leq \mu_x(y)$ et $\mu_x(y) \mathbf{P}^{m_2}(y, x) \leq \mu_x(x)$. Il suffit de choisir m_1, m_2 t.q. $\mathbf{P}^{m_1}(x, y) > 0$ et $\mathbf{P}^{m_2}(y, x) > 0$ par irréductibilité de \mathbf{P} , et d'utiliser $\mu_x(x) = 1$. Enfin (iii) vient de $T_x^* = \sum_{y \in E} \sum_{n=0}^{T_x^* - 1} \mathbb{1}_{\{X_n=y\}}$. Nous avons utilisé plusieurs fois le théorème de Fubini–Tonelli dans $[0, \infty]$ pour commuter sommes ou espérances. ■

Théorème 5.7.3. Mesures invariantes des noyaux irréductibles.

Soit \mathbf{P} un noyau de transition sur E , irréductible. Alors :

- toute mesure invariante μ charge tous les états et vérifie $\mu(y) \geq \mu(x)\mu_x(y)$ pour tous x et y .

2. Dans le cas transitoire, il y a égalité pour $y \neq x$. Dans la même veine, si \mathbf{P} est transitoire, alors le potentiel vérifie $\mathbf{G}(x, y) = \sum_{n=0}^{\infty} \mathbf{P}^n(x, y) < \infty$ pour tous x, y , et l'identité $\mathbf{G}\mathbf{P} = \mathbf{I} + \mathbf{G}$ donne $(v_x \mathbf{P})(y) = \mathbb{1}_{x=y} + v_x(y)$ où $v_x(y) := \sum_{n=0}^{\infty} \mathbf{P}^n(x, y) = \mathbb{E}^x \left(\sum_{n=0}^{\infty} \mathbb{1}_{\{X_n=y\}} \right)$, et $v_x = \mu_x$ si $\mathbb{P}^x(T_x^* = \infty) = 1$.

— il y a deux cas distincts :

(i) **P** est transitoire. Alors :

- $\mu(E) = \infty$ pour toute mesure invariante μ .
- E est nécessairement infini et il n'y a pas de loi invariante.

(ii) **P** est récurrent. Alors :

- μ_x est invariante pour tout x .
- les mesures invariantes sont toutes proportionnelles.
- si μ est invariante alors $\mu(y) = \mu(x)\mu_x(y)$ pour tous x et y , en particulier, $\mu_x(y)\mu_y(x) = 1$.

En particulier, un noyau irréductible qui admet une loi invariante est forcément récurrent positif.

Démonstration. Soit $(X_n)_{n \in \mathbb{N}}$ une CM(E, \mathbf{P}). Soit μ une mesure invariante pour **P**. Il existe au moins un état z tel que $\mu(z) > 0$. Pour tout y , il existe un entier n tel que $\mathbf{P}^n(z, y) > 0$ car **P** est irréductible. Il en découle que

$$\mu(y) = \sum_x \mu(x)\mathbf{P}^n(x, y) \geq \mu(z)\mathbf{P}^n(z, y) > 0.$$

Ainsi, μ charge tous les états. Soit μ invariante avec $\mu(x) = 1$. La formule $\mu(y) = \mu(x)\mathbf{P}(x, y) + \sum_{z \neq x} \mu(z)\mathbf{P}(z, y)$ utilisée récursivement donne, si $y \neq x$, $\mu(y) \geq \mathbb{P}^x(X_1 = y, T_x^* > 1) + \dots + \mathbb{P}^x(X_n = y, T_x^* > n)$. D'où $\mu(y) \geq \mu_x(y)$ quand $n \rightarrow \infty$.

Cas où **P** est récurrent. Si μ est invariante et $\mu(x) = 1$, alors nous savons que $\mu \geq \mu_x$, donc $\mu - \mu_x$ est une mesure sur E . Comme μ_x est invariante (lemme 5.7.2) et $\mu(x) = \mu_x(x) = 1$, il vient que pour tout n et tout y ,

$$0 = (\mu - \mu_x)(x) = ((\mu - \mu_x)\mathbf{P}^n)(x) \geq (\mu - \mu_x)(y)\mathbf{P}^n(y, x),$$

d'où $(\mu - \mu_x)(y) = 0$ en choisissant, grâce à l'irréductibilité de **P**, un entier n tel que $\mathbf{P}^n(y, x) > 0$.

Cas où **P** est transitoire. Dans ce cas $\mathbf{G}(x, y) < \infty$ pour tous x et y , d'où $\lim_{n \rightarrow \infty} \mathbf{P}^n(x, y) = 0$. Si $\mu(E) < \infty$ alors par convergence dominée, $\mu(y) = \lim_n \sum_x \mu(x)\mathbf{P}^n(x, y) = 0$ pour tout y , ce qui est impossible, d'où $\mu(E) = \infty$. ■

Plus généralement, pour un noyau quelconque, l'argument utilisé dans la preuve permet d'établir que les loi invariantes ne chargent jamais les états transitoires, ce qui couvre les noyaux non-irréductibles possédant une classe transitoire non-close, comme par exemple une marche aléatoire sur \mathbb{Z}^3 augmentée d'un point absorbant.

Un état récurrent x est dit . . .

- récurrent positif si $\mathbb{E}^x(T_x^*) = \mu_x(E) < \infty$ (c'est toujours le cas lorsque x est absorbant).
- récurrent nul si $\mathbb{E}^x(T_x^*) = \mu_x(E) = \infty$ (dans ce cas T_x^* est à queue lourde).

Corollaire 5.7.4. Récurrence positive.

Pour tout noyau de transition **P** sur E , irréductible récurrent, l'un des deux cas suivants est réalisé :

(i) **P** est récurrent positif : tous les états sont récurrents positifs. Alors :

- $\mu(E) < \infty$ pour toute mesure invariante et il existe une unique loi invariante : $\mu(x) = \frac{1}{\mathbb{E}^x(T_x^*)}$, $x \in E$.
- toujours le cas si E est fini!

(ii) **P** est récurrent nul : tous les états sont récurrents nuls. Alors :

- $\mu(E) = \infty$ pour toute mesure invariante μ .
- E est nécessairement infini, et il n'existe pas de loi invariante.

Démonstration. Par le théorème 5.7.3, pour tous $x, y \in E$, les mesures invariantes μ_x et μ_y sont proportionnelles, donc $\mathbb{E}^x(T_x^*) = \mu_x(E)$ et $\mathbb{E}^y(T_y^*) = \mu_y(E)$ aussi. Donc x et y sont simultanément récurrents nuls ou récurrents positifs. Si μ est une loi invariante alors en sommant sur y dans $\mu(y)/\mu(x) = \mu_x(y)$ il vient $\mu(x) = 1/\mu_x(E) = 1/\mathbb{E}^x(T_x^*)$. ■

Ainsi, la récurrence positive est une propriété de classe.

$$\text{classe de récurrence} \begin{cases} (1) \text{ la classe est récurrente positive} \\ (2) \text{ la classe est récurrente nulle} \end{cases}$$

En particulier, les classes de récurrence positives sont les seules classes irréductibles closes qui peuvent être finies (elles ne le sont pas toujours). Ce sont les seules classes irréductibles closes qui portent une loi invariante.

Les mesures invariantes liées aux classes de récurrence s'expriment au moyen des quantités trajectorielles moyennes $\mu_x(y)$ et $1/\mathbb{E}^x(T_x^*)$. Réciproquement, la résolution du système linéaire $\mu\mathbf{P} = \mu$ fournit les quantités $\mu_x(y)$ et $1/\mathbb{E}^x(T_x^*)$. La section suivante renforce ce lien entre trajectoires et mesures invariantes. Pour certains noyaux, le calcul de $\mathbb{E}^x(T_x^*)$ est plus simple que la résolution directe de l'équation $\mu\mathbf{P} = \mu$.

Exemple 5.7.5. Mesures invariantes.

- Toute chaîne irréductible sur un espace d'état fini est récurrente positive : elle est récurrente car toute trajectoire passe une infinité de fois par au moins un état, et récurrente positive car toute mesure est finie. Ainsi les propriétés des mesures donnent des informations sur les temps de retour.
- La marche aléatoire simple sur \mathbb{Z} est irréductible, et la mesure de comptage est invariante ^a, comme cette mesure est de masse infinie, la chaîne est récurrente nulle ou bien transitoire. Nous savons par ailleurs qu'elle est récurrente car $\mathbf{G}(x, x) = \mathbf{G}(0, 0) = \infty$ pour tout x . Elle est donc récurrente nulle.
- Pour tout $0 < p < 1$, la marche aléatoire sur $E = \mathbb{N}$ réfléchiée en zéro de noyau donné par $\mathbf{P}(x, x+1) = p$, $\mathbf{P}(x+1, x) = 1 - p$, $x \in \mathbb{N}$, $\mathbf{P}(0, 0) = 1 - p$, et $\mathbf{P}(0, 1) = p$, est irréductible. La mesure $\mu(x) = \rho^x$, $x \in E$, où $\rho = p/(1 - p)$ est invariante. Elle est de masse finie ssi $p < 1 - p$ c'est-à-dire $p < 1/2$. Il en découle que si $p < 1/2$ alors la chaîne est récurrente positive de loi invariante $\text{Geo}_{\mathbb{N}}(p/(1 - p))$. Cela illustre également le critère de récurrence positive de Foster (théorème 5.8.2). De plus, il est possible d'établir que pour $p = 1/2$, la chaîne est récurrente nulle, tandis que pour $p > 1/2$, elle est transitoire. Plus généralement, des critères sont disponibles pour les processus de vie et de mort sur \mathbb{N} .

^a. Plus généralement, sur un groupe, si le noyau de transition est invariant par translation, alors la mesure de Haar est invariante.

Ceci n'est pas abordé lors du cours oral.

5.8 Critère de Foster et fonction de Liapounov

Lemme 5.8.1: Critère de récurrence positive

Si $(X_n)_{n \in \mathbb{N}}$ est une CM(E) irréductible, et s'il existe $F \subset E$ fini non-vidé tel que $\mathbb{E}^x(T_F^*) < \infty$ pour tout $x \in F$, où $T_F^* := \inf\{n \geq 1 : X_n \in F\}$, alors X est récurrente positive.

Démonstration. Notons $(T^n)_{n \in \mathbb{N}}$ les temps de retours successifs de X en F , avec $T^0 := 0$, et $T^1 = T_F^*$. Soit $x \in F$. Par la propriété de Markov forte, $(Y_n)_{n \in \mathbb{N}} := (X_{T^n})_{n \in \mathbb{N}}$ est une CM(F) irréductible car X est irréductible. De plus, comme F est fini, Y est récurrente positive, et en particulier $\mathbb{E}^x(S_x) < \infty$ où S_x est le temps de retour de Y en $x \in F$. Si T_x^* est le temps de retour en x de X , le but à présent est de contrôler $\mathbb{E}^x(T_x^*)$ avec $\mathbb{E}^x(S_x) < \infty$. Pour cela, on observe que $T_x^* = \sum_{k=0}^{\infty} U_k \mathbb{1}_{\{k < S_x\}}$ où $U_k := T^{k+1} - T^k$, d'où

$$\mathbb{E}^x(T_x^*) = \sum_{k=0}^{\infty} \mathbb{E}^x(U_k \mathbb{1}_{\{k < S_x\}}), \quad \text{et} \quad \mathbb{E}^x(U_k \mathbb{1}_{\{k < S_x\}}) = \sum_{y \in F} \mathbb{E}^x(U_k \mathbb{1}_{\{k < S_x\}} \mathbb{1}_{\{X_{T^k} = y\}}).$$

Par la propriété de Markov forte pour T^k , en observant que $\{k < S_x\}$ appartient au passé de X au temps T^k ,

$$\mathbb{E}^x(U_k \mathbb{1}_{\{k < S_x\}} \mathbb{1}_{\{X_{T^k} = y\}}) = \mathbb{E}^x(U_k \mid k < S_x, X_{T^k} = y) \mathbb{P}^x(k < S_x, X_{T^k} = y) = \mathbb{E}^x(U_k \mid X_{T^k} = y) \mathbb{P}^x(k < S_x, X_{T^k} = y).$$

Or $\mathbb{E}^x(U_k \mid X_{T^k} = y) = \mathbb{E}^y(T_F^*)$, d'où $\mathbb{E}^x(T_x^*) \leq \left(\max_{y \in F} \mathbb{E}^y(T_F^*) \right) \sum_{k=0}^{\infty} \mathbb{P}^x(S_x > k) = \left(\max_{y \in F} \mathbb{E}^y(T_F^*) \right) \mathbb{E}^x(S_x) < \infty$. ■

Théorème 5.8.2: Critère de récurrence positive de Foster et fonction de Liapounov

Soit \mathbf{P} un noyau de transition irréductible sur E . Alors \mathbf{P} est récurrent positif s'il admet une fonction de Liapounov, c'est-à-dire $V : E \rightarrow \mathbb{R}$, $F \subset E$ fini non-vidé, et $\varepsilon > 0$, tels que :

- (i) $\inf_E V > -\infty$.
- (ii) $\mathbf{P}V < \infty$ sur F .
- (iii) $\mathbf{P}V \leq V - \varepsilon$ sur F^c .

Ce théorème est dû à F. G. Foster vers 1953. Il est inspiré de la stabilité des EDO au sens de Liapounov. La propriété (iii) s'écrit également $\mathbf{L}V \leq -\varepsilon$ où $\mathbf{L} = \mathbf{P} - \mathbf{I}$ est le générateur. En quelque sorte, V est uniformément strictement surharmonique en dehors d'un ensemble fini. La preuve consiste à m.q. $\mathbb{E}^x(T_F^*) < \infty$ pour tout $x \in F$, où $T_F^* := \inf\{n \geq 1 : X_n \in F\}$ et à utiliser le lemme 5.8.1.

Exemple 5.8.3: Application du critère de Foster

Soit $(X_n)_{n \in \mathbb{N}}$ une CM(\mathbb{N}) irréductible telle que pour tout $x \in E$,

$$\mathbb{E}(X_{n+1} | X_n = x) < \infty \quad \text{et} \quad d := \overline{\lim}_{x \rightarrow \infty} \mathbb{E}(X_{n+1} - X_n | X_n = x) < 0,$$

comme pour la marche biaisée de l'exemple 5.7.5, qui est une file d'attente discrète à un serveur et un processus de vie et de mort. Alors X est récurrente positive. En effet, le théorème 5.8.2 s'applique avec $\varepsilon = -d$, $V(x) = x$, $F = \{x : x \leq x_0\}$, avec x_0 assez grand pour que $\mathbb{E}(X_{n+1} - X_n | X_n = x) \leq -\varepsilon$.

Démonstration. Grâce à (i) on peut supposer sans perte de généralité que $V \geq 0$, car (ii) et (iii) sont invariantes par ajout d'une constante à V . Considérons $T_F^* = \inf\{n \geq 1 : X_n \in F\}$. Comme F est fini non-vidé, d'après le lemme 5.8.1, il suffit d'établir que $\mathbb{E}^x(T_F^*) < \infty$ pour tout $x \in F$. Soit $(X_n)_{n \in \mathbb{N}}$ une CM(E, \mathbf{P}) et $Y_n := V(X_n) \mathbb{1}_{\{T_F^* \geq n\}}$. La propriété (iii) s'écrit $\mathbb{E}(V(X_{n+1}) | X_n = x) \leq V(x) - \varepsilon$ pour tout $x \notin F$. Si $x \notin F$, alors

$$\begin{aligned} \mathbb{E}^x(Y_{n+1} | X_0, \dots, X_n) &= \mathbb{E}^x(V_{n+1} \mathbb{1}_{\{T_F^* > n\}} | X_0, \dots, X_n) = \mathbb{1}_{\{T_F^* > n\}} \mathbb{E}^x(V(X_{n+1}) | X_n) \\ &\leq \mathbb{1}_{\{T_F^* > n\}} (V(X_n) - \varepsilon) \leq Y_n - \varepsilon \mathbb{1}_{\{T_F^* > n\}} \end{aligned}$$

où la première inégalité utilise le fait que $X_n \notin F$ sur $\{T_F^* > n\}$, et la seconde le fait que $V \geq 0$ sur F . En prenant l'espérance et en utilisant $V \geq 0$, cela donne

$$0 \leq \mathbb{E}^x(Y_{n+1}) \leq \mathbb{E}^x(Y_n) - \varepsilon \mathbb{P}^x(T_F^* > n),$$

d'où, en itérant,

$$0 \leq \mathbb{E}^x(Y_0) - \varepsilon \sum_{k=0}^n \mathbb{P}^x(T_F^* > k).$$

Or comme $Y_0 = V(X_0) \mathbb{1}_{\{T_F^* > 0\}} = V(X_0)$, et comme $\sum_{k=0}^{\infty} \mathbb{P}^x(T_F^* > k) = \mathbb{E}^x(T_F^*)$, il vient, pour tout $x \notin F$,

$$\mathbb{E}^x(T_F^*) \leq \varepsilon^{-1} V(x) < \infty.$$

D'autre part, pour tout $x \in F$, le découpage selon X_1 , la propriété de Markov faible, et (ii) donnent

$$\mathbb{E}^x(T_F^*) \leq 1 + \sum_{y \notin F} \mathbf{P}(x, y) \mathbb{E}^y(T_F^*) \leq 1 + \varepsilon^{-1} \sum_{y \notin F} \mathbf{P}(x, y) V(y) \leq 1 + \varepsilon^{-1} \sum_{y \in E} \mathbf{P}(x, y) V(y) < \infty.$$

■

5.9 Loi des grands nombres ou théorème ergodique

Chaque trajectoire partant d'un état récurrent se décompose en une suite infinie d'excursions enracinées en cet état. La propriété de Markov forte indique que ces excursions sont indépendantes et identiquement distribuées. Leur longueur, aléatoire, est intégrable ssi l'état est récurrent positif. Cette observation mène au théorème suivant.

Lemme 5.9.1. Nombre de passages et temps de retour.

Soit $(X_n)_{n \in \mathbb{N}}$ une CM(E) irréductible, et $x \in E$, alors la fréquence de passages en x avant le temps n vérifie

$$\frac{\sum_{k=0}^{n-1} \mathbb{1}_{\{X_k=x\}}}{n} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \frac{1}{\mathbb{E}^x(T_x^*)} \in [0, 1].$$

La preuve ci-dessous révèle que l'hypothèse d'irréductibilité est inutile si x est transitoire, tandis que si x est récurrent, elle peut être affaiblie en $\mathbb{P}(T_x^* < \infty) = 1$, ce qui est le cas si X_0 prend ses valeurs dans la classe de x .

Démonstration. Notons $N_x^n := \sum_{k=0}^{n-1} \mathbb{1}_{\{X_k=x\}}$ le nombre de passages en x avant le temps n . Si x est transitoire, alors p.s. le nombre de passages en x est fini, et donc $N_x^n/n \rightarrow 0$, tandis que $\mathbb{E}^x(T_x^*) = \infty$ car T_x^* n'est pas fini p.s.

Si x est récurrent, comme le temps d'atteinte de x est fini p.s., on peut donc supposer que la chaîne part de x . De plus, p.s., la suite strictement croissante $(T_x^n)_{n \geq 1}$ des temps de retour successifs en x est bien définie et tend vers

∞ . On convient que $T_x^0 = 0$ et que $T_x^1 = T_x^*$. Les définitions de N_x^n et T_x^n donnent $T_x^{N_x^n} \leq n < T_x^{N_x^n+1}$, d'où

$$\frac{T_x^{N_x^n}}{N_x^n} \leq \frac{n}{N_x^n} \leq \frac{T_x^{N_x^n+1}}{N_x^n+1} \frac{N_x^n+1}{N_x^n}.$$

Par la propriété de Markov forte, $(T_x^{n+1} - T_x^n)_{n \geq 1}$ est formée de v.a. i.i.d. de loi $\text{Loi}(T_x | X_0 = x) = \text{Loi}(T_x^2 - T_x^1)$, d'où

$$\frac{T_x^n}{n} = \frac{T_x^1 + T_x^2 - T_x^1 + \dots + T_x^n - T_x^{n-1}}{n} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \mathbb{E}^x(T_x^1)$$

par la loi forte des grands nombres (somme télescopique). D'autre part, la récurrence et l'irréductibilité entraînent que $N_x^n \rightarrow \infty$ p.s. Ces deux convergences et l'encadrement précédent entraînent que $n/N_x^n \rightarrow \mathbb{E}^x(T_x^1)$ p.s. ■

En particulier, le lemme 5.9.1 indique que si x et y sont récurrents positifs, alors

$$\frac{\sum_{k=0}^{n-1} \mathbb{1}_{\{X_k=x\}}}{\sum_{k=0}^{n-1} \mathbb{1}_{\{X_k=y\}}} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \frac{\mathbb{E}^y(T_y^*)}{\mathbb{E}^x(T_x^*)} \in \mathbb{R}.$$

De plus, s'ils sont dans la même classe de récurrence, alors, en notant μ la loi invariante portée par cette classe,

$$\frac{\sum_{k=0}^{n-1} \mathbb{1}_{\{X_k=x\}}}{\sum_{k=0}^{n-1} \mathbb{1}_{\{X_k=y\}}} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \frac{\mathbb{E}^y(T_y^*)}{\mathbb{E}^x(T_x^*)} = \frac{\mu(x)}{\mu(y)} \in \mathbb{R}.$$

Dans cet esprit, on dispose plus généralement du théorème suivant.

Théorème 5.9.2. Loi des grands nombres ou théorème ergodique de Chacon – Ornstein – Hopf.

Soit $(X_n)_{n \in \mathbb{N}}$ une chaîne de Markov sur E récurrente irréductible, de loi initiale quelconque. Alors :

(i) pour toute mesure invariante μ et pour toutes fonctions $f, g : E \rightarrow \mathbb{R}$ μ -intégrables avec $\mu g \neq 0$, on a

$$\frac{f(X_0) + \dots + f(X_{n-1})}{g(X_0) + \dots + g(X_{n-1})} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \frac{\sum_{z \in E} f(z) \mu(z)}{\sum_{z \in E} g(z) \mu(z)} = \frac{\mu f}{\mu g} \in \mathbb{R}.$$

(ii) si la chaîne est récurrente nulle, alors pour toute fonction $f : E \rightarrow \mathbb{R}$ μ -intégrable,

$$\frac{f(X_0) + \dots + f(X_{n-1})}{n} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} 0.$$

(iii) si la chaîne est récurrente positive, de loi invariante μ , alors pour toute $f : E \rightarrow \mathbb{R}$ μ -intégrable,

$$\frac{f(X_0) + \dots + f(X_{n-1})}{n} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \sum_{z \in E} f(z) \mu(z) \in \mathbb{R}.$$

Toute suite $(X_n)_{n \in \mathbb{N}}$ de v.a. i.i.d. de loi μ sur E chargeant tous les états est une chaîne de Markov irréductible récurrente positive de loi invariante μ et le théorème 5.9.2 coïncide avec la loi des grands nombres pour la suite de v.a. i.i.d. $(f(X_n))_{n \in \mathbb{N}}$. Il s'agit donc bien d'une généralisation de la loi des grands nombres classique.

Démonstration. Raisonnons conditionnellement à $\{X_0 = x\}$, où $x \in E$ est arbitraire. Les ensembles p.s. construits dépendront de x , mais leur intersection sur $x \in E$ sera également p.s. car E est au plus dénombrable.

Pour établir (i), considérons tout d'abord le cas $g = \mathbb{1}_{\{x\}}$ et $f \geq 0$. Comme le noyau est irréductible récurrent, le théorème 5.7.3 indique que toutes les mesures invariantes sont proportionnelles, il suffit de n'en considérer qu'une, par exemple $\mu = \mu_x$, la mesure invariante valant 1 en x étudiée dans le lemme 5.7.2. La méthode du découpage en excursions i.i.d. de la preuve du lemme 5.9.1 donne alors, avec les mêmes notations, avec $T := T_x^*$ et $N^n := N_x^n$,

$$\frac{f(X_0) + \dots + f(X_{T^{N^n}-1})}{N^n} \leq \frac{f(X_0) + \dots + f(X_n)}{N^n} \leq \frac{f(X_0) + \dots + f(X_{T^{N^n+1}-1})}{N^n+1} \frac{N^n+1}{N^n}.$$

La propriété de Markov forte donne $f(X_0) + \dots + f(X_{T^m-1}) = Z_1 + \dots + Z_m$ où $(Z_k)_{k \geq 1}$ est une suite de v.a. i.i.d. De plus $\mathbb{E}(Z_1) = \mathbb{E}(f(X_0) + \dots + f(X_{T-1})) = \mu_x f$ par définition de μ_x et le théorème de Fubini – Tonelli. Comme par ailleurs $N^n \rightarrow \infty$ p.s., la loi forte des grands nombres pour $(Z_k)_{k \geq 1}$ et l'encadrement précédent donnent

$$R_n(f) := \frac{f(X_0) + \dots + f(X_{T^{N^n}-1})}{N^n} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \mu_x f.$$

À présent, le résultat pour tout $f \in L^1(\mu_x)$ s'en déduit en posant $f = f^+ - f^-$ avec $f^\pm := \max(0, \pm f)$. Enfin, pour tout $g \in L^1(\mu_x)$ telle que $\mu_x g \neq 0$, il vient que $R_n(f)/R_n(g) \rightarrow \mu_x f / \mu_x g$ p.s. ce qui est le résultat voulu.

La propriété (iii) s'obtient à partir de (i) avec $g \equiv 1$ et μ loi invariante.

La propriété (ii) s'obtient à partir de (i) en se ramenant à $f \geq 0$, puis en observant que si $g_k = \mathbb{1}_{E_k}$ avec E_k fini, (E_k) croissante avec $E = \cup_k E_k$, alors

$$0 \leq \frac{f(X_0) + \dots + f(X_n)}{n+1} \leq \frac{f(X_0) + \dots + f(X_n)}{g_k(X_0) + \dots + g_k(X_n)} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \frac{\mu f}{\mu g_k} \xrightarrow[k \rightarrow \infty]{} \frac{\mu f}{\mu 1} = 0$$

où la dernière limite s'obtient par convergence monotone et $\mu 1 = \mu(E) = \infty$ (corollaire 5.7.4). Alternativement, nous pouvons utiliser $g = \mathbb{1}_{\{x\}}$, multiplier par N^n/n , et utiliser le lemme 5.9.1 qui donne $N^n/n \rightarrow 0$ p.s. ■

Le théorème 5.9.2 affirme qu'une moyenne en temps $\frac{1}{n}(f(X_0) + \dots + f(X_{n-1})) = \frac{1}{n} \int_0^n f(X_k) dk$ converge vers une moyenne en espace $\sum_{z \in E} f(z) \mu(z) = \int f d\mu$. Le long des trajectoires, la fraction de temps passée sur un état (fréquence de passage), converge vers la masse affecté par la loi invariante à cet état, égale à l'inverse du temps de retour moyen à cet état. Le lien temps-espace est déjà visible dans la formule de la loi invariante : $\mu(x) \mathbb{E}^x(T_x^*) = 1$.

La variable aléatoire $\frac{1}{n}(f(X_0) + \dots + f(X_{n-1}))$ est la moyenne de f pour la mesure empirique \mathbb{P}_n , définie par $\mathbb{P}_n = \frac{1}{n}(\delta_{X_0} + \dots + \delta_{X_{n-1}})$. Une autre façon d'exprimer le résultat de convergence consiste à dire que $\mathbb{P}_n \rightarrow \mu$ p.s.

La méthode des excursions permet également d'établir un théorème limite central.

5.10 Convergence en loi, période, couplage

Soit $(X_n)_{n \in \mathbb{N}}$ une $CM(E, \nu, \mathbf{P})$ récurrente irréductible positive et μ sa loi invariante. Le théorème 5.9.2 donne, par convergence dominée, pour toute $f : E \rightarrow \mathbb{R}$ (continue) bornée,

$$\frac{\nu \mathbf{P}^0 + \dots + \nu \mathbf{P}^{n-1}}{n} f = \frac{\mathbb{E}(f(X_0)) + \dots + \mathbb{E}(f(X_{n-1}))}{n} \xrightarrow[n \rightarrow \infty]{} \mu f.$$

Ainsi $\nu \mathbf{P}^n \rightarrow \mu$ étroitement au sens de Cesàro. Peut-elle converger au sens usuel? Considérons le cas du processus de Bernoulli modulo d avec $p = 1$, c'est-à-dire la chaîne sur $\mathbb{Z}/d\mathbb{Z}$ de noyau de transition \mathbf{P} défini par $\mathbf{P}(x, y) = 1$ si $y = x + 1$ dans $\mathbb{Z}/d\mathbb{Z}$ et $\mathbf{P}(x, y) = 0$ sinon. Ce noyau est récurrent irréductible positif, et la loi uniforme sur $\mathbb{Z}/d\mathbb{Z}$ est invariante. Cependant, $\text{Loi}(X_{dn+i} | X_0) = \delta_{X_0+i}$ pour tout $n \in \mathbb{N}$ et tout $0 \leq i < d$. Cela s'écrit également $\mathbf{P}^{dn+i} = \mathbf{P}^i$. Ainsi, bien que $\nu \mathbf{P}^n \rightarrow \mu$ au sens de Cesàro, des phénomènes périodiques empêchent sa convergence au sens usuel.

Théodéf 5.10.1. Période.

Si \mathbf{P} est un noyau irréductible sur E au plus dénombrable, alors il existe un entier $d > 0$ appelé période et une partition $E = C_0 \cup \dots \cup C_{d-1}$ tels que, avec la convention $C_{nd+i} = C_i$, les propriétés suivantes ont lieu :

- (i) Si $\mathbf{P}^n(x, y) > 0$, alors $(x, y) \in C_i \times C_{i+n}$ pour un certain i .
- (ii) Pour tous $x, y \in C_i$, et tout i , il existe r tel que $\mathbf{P}^{nd}(x, y) > 0$ pour tout $n \geq r$.
- (iii) Pour tout $z \in E$, $d = \text{PGCD}\{n \geq 1 : \mathbf{P}^n(z, z) > 0\}$.

En particulier, le noyau \mathbf{P}^d possède d classes irréductibles, qui sont C_0, \dots, C_{d-1} . Enfin on dit que \mathbf{P} est apériodique lorsque $d = 1$.

Il en découle les faits utiles suivants :

- Une chaîne irréductible traverse les ensembles C_0, \dots, C_{d-1} successivement et de façon cyclique. Si le support de $\text{Loi}(X_0)$ est inclus dans C_i , alors le support de $\text{Loi}(X_1)$ est inclus dans C_{i+1} , etc., et ce phénomène rend la convergence en loi de la chaîne impossible lorsque $d > 1$. Notons qu'on a toujours $1 \leq d \leq |E|$.
- S'il existe $x \in E$ tel que $\mathbf{P}(x, x) > 0$, alors \mathbf{P} est apériodique.
- Si \mathbf{P} est apériodique, alors pour tous $x, y \in E$, il existe r tel que $\mathbf{P}^n(x, y) > 0$ pour tous $n \geq r$.
- La période de \mathbf{P} ne dépend que du squelette de son graphe des transitions : $\{(x, y) \in E \times E : \mathbf{P}(x, y) > 0\}$.

Démonstration. Démontrons tout d'abord que $d_z := \text{PGCD}\{n \geq 1 : \mathbf{P}^n(z, z) > 0\}$ ne dépend pas de z . Soient $z, z' \in E$, et n, n' tels que $\mathbf{P}^n(z, z') > 0$ et $\mathbf{P}^{n'}(z', z) > 0$. Par définition, d_z divise $n + n'$. D'autre part, pour tout m tel que $\mathbf{P}^m(z', z') > 0$, d_z divise $n + m + n'$ car $\mathbf{P}^{n+m+n'}(z, z) > 0$. Or comme d_z divise $n + n'$, il divise m , donc il divise $d_{z'}$.

Fixons z , et considérons $n_2 > n_1$ dans $S_z := \{n \geq 1 : \mathbf{P}^n(z, z) > 0\}$ tels que la différence $d = n_2 - n_1$ soit la plus petite possible. On définit alors l'ensemble C_i pour $0 \leq i \leq d-1$ par

$$C_i := \{x \in E : \mathbf{P}^{nd+i}(z, x) > 0 \text{ pour un entier } n \geq 0\}.$$

L'irréductibilité de \mathbf{P} donne $C_0 \cup \dots \cup C_{d-1} = E$. Remettons à plus tard la preuve du fait qu'ils sont disjoints.

Démontrons (i). Soit $x \in C_i$ et supposons que $\mathbf{P}^n(x, y) > 0$. Alors $\mathbf{P}^{md+i}(z, x) > 0$ pour un entier m , et donc $\mathbf{P}^{md+n+i}(z, y) > 0$, d'où $y \in C_{n+i}$. En prenant $x = y = z$, on obtient que $md + n + i$ et i sont égaux modulo d c'est-à-dire que n est divisible par d , et donc d divise tout élément de S_z , et donc divise n_1 .

Si $nd \geq n_1^2$, alors $nd = qn_1 + r$ avec $0 \leq r < n_1 \leq q$. Comme d divise n_1 , il divise aussi r . En écrivant $r = dm = (n_2 - n_1)m$, on obtient $nd = (q - m)n_1 + mn_2$. Ainsi $\mathbf{P}^{nd}(z, z) \geq (\mathbf{P}^{n_1}(z, z))^{q-m} (\mathbf{P}^{n_2}(z, z))^m > 0$, et donc $nd \in S_z$.

Démontrons (ii). Soient $x, y \in C_i$. L'irréductibilité de \mathbf{P} entraîne que $\mathbf{P}^{m_1}(x, z) > 0$ et $\mathbf{P}^{m_2}(z, y) > 0$ pour des entiers m_1 et m_2 . Par ailleurs si $nd \geq n_1^2$ alors nous avons déjà démontré que $nd \in S_z$, c'est-à-dire que $\mathbf{P}^{nd}(z, z) > 0$. Donc $\mathbf{P}^{m_1+nd+m_2}(x, y) \geq \mathbf{P}^{m_1}(x, z) \mathbf{P}^{nd}(z, z) \mathbf{P}^{m_2}(z, y) > 0$. Or $m_1 + m_2$ est alors un multiple de d par (i), d'où (ii).

Supposons que $x \in C_i \cap C_j$. Alors $\mathbf{P}^{nd+i}(z, x) > 0$ et $\mathbf{P}^{n'+j}(z, x) > 0$ pour des entiers n et n' , et (ii) indique qu'on peut prendre $n = n'$ (assez grands). Par irréductibilité de \mathbf{P} , $\mathbf{P}^m(x, z) > 0$ pour un entier m . D'où $\mathbf{P}^{nd+i+m}(z, z) > 0$ et $\mathbf{P}^{nd+j+m}(z, z) > 0$, ce qui n'est possible que si $i = j$, par minimalité de d . Donc C_0, \dots, C_{d-1} partitionnent E .

Démontrons (iii). Nous avons déjà démontré que d divise tous les éléments de S_z . À présent, si k divise tous les éléments de S_z , alors il divise n_1 et n_2 et donc divise $n_2 - n_1 = d$. Par conséquent $d = \text{PGCD}(S_z)$. ■

Remarque 5.10.2. Apériodicité par perturbation et chaînes paresseuses.

Si \mathbf{P} est un noyau de transition, alors pour tout $0 \leq \varepsilon < 1$, la combinaison convexe $\mathbf{P}_\varepsilon = (1 - \varepsilon)\mathbf{P} + \varepsilon\mathbf{I}$ définit un noyau de transition, irréductible si \mathbf{P} est irréductible, qui possède les mêmes mesures invariantes que \mathbf{P} . De plus, si $0 < \varepsilon < 1$ alors ce noyau est apériodique car $\mathbf{P}_\varepsilon(x, x) \geq \varepsilon > 0$ pour tout x (un seul suffit). La chaîne de noyau \mathbf{P}_ε est paresseuse : à chaque pas de temps, elle fait du surplace avec probabilité ε . Enfin $\mathbf{P}_\varepsilon \rightarrow \mathbf{P}$ quand $\varepsilon \rightarrow 0$, par exemple pour la topologie de la convergence simple, d'où la densité des noyaux apériodiques dans l'ensemble des noyaux pour cette topologie. Cette genericité topologique est utile pour évaluer la robustesse des modèles markoviens et la stabilité de leurs propriétés par perturbation. Cependant, en pratique, certains modèles sont périodiques pour de bonnes raisons liées à la physique modélisée. Enfin, pour rendre apériodique un noyau irréductible, il suffit d'effectuer la perturbation précédente sur au moins une ligne, c'est-à-dire de rendre la chaîne paresseuse sur un seul site.

Notre but à présent est d'établir que les chaînes de Markov irréductibles récurrentes positives apériodiques, sur un espace d'état au plus dénombrable E , convergent en loi vers leur unique loi invariante, quelle que soit la loi initiale. Il y a donc oubli de la loi initiale au profit de la loi invariante, qui représente un équilibre, et qui ne dépend que du noyau de transition. Pour cela, il est commode d'utiliser la distance en variation totale :

$$d_{\text{VT}}(\mu, \nu) = \frac{1}{2} \sup_{\|f\|_\infty \leq 1} (\mu - \nu)f.$$

Le préfacteur $1/2$ est peu important, il assure que les valeurs prises sont dans $[0, 1]$, la valeur 1 étant atteinte quand μ et ν sont à supports disjoints. Il est clair que dans l'ensemble des mesures de probabilités sur E , la convergence pour cette distance entraîne la convergence étroite, c'est-à-dire la convergence en loi³.

Une chaîne de Markov ne s'éloigne jamais d'une loi invariante en variation totale :

Lemme 5.10.3. Contraction markovienne.

Soit \mathbf{P} un noyau de transition sur E . Alors pour toutes mesures de probabilité ν_1 et ν_2 sur E ,

$$d_{\text{VT}}(\nu_1 \mathbf{P}, \nu_2 \mathbf{P}) \leq d_{\text{VT}}(\nu_1, \nu_2).$$

En particulier, si μ est une loi invariante pour \mathbf{P} alors pour toute mesure de probabilité ν sur E et tout $n \in \mathbb{N}$,

$$d_{\text{VT}}(\nu \mathbf{P}^{n+1}, \mu) \leq d_{\text{VT}}(\nu \mathbf{P}^n, \mu).$$

Démonstration. Si $f : E \rightarrow [-1, +1]$, alors $\mathbf{P}f : E \rightarrow [-1, +1]$, d'où $|(\nu_1 - \nu_2)\mathbf{P}f| \leq \sup_{\|f\|_\infty \leq 1} |(\nu_1 - \nu_2)f| = 2d_{\text{VT}}(\nu_1, \nu_2)$. ■

Définition 5.10.4. Couplage de chaîne de Markov.

Soit \mathbf{P} un noyau de transition sur E . Un couplage de \mathbf{P} est une suite $(Z_n)_{n \in \mathbb{N}} = ((X_n, Y_n))_{n \in \mathbb{N}}$ de v.a. sur $E \times E$ telle que ses deux composantes $(X_n)_{n \in \mathbb{N}}$ et $(Y_n)_{n \in \mathbb{N}}$ sont toutes deux des chaînes de Markov de noyau \mathbf{P} . Le temps de couplage est la v.a. à valeurs dans $\mathbb{N} \cup \{\infty\}$ définie par

$$T_C = \inf\{n \geq 0 : X_n = Y_n\}.$$

3. Il se trouve qu'il y a équivalence (lemme A.2.1). Le chapitre A en appendice est consacré aux propriétés de cette distance dans ce cadre.

Le couplage est coalescent lorsque $X_n = Y_n$ pour tout $n \geq T_C$.

Les suites marginales X et Y ne sont pas forcément indépendantes, n'ont pas forcément la même loi initiale, et la suite couplée Z n'est pas forcément une chaîne de Markov sur $E \times E$. Notons que $\{T_C > n\} \subset \{X_n \neq Y_n\}$ pour tout n , avec égalité lorsque le couplage est coalescent. Le temps de couplage T_C s'interprète également comme le temps d'atteinte de la diagonale $T_C = \inf\{n \geq 0 : Z_n \in D\}$ où $D = \{(x, x) : x \in E\} \subset E \times E$.

Lemme 5.10.5. Distance en temps long et couplage.

Soit \mathbf{P} un noyau de transition sur E et $Z = (X, Y)$ un couplage de \mathbf{P} de temps de couplage T_C . Supposons que pour tout n , tout $0 \leq k \leq n$, et toute $f : E \rightarrow \mathbb{R}$ bornée, sur $\{T_C = k\}$:

$$\mathbb{E}(f(X_n) \mid Z_0, \dots, Z_k) = \mathbb{E}(f(X_n) \mid X_k) \quad \text{et} \quad \mathbb{E}(f(Y_n) \mid Z_0, \dots, Z_k) = \mathbb{E}(f(Y_n) \mid Y_k).$$

Alors pour tout n ,

$$d_{\text{VT}}(\text{Loi}(X_n), \text{Loi}(Y_n)) = d_{\text{VT}}(\text{Loi}(X_0)\mathbf{P}^n, \text{Loi}(Y_0)\mathbf{P}^n) \leq \mathbb{P}(T_C > n).$$

Le couplage des chaînes de Markov a été introduit par Wolfgang Döblin (Vincent Doblin) vers 1938.

Démonstration. Les suites X et Y sont des chaînes de Markov sur E de noyau \mathbf{P} . Soit $\nu = \text{Loi}(X_0)$ et $\mu = \text{Loi}(Y_0)$. Pour tout $f : E \rightarrow [-1, +1]$,

$$\nu\mathbf{P}^n f = \mathbb{E}(f(X_n)) = \mathbb{E}(f(X_n)\mathbb{1}_{\{T_C > n\}}) + \sum_{k=0}^n \mathbb{E}(f(X_n)\mathbb{1}_{\{T_C = k\}}).$$

Or $\mathbb{E}(f(X_n)\mathbb{1}_{\{T_C = k\}}) = \mathbb{E}(\mathbb{1}_{\{T_C = k\}}\mathbb{E}(f(X_n) \mid Z_0, \dots, Z_k))$, et par hypothèse sur X et Y , on a, sur $\{T_C = k\}$,

$$\mathbb{E}(f(X_n) \mid Z_0, \dots, Z_k) = (\mathbf{P}^{n-k} f)(X_k).$$

Ainsi, $\mathbb{E}(f(X_n)\mathbb{1}_{\{T_C = k\}}) = \mathbb{E}(f(Y_n)\mathbb{1}_{\{T_C = k\}})$ car $X_k = Y_k$ sur $\{T_C = k\}$. Par conséquent, on a

$$|\nu\mathbf{P}^n f - \mu\mathbf{P}^n f| \leq \mathbb{E}(|f(X_n) - f(Y_n)|\mathbb{1}_{\{T_C > n\}}) \leq 2 \left(\sup_{x \in E} |f(x)| \right) \mathbb{P}(T_C > n).$$

Notons que $\mathbb{P}(T_C > n)$ dépend de $\text{Loi}(X_0)$ et $\text{Loi}(Y_0)$ via la loi de Z . ■

Théorème 5.10.6. Convergence en loi vers l'équilibre ou théorème de Döblin.

Soit \mathbf{P} un noyau de transition sur E , irréductible, récurrent positif, apériodique, de loi invariante μ .

Alors $\lim_{n \rightarrow \infty} d_{\text{VT}}(\nu\mathbf{P}^n, \mu) = 0$ pour toute mesure de probabilité ν sur E .

Ainsi, pour toute chaîne de Markov $(X_n)_{n \in \mathbb{N}}$ de noyau \mathbf{P} , et quelle que soit sa loi initiale $\text{Loi}(X_0)$,

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

Autrement dit, toutes les lignes de \mathbf{P} convergent vers μ : $\lim_{n \rightarrow \infty} \mathbf{P}^n(x, y) = \mu(y)$ pour tous x et y .

Nous avons vu que la loi des grands nombres ou théorème ergodique (théorème 5.9.2) entraîne, par convergence dominée, que $\nu\mathbf{P}^n \rightarrow \mu$ au sens de Cesàro, dès que \mathbf{P} est récurrent irréductible positif. Le théorème 5.10.6 précise que $\nu\mathbf{P}^n \rightarrow \mu$ au sens usuel lorsque \mathbf{P} est de plus apériodique. Ainsi, bien que les théorèmes 5.9.2 et 5.10.6 soient de natures différentes, une conséquence du premier est renforcée par le second. Le théorème 5.10.6 énonce une convergence en loi, et cela ne dit rien pour une trajectoire donnée, contrairement au théorème 5.9.2.

Démonstration. Soient X et Y deux chaînes de Markov indépendantes de même noyau \mathbf{P} , et de lois initiales respectives ν_1 et ν_2 . Alors $Z = (X, Y)$ est une CM($E \times E, \nu_1 \otimes \nu_2, \mathbf{P} \otimes \mathbf{P}$) avec $(\nu_1 \otimes \nu_2)(x, y) = \nu_1(x)\nu_2(y)$ et

$$(\mathbf{P} \otimes \mathbf{P})((x, y), (x', y')) = \mathbf{P}(x, x')\mathbf{P}(y, y').$$

De plus, les hypothèses du lemme 5.10.5 sont satisfaites. Soit $T_C = \inf\{n \geq 0 : X_n = Y_n\} = \inf\{n \geq 0 : Z_n \in D\}$ le temps de couplage de Z , qui est le temps d'atteinte de la diagonale de $E \times E$. Comme \mathbf{P} est irréductible apériodique, par le théorème 5.10.1, pour tous (x, y) et (x', y') dans $E \times E$, il existe des entiers r_1 et r_2 tels que $\mathbf{P}^n(x, x') > 0$ pour tout $n \geq r_1$ et $\mathbf{P}^n(y, y') > 0$ pour tout $n \geq r_2$. Ainsi, $(\mathbf{P} \otimes \mathbf{P})^n((x, y), (x', y')) > 0$ pour tout $n \geq \max(r_1, r_2)$, donc $\mathbf{P} \otimes \mathbf{P}$ est irréductible. Or $\mu \otimes \mu$ est une loi invariante pour $\mathbf{P} \otimes \mathbf{P}$, donc $\mathbf{P} \otimes \mathbf{P}$ est récurrent positif (théorème 5.7.3). En particulier $\mathbb{P}(T_C < \infty) = 1$ (atteinte de la diagonale), d'où $\lim_{n \rightarrow \infty} \mathbb{P}(T_C > n) = \mathbb{P}(T_C = \infty) = 0$. Le lemme 5.10.5 entraîne alors que $\lim_{n \rightarrow \infty} d_{\text{VT}}(\nu_1\mathbf{P}^n, \nu_2\mathbf{P}^n) = 0$ pour toutes lois ν_1 et ν_2 sur E . Il ne reste plus qu'à prendre $\nu_1 = \nu$ et $\nu_2 = \mu$. ■

Corollaire 5.10.7. Convergence en loi vers l'équilibre dans le cas périodique.

Soit \mathbf{P} un noyau de transition sur E , irréductible, récurrent positif, de période d , et de loi invariante μ . Soit $E = C_0 \cup \dots \cup C_{d-1}$ la partition de E en classe de récurrence de \mathbf{P}^d , et $\mu_i = \mu(C_i)^{-1} \mu(\cdot \cap C_i)$. Alors $\lim_{n \rightarrow \infty} d_{VT}(\nu \mathbf{P}^{nd}, \mu_i) = 0$ pour tout i et toute loi de probabilité ν portée par C_i . Ainsi, pour toute chaîne de Markov $(X_n)_{n \in \mathbb{N}}$ de noyau \mathbf{P} , si sa loi initiale $\text{Loi}(X_0)$ est portée par C_i , alors

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

5.11 Estimation du noyau et de la mesure invariante

Théorème 5.11.1. Estimation du noyau de transition.

Soit $(X_n)_{n \in \mathbb{N}}$ une CM(E, ν, \mathbf{P}) irréductible, récurrente positive, de loi invariante μ . Alors pour tous x, y ,

$$\hat{\mathbf{P}}_n(x, y) := \frac{\sum_{k=0}^{n-1} \mathbb{1}_{\{(X_k, X_{k+1})=(x, y)\}}}{\sum_{k=0}^{n-1} \mathbb{1}_{\{X_k=x\}}} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \mathbf{P}(x, y).$$

Démonstration. La suite $X' = ((X_n, X_{n+1}))_{n \in \mathbb{N}}$ est une CM sur $F = \{(x, x') \in E \times E : \mathbf{P}(x, x') > 0\}$, de noyau de transition $\mathbf{Q}((x, x'), (y, y')) = \mathbf{P}(y, y') \mathbb{1}_{x'=y}$. Ce noyau est irréductible récurrent, la mesure $(x, x') \mapsto \mu(x) \mathbf{P}(x, x')$ est invariante, et c'est une loi, donc \mathbf{Q} est positif. Le théorème 5.9.2 utilisé pour X et X' donne $\hat{\mathbf{P}}_n(x, y) \rightarrow \mu(x) \mathbf{P}(x, y) / \mu(x) = \mathbf{P}(x, y)$ p.s. quand $n \rightarrow \infty$, pour tous $(x, y) \in F$, tandis que $\hat{\mathbf{P}}_n(x, y) = 0 = \mathbf{P}(x, y)$ si $(x, y) \notin F$. ■

Il se trouve qu'il s'agit de l'estimateur du maximum de vraisemblance qui maximise la log-vraisemblance

$$\mathbf{P} \mapsto \log\left(\nu(X_0) \prod_{k=1}^n \mathbf{P}(X_{k-1}, X_k)\right) = \log(\nu(X_0)) + \sum_{k=1}^n \log(\mathbf{P}(X_{k-1}, X_k))$$

sur l'ensemble convexe des noyaux de transition sur E (compact si E est fini). Lorsque le noyau à estimer est défini de manière paramétrique, le maximum de vraisemblance reste un estimateur tout à fait intéressant.

Estimation de la mesure invariante

Pour estimer μ , il est possible d'exploiter la formule $\mu(x) = 1/\mathbb{E}^x(T_x^*)$, et d'utiliser les excursions le long d'une ou plusieurs trajectoires de la chaîne. Alternativement, le théorème 5.9.2 fournit un estimateur direct $\frac{1}{n} \sum_{k=0}^{n-1} \mathbb{1}_{\{X_k=x\}}$ de $\mu(x)$ à partir d'une trajectoire de la chaîne. Plus généralement, ce théorème montre que l'histogramme des valeurs prises par la chaîne le long d'une trajectoire fournit un estimateur de la mesure μ .

Lorsque \mathbf{P} est irréductible récurrent nul, la mesure invariante μ n'est définie qu'à une constante multiplicative près, mais un estimateur de $\mu(y)/\mu(x) = \mu_x(y)$ s'obtient de la même manière grâce au théorème 5.9.2.

Lorsque \mathbf{P} est apériodique, toutes les lignes de \mathbf{P}^n convergent vers μ d'après le théorème 5.10.6. La suite récurrente définie par $\nu_0 = \nu$ et $\nu_{n+1} = \nu_n \mathbf{P}$ converge vers μ . Il est également possible d'estimer μ en utilisant un échantillon de la loi $\text{Loi}(X_n)$ lorsque cela est plus commode, et c'est précisément ce qui fonde les méthodes MCMC.

5.12 Réversibilité et algorithme de Metropolis–Hastings

Mesures réversibles

Une mesure μ sur E est réversible⁴ pour le noyau de transition \mathbf{P} lorsqu'elle vérifie (équation de bilan détaillé)

$$\mu(x) \mathbf{P}(x, y) = \mu(y) \mathbf{P}(y, x) \quad \text{pour tous } x \text{ et } y \text{ dans } E.$$

De manière équivalente, cela revient à dire que pour tout n et toute suite x_0, \dots, x_n dans E :

$$\mu(x_0) \mathbf{P}(x_0, x_1) \cdots \mathbf{P}(x_{n-1}, x_n) = \mu(x_n) \mathbf{P}(x_n, x_{n-1}) \cdots \mathbf{P}(x_1, x_0).$$

Il en découle qu'une mesure de probabilité μ est réversible ssi pour toute CM $(X_n)_{n \in \mathbb{N}}$ de noyau \mathbf{P} et de loi initiale μ , la loi des trajectoires de X à horizon fini ne change pas par renversement du temps :

$$\text{Loi}((X_0, X_1, \dots, X_n)) = \text{Loi}((X_n, X_{n-1}, \dots, X_0)) \quad \text{pour tout } n \in \mathbb{N}.$$

4. Le terme symétrique est également utilisé.

Si une mesure μ est réversible et vérifie $\mu(E) < \infty$, alors la mesure de probabilité $\mu(E)^{-1}\mu$ est réversible.

Si une mesure μ est réversible, alors elle est invariante car

$$\sum_{x \in E} \mu(x) \mathbf{P}(x, y) = \mu(y) \sum_{x \in E} \mathbf{P}(y, x) = \mu(y) \quad \text{pour tout } y \in E.$$

La réciproque est fautive. Il est plus facile de rechercher une mesure réversible qu'une mesure invariante.

Exemple 5.12.1. Mesure uniforme, invariance, et réversibilité.

La loi uniforme sur $\mathbb{Z}/d\mathbb{Z}$ est réversible pour la marche aléatoire simple sur $\mathbb{Z}/d\mathbb{Z}$. La mesure de comptage sur \mathbb{Z} est réversible pour la marche aléatoire simple sur \mathbb{Z} . Plus généralement $\mathbf{P}^\top : E \times E \rightarrow [0, 1]$ définie par $\mathbf{P}^\top(x, y) := \mathbf{P}(y, x)$ n'est pas forcément un noyau de transition sur E . Cependant, la mesure de comptage est invariante pour \mathbf{P} ssi \mathbf{P}^\top est un noyau de transition, et elle est réversible pour \mathbf{P} ssi $\mathbf{P}^\top = \mathbf{P}$. Cela permet de construire des noyaux dont les mesures invariantes ne sont pas réversibles. En particulier, la mesure uniforme est par exemple invariante pour toute matrice de permutation, car la matrice transposée est associée à la permutation inverse. De plus il en découle que cette mesure est réversible ssi la permutation est son propre inverse, c'est-à-dire que ses cycles sont tous de longueur 2 (transpositions à supports disjoints). Une matrice de transition dont la transposée est aussi une matrice de transition est dite bistochastique ou doublement stochastique. L'ensemble des matrices bistochastiques est un polytope convexe et compact de $\mathbb{R}^{n \times n}$ dont les points extrémaux sont les matrices de permutation (théorème de Birkhoff-von-Neumann).

19 – 20

Ceci n'est pas abordé lors du cours oral.

Théorème 5.12.2: Caractérisations matricielle et fonctionnelle de la réversibilité

Soient μ et \mathbf{P} une mesure et un noyau de transition sur E . Les propriétés suivantes sont équivalentes :

- (i) μ est réversible pour \mathbf{P} .
- (ii) $\mathbf{D}\mathbf{P} = \mathbf{P}^\top \mathbf{D}$, où $\mathbf{D} = \text{diag}(\mu)$, c'est-à-dire $\mathbf{D}(x, y) := \mu(x)\mathbb{1}_{x=y}$, tandis que $\mathbf{P}^\top(x, y) := \mathbf{P}(y, x)$.
- (iii) $\mathbf{P}f \in \ell_{\mathbb{C}}^2(\mu)$ pour tout $f \in \ell_{\mathbb{C}}^2(\mu)$, et l'opérateur linéaire $\mathbf{P} : \ell_{\mathbb{C}}^2(\mu) \rightarrow \ell_{\mathbb{C}}^2(\mu)$ est auto-adjoint :

$$\langle \mathbf{P}f, g \rangle = \langle f, \mathbf{P}g \rangle \quad \text{pour tous } f, g \in \ell_{\mathbb{C}}^2(\mu).$$

— L'ensemble $\ell_{\mathbb{C}}^2(\mu)$ des $f : E \rightarrow \mathbb{C}$ t.q. $\sum_{x \in E} |f(x)|^2 \mu(x) < \infty$ est un espace de Hilbert complexe pour le produit scalaire $\langle f, g \rangle := \sum_{x, y \in E} f(x)g(x)\mu(x)$, qui s'avère utile pour l'analyse spectrale de \mathbf{P} .

— L'application $f(x) \mapsto \sqrt{\mu(x)}f(x)$ définit une isométrie linéaire $\ell_{\mathbb{C}}^2(\mu) \rightarrow \ell_{\mathbb{C}}^2(\lambda)$ où λ est la mesure de comptage, de matrice $\sqrt{\mathbf{D}}$. Elle est bijective (\mathbf{D} est inversible) ssi μ charge tous les états.

Démonstration. L'équivalence (i) \Leftrightarrow (ii) est immédiate. L'implication (iii) \Rightarrow (i) s'obtient en spécialisant à $f = \mathbb{1}_{\{x\}}$ et $g = \mathbb{1}_{\{y\}}$. Démontrons que (i) \Rightarrow (iii). Soit $f \in \ell_{\mathbb{C}}^2(\mu)$, alors $\mathbf{P}f \in \ell_{\mathbb{C}}^2(\mu)$ car par l'inégalité de Jensen pour $\mathbf{P}(x, \cdot)$, la propriété (i), et le théorème de Fubini-Tonelli,

$$\sum_{x \in E} |\mathbf{P}f(x)|^2 \mu(x) \leq \sum_{x, y \in E} \mathbf{P}(x, y) |f(y)|^2 \mu(x) = \sum_{x, y \in E} \mathbf{P}(y, x) |f(y)|^2 \mu(y) = \sum_{y \in E} |f(y)|^2 \mu(y) \sum_{x \in E} \mathbf{P}(y, x) = \|f\|_2^2 < \infty.$$

Enfin (i) donne, par linéarité, $\langle \mathbf{P}f, g \rangle = \langle f, \mathbf{P}g \rangle$ pour tous $f, g : E \rightarrow \mathbb{C}$ à support fini, et la formule se généralise ensuite à tous $f, g \in \ell_{\mathbb{C}}^2(\mu)$ par approximation. Plus précisément, pour tout $k \in \mathbb{N}$, soit $f_k := f\mathbb{1}_{E_k}$ et $g_k := g\mathbb{1}_{E_k}$ avec E_k fini et $\cup_k E_k = E$. Alors $\langle \mathbf{P}f_k, g_k \rangle \rightarrow \langle \mathbf{P}f, g \rangle$ et $\langle f_k, \mathbf{P}g_k \rangle \rightarrow \langle f, \mathbf{P}g \rangle$ quand $k \rightarrow \infty$, par convergence dominée dans $\ell_{\mathbb{C}}^1(\pi)$ où π est la mesure $\pi(x, y) = \mathbf{P}(x, y)\mu(y) = \mathbf{P}(y, x)\mu(x)$ sur $E \times E$. ■

Théorème 5.12.3. Caractérisation de Kolmogorov de la réversibilité.

Soit \mathbf{P} un noyau de transition irréductible sur E vérifiant $\mathbf{P}(x, y) > 0$ ssi $\mathbf{P}(y, x) > 0$.

Alors \mathbf{P} admet une mesure réversible ssi pour tout $n \geq 0$ et toute suite x_0, \dots, x_{n+1} dans E avec $x_{n+1} = x_0$,

$$\prod_{i=0}^n \mathbf{P}(x_i, x_{i+1}) = \prod_{i=0}^n \mathbf{P}(x_{i+1}, x_i).$$

Cela signifie que le poids d'un cycle ne dépend pas du sens de parcours.

De plus, dans ce cas, les mesures réversibles sont toutes proportionnelles.

Un noyau irréversible possède donc en quelque sorte un courant (asymétrie) le long de certains cycles.

Démonstration. Si μ est une mesure réversible, alors elle est invariante, donc $\mu(x) > 0$ pour tout x car \mathbf{P} est irréductible, et $\mu(x)/\mu(y) = \mathbf{P}(y, x)/\mathbf{P}(x, y)$ pour tous x et y . Ainsi, lorsque \mathbf{P} admet une mesure réversible, elle est unique à une constante multiplicative près. Si μ est réversible, alors la propriété de cocycle s'obtient en utilisant la définition de la réversibilité. Réciproquement, supposons que la propriété de cocycle est vérifiée par \mathbf{P} . Fixons un état y arbitraire, et choisissons $\mu(y) > 0$. Pour tout x , comme \mathbf{P} est irréductible, il existe une suite x_0, \dots, x_n , $n \geq 1$, telle que $x_0 = x$, $x_n = y$, et $\mathbf{P}(x_i, x_{i+1}) > 0$ pour tout $0 \leq i \leq n-1$. Posons alors

$$\mu(x) := \mu(y) \prod_{i=0}^{n-1} \frac{\mathbf{P}(x_{i+1}, x_i)}{\mathbf{P}(x_i, x_{i+1})}.$$

Si à présent x'_0, \dots, x'_m est un autre chemin reliant x à y , de longueur $m \geq 1$, alors les deux définitions de $\mu(x)$ coïncident grâce à la propriété de cocycle pour le cycle $x_0, \dots, x_n, x'_{m-1}, \dots, x'_0$. Enfin, un raisonnement du même type donne la réversibilité de μ : $\mu(x)\mathbf{P}(x, z) = \mu(z)\mathbf{P}(z, x)$ pour tous x, z . ■

La condition se réduit aux cycles triangulaires lorsqu'il existe un état accessible en une étape par tous les autres.

Cette caractérisation de la réversibilité a l'avantage de ne faire intervenir que le noyau \mathbf{P} , et montre en particulier que la réversibilité de \mathbf{P} constitue une contrainte polynomiale sur les coefficients de \mathbf{P} . Elle a été découverte par Andrei Kolmogorov vers 1936. Réciproquement, partant d'une mesure μ , la construction de Metropolis–Hastings ci-dessous fournit un noyau qui admet μ comme mesure réversible.

Mesures de Boltzmann–Gibbs et méthode MCMC

Une classe importante de mesures de probabilité est donnée par les mesures de Boltzmann–Gibbs, de la forme

$$\mu(x) = \frac{e^{-V(x)}}{Z}, \quad x \in E,$$

où $V : E \rightarrow \mathbb{R}$ et où $Z = \sum_{x \in E} e^{-V(x)} < \infty$ est la normalisation. La probabilité $\mu(x)$ est d'autant plus élevée que $V(x)$ est basse. Typiquement x s'interprète comme une configuration d'un système complexe, et $V(x)$ comme son énergie.

Un exemple emblématique issu de la physique statistique théorique est le modèle d'Ising à spins ± 1 , donné par $E = \{-1, 1\}^\Lambda$, $\Lambda \subset \mathbb{Z}^d$, et, en notant Λ' les arêtes de \mathbb{Z}^d dont les extrémités sont dans Λ et h un paramètre réel,

$$V(x) = \sum_{(i,j) \in \Lambda'} x_i x_j - h \sum_{i \in \Lambda} x_i.$$

Une classe importante d'exemples correspond au cas où V est constante, donc μ uniforme, et E complexe, obtenu comme intersection de contraintes dans un espace produit, dans lequel des mouvements locaux sont possibles.

Dans ces exemples, les différences $V(y) - V(x)$ sont faciles à évaluer quand x et y diffèrent peu, tandis que Z ne l'est pas facilement. Les rapports $\mu(x)/\mu(y) = e^{V(y)-V(x)}$ sont donc évaluables quand x et y diffèrent peu, tandis que $\mu(x)$ ne l'est pas facilement. Une question importante est alors d'approcher numériquement la quantité

$$\mathbb{E}(f(X)) = \int f d\mu = \mu f = \sum_x f(x)\mu(x), \quad X \sim \mu,$$

pour $f : E \rightarrow \mathbb{R}$ évaluable. La méthode Markov Chain Monte Carlo (MCMC) consiste à construire, à partir des rapports $\mu(x)/\mu(y)$, une chaîne $(X_n)_{n \in \mathbb{N}}$ irréductible récurrente positive de loi invariante μ , simulable, et d'approcher $\mathbb{E}(f(X))$ par $\frac{1}{n}(f(X_0) + \dots + f(X_{n-1}))$ (horizontal), ou encore, lorsque X est apériodique, par $\frac{1}{r}(f(Z_1) + \dots + f(Z_r))$ où Z_1, \dots, Z_r est un échantillon de la loi de X_n pour n assez grand (vertical). Il est possible de combiner ces approches horizontales et verticales en un rectangle spatio-temporel. La construction de Metropolis–Hastings fournit même une chaîne irréductible récurrente positive apériodique explicite qui admet μ comme mesure réversible!

Algorithme de Metropolis–Hastings

Soit μ une mesure de probabilité sur E qui charge tous les états, connue à normalisation près. Soit \mathbf{R} un noyau de transition auxiliaire sur E vérifiant $\mathbf{R}(x, y) = 0$ ssi $\mathbf{R}(y, x) = 0$, appelé noyau d'exploration. Ce noyau est typiquement une marche aléatoire sur E lorsque E s'y prête, ou toute autre exploration facile à simuler. Pour tous $x \neq y$, soit

$$\alpha(x, y) = \min\left(1, \frac{\mu(y)\mathbf{R}(y, x)}{\mu(x)\mathbf{R}(x, y)}\right) \quad \text{si } \mathbf{R}(x, y) > 0, \quad \text{et } \alpha(x, y) = 0 \quad \text{sinon}$$

appelée fonction d'acceptation-rejet. Elle quantifie en quelque sorte le défaut de réversibilité de μ pour \mathbf{R} .

Théodéf 5.12.4. Noyau de Metropolis–Hastings.

Soit $\mathbf{P} : E \times E \rightarrow \mathbb{R}$ définie pour tous $x, y \in E$ par $\mathbf{P}(x, y) = \begin{cases} \mathbf{R}(x, y)\alpha(x, y) & \text{si } x \neq y \\ 1 - \sum_{z \neq x} \mathbf{P}(x, z) & \text{si } x = y \end{cases}$. Alors :

- (i) \mathbf{P} est un noyau de transition sur E appelé noyau de Metropolis–Hastings.
- (ii) μ est réversible pour \mathbf{P} et en particulier invariante.
- (iii) \mathbf{P} est irréductible (respectivement apériodique) si \mathbf{R} est irréductible (respectivement apériodique).

En particulier, en combinant (ii) et (iii), il vient que \mathbf{P} est irréductible récurrent positif lorsque \mathbf{R} est irréductible, apériodique si \mathbf{R} l'est, ce qui permet d'utiliser aussi bien la LGN que la convergence en loi pour approcher μ !

Démonstration.

- (i) Il s'agit bien d'un noyau de transition sur E car $0 \leq \mathbf{P}(x, y) \leq \mathbf{R}(x, y)$ pour tout $x \neq y$.
- (ii) La loi μ est réversible et en particulier invariante pour \mathbf{P} . En effet, si $x \neq y$ et $\mathbf{R}(x, y) > 0$, alors

$$\mu(x)\mathbf{P}(x, y) = \min(\mu(x)\mathbf{R}(x, y), \mu(y)\mathbf{R}(y, x)) = \mu(y)\mathbf{P}(y, x).$$

- (iii) L'héritage de l'irréductibilité et de l'apériodicité découle du fait que les transitions possibles entre états distincts sont les mêmes pour \mathbf{P} et \mathbf{R} (autrement dit, les squelettes sont les mêmes hors diagonale). ■

Les trajectoires de la chaîne de Metropolis–Hastings se simulent comme suit : partant de l'état x , on génère une proposition $Y \sim \mathbf{R}(x, \cdot)$, et on compare $\alpha(x, Y)$ à $U \sim \text{Unif}([0, 1])$ indépendante de Y . Si $\alpha(x, Y) > U$, on accepte la transition et on pose $Z = Y$, sinon on rejette la transition et on pose $Z = x$. Ainsi, pour tous $x \neq y$,

$$\mathbb{P}(Z = y) = \mathbb{P}(U < \alpha(x, Y), Y = y) = \alpha(x, y)\mathbf{R}(x, y) = \mathbf{P}(x, y),$$

ce qui donne bien $Z \sim \mathbf{P}(x, \cdot)$. Il est remarquable que l'évaluation de $\mathbf{P}(x, x)$ soit inutile dans la simulation.

Lorsque E est muni d'une distance d dont les boules fermées de rayon unité sont finies, il est commode de prendre $\mathbf{R}(x, y) = 1/n_x$ si $d(x, y) = 1$ et $\mathbf{R}(x, y) = 0$ sinon, où $n_x := |\{y \in E : d(x, y) = 1\}|$.

Remarque 5.12.5. Fonctions d'acceptation-rejet alternatives.

En posant $\beta(x, y) = \mu(x)\mathbf{R}(x, y)$, nous avons $\alpha(x, y) = h(\beta(y, x)/\beta(x, y))$, $0 \leq h \leq 1$, tandis que la condition de réversibilité s'écrit $\beta(x, y)h(\beta(y, x)/\beta(x, y)) = \beta(y, x)h(\beta(x, y)/\beta(y, x))$, vérifiée dès que $h(u) = uh(1/u)$. C'est le cas pour notre fonction $h(u) = \min(1, u)$, mais aussi par exemple pour $h(u) = u/(u+1)$ (noyau de Barker).

Remarque 5.12.6. Algorithme d'optimisation stochastique du recuit simulé.

L'algorithme de Metropolis–Hastings permet de simuler approximativement la loi μ , en ne faisant appel qu'à l'évaluation des rapports $\mu(x)/\mu(y)$ c'est-à-dire des différences $V(y) - V(x)$ lorsque μ est une mesure de Boltzmann–Gibbs. En remplaçant V par βV où $\beta > 0$ est un paramètre (température inverse), et en faisant décroître la température par paliers le long de l'algorithme, il est possible d'obtenir une approximation du minimum global de V , qui évite d'être piégée dans un minimum local trop grossier. C'est l'algorithme du recuit simulé, inspiré de la métallurgie. La loi $\mu = \mu_\beta = Z_\beta^{-1}e^{-\beta V}$ est d'autant plus concentrée sur le minimum global de V que β est grand. Il s'agit d'un usage probabiliste de la méthode de Laplace.

5.13 Cas des espaces d'état finis

Lorsque E est fini, de cardinal d , les noyaux sur E sont des matrices $d \times d$, on identifie les mesures sur E aux vecteurs lignes non nuls de \mathbb{R}^d , et les fonctions $E \rightarrow \mathbb{R}$ à des vecteurs colonnes de \mathbb{R}^d . Le terme noyau de transition devient matrice de transition ou matrice stochastique ou markovienne. Le noyau spécial \mathbf{I} est la matrice identité. Toutes les mesures sont finies, toutes les fonctions sont bornées et intégrables pour toutes les mesures.

Soit \mathbf{P} une matrice de transition sur E de cardinal d . Une mesure μ est invariante ssi le vecteur colonne μ^\top est un vecteur propre de \mathbf{P}^\top associé à la valeur propre 1. De même, une fonction f est harmonique ssi le vecteur colonne f est un vecteur propre de \mathbf{P} associé à la valeur propre 1, c'est-à-dire $f \in \ker(\mathbf{L})$ où $\mathbf{L} = \mathbf{P} - \mathbf{I}$.

Existence de loi invariante par un argument spectral de Perron – Frobenius

Les fonctions constantes sont harmoniques. Autrement dit 1 est toujours valeur propre de \mathbf{P} , associée au vecteur propre $(1, \dots, 1)^\top$. Donc 1 est aussi valeur propre de \mathbf{P}^\top , mais les espaces propres ne sont pas ceux de \mathbf{P} en général. Cependant, si u est un vecteur propre de \mathbf{P}^\top associé à la valeur propre 1, alors il en est de même du vecteur $(|u_1|, \dots, |u_d|)^\top$ où $d = |E|$, et le vecteur ligne $(\sum_{i=1}^d |u_i|)^{-1} |u_i|$ est une loi invariante pour \mathbf{P} . En effet, si $v_i = |u_i|$ et $w = \mathbf{P}^\top v - v$, alors $w_i \geq |(\mathbf{P}^\top u)_i| - |u_i| = 0$ pour tout i par l'inégalité triangulaire car \mathbf{P} est à coefficients réels positifs. Or $w_1 + \dots + w_d = 0$ car \mathbf{P} est une matrice de transition, d'où $w = 0$.

Cet argument, publié indépendamment par Oskar Perron (1907) et Ferdinand Georg Frobenius (1912), et indépendamment de A. Markov (1906), possède une généralisation en dimension infinie (théorème de Krein – Rutman).

Unicité de la loi invariante via une propriété de Liouville

Soit f harmonique pour \mathbf{P} irréductible sur E fini. Comme E est fini, il existe $x \in E$ tel que $f(x) = M = \max_{z \in E} f(z)$. Supposons qu'il existe y tel que $f(y) < M$. Comme \mathbf{P} est irréductible, $\mathbf{P}^n(x, y) > 0$ pour un $n \geq 1$. Or f est harmonique pour \mathbf{P}^n , d'où $f(x) = \sum_{z \in E} \mathbf{P}^n(x, z) f(z) = \mathbf{P}^n(x, y) f(y) + \sum_{z \neq y} \mathbf{P}^n(x, z) f(z) < M$, ce qui contredit la définition de x . Ainsi, les fonctions harmoniques d'un noyau irréductible sur E fini sont les fonctions constantes, et on parle de propriété de Liouville⁵. Cela s'écrit également $\ker(\mathbf{L}) = \mathbb{R}(1, \dots, 1)^\top$. Par conséquent, l'espace propre de \mathbf{P} associé à la valeur propre 1 est de dimension 1, et il en est donc de même pour \mathbf{P}^\top . Ainsi, une matrice de transition irréductible admet au plus une loi invariante.

Existence de loi invariante par un argument de convexité-compacité

L'ensemble Λ_E des lois sur E , appelé parfois simplexe, est le sous-ensemble de \mathbb{R}^d formé par les vecteurs ligne (p_1, \dots, p_d) avec $p_1 \geq 0, \dots, p_d \geq 0$ et $p_1 + \dots + p_d = 1$. Soit $\mu_0 \in \Lambda_E$. Comme Λ_E est convexe,

$$\mu_n = \frac{\mu_0 + \mu_0 \mathbf{P} + \dots + \mu_0 \mathbf{P}^{n-1}}{n} \in \Lambda_E$$

pour tout $n \geq 1$. D'autre part, comme Λ_E est compact dans \mathbb{R}^d , le théorème de Bolzano – Weierstrass assure l'existence d'une sous-suite convergente (μ_{n_k}) de limite $\mu \in \Lambda_E$. On a alors $\mu \mathbf{P} = \mu$ car

$$\mu_{n_k} \mathbf{P} = \frac{\mu_0 \mathbf{P} + \dots + \mu_0 \mathbf{P}^{n_k}}{n_k} = \mu_{n_k} - \frac{\mu_0 - \mu_0 \mathbf{P}^{n_k}}{n_k}.$$

Toute valeur d'adhérence de $(\mu_n)_{n \geq 1}$ est invariante pour \mathbf{P} . La réciproque est vraie car si μ_0 est invariante pour \mathbf{P} alors $(\mu_n)_{n \geq 1}$ est constante et égale à μ_0 . Pour E infini, la convexité subsiste, mais pas la compacité.

Résolution du problème de Poisson par une propriété de Liouville

Le problème de Poisson consiste à inverser le générateur $\mathbf{L} = \mathbf{P} - \mathbf{I}$. Si \mathbf{P} est irréductible sur E fini, de loi invariante μ , alors pour toute $f : E \rightarrow \mathbb{R}$, les deux propriétés suivantes sont équivalentes :

- (i) f est de μ -moyenne nulle : $\mu f = \sum_{x \in E} f(x) \mu(x) = 0$.
- (ii) il existe une fonction $g : E \rightarrow \mathbb{R}$ vérifiant $\mathbf{L}g = f$ (équation de Poisson).

De plus, g est unique à une constante additive près.

En effet, comme μ est invariante, on a $\mu \mathbf{L} = 0$. Donc si $\mathbf{L}g = f$, alors $\mu f = 0$. Réciproquement, supposons que $\mu f = 0$. Comme \mathbf{P} est irréductible, la propriété de Liouville donne $\dim(\ker(\mathbf{L})) = 1$, donc $\dim(\text{Im}(\mathbf{L})) = |E| - 1$. Or comme $\mu \mathbf{L} = 0$, on a $\text{Im}(\mathbf{L}) = (\ker(\mathbf{L}^\top))^\perp \subset (\mathbb{R}\mu)^\perp$ espace des fonctions de μ -moyenne nulle, qui est de dimension $|E| - 1$. Ainsi, $\text{Im}(\mathbf{L}) = (\mathbb{R}\mu)^\perp$, ce qui montre l'existence de g , unique à une constante additive près, telle que $\mathbf{L}g = f$.

Le problème de Poisson est plus délicat lorsque l'espace d'état est infini. Lorsque μ est réversible, cela est lié à l'existence d'un trou dans le spectre de \mathbf{L} , vu comme un opérateur linéaire sur l'espace de Hilbert $\ell^2(\mu)$.

Théorème limite central

Théorème 5.13.1. Théorème limite central pour les chaînes de Markov.

Soit $(X_n)_{n \in \mathbb{N}}$ une CM(E, ν, \mathbf{P}) sur E fini, irréductible, de loi invariante μ .

5. En analyse complexe, une fonction entière (holomorphe sur \mathbb{C}) bornée est constante. Ses parties réelles et imaginaires sont harmoniques.

Alors pour toute fonction $f : E \rightarrow \mathbb{R}$, en notant g la solution de l'équation de Poisson $\mathbf{L}g = f - \mu f$,

$$\sqrt{n} \left(\frac{f(X_0) + \dots + f(X_{n-1})}{n} - \mu f \right) \xrightarrow[n \rightarrow \infty]{\text{loi}} \mathcal{N}(0, \sigma^2(f)) \quad \text{où} \quad \sigma^2(f) := \mu(\mathbf{P}(g^2) - (\mathbf{P}g)^2).$$

La quantité $(\mathbf{P}(g^2) - (\mathbf{P}g)^2)(x)$ est la variance de g pour $\mathbf{P}(x, \cdot)$.

Le théorème 5.13.1 généralise le TLC des suites i.i.d. En effet, lorsque $(X_n)_{n \in \mathbb{N}}$ est une suite i.i.d. de loi μ , alors $\mathbf{P}(x, \cdot) = \mu$ pour tout $x \in E$, en particulier, on peut prendre $g = f$ et $\mu(\mathbf{P}(g^2) - (\mathbf{P}g)^2)$ est la variance de f pour la loi μ .

Démonstration. Par translation, on peut supposer sans perte de généralité que $\mu f = 0$. Comme E est fini, toutes les fonctions sont bornées et intégrables pour toutes les mesures, et le théorème 5.6.3 indique alors que la suite $(M_n)_{n \in \mathbb{N}}$ définie par $M_0 = 0$ et $M_n = g(X_n) - g(X_0) - \sum_{k=1}^n (\mathbf{L}g)(X_{k-1})$ pour $n \geq 1$ est une martingale de carré intégrable, de processus croissant (théorème 4.7.1) donné par $\langle M \rangle_0 = 0$ et, pour $n \geq 1$,

$$\begin{aligned} \langle M \rangle_n &= \sum_{k=1}^n \mathbb{E}((M_k - M_{k-1})^2 \mid \mathcal{F}_{k-1}) = \sum_{k=1}^n \mathbb{E}((g(X_k) - g(X_{k-1}) - \mathbf{L}(g)(X_{k-1}))^2 \mid X_{k-1}) \\ &= \sum_{k=1}^n \mathbb{E}((g(X_k) - \mathbf{P}(g)(X_{k-1}))^2 \mid X_{k-1}) = \sum_{k=1}^n h(X_{k-1}), \end{aligned}$$

où $h(x) = \mathbf{P}((g - (\mathbf{P}g)(x))^2)(x) = \mathbf{P}(g^2)(x) - ((\mathbf{P}g)(x))^2$ est la variance de g pour $\mathbf{P}(x, \cdot)$. Donc $\frac{1}{n} \langle M \rangle_n \rightarrow \sigma^2(f)$ p.s. (th. 5.9.2, LGN pour la CM X , irréductible récurrente positive car E fini). D'autre part, la condition de Lindeberg du TLC martingale (th. 4.12.1) est vérifiée car toutes les fonctions sont bornées. Donc $\frac{1}{\sqrt{n}} M_n \rightarrow \mathcal{N}(0, \sigma^2(f))$ en loi. Enfin, comme g est bornée, il vient que $\frac{1}{\sqrt{n}}(g(X_n) - g(X_0)) \rightarrow 0$ en \mathbb{P} , donc $-\frac{1}{\sqrt{n}} \sum_{k=1}^n (\mathbf{L}g)(X_{k-1}) \rightarrow \mathcal{N}(0, \sigma^2(f))$ en loi, d'où le résultat en multipliant par -1 , car cette loi gaussienne est symétrique, et $\mathbf{L}g = f$. ■

Ceci n'est pas abordé lors du cours oral.

Réurrence et invariance

Théorème 5.13.2: Réurrence et invariance en espace d'état fini

Soit $(X_n)_{n \in \mathbb{N}}$ une CM(E, ν, \mathbf{P}) sur E fini, E_R les états récurrents, et E_T les états transitoires. Alors :

- (i) L'ensemble E_R est non vide, et tous les états récurrents sont récurrents positifs. Si \mathbf{P} est irréductible, alors il est irréductible récurrent positif.
- (ii) La chaîne est presque sûrement capturée par une classe de récurrence : $\mathbb{P}(T_{E_R} < \infty) = 1$. Quand tous les états récurrents sont absorbants, la chaîne finit p.s. par être constante.
- (iii) Si $(\mathbf{P}^n(x, \cdot))_{n \in \mathbb{N}}$ converge, sa limite est une loi de probabilité invariante.
- (iv) Les mesures invariantes ne chargent pas E_T , et si R_1, \dots, R_r sont les classes de récurrence de \mathbf{P} , alors chaque R_i porte une unique loi invariante μ_i . Les mesures invariantes pour \mathbf{P} sont les combinaisons convexes de la forme $\alpha_1 \mu_1 + \dots + \alpha_r \mu_r$ où $\alpha_1, \dots, \alpha_r$ sont des réels positifs non tous nuls. Les lois invariantes correspondent au cas où $\alpha_1 + \dots + \alpha_r = 1$.
- (v) L'espace propre de $(\mathbf{P}_{R_i \times R_i})^\top$ associé à la valeur propre 1 est de dimension 1, et contient un unique vecteur à composantes positives dont la somme vaut 1, qui est précisément μ_i .
- (vi) La matrice \mathbf{P} est irréductible ssi \mathbf{P} admet une unique loi invariante.
- (vii) Une classe irréductible est close ssi elle est récurrente. Les classes irréductibles dont aucune flèche ne sort sont récurrentes, tandis que les autres sont transitoires. Un état x est transitoire ssi il existe un état y tel que $x \rightarrow y$ et $y \not\rightarrow x$. La décomposition de E ne dépend que de la connectivité du squelette du graphe des transitions, c'est-à-dire des flèches mais pas de leur poids.

Éléments de démonstration. Si $x \in E$ et $y \in E_T$, alors $\sum_{n=0}^{\infty} \mathbf{P}^n(x, y) = \mathbf{G}(x, y) < \infty$ donc $\lim_{n \rightarrow \infty} \mathbf{P}^n(x, y) = 0$. De plus, $\sum_{y \in E} \mathbf{P}^n(x, y) = 1$ donne $\lim_{n \rightarrow \infty} \mathbb{P}^x(X_n \in E_R) = 1$, d'où $E_R \neq \emptyset$. De plus, pour tout x et tout n ,

$$\mathbb{P}^x(T_{E_R} = \infty) = \mathbb{P}^x(N_{E_R} = 0) \leq \mathbb{P}^x(X_n \notin E_R) \quad \text{donc} \quad \mathbb{P}^x(T_{E_R} = \infty) = 0.$$

Enfin, comme les classes de récurrence sont finies, elles sont récurrentes positives (cor. 5.7.4). La troisième assertion est immédiate en écrivant $\mathbf{P}^{n+1}(x, \cdot) = \mathbf{P}^n(x, \cdot)\mathbf{P}$. Pour terminer, si μ est invariante, alors comme

$\lim_{n \rightarrow \infty} \mathbf{P}^n(x, y) = 0$ pour tout $y \in E_T$ et tout $x \in E$, il vient que $\mu(y) = \sum_{x \in E} \mu(x) \mathbf{P}^n(x, y) \rightarrow 0$ quand $n \rightarrow \infty$ par convergence dominée, et donc $\mu(E_T) = 0$. ■

Spectre et période

Théorème 5.13.3: Spectre et période

Soit \mathbf{P} irréductible sur E fini. Alors :

- (i) La période est toujours $\leq |E|$.
- (ii) Le spectre de \mathbf{P} est inclus dans $\{z \in \mathbb{C} : |z| \leq 1\}$, est invariant par $z \mapsto \bar{z}$, contient 1.
- (iii) Les propriétés suivantes sont équivalentes :
 - (a) \mathbf{P} est apériodique.
 - (b) il existe un entier r tel que $\mathbf{P}^n(x, y) > 0$ pour tous x et y et tout $n \geq r$.
 - (c) 1 est la seule valeur propre de \mathbf{P} de module 1.
- (iv) Si \mathbf{P} est de période d , alors son spectre contient des racines de l'unité $\{e^{\frac{2\pi k}{d}i} : 0 \leq k \leq d-1\}$.
- (v) Si \mathbf{P} est réversible alors son spectre est réel.

Notons que comme \mathbf{P} est irréductible, nous savons que 1 est une valeur propre simple.

Si \mathbf{P} est une matrice de permutation, alors son spectre est l'union des racines de l'unité correspondant aux longueurs des cycles de la permutation.

Démonstration.

- (i) Découle du partitionnement fourni par le théorème 5.10.1.
- (ii) Nous savons déjà que 1 est dans le spectre, associé au vecteur propre $(1, \dots, 1)$. Si λ est valeur propre de \mathbf{P} , associé au vecteur propre $v \neq 0$, alors $\mathbf{P}v = \lambda v$, et il existe x tel que $v_x \neq 0$, d'où $|\lambda| \leq \sum_j \mathbf{P}(i, j) |v_j| / |v_i| = 1$. L'invariance du spectre par rapport à $z \mapsto \bar{z}$ vient du fait que \mathbf{P} est réelle.
- (iii) (a) \Rightarrow (b). Pour tout $x \in E$, soit n_x tel que $\mathbf{P}^{n_x}(x, x) > 0$ pour tout $n \geq n_x$ (th. 5.10.1), et $N := \max_{x \in E} n_x$. D'autre part, pour tous $x, y \in E$, il existe $n_{x,y}$ tel que $\mathbf{P}^{n_{x,y}}(x, y) > 0$. Soit $N' := \max_{x,y \in E} n_{x,y}$. Alors pour tous $x, y \in E$, $\mathbf{P}^{N+N'}(x, y) \geq \mathbf{P}^{n_{x,y}}(x, y) \mathbf{P}^{N+N'-n_{x,y}}(y, y) > 0$.
(b) \Rightarrow (c). Si $e^{i\theta}$, $\theta \in \mathbb{R}$, est valeur propre de module 1 de \mathbf{P} , donc valeur propre de \mathbf{P}^\top , alors il existe un vecteur ligne v tel que $v\mathbf{P} = e^{i\theta}v$, et donc $v\mathbf{P}^n = e^{in\theta}v$ pour tout $n \in \mathbb{N}$. Par l'argument de Perron-Frobenius, $|v|\mathbf{P}^n = |v|$, donc pour tout x , $|\sum_y \mathbf{P}^n(y, x)v_y| = \sum_y \mathbf{P}^n(y, x)|v_y|$. Si $n \geq r$ alors $\mathbf{P}^n(x, y) > 0$ pour tout y , donc il existe $\alpha \in \mathbb{R}$ tel que pour tout $y \in E$, $v_y = e^{i\alpha}|v_y|$. Il vient qu'à la fois $v\mathbf{P}^n = e^{in\theta}v$ et $e^{in\theta}v = e^{i(n\theta+\alpha)}|v| = e^{i(n\theta+\alpha)}|v|\mathbf{P}^n = e^{2in\theta}v\mathbf{P}^n$. Donc $e^{in\theta} = 1$, pour une infinité d'entiers n , d'où $\theta = 0$.
(c) \Rightarrow (a). Supposons par l'absurde que la période d vérifie $d > 1$. Soient C_0, \dots, C_{d-1} les classes irréductibles de \mathbf{P}^d , comme dans le th. 5.10.1. Soit μ l'unique loi invariante de la chaîne sur C_0 de noyau égal à la restriction de \mathbf{P}^d sur C_0 . Le th. 5.10.1 donne aussi que $\mu\mathbf{P}^r$ est portée par C_r pour tout $0 \leq r \leq d-1$. Il en découle que $v := \sum_{r=0}^{d-1} e^{\frac{2\pi r i}{d}} \mu\mathbf{P}^r \neq 0$ tandis que $v\mathbf{P} = e^{-\frac{2\pi i}{d}}v$, d'où la contradiction.
- (iv) L'argument de la preuve du (iii) permet de conclure.
- (v) Comme \mathbf{P} est irréductible et E fini, par le théorème 5.13.2 (i) et (vi), il existe une unique loi invariante μ , qui charge tous les états. En particulier la matrice $\mathbf{D} = \text{diag}(\mu)$ est inversible. À présent, comme l'affirme le théorème 5.12.2, la réversibilité de μ donne $\mathbf{S} := \sqrt{\mathbf{D}\mathbf{P}\mathbf{D}^{-1}} = \sqrt{\mathbf{D}^{-1}\mathbf{P}\mathbf{D}}$. Les matrices \mathbf{P} et \mathbf{S} sont semblables et ont donc le même spectre. Or \mathbf{S} est symétrique réelle, et a donc un spectre réel. ■

Convergence sous-exponentielle vers l'équilibre et couplage

Théorème 5.13.4: Convergence sous-exponentielle vers l'équilibre

Soit \mathbf{P} un noyau de transition sur E fini, irréductible apériodique, et soit μ sa loi invariante. Alors il existe un réel $0 < \rho < 1$ et un entier $r > 0$ qui dépendent de \mathbf{P} tels que

$$\max_{x \in E} \|\mathbf{P}^n(x, \cdot) - \mu\|_{\text{VT}} \leq \rho^{\lfloor n/r \rfloor} \xrightarrow{n \rightarrow \infty} 0.$$

Pour certaines chaînes comme celle de l'urne d'Ehrenfest et celle de la marche aléatoire sur l'hypercube discret, la convergence avec pire condition initiale se produit de manière abrupte après un temps critique (cutoff phenomenon) et ce phénomène est d'autant plus marqué que le modèle est en grande dimension.

Démonstration. Comme \mathbf{P} est irréductible apériodique sur E fini, il existe un entier r tel que $\varepsilon := \min_{x,y} \mathbf{P}^r(x,y) > 0$. En particulier $\mathbf{P}^r(x,y) \geq \varepsilon \mu(y)$ pour tous x et y . Posons $\rho = 1 - \varepsilon$. D'autre part, μ charge tous les états. Soit \mathbf{M} le noyau de transition sur E dont toutes les lignes sont identiques à μ . Grâce à la définition de ε , la matrice $\mathbf{Q} = \rho^{-1}(\mathbf{P}^r - (1 - \rho)\mathbf{M})$ est un noyau de transition sur E , vérifiant $\mu\mathbf{Q} = \mu$. Notons que $\mathbf{P}^r = (1 - \rho)\mathbf{M} + \rho\mathbf{Q}$. Pour tout entier k , on a

$$\mathbf{P}^{rk} = (1 - \rho^k)\mathbf{M} + \rho^k\mathbf{Q}^k.$$

Pour le voir, on observe que $\mathbf{R}\mathbf{M} = \mathbf{M}$ pour tout noyau de transition \mathbf{R} (car les lignes de \mathbf{M} sont identiques), et que $\mathbf{M}\mathbf{R} = \mathbf{M}$ pour tout noyau \mathbf{R} qui admet μ comme mesure invariante. Il en découle que pour tout entier i ,

$$\mathbf{P}^{rk+i} - \mathbf{M} = \rho^k(\mathbf{Q}^k\mathbf{P}^i - \mathbf{M}).$$

Donc $\mathbf{P}^{rk+i}(x,y) - \mu(y) = (\mathbf{P}^{rk+i} - \mathbf{M})(x,y) = \rho^k(\mathbf{Q}^k\mathbf{P}^i - \mathbf{M})(x,y)$ pour tous x et y . Comme $\mathbf{Q}^k\mathbf{P}^i$ et \mathbf{M} sont des noyaux de transition,

$$\sum_y |\mathbf{P}^{rk+i}(x,y) - \mu(y)| \leq \rho^k \sum_y (\mathbf{Q}^k\mathbf{P}^i(x,y) + \mathbf{M}(x,y)) = 2\rho^k.$$

Cela entraîne $\|\mathbf{P}^{rk+i}(x,\cdot) - \mu\|_{\text{VT}} \leq \rho^k$. Le résultat final s'obtient en écrivant $n = kr + i$ avec $k = \lfloor n/r \rfloor$. La méthode est valide pour tout $\rho > 1 - \min_{x,y} (\mathbf{P}^r(x,y)/\mu(y))$.

Examinons une preuve alternative, qui reprend la preuve du théorème 5.10.6 basée sur un couplage indépendant. Soient X et Y deux CM indépendantes de noyau \mathbf{P}^r et de lois initiales δ_x et μ . La suite $Z = (X, Y)$ est une CM sur $E \times E$ de loi initiale $\delta_x \otimes \mu$ et de noyau $\mathbf{C}((x,y), (x',y')) = \mathbf{P}^r(x,x')\mathbf{P}^r(y,y')$. En particulier, $\min_{z,z'} \mathbf{C}(z,z') \geq \varepsilon^2$. Si T_C est le temps de couplage de Z , alors pour tout entier k , en notant $z_* = (x,x)$,

$$\mathbb{P}(T_C > k) \leq \mathbb{P}(Z_1 \neq z_*, \dots, Z_k \neq z_*) = \sum_{y, z_1 \neq z_*, \dots, z_k \neq z_*} \mu(y)\mathbf{C}((x,y), z_1) \cdots \mathbf{C}(z_{k-1}, z_k) \leq (1 - \varepsilon^2)^k.$$

Comme $\text{Loi}(Y_k) = \mu\mathbf{P}^{kr} = \mu$ et $\text{Loi}(X_k) = \delta_x\mathbf{P}^{kr} = \mathbf{P}^{kr}(x,\cdot)$, le lemme 5.10.5 conduit à

$$d_{\text{VT}}(\mathbf{P}^{kr}(x,\cdot), \mu) \leq (1 - \varepsilon^2)^k.$$

À présent, tout entier n s'écrit $n = kr + i$ avec $k = \lfloor n/r \rfloor$ et $0 \leq i < r$. Enfin, par le lemme 5.10.3,

$$d_{\text{VT}}(\mathbf{P}^n(x,\cdot), \mu) = d_{\text{VT}}(\mathbf{P}^{kr}(x,\cdot)\mathbf{P}^i, \mu) \leq d_{\text{VT}}(\mathbf{P}^{kr}(x,\cdot), \mu) \leq (1 - \varepsilon^2)^{\lfloor n/r \rfloor}.$$

■

La preuve alternative ci-dessus mène au résultat général suivant, au-delà des espaces d'états finis :

Remarque 5.13.5: Critère de Döblin pour la convergence sous-exponentielle à l'équilibre

Pour tout noyau de transition \mathbf{P} irréductible récurrent positif apériodique sur un espace d'état au plus dénombrable, s'il existe une loi ν , $\varepsilon > 0$, et $r \geq 1$ tels que $\mathbf{P}^r(x,y) \geq \varepsilon\nu(y) > 0$ pour tous x, y , alors il y a convergence sous-exponentielle à l'équilibre, comme dans le théorème 5.13.4.

5.14 Quelques exemples classiques

Ceci n'est pas abordé lors du cours oral.

5.14.1 Processus de vie ou de mort sur \mathbb{N}

Il s'agit des CM sur \mathbb{N} dont le noyau vérifie $\mathbf{P}(x,y) = 0$ si $|x - y| > 1$. Elles permettent de modéliser des phénomènes discrets qui ne peuvent varier que d'une unité à chaque étape. Soient $p, q, r : \mathbb{N} \rightarrow [0, 1]$ trois fonctions telles que $p + q + r = 1$ sur \mathbb{N} et $q(0) = 0$. Le processus de vie ou de mort associé est la chaîne de

Markov sur \mathbb{N} de noyau de transition \mathbf{P} donné par $\mathbf{P}(x, \cdot) = q(x)\delta_{x-1} + r(x)\delta_x + p(x)\delta_{x+1}$, c'est-à-dire par

$$\mathbf{P}(x, y) = \begin{cases} q(x) & \text{si } y = x - 1 \\ r(x) & \text{si } y = x \\ p(x) & \text{si } y = x + 1 \end{cases}.$$

Cas irréductible général

Supposons que $p > 0$ sur \mathbb{N} , et $q > 0$ sur \mathbb{N}^* , ce qui garantit que \mathbf{P} est irréductible. Considérons le système d'équations associé à la probabilité d'atteinte de 0 (théorème 5.6.1). Recherchons la solution positive minimale $a_{\{0\}}(x) = \mathbb{P}^x(T_0 < \infty)$. Si a est solution, alors $(a(x+1) - a(x))p(x) = (a(x) - a(x-1))q(x)$ pour tout $x > 0$, et $a(0) = 1$. Pour tout $x > 0$, il vient que $a(x+1) = 1 + (\gamma(0) + \dots + \gamma(x))(a(1) - 1)$ où $\gamma(0) = 1$ et $\gamma(x) = \frac{q(1)\dots q(x)}{p(1)\dots p(x)}$.

Si $\sum_{x=0}^{\infty} \gamma(x) = \infty$, alors $0 \leq a(x+1) \leq 1$ entraîne $a(1) = 1$, et donc $a_{\{0\}}(x) = 1$ pour tout x .

Si $\sum_{x=0}^{\infty} \gamma(x) < \infty$, alors la solution positive minimale est donnée par $a_{\{0\}}(x) = \frac{\sum_{y=x}^{\infty} \gamma(y)}{\sum_{y=0}^{\infty} \gamma(y)}$.

La propriété de Markov forte donne $\mathbb{P}^0(T_0^* < \infty) = \mathbf{P}(0, 0) + \mathbf{P}(0, 1)\mathbb{P}^1(T_0^* < \infty)$. Or $\mathbf{P}(0, 0) + \mathbf{P}(0, 1) = 1$ avec $\mathbf{P}(0, 1) = p(1) > 0$, et $\mathbb{P}^1(T_0^* < \infty) = \mathbb{P}^1(T_0 < \infty) = a_{\{0\}}(1)$. Ainsi, \mathbf{P} est récurrent ssi $a_{\{0\}}(1) = 1$, c'est-à-dire ssi

$$\sum_{x=0}^{\infty} \gamma(x) = \infty.$$

En particulier \mathbf{P} est récurrent lorsque $\overline{\lim}_{x \rightarrow \infty} (q(x)/p(x)) > 1$, et transitoire lorsque $\underline{\lim}_{x \rightarrow \infty} (q(x)/p(x)) < 1$. Un calcul direct montre qu'une mesure μ sur \mathbb{N} est invariante ssi $\mu(0)p(0)\dots p(x) = \mu(x+1)q(1)\dots q(x+1)$ pour tout $x > 0$. Le corollaire 5.7.4 entraîne alors que \mathbf{P} est récurrent positif ssi

$$\sum_{x=0}^{\infty} \frac{1}{\gamma(x)p(x)} < \infty.$$

Marche aléatoire simple sur \mathbb{N} réfléchi en 0

Correspond au cas $r \equiv 0$ et $p(x) = q(x) = 1/2$ pour tout $x > 0$. Dans ce cas, la chaîne est récurrente nulle, de période 2, et les mesures invariantes sont les mesures proportionnelles à la mesure de comptage sur \mathbb{N} .

Processus de mort et processus de vie

Le processus de mort correspond au cas où la fonction p est identiquement nulle et où q est strictement positive sur \mathbb{N}^* . L'état 0 est absorbant, et tous les autres états sont transitoires. Les classes irréductibles sont des singletons. Les trajectoires sont décroissantes, et la chaîne est presque sûrement absorbée par 0. Le processus de vie correspond au cas où la fonction q est identiquement nulle et où p est strictement positive. Les trajectoires sont croissantes. Tous les états sont transitoires, et les classes irréductibles sont des singletons. Lorsque p est constante, on retrouve le processus de Bernoulli de l'exemple 5.2.2.

Files d'attente à temps discret

Dans cette interprétation, la v.a. X_n est le nombre de clients dans la file à l'instant n . Conditionnellement à $\{X_n = x\}$, $p(x)$ (respectivement $q(x)$) est la probabilité que la file perde (respectivement gagne) un client.

5.14.2 Marche aléatoire simple sur \mathbb{Z}^d

Soit $(X_n)_{n \in \mathbb{N}}$ la marche aléatoire simple sur \mathbb{Z}^d , $d \geq 1$, de l'exemple 5.2.1. Tous les états communiquent et la chaîne est irréductible. Les seules mesures invariantes sont les multiples de la mesure de comptage sur \mathbb{Z}^d . Elles sont de plus réversibles. Il n'y a pas de loi invariante. Il en découle que les états sont soit tous transitoires, soit tous récurrents nuls. La période vaut 2. Lorsque $d = 1$, les entiers relatifs $X_n - X_0$ et n sont de même parité, et la suite $(\frac{1}{2}(X_n - X_0 + n))$ est un processus de Bernoulli de paramètre $1/2$. Plus généralement, pour tout $d \geq 1$ et tout $n \geq 1$, la loi $\text{Loi}(X_n - X_0) = \text{Loi}(X_n | X_0 = 0) = \mathbf{P}^n(0, \cdot)$ est la loi image par l'application

$$(k_1, \dots, k_{2d}) \in \mathbb{N}^{2d} \mapsto (k_1 - k_2, \dots, k_{2d-1} - k_{2d}) \in \mathbb{Z}^d$$

de la loi multinomiale symétrique sur \mathbb{N}^{2d} (n jets indépendants d'un dé à $2d$ faces, pile ou face si $d = 1$) :

$$\left(\frac{\delta_{e_1} + \dots + \delta_{e_{2d}}}{2d} \right)^* n = \frac{1}{(2d)^n} \sum_{k_1 + \dots + k_{2d} = n} \frac{n!}{k_1! \dots k_{2d}!} \delta_{(k_1, \dots, k_{2d})}.$$

Cette formulation multinomiale fournit immédiatement pour tout d et tout $n \geq 1$,

$$\mathbf{P}^n(0,0) = \frac{1}{(2d)^n} \sum_{2r_1+\dots+2r_d=n} \frac{n!}{r_1!^2 \dots r_d!^2}.$$

En particulier, $\mathbf{P}^{2n+1}(0,0) = 0$ pour tous n, d , ce qui n'est pas surprenant car la période vaut 2. Nous avons déjà vu dans l'exemple 5.4.2 la récurrence de \mathbf{P} lorsque $d = 1$, en démontrant que $\mathbf{G}(0,0) = \sum_{n=0}^{\infty} \mathbf{P}^n(0,0) = \infty$ grâce à la formule de Stirling. Lorsque $d = 2$, par la formule de Vandermonde $\sum_{k=0}^n \binom{n}{k}^2 = \binom{2n}{n}$,

$$\mathbf{P}^{2n}(0,0) = \frac{1}{4^{2n}} \sum_{r_1+r_2=n} \frac{(2n)!}{r_1!^2 r_2!^2} = \frac{(2n)!}{4^{2n} n!^2} \sum_{r_1+r_2=n} \left(\frac{n!}{r_1! r_2!} \right)^2 = \frac{(2n)!}{4^{2n} n!^2} \frac{(2n)!}{n!^2},$$

de sorte que $\mathbf{G}(0,0) = \sum_{n=0}^{\infty} \mathbf{P}^n(0,0) = \infty$ par la formule de Stirling. Ainsi, \mathbf{P} est récurrent pour $d = 2$. Il est possible de procéder ainsi pour établir que \mathbf{P} est transitoire pour $d = 3$. Pour établir que \mathbf{P} est transitoire pour $d > 3$, il est possible de se ramener au cas $d = 3$, en considérant les trois premières coordonnées, qui forment une chaîne paresseuse, et une chaîne incluse non paresseuse. Alternativement à cette approche combinatoire, il est possible d'utiliser l'analyse de Fourier. Rappelons que pour toute v.a. Z sur \mathbb{Z}^d ,

$$\mathbb{P}(Z = z) = \int_{(0,1)^d} e^{2\pi i \langle z, x \rangle} \varphi_Z(x) dx \quad \text{où} \quad \varphi_Z(x) := \mathbb{E}(e^{-2\pi i \langle x, Z \rangle}) = \sum_{z \in \mathbb{Z}^d} e^{-2\pi i \langle z, x \rangle} \mathbb{P}(Z = z).$$

Ensuite, en introduisant un paramètre $\rho \in [0, 1)$ pour permuter série et intégrale, il vient que

$$\begin{aligned} \sum_{n=0}^{\infty} \mathbf{P}^n(0,0) &= \lim_{\rho \uparrow 1} \sum_{n=0}^{\infty} \rho^n \mathbb{P}(\varepsilon_1 + \dots + \varepsilon_n = 0) = \lim_{\rho \uparrow 1} \sum_{n=0}^{\infty} \int_{(0,1)^d} \rho^n e^{2\pi i \langle 0, x \rangle} \varphi_{\varepsilon_1 + \dots + \varepsilon_n}(x) dx \\ &= \lim_{\rho \uparrow 1} \sum_{n=0}^{\infty} \int_{(0,1)^d} \rho^n (\varphi_{\varepsilon_1}(x))^n dx \\ &= \lim_{\rho \uparrow 1} \int_{(0,1)^d} \sum_{n=0}^{\infty} \rho^n (\varphi_{\varepsilon_1}(x))^n dx \\ &= \lim_{\rho \uparrow 1} \int_{(0,1)^d} \frac{1}{1 - \rho \varphi_{\varepsilon_1}(x)} dx \\ &= \lim_{\rho \uparrow 1} \int_{(0,1)^d} \frac{d}{\sum_{k=1}^d (1 - \rho \cos(2\pi x_k))} dx \\ &= \int_{(0,1)^d} \frac{d}{\sum_{k=1}^d (1 - \cos(2\pi x_k))} dx, \end{aligned}$$

où la commutation de la limite quand $\rho \uparrow 1$ et de l'intégrale s'obtient par convergence monotone ou dominée. Le comportement de l'intégrale obtenue dépend uniquement du comportement de la fonction sous le signe somme au voisinage de 0. Un passage en coordonnées sphériques ainsi qu'un développement limité en 0 montrent que l'intégrale est finie ssi $r \geq 0 \mapsto r^{d-3}$ est intégrable en 0, c'est-à-dire $d > 2$, c'est-à-dire $d \geq 3$ car d est entier! Ainsi, tous les états sont récurrents nuls si $d \leq 2$, et tous transitoires si $d \geq 3$.

5.14.3 Urne d'Ehrenfest

Il s'agit de la chaîne de Markov $(X_n)_{n \in \mathbb{N}}$ sur $\{0, 1, \dots, d\}$ introduite dans l'exemple 5.2.3, de matrice de transition

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 & & & & 0 \\ \frac{1}{d} & 0 & \frac{d-1}{d} & & & & \\ 0 & \frac{2}{d} & 0 & \frac{d-2}{d} & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & \frac{d-2}{d} & 0 & \frac{2}{d} & 0 \\ & & & & \frac{d-1}{d} & 0 & \frac{1}{d} \\ 0 & & & & 0 & 1 & 0 \end{pmatrix} \quad \text{autrement dit} \quad \text{Loi}(X_{n+1} | X_n) = \frac{d - X_n}{d} \delta_{X_n+1} + \frac{X_n}{d} \delta_{X_n-1}.$$

Tous les états communiquent et la chaîne est donc irréductible récurrente positive car l'espace d'état est fini. La loi binomiale $\mu = \text{Bin}(d, 1/2) = 2^{-d} \sum_{k=0}^d \binom{d}{k} \delta_k$ est réversible donc invariante. De plus, $\mathbb{E}^x(T_x^*) = 1/\mu(x) = 2^d / \binom{d}{x}$ pour tout x . Lorsque d est pair, cette espérance est minimale pour $x = d/2$ et maximale pour $x = 0$ et $x = d$. La formule de Stirling donne $\mathbb{E}^{d/2}(T_{d/2}^*) \sim \sqrt{\pi d/2}$ pour $d \rightarrow \infty$. De même, $\mathbb{E}^d(T_d^*) = \mathbb{E}^0(T_0^*) = 2^d$, et cette quantité est

considérable lorsque d est de l'ordre du nombre d'Avogadro⁶ et ce fait sur ce modèle élémentaire a été mis en avant par Ehrenfest, ancien élève de Boltzmann, pour dépasser la contradiction apparente entre l'irréversibilité macroscopique en théorie cinétique des gaz (Boltzmann) et la récurrence des systèmes dynamiques (Poincaré). L'expression de $\text{Loi}(X_{n+1} | X_n)$ fournit $\mathbb{E}(X_{n+1} | X_n) = (1 - \frac{2}{d})X_n + 1$, et par récurrence sur n , on obtient

$$\mathbb{E}(X_n | X_0 = x) = \frac{d}{2} + (1 - \frac{2}{d})^n (x - \frac{d}{2}) \xrightarrow{n \rightarrow \infty} \frac{d}{2}.$$

En moyenne, la chaîne oublie l'état initial. Cependant, la période vaut 2, et la chaîne ne converge pas en loi. Mais d'après la remarque 5.10.2, le modèle peut facilement être rendu apériodique : une fois la boule choisie, on tire à pile ou face pour décider de l'urne qui va l'accueillir. Cela ne change pas la mesure invariante, qui reste réversible.

Ceci n'est pas abordé lors du cours oral.

5.14.4 Modèle de Wright – Fisher

Il s'agit de la chaîne de Markov sur $\{0, 1, \dots, d\}$ considérée dans l'exemple 5.2.3 et la section 4.2. Comme par construction $\text{Loi}(X_{n+1} | X_n = x) = \text{Bin}(d, \frac{x}{d})$, sa matrice de transition est donnée par

$$\mathbf{P}(x, y) = \binom{d}{y} \left(\frac{x}{d}\right)^y \left(1 - \frac{x}{d}\right)^{d-y}.$$

Les états 0 et d sont absorbants donc récurrents, tandis que les états $1, \dots, d-1$ communiquent et sont tous transitoires. Les lois $\delta_0 = \text{Bin}(d, 0)$ et $\delta_d = \text{Bin}(d, 1)$ sont donc les seules mesures invariantes. L'absorption par 0 (respectivement d) correspond à l'extinction du type A (respectivement B). La suite X/d est une chaîne de Markov sur $\{0/d, 1/d, \dots, d/d\} \subset [0, 1]$, qui décrit l'évolution du paramètre $p_n = X_n/d$ d'une loi binomiale de taille d . Plus généralement, la version à r types de la chaîne de Wright – Fisher décrit l'évolution du paramètre $p_n = (p_{n,1}, \dots, p_{n,r})$ d'une loi multinomiale de taille d (pour $r = 2$, c'est une loi binomiale).

Probabilité d'extinction

Le théorème 5.13.2 entraîne que le temps d'extinction $T = \inf\{n \in \mathbb{N} : X_n \in \{0, d\}\}$ est un temps d'arrêt presque sûrement fini. Le théorème 5.6.1 indique que la probabilité $a_A(x)$ que la chaîne partant de x soit absorbée par d (extinction de B) est donnée par $a_A(x) = x/d$. Par symétrie $a_B(x) = (d-x)/d$.

Martingale

Comme observé dans la section 4.2, la suite X est une martingale, bornée, elle converge donc p.s. et dans tout $L^{p \geq 1}$ vers une v.a. X_∞ bornée. On a $\text{Loi}(X_\infty | X_0 = x) = a_A(x)\delta_d + (1 - a_A(x))\delta_0$, et par conséquent on obtient $x = \mathbb{E}^x(X_0) = \mathbb{E}^x(X_\infty) = a_A(x)d$. On retrouve bien $a_A(x) = x/d$, plus simplement que via le th. 5.6.1.

Temps moyen d'extinction

Par le théorème 5.6.1, le temps d'extinction moyen $m(x) = \mathbb{E}^x(T)$ vérifie, $m(x) = 1 + \sum_{y \in E} \mathbf{P}(x, y)m(y)$ pour tout $x \notin \{0, d\}$. Il est difficile d'en déduire une expression explicite pour m en fonction de x et d . Cependant, lorsque $x/d \rightarrow p \in (0, 1)$ quand $d \rightarrow \infty$, le vecteur m/d s'approche d'une fonction $m_* : [0, 1] \rightarrow \mathbb{R}_+$ solution de EDO $p(1-p)m''(p) = -2$ pour $0 < p < 1$ et $m_*(0) = m_*(1) = 0$. Il en découle que

$$m_*(p) = -2(p \log(p) + (1-p) \log(1-p)) \quad \text{et donc} \quad m(x) \approx dm_*\left(\frac{x}{d}\right) \quad \text{quand } d \gg 1.$$

Pour une grande population, lorsque $x \approx d/2$, on a $m(x) \approx 2 \log(2)d \approx 1.39d$, tandis que $m(x) \approx 2 \log(d)$ lorsque $x \approx 1$. Ainsi, lorsque d est de l'ordre de k milliards, le temps moyen d'extinction est de l'ordre de d lorsque l'état initial est équilibré, tandis qu'il est de l'ordre de k lorsque l'état initial est très déséquilibré.

Mutations

Afin de rendre la chaîne de Wright – Fisher plus réaliste, on introduit la possibilité de mutations. Le passage de la génération n à la génération $n+1$ se fait d'abord en provoquant de façon indépendante pour chacun des d individus de la génération n une mutation de A à B avec probabilité p_{AB} et une mutation de B à A avec probabilité p_{BA} . Ensuite, on procède comme pour le cas sans mutation pour obtenir les d individus de la génération $n+1$. La nouvelle chaîne $(X'_n)_{n \in \mathbb{N}}$ est donnée par

$$\text{Loi}(X'_{n+1} | X'_n = x') = \text{Bin}\left(d, q_{AB} \frac{x'}{d} + p_{BA} \left(1 - \frac{x'}{d}\right)\right) = \text{Bin}\left(d, p_{BA} + (q_{AB} - p_{BA}) \frac{x'}{d}\right),$$

6. Le nombre de Amedeo Avogadro a vaut environ 6.02×10^{23} , ce qui est énorme. Il est défini dans le système international comme étant égal au nombre d'atomes dans 12 grammes de carbone-12. Par définition, une mole d'un gaz parfait contient a molécules. Aux conditions normales de température et de pression (0 °C et 101325 Pa), une mole d'un gaz parfait occupe un volume d'environ 22.41 litres.

où $q_{AB} = 1 - p_{AB}$ et $q_{BA} = 1 - p_{BA}$. L'état 0 (respectivement d) n'est plus absorbant dès que $p_{BA} > 0$ (respectivement $p_{AB} > 0$). La chaîne est irréductible récurrente positive apériodique si $0 < p_{AB}p_{BA} < 1$. Dans ce cas, la suite X' converge en loi vers l'unique loi invariante μ . Cependant, elle n'est pas une sous-martingale ni une surmartingale puisque $\mathbb{E}(X'_{n+1} | X'_n) = dp_{BA} + (q_{AB} - p_{BA})X'_n$. Bien qu'il n'existe pas de formule explicite connue pour μ , il est possible de calculer sa moyenne et sa variance. Pour la moyenne m , on écrit :

$$m = \sum_{x \in E} x\mu(x) = \sum_{x \in E} x \sum_{y \in E} \mu(y)\mathbf{P}'(y, x) = \sum_{y \in E} \mu(y) \sum_{x \in E} x\mathbf{P}'(y, x)$$

et comme $\sum_{x \in E} x\mathbf{P}'(y, x) = dp_{BA} + (q_{AB} - p_{BA})y$, on obtient la formule simple $m = \frac{dp_{BA}}{p_{AB} + p_{BA}}$. Lorsque $q_{AB} = p_{BA}$, la suite X' est i.i.d. de loi Bin(d, p_{AB}). Cette étude du modèle de Wright–Fisher illustre une démarche générale : lorsque tous les états récurrents sont absorbants, l'étude porte naturellement sur le temps d'absorption, en revanche, lorsque tous les états sont récurrents, l'étude porte sur les lois invariantes.

5.14.5 Processus de Galton–Watson

Le processus de Galton–Watson est un modèle de base qui joue un rôle presque aussi important que les marches aléatoires en modélisation. Il a été étudié notamment par Irénée-Jules Bienaymé vers 1845, puis par Francis Galton et Henri William Watson, vers 1875. Il décrit l'évolution d'une population asexuée au fil de générations qui ne se recourent pas. Si Z_n désigne le nombre d'individus de la génération $n \in \mathbb{N}$, alors

$$Z_{n+1} = \sum_{k=1}^{Z_n} X_{n+1,k}$$

où $(X_{n,k})_{n \geq 1, k \geq 1}$ sont i.i.d. de loi $P := p_0\delta_0 + p_1\delta_1 + \dots$ sur \mathbb{N} , appelée loi de reproduction. Par convention, $\sum_{\emptyset} = 0$, de sorte que si $Z_n = 0$ alors $Z_{n+1} = 0$. La suite $(Z_n)_{n \geq 0}$ est une CM sur $E = \mathbb{N}$, de noyau de transition \mathbf{P} donné par $\mathbf{P}(z, \cdot) = P^{*z}$. L'état 0 est absorbant (extinction de population), et le temps d'extinction est

$$T := \inf\{n \geq 0 : Z_n = 0\} \in \mathbb{N} \cup \{\infty\}.$$

La suite $(Z_n)_{n \geq 0}$ issue de $Z_0 = 1$ se représente comme un arbre aléatoire dont la loi de branchement est P , de sorte que la variable Z_n est le nombre de nœuds de profondeur n . L'arbre a pour profondeur ou hauteur totale $T = \inf\{n \in \mathbb{N} : Z_n = 0\}$, pour largeur totale $\max_{n \in \mathbb{N}} Z_n$, et pour taille totale $N = Z_0 + \dots + Z_{T-1}$.

- Le comportement de $(Z_n)_{n \geq 0}$ se ramène au cas $Z_0 = 1$ conditionnellement à $\{Z_0 = z\}$, la suite $(Z_n)_{n \geq 0}$ a la loi de la somme de z copies i.i.d. partant de 1.
- Si $p_z = 1$ pour un $z \in \mathbb{N}$ alors $P = \delta_z$ d'où, pour tout $n \geq 0$, $Z_{n+1} = zZ_n = \dots = z^{n+1}Z_0$.
- Si $p_0 = 0$ et $p_1 < 1$ alors $\mathbb{P}(Z_n \nearrow \infty | Z_0 = 1) = 1$ en comparant à un jeu de pile ou face.
- Si $p_0 + p_1 = 1$ et $p_0 \neq 0$, alors $\mathbb{P}(Z_n \searrow 0 | Z_0 = 1) = 1$ en comparant à un jeu de pile ou face, et le temps d'extinction T est alors la loi géométrique $\text{Geo}_{\mathbb{N}^*}(p_0)$.

Ces observations nous conduisent à supposer que dans toute la suite : $Z_0 = 1$ et $0 < p_0 \leq p_0 + p_1 < 1$.

Ceci implique que $p_z < 1$ pour tout $z \in \mathbb{N}$. Ceci implique aussi que $Z_1 \sim P$ car $Z_0 = 1$.

Nous supposons également dans toute la suite que la moyenne $m := p_1 + 2p_2 + \dots$ de P existe.

Trichotomie

Comme $\mathbb{E}(Z_n) = m^n$ pour tout n , on obtient la trichotomie suivante :

- cas sous-critique : $m < 1$, $\mathbb{E}(Z_n) \searrow 0$.
- cas critique : $m = 1$, $\mathbb{E}(Z_n) = 1$ pour tout n .
- cas sur-critique : $m > 1$, $\mathbb{E}(Z_n) \nearrow \infty$.

Fonction génératrice

La fonction génératrice $g : [0, 1] \mapsto [0, 1]$ de la loi P est la série entière réelle $g(s) = \mathbb{E}(s^{X_{1,1}}) = \sum_{z=0}^{\infty} p_z s^z$. Elle vérifie $g'(1^-) = m$. Pour tout $n \geq 1$, la fonction génératrice g_n de Z_n vérifie

$$g_n = g^{\circ n} := g \circ \dots \circ g \quad (\text{composée } n \text{ fois}).$$

En effet, $g_1 = g$ car $Z_0 = 1$, et $g_{n+1}(s) = \mathbb{E}(s^{Z_{n+1}}) = \mathbb{E}(\mathbb{E}(s^{Z_{n+1}} | Z_n)) = \mathbb{E}(\mathbb{E}(s^{X_{1,1}})^{Z_n}) = \mathbb{E}(g(s)^{Z_n}) = g_n(g(s))$.

Extinction et phénomène de seuil

Comme $p_0 > 0$, on a $\mathbf{P}(z, 0) = p_0^z > 0$ pour tout $z \in \mathbb{N}^*$, et comme 0 est absorbant, on en déduit que tout $z \in \mathbb{N}^*$ est transitoire. Or presque sûrement, la chaîne ne visite qu'un nombre fini de fois chaque état transitoire, et donc presque sûrement, soit la chaîne est capturée par l'état absorbant 0 soit elle diverge vers l'infini. Autrement dit, p.s. le processus finit par sortir, en temps suffisamment grand, de tout intervalle fini de \mathbb{N}^* . Ainsi, presque sûrement, soit le processus s'éteint, soit il tend vers l'infini (explosion) :

$$\lim_{n \rightarrow \infty} Z_n = \begin{cases} \infty & \text{si } T = \infty \\ 0 & \text{si } T < \infty \end{cases}.$$

Probabilité d'extinction

Comme Z_n est entière, il vient que $\{T < \infty\} = \{Z_n \rightarrow 0\} = \cup_{n \geq 0} \{Z_n = 0\}$. La suite $(\{Z_n = 0\})_{n \geq 0}$ est croissante et la probabilité d'extinction vérifie $\mathbb{P}(T < \infty) = \mathbb{P}(Z_n \rightarrow 0) = \mathbb{P}(\cup_{n \geq 0} \{Z_n = 0\}) = \lim_{n \rightarrow \infty} \mathbb{P}(Z_n = 0)$. Or nous avons $\mathbb{P}(Z_n = 0) = g_n(0) = g^{\circ n}(0)$, et donc

$$\mathbb{P}(T < \infty) = \lim_{n \rightarrow \infty} g_n(0) = \lim_{n \rightarrow \infty} g^{\circ n}(0).$$

Le réel $\rho = \mathbb{P}(T < \infty) = \lim_{n \rightarrow \infty} g^{\circ n}(0)$ est un point fixe : il vérifie $g(\rho) = \rho$ car $g : [0, 1] \rightarrow [0, 1]$ est continue. La fonction g est convexe car $g''(s) = \sum_{z=0}^{\infty} (z+2)(z+1)s^z p_{z+2} \geq 0$. Comme $p_0 + p_1 < 1$, la fonction g est en fait strictement convexe (elle est affine lorsque $p_0 + p_1 = 1$). D'autre part, $g(0) = p_0 \in (0, 1)$ et $g(1) = 1$. Par conséquent, si $g'(1^-) = m \leq 1$ alors le graphe de g est au-dessus de la première bissectrice et 1 est le seul point fixe. Si $g'(1^-) = m > 1$ alors le graphe de g traverse une et une seule fois la première bissectrice sur l'intervalle $(0, 1)$ et g admet un second point fixe $s_* \in (0, 1)$.

- Si $m > 1$ alors $g'(1^-) = m > 1$ et donc 1 n'est pas un point fixe attractif.
- Si $m \leq 1$ alors g possède un unique point fixe, égal à 1, et $\mathbb{P}(T < \infty) = 1$.
- Si $m > 1$ alors g possède deux points fixes : 1 (répulsif), et $s_* \in (0, 1)$ (attractif), et $\mathbb{P}(T < \infty) = s_*$.

Cela permet de préciser la trichotomie précédente :

- cas sous-critique ($m < 1$) : la population s'éteint presque sûrement.
- cas critique ($m = 1$) : la population s'éteint presque sûrement.
- cas sur-critique ($m > 1$) : la population s'éteint avec probabilité $s_* \in (0, 1)$.

Dans le cas d'une loi de reproduction poissonnienne $P = \text{Poi}(\lambda)$, il vient que $m = \lambda$ et $g(s) = e^{\lambda(s-1)}$, et donc si $\lambda \leq 1$, alors la population s'éteint presque sûrement tandis que si $\lambda > 1$, alors la probabilité d'extinction s_* est l'unique solution sur $(0, 1)$ de l'équation $\lambda = \log(s)/(s-1)$ en la variable s .

Dans le cas d'une loi de reproduction géométrique, $P = \text{Geo}_{\mathbb{N}}(p) = p \sum_{n \geq 0} q^n \delta_n$, il vient que $m = q/p$ et $g(s) = p/(1-qs)$ et donc si $p \geq 1/2$ alors la population s'éteint presque sûrement tandis que si $p < 1/2$ alors la probabilité d'extinction s_* est l'unique solution sur $(0, 1)$ de l'équation $qs^2 - s + p = 0$, d'où $s_* = p/q$.

Martingales

La suite $(Y_n = \frac{Z_n}{m^n})_{n \geq 0}$ est une martingale positive pour la filtration naturelle de Z . Elle converge donc p.s. vers une v.a.r. positive et intégrable Y_∞ , et en particulier $Y_\infty < \infty$ p.s. Aussi $Z_n \sim_{n \rightarrow \infty} m^n Y_\infty$ p.s. sur $\{Y_\infty > 0\}$. Comme $Y_n = 0$ si $n \geq T$, on obtient $\{T < \infty\} \subset \{Y_\infty = 0\}$ p.s. Dans le cas sous-critique ($m < 1$) et dans le cas critique ($m = 1$), comme vu précédemment, $\mathbb{P}(T < \infty) = 1$, d'où $\mathbb{P}(Y_\infty > 0) = 0$. Néanmoins, si $m \leq 1$ alors $\sup_{n \geq 0} \frac{Z_n}{m^n} \notin L^1$, car cela donnerait par convergence dominée $0 = \mathbb{E}(Y_T) = \mathbb{E}(\lim_{n \rightarrow \infty} Y_n) = \lim_{n \rightarrow \infty} \mathbb{E}(Y_n) = 1$.

Il est possible d'étudier la probabilité d'extinction avec l'outil martingale.

Distribution quasi-stationnaire

Considérons le cas sous-critique $m < 1$. Nous savons que $\mathbb{P}(T < \infty) = 1$ et $Z_n \rightarrow 0$ p.s. et un théorème de Yaglom affirme que $\text{Loi}(Z_n | Z_n > 0)$ converge étroitement lorsque $n \rightarrow \infty$ vers une loi μ sur \mathbb{N}^* , appelée distribution quasi-stationnaire, dont la fonction génératrice h vérifie $(h-1) \circ g = m(h-1)$. Les populations sous-critiques conditionnées à ne pas mourir sont typiquement celles qu'on observe dans la nature.

L'étude du processus peut être raffinée, notamment lorsque la variance de P est finie.

Pour aller plus loin 5.14.1.

- Échantillonneur de Gibbs, modèle d'Ising.
- Temps forts de stationnarité.
- Temps de couverture ou de recouvrement.
- Temps de mélange et phénomène de convergence abrupte à l'équilibre (cutoff).
- Marches aléatoires auto-évitantes, en paysage aléatoire, en environnement aléatoire.
- Phénomène de convergence abrupte à l'équilibre (cutoff).
- Renouvellement, fiabilité des systèmes, et processus semi-markoviens.
- Distributions quasi-stationnaires.
- Réseaux électriques.
- Coalescence, couplage par le passé, algorithme de Propp–Wilson, simulation parfaite.
- Récurrence au sens de Harris et chaîne de Harris.

ébauche

Chapitre 6

Processus de Poisson et chaînes de Markov à temps continu et espace discret



FIGURE 6.1 – Un central téléphonique vers 1904. Agner Krarup Erlang (1878–1929), à la fois mathématicien, ingénieur, et statisticien, a été membre de l’association danoise des mathématiciens, ce qui lui a fait rencontrer le mathématicien amateur Johan Jensen (qui a donné son nom à l’inégalité bien connue), ingénieur en chef de la compagnie du téléphone de Copenhague, filiale de Bell. Erlang a travaillé pour cette compagnie à partir de 1908, jusqu’à la fin de sa vie. Son étude mathématique des réseaux téléphoniques a contribué à leur automatisation. Aujourd’hui les processus de Poisson sont incontournables dans tous les champs de la modélisation stochastique.

Le processus de Poisson semble avoir été introduit vers 1909 par Agner Krarup Erlang pour l’étude de réseaux téléphoniques, et par Ernest Rutherford et Hans Geiger pour l’étude des désintégrations radioactives. Mais on le trouve également dans la thèse de Filip Lundberg vers 1903 pour l’étude des compagnies d’assurances. Tout cela se passe bien avant la formalisation moderne des probabilités, et en particulier des processus stochastiques.

Le processus de Poisson est l’analogie à temps continu du processus de Bernoulli. Il ouvre la voie à la fois aux processus ponctuels, aux processus de Lévy, et aux chaînes de Markov à temps continu, dont les files d’attente.

6.1 Processus de Bernoulli et lois géométriques

Le jeu de pile ou face avec probabilité de gain $p \in (0, 1]$ est modélisé par le processus de Bernoulli $(B_n)_{n \in \mathbb{N}}$ de l’exemple 5.2.2, défini par $B_0 = 0$ et $B_n = Y_1 + \dots + Y_n$ où $(Y_n)_{n \geq 1}$ est une suite de v.a. i.i.d. de loi de Bernoulli $\text{Bin}(1, p) = p\delta_1 + (1-p)\delta_0$. La suite des temps de saut de B est définie par

$$T_0 = 0 \quad \text{et} \quad T_{n+1} = \inf\{k > T_n : B_k \neq B_{T_n}\}.$$

Pour $n \geq 1$, la v.a. T_n est l’instant du n^{e} succès à pile ou face. Les durées inter-sauts $(E_n)_{n \geq 1}$ définies par $E_n = T_n - T_{n-1}$ sont i.i.d. de loi géométrique $\text{Geo}_{\mathbb{N}^*}(p) = \sum_{n=1}^{\infty} p(1-p)^{n-1} \delta_n$ sur \mathbb{N}^* de moyenne $1/p$. Pour tout $n \geq 1$,

$$T_n = E_1 + \dots + E_n \quad \text{et} \quad B_n = \max\{m \in \mathbb{N} : T_m \leq n\}.$$

Ainsi B est aussi le processus de comptage de tops espacés par des durées i.i.d. de loi géométrique.

Lemme 6.1.1. Caractérisation des lois géométriques par l'absence de mémoire ou amnésie.

Pour toute v.a. X sur \mathbb{N}^* , les deux propriétés suivantes sont équivalentes :

- (i) Loi(X) est une loi géométrique sur \mathbb{N}^* .
- (ii) $\text{Loi}(X - n \mid X > n) = \text{Loi}(X)$ pour tout $n \in \mathbb{N}$ (propriété d'absence de mémoire, ou d'amnésie).

Démonstration. La propriété (ii) s'écrit $G(n + m) = G(n)G(m)$ pour tous $n, m \in \mathbb{N}$, où $G(n) = \mathbb{P}(X > n)$, ce qui donne $G(n) = G(1)^n$ pour tout $n \in \mathbb{N}$, d'où $X \sim \text{Geo}_{\mathbb{N}^*}(1 - G(1))$. Réciproquement, si $X \sim \text{Geo}_{\mathbb{N}^*}(p)$ avec $0 < p \leq 1$, alors $\mathbb{P}(X - n > m \mid X > n) = G(n + m)/G(n) = (1 - p)^m$ pour tous $n, m \in \mathbb{N}$, et donc $\text{Loi}(X - n \mid X > n) = \text{Loi}(X)$. ■

6.2 Lois géométriques et lois exponentielles**Lemme 6.2.1. Caractérisation des lois exponentielles par l'absence de mémoire ou amnésie.**

Pour toute v.a. X sur \mathbb{R}_+ telle que $\mathbb{P}(X > 0) > 0$, les deux propriétés suivantes sont équivalentes :

- (i) Loi(X) est une loi exponentielle .
- (ii) $\text{Loi}(X - t \mid X > t) = \text{Loi}(X)$ pour tout $t \in \mathbb{R}_+$ (propriété d'absence de mémoire, ou d'amnésie).

Démonstration. Si $X \sim \text{Exp}(\lambda)$, alors $\mathbb{P}(X - t > s \mid X > t) = \mathbb{P}(X > t + s)/\mathbb{P}(X > t) = e^{-\lambda s}$ pour tous $s, t \in \mathbb{R}_+$, ce qui donne $\text{Loi}(X - t \mid X > t) = \text{Loi}(X)$. Réciproquement, la propriété (ii) s'écrit $G(t)G(s) = G(t + s)$ pour tous $s, t \in \mathbb{R}_+$, où $G(t) = \mathbb{P}(X > t)$. Par continuité inférieure, $G(\varepsilon) = \mathbb{P}(X > \varepsilon) \rightarrow \mathbb{P}(X > 0) > 0$ lorsque $\varepsilon \rightarrow 0^+$. Ainsi, $G(\varepsilon) > 0$ pour $\varepsilon > 0$ assez petit. Si $t > 0$ alors $t \leq n\varepsilon$ avec $n \in \mathbb{N}$ et donc $G(t) = \mathbb{P}(X > t) \geq \mathbb{P}(X > n\varepsilon) = G(\varepsilon)^n$. Ainsi, $G(t) > 0$ pour tout $t \geq 0$. D'autre part, les solutions non identiquement nulles de l'équation fonctionnelle $G(t + s) = G(t)G(s)$, $s, t \in \mathbb{R}_+$, sont de la forme $G(t) = G(1)^t$ (considérer d'abord $t \in \mathbb{Q}$ puis utiliser la continuité à droite de G pour $t \in \mathbb{R}_+$). ■

Lemme 6.2.2. Lois exponentielles et géométriques.

- (i) *Loi exponentielle comme contraction de la loi géométrique.* Soit $(p_k)_{k \geq 1}$ une suite dans $(0, 1)$ telle que $kp_k \rightarrow \lambda > 0$. Alors, si $X_k \sim \text{Geo}_{\mathbb{N}^*}(p_k)$, il vient que $\frac{1}{k}X_k \rightarrow \text{Exp}(\lambda)$ en loi.
- (ii) *Loi géométrique comme discrétisation de la loi exponentielle.* Si $Y \sim \text{Exp}(\lambda)$ et si $\lfloor Y \rfloor$ désigne la partie entière de Y , alors $\lfloor Y \rfloor$ et $Y - \lfloor Y \rfloor$ sont indépendantes et $1 + \lfloor Y \rfloor \sim \text{Geo}_{\mathbb{N}^*}(1 - e^{-\lambda})$.

Démonstration. La propriété (i) s'obtient par transformée de Fourier ou de Laplace¹ : pour tout $\theta > 0$,

$$\mathbb{E}(e^{-\theta \frac{X_k}{k}}) = \sum_{n=1}^{\infty} p_k e^{-n \frac{\theta}{k}} (1 - p_k)^{n-1} = \frac{p_k e^{-\frac{\theta}{k}}}{1 - e^{-\frac{\theta}{k}} (1 - p_k)} = \frac{kp_k e^{-\frac{\theta}{k}}}{k(1 - e^{-\frac{\theta}{k}}) + kp_k e^{-\frac{\theta}{k}}} \xrightarrow{k \rightarrow \infty} \frac{\lambda}{\theta + \lambda}.$$

La propriété (ii) découle de $\mathbb{P}(Y > t + s \mid Y > s) = \mathbb{P}(Y > t) = e^{-\lambda t}$, $s, t \in \mathbb{R}_+$. Alternativement, pour $n \in \mathbb{N}$ et $u \in [0, 1)$,

$$\mathbb{P}(\lfloor Y \rfloor = n, Y - \lfloor Y \rfloor \leq u) = \mathbb{P}(Y \in [n, n + u]) = \int_n^{n+u} \lambda e^{-\lambda s} ds = e^{-\lambda n} (1 - e^{-\lambda u}) = \mathbb{P}(\lfloor Y \rfloor = n) \mathbb{P}(Y - \lfloor Y \rfloor \leq u),$$

la dernière égalité provenant de la séparation produit de n et u , qui donne indépendance et lois marginales. ■

L'absence de mémoire des lois géométriques est l'empreinte de la propriété de Markov du processus de Bernoulli sur les temps de saut. La version à temps continu des chaînes de Markov à espace d'état au plus dénombrable s'obtient en remplaçant les temps géométriques par des temps exponentiels. Nous commençons par construire la version à temps continu du processus de Bernoulli : le processus de Poisson.

Remarque 6.2.3. Durées, durées de vie, rodage et vieillissement, fiabilité, processus semi-markoviens.

En raison de leur absence de mémoire, les lois géométriques et exponentielles sont adaptées à la modélisation des durées sans phénomène de rodage ni de vieillissement : files d'attente, désintégration radioactives, etc. Pour la modélisation des durées de vie avec rodage et vieillissement, on leur préfère les lois de type Weibull par exemple. Les processus de comptage associés ne sont plus markoviens, mais semi-markoviens, et jouent un rôle important pour la modélisation stochastique de la fiabilité des systèmes [14].

1. Plus généralement, $\text{Gamma}(m, \lambda) = \text{Exp}(\lambda)^{*m}$ est limite de lois binomiales négatives $\text{NegBin}(m, p_k) = \text{Geo}_{\mathbb{N}^*}(p_k)^{*m}$ renormalisées.

6.3 Lois exponentielles et processus de Poisson

Le processus de Bernoulli de paramètre p est le processus de comptage de tops espacés dans le temps par des durées i.i.d. de loi géométrique sur \mathbb{N}^* de paramètre p . Le processus de Poisson simple est le processus de comptage de tops espacés dans le temps par des durées i.i.d. de loi exponentielle. Il est au temps continu ce que le processus de Bernoulli est au temps discret. Soit $(E_n)_{n \geq 1}$ une suite de v.a. i.i.d. de loi exponentielle $\text{Exp}(\lambda)$, $\lambda > 0$. Pour tout $n \in \mathbb{N}$ et tout $t \in \mathbb{R}_+$, on définit les v.a. T_n et N_t à valeurs dans \mathbb{R}_+ et \mathbb{N} par

$$T_0 = 0, \quad T_n = E_1 + \dots + E_n, \quad \text{et} \quad N_t = \sup\{n \in \mathbb{N} : T_n \leq t\}.$$

En particulier, $N_0 = 0$. Le lemme suivant assure que $T_n \nearrow \infty$ p.s. de sorte que $N_t < \infty$ pour tout $t \in \mathbb{R}_+$ p.s. Autrement dit, il n'y a pas d'accumulation ou d'explosion, la trajectoire aléatoire $t \mapsto N_t$ est bien définie de \mathbb{R}_+ dans \mathbb{N} , p.s.

Lemme 6.3.1. Loi du zéro-un pour les horloges exponentielles cumulées.

Soit $(E_n)_{n \geq 1}$ une suite de v.a. indépendantes telles que $E_n \sim \text{Exp}(\lambda_n)$, $0 < \lambda_n < \infty$, pour tout n . Alors $\mathbb{P}(\sum_{n=1}^{\infty} E_n = \infty) \in \{0, 1\}$, et $\mathbb{P}(\sum_{n=1}^{\infty} E_n = \infty) = 1$ ssi $\sum_{n=1}^{\infty} \frac{1}{\lambda_n} = \infty$. Autrement dit, la série $\sum_{n=1}^{\infty} E_n$ diverge en moyenne ssi elle diverge p.s.

Démonstration. La v.a. $T_{\infty} = \sum_{n=1}^{\infty} E_n$ prend ses valeurs dans $\mathbb{R}_+ \cup \{\infty\}$, et par convergence monotone, on obtient $\mathbb{E}(T_{\infty}) = \sum_{n=1}^{\infty} 1/\lambda_n$ dans $\mathbb{R}_+ \cup \{\infty\}$. Par conséquent, si $\sum_{n=1}^{\infty} 1/\lambda_n < \infty$ alors $\mathbb{E}(T_{\infty}) < \infty$ et donc $\mathbb{P}(T_{\infty} = \infty) = 0$. L'indépendance n'a pas été utilisée. Réciproquement, par convergence monotone et indépendance,

$$\mathbb{E}(e^{-T_{\infty}}) = \prod_{n=1}^{\infty} \mathbb{E}(e^{-E_n}) = \prod_{n=1}^{\infty} \frac{\lambda_n}{1 + \lambda_n} = \left(\prod_{n=1}^{\infty} \left(1 + \frac{1}{\lambda_n}\right) \right)^{-1}.$$

Or $\prod_{n=1}^{\infty} (1 + 1/\lambda_n) = \infty$ ssi $\sum_{n=1}^{\infty} 1/\lambda_n = \infty$. Ainsi, si $\sum_{n=1}^{\infty} 1/\lambda_n = \infty$, alors $\mathbb{E}(e^{-T_{\infty}}) = 0$, d'où $\mathbb{P}(T_{\infty} = \infty) = 1$. ■

Si $(T_n)_{n \in \mathbb{N}}$ modélise les tops espacés par des durées i.i.d. de loi exponentielle $\text{Exp}(\lambda)$, alors la v.a. N_t donne le nombre de tops dans l'intervalle de temps $[0, t]$. Le processus de comptage $(N_t)_{t \in \mathbb{R}_+}$ de la suite croissante $(T_n)_{n \in \mathbb{N}}$ est appelé processus de Poisson simple issu de 0 et d'intensité λ . En abrégé PPS(0, λ). Comme $\mathbb{E}(E_n) = 1/\lambda$, plus λ est élevée, plus les durées sont courtes, et plus il y a de tops par unité de temps, d'où la terminologie.

Notons que pour tout entier $n \in \mathbb{N}$ et tout réel $t \in \mathbb{R}_+$,

$$N_{T_n} = n, \quad T_{N_t} \leq t < T_{N_t+1}, \quad \{N_t = n\} = \{T_n \leq t < T_{n+1}\} = \{T_n \leq t\} \setminus \{T_{n+1} \leq t\}.$$

On dispose également des expressions alternatives suivantes de N_t , pour tout réel $t > 0$:

$$N_t = \inf\{m \in \mathbb{N} : T_{m+1} > t\} = \sum_{n=0}^{\infty} n \mathbb{1}_{[T_n, T_{n+1})}(t) = \sum_{n=1}^{\infty} \mathbb{1}_{[T_n, +\infty)}(t) = \sum_{n=1}^{\infty} \mathbb{1}_{[0, t]}(T_n) = |\{n \in \mathbb{N}^* : T_n \in [0, t]\}|.$$

Presque sûrement, les trajectoires $t \in \mathbb{R}_+ \mapsto N_t \in \mathbb{N}$ de $(N_t)_{t \geq 0}$ sont issues de 0, croissantes, constantes par morceaux, continues à droite avec limites à gauche (càdlàg), les sauts valent +1 et T_n est le n^{e} temps de saut, cf. figure 6.2.

6.4 Propriété de Markov du processus de Poisson

Nous savons que le processus de Bernoulli est une chaîne de Markov à temps discret, et que la propriété de Markov est reliée à la nature géométrique des temps de saut. Il en va de même du processus de Poisson simple.

Théorème 6.4.1. Propriété de Markov faible du processus de Poisson.

Soit $(N_t)_{t \in \mathbb{R}_+}$ un processus de Poisson simple issu de 0 et d'intensité λ . Alors pour tout $s \in \mathbb{R}_+$, le processus $(N_{s+t} - N_s)_{t \in \mathbb{R}_+}$ est un processus de Poisson simple issu de 0 et d'intensité λ , indépendant de $\sigma((N_u)_{u \in [0, s]})$.

Démonstration. Soient $(T_n)_{n \in \mathbb{N}}$ les temps de saut et $(E_n)_{n \geq 1}$ les durées inter-sauts de N . Plaçons-nous sur

$$\{N_s = m\} = \{T_m \leq s < T_{m+1}\} = \{T_m \leq s\} \cap \{s < T_m + E_{m+1}\}.$$

Soient $(E'_n)_{n \geq 1}$ les durées inter-sauts de $(N_{s+t} - N_s)_{t \in \mathbb{R}_+}$. Nous avons

$$E'_1 = E_{m+1} - (s - T_m) \quad \text{et} \quad E'_n = E_{m+n}, \quad n \geq 2.$$

2. Pour toute suite $(a_n)_{n \geq 1}$ de réels positifs, le produit infini $\prod_{n=1}^{\infty} (1 + a_n)$ converge ssi la série $\sum_{n=1}^{\infty} a_n$ converge.

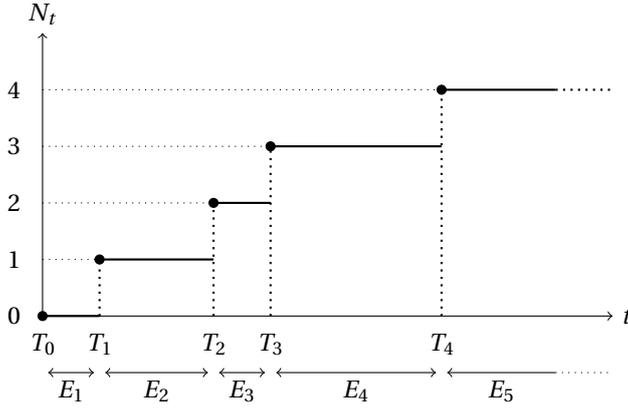


FIGURE 6.2 – Le début d’une trajectoire d’un processus de Poisson simple. La simulation d’une trajectoire est immédiate en utilisant la définition comme processus de comptage : $\text{cumsum}(-\log(\text{rand}(1, n)))$.

Conditionnellement à $\{N_s = m\}$ et $\{E_1 = e_1, \dots, E_m = e_m\}$, la variable aléatoire E'_1 suit la loi exponentielle $\text{Exp}(\lambda)$ car

$$\text{Loi}(E_{m+1} - (s - t_m) \mid E_{m+1} > s - t_m) = \text{Loi}(E_{m+1}) = \text{Exp}(\lambda)$$

où $t_m = e_1 + \dots + e_m$, grâce à la propriété d’absence de mémoire des lois exponentielles (lemme 6.2.1). Ainsi, conditionnellement à l’événement $\{N_s = m\}$ et $\{E_1 = e_1, \dots, E_m = e_m\}$, les variables aléatoires $(E'_n)_{n \geq 1}$ sont i.i.d. de loi exponentielle $\text{Exp}(\lambda)$. Cette loi ne dépend pas de m ni de e_1, \dots, e_m , et en particulier ne dépend pas de N_u avec $0 \leq u \leq s$. Il en découle que $(N_{t+s} - N_s)_{t \in \mathbb{R}_+}$ est le processus de comptage d’une suite de v.a. séparées par des durées i.i.d. de loi $\text{Exp}(\lambda)$, indépendant de N_u , pour tout $0 \leq u \leq s$, donc indépendant de $\sigma(N_u, 0 \leq u \leq s)$. ■

Théorème 6.4.2. Propriété de Markov forte du processus de Poisson.

Soit $N = (N_t)_{t \in \mathbb{R}_+}$ un PPS(0, λ) et $(\mathcal{F}_t)_{t \in \mathbb{R}_+}$ sa filtration naturelle $\mathcal{F}_t = \sigma(N_s : s \in [0, t])$.

Soit T un temps d’arrêt^a et $\mathcal{F}_T := \{A \in \mathcal{F} : \forall t \in \mathbb{R}_+, A \cap \{T \leq t\} \in \mathcal{F}_t\}$ sa tribu d’arrêt. Alors :

- (i) Conditionnellement à $\{T < \infty\}$, $(N_{T+t} - N_T)_{t \in \mathbb{R}_+}$ est également un PPS(0, λ), indépendant de \mathcal{F}_T .
- (ii) Pour tous $n \geq 1$, $f : \mathbb{R}^n \mapsto \mathbb{R}$ mesurable et bornée, et $0 \leq t_1 \leq \dots \leq t_n$,

$$\mathbb{E}(f(N_{T+t_1}, \dots, N_{T+t_n}) \mathbb{1}_{\{T < \infty\}} \mid \mathcal{F}_T) = F(N_T) \mathbb{1}_{\{T < \infty\}} \quad \text{où} \quad F(x) = \mathbb{E}(f(x + N_{t_1}, \dots, x + N_{t_n})).$$

en particulier $\mathbb{E}(f(N_{T+t_1}, \dots, N_{T+t_n}) \mid T < \infty, N_T = x) = \mathbb{E}(f(x + N_{t_1}, \dots, x + N_{t_n}))$.

a. C’est-à-dire une variable aléatoire T à valeurs dans $[0, \infty]$ telle que $\{T \leq t\} \in \mathcal{F}_t$ pour tout $t \in \mathbb{R}_+$.

La formule $\{T_n \leq t\} = \{N_t \geq n\} \in \mathcal{F}_t$ montre que $T_n := E_1 + \dots + E_n = \inf\{t \geq 0 : N_t = n\}$ est bien un temps d’arrêt.

Démonstration.

- (i) Supposons tout d’abord que $\mathbb{P}(T < \infty) = 1$. Posons $N^* := (N_{T+t} - N_T)_{t \in \mathbb{R}_+}$. Approchons T par des temps d’arrêt discrets, par excès pour bénéficier de la continuité à droite des trajectoires de N . Pour tout $n \in \mathbb{N}$, soit

$$T_n := \sum_{k=0}^{\infty} \frac{k+1}{2^n} \mathbb{1}_{T \in \left[\frac{k}{2^n}, \frac{k+1}{2^n}\right)}.$$

À ne pas confondre avec les temps de sauts de N . Nous avons $T \leq T_n$ et T_n est à valeurs dans l’ensemble dyadique $D_n := \{k/2^n : k \in \mathbb{N}\}$. De plus T_n est un temps d’arrêt, et $T_n \searrow T$ quand $n \rightarrow \infty$. Observons au passage que comme N est p.s. à trajectoires continues à droite, nous avons $N_{T+t} = \lim_{n \rightarrow \infty} N_{T_n+t}$ p.s. et donc N_{T+t} est bien mesurable. Soit $A \in \mathcal{F}_T$, $m \in \mathbb{N}$, et $0 = t_0 < \dots < t_m < \infty$. Par convergence dominée, pour toute $\varphi : \mathbb{N}^m \rightarrow \mathbb{R}$ bornée (forcément continue), grâce au fait que les trajectoires de N sont p.s. continues à droite,

$$\begin{aligned} \mathbb{E}(\mathbb{1}_A \varphi(N_{t_1}^*, \dots, N_{t_m}^*)) &= \mathbb{E}(\mathbb{1}_A \varphi(N_{T+t_1} - N_T, \dots, N_{T+t_m} - N_T)) \\ &= \lim_{n \rightarrow \infty} \mathbb{E}(\mathbb{1}_A \varphi(N_{T_n+t_1} - N_{T_n}, \dots, N_{T_n+t_m} - N_{T_n})). \end{aligned}$$

De plus $A \in \mathcal{F}_T \subset \mathcal{F}_{T_n}$ car $T \leq T_n$ et, en utilisant $A \in \mathcal{F}_{T_n}$ et la propriété de Markov faible (th. 6.4.1),

$$\mathbb{E}(\mathbb{1}_A \varphi(N_{T_n+t_1} - N_{T_n}, \dots, N_{T_n+t_m} - N_{T_n})) = \sum_{r \in D_n} \mathbb{E}(\mathbb{1}_{A \cap \{T_n=r\}} \varphi(N_{r+t_1} - N_r, \dots, N_{r+t_m} - N_r))$$

$$\begin{aligned}
&= \sum_{r \in \mathcal{D}_n} \mathbb{E}(\mathbb{1}_{A \cap \{T_n = r\}} \mathbb{E}(\varphi(N_{r+t_1} - N_r, \dots, N_{r+t_m} - N_r) \mid \mathcal{F}_r)) \\
&= \sum_{r \in \mathcal{D}_n} \mathbb{P}(A \cap \{T_n = r\}) \mathbb{E}(\varphi(N_{t_1}, \dots, N_{t_m})) \\
&= \mathbb{P}(A) \mathbb{E}(\varphi(N_{t_1}, \dots, N_{t_m})).
\end{aligned}$$

Donc N^* est indépendant de \mathcal{F}_T , et en prenant $A = \Omega$, N^* a les mêmes lois fini-dimensionnelles que N . Or la nature des trajectoires de N^* est héritée de celles de N : p.s. issues de 0, constantes par morceaux, càdlàg, avec des sauts de +1. Pour en déduire que les temps de saut de N et N^* ont même loi, il suffit de se ramener aux lois fini-dimensionnelles en observant que pour tout entier $n \in \mathbb{N}$ et toute suite de réels $0 =: t_0 < \dots < t_{n+1}$,

$$\{T_1^* \leq t_1, \dots, T_n^* \leq t_n, T_{n+1}^* > t_{n+1}\} = \{N_{t_1}^* - N_{t_0}^* \geq 1, \dots, N_{t_n}^* - N_{t_{n-1}}^* \geq 1, N_{t_{n+1}}^* = n\}$$

où cette fois-ci les T_n^* sont les temps de sauts de N^* . Enfin, plus généralement, pour un temps d'arrêt T pouvant prendre la valeur ∞ , le même argument fonctionne avec A remplacé par $A \cap \{T < \infty\}$.

(ii) Il suffit d'écrire, sur $\{T < \infty\}$, la décomposition

$$(N_{T+t_1}, \dots, N_{T+t_n}) = (N_{T+t_1} - N_T, \dots, N_{T+t_n} - N_T) + (N_T, \dots, N_T),$$

où $(N_{T+t_1} - N_T, \dots, N_{T+t_n} - N_T)$ est indépendant de \mathcal{F}_T et (N_T, \dots, N_T) est \mathcal{F}_T -mesurable, puis d'exploiter une propriété adéquate de l'espérance conditionnelle (aléa résiduel après moyennisation de l'aléa indépendant). Enfin, le « en particulier » vient du fait que $\sigma(\{T < \infty\}, N_T) \subset \mathcal{F}_T$. ■

Il est possible de concevoir un PPS comme une v.a. à valeurs dans un espace de fonctions càdlàg muni d'une tribu contenant la tribu cylindrique. Cette loi est alors caractérisée par la donnée des lois fini-dimensionnelles, c'est-à-dire la loi jointe des accroissements. Cette tribu est la tribu borélienne associée à la topologie de la convergence simple. Cela suffit pour obtenir la loi jointe des temps de saut d'un PPS, mais ne suffit plus pour des processus à temps continu plus généraux, pour lesquels il est alors opportun de choisir une tribu borélienne associée à une topologie plus forte, uniforme par exemple. Ce type de considérations est en général abordé en M2.

6.5 Structure des temps de saut et des accroissements du processus de Poisson

Si B est un processus de Bernoulli de paramètre p issu de $B_0 = 0$, alors pour tous $0 \leq n_0 \leq n_1 \leq \dots \leq n_k$, les v.a. $B_{n_k} - B_{n_{k-1}}, \dots, B_{n_1} - B_{n_0}$ sont indépendantes et suivent les lois binomiales $\text{Bin}(n_k - n_{k-1}, p), \dots, \text{Bin}(n_1 - n_0, p)$. Une propriété similaire a lieu pour le processus de Poisson simple, avec la loi de Poisson.

Commençons par l'étude de la structure des temps de saut. Si U_1, \dots, U_n sont des v.a. i.i.d. de loi uniforme sur $[0, t]$, alors la loi du vecteur (U_1, \dots, U_n) admet une densité donnée par

$$(t_1, \dots, t_n) \mapsto \frac{1}{t^n} \mathbb{1}_{\{0 < t_1, \dots, t_n < t\}}.$$

Il existe p.s. une unique permutation σ de $\{1, \dots, n\}$ t.q. $U_{\sigma(1)} < \dots < U_{\sigma(n)}$. La loi du vecteur aléatoire $(U_{\sigma(1)}, \dots, U_{\sigma(n)})$ est la statistique d'ordre de taille n de la loi uniforme sur $[0, t]$, notée $\text{STU}(n, [0, t])$. Sa densité sur \mathbb{R}^n est donnée par

$$(t_1, \dots, t_n) \mapsto \frac{n!}{t^n} \mathbb{1}_{\{0 < t_1 < \dots < t_n < t\}}.$$

Pour le concevoir sans calcul, il suffit d'observer que le cube $[0, t]^n$ de mesure de Lebesgue t^n est découpé en $n!$ morceaux isométriques à $\{0 < t_1 < \dots < t_n < t\}$, qui correspondent à l'action du groupe symétrique.

Lemme 6.5.1. Structure des temps de saut.

Si $(N_t)_{t \in \mathbb{R}_+}$ est un processus de Poisson simple issu de 0 et d'intensité λ , de temps de saut $(T_n)_{n \in \mathbb{N}}$, alors :

- (i) Loi $((T_1, \dots, T_n))$ a pour densité $(t_1, \dots, t_n) \mapsto \lambda^n e^{-\lambda t_n} \mathbb{1}_{\{0 < t_1 < \dots < t_n\}}$.
- (ii) Loi (T_n) est la loi Gamma (n, λ) de densité $t \mapsto \frac{\lambda^n}{(n-1)!} t^{n-1} e^{-\lambda t} \mathbb{1}_{\{0 < t\}}$.
- (iii) Loi $((T_1, \dots, T_n) \mid T_{n+1} = t) = \text{STU}(n, [0, t])$.
- (iv) Loi (N_t) est la loi de Poisson $\text{Poi}(\lambda t)$.
- (v) Loi $((T_1, \dots, T_n) \mid N_t = n) = \text{STU}(n, [0, t])$.

Il est commode de simuler une trajectoire d'un processus de Poisson simple en simulant un nombre n prescrit de temps de saut. Lorsque c'est le temps final t qui est prescrit plutôt que le nombre de sauts n , il est possible de procéder par simulation de $N_t \sim \text{Poi}(\lambda t)$ puis conditionnellement par simulation de STU du lemme 6.5.1.

Démonstration.

- (i) S'obtient par le changement de variable linéaire triangulaire $(s_1, s_2, \dots, s_n) \mapsto (s_1, s_1 + s_2, \dots, s_1 + \dots + s_n)$ à partir de la loi de $(T_1, T_2 - T_1, \dots, T_n - T_{n-1})$ de densité $(s_1, \dots, s_n) \mapsto \prod_{i=1}^n \lambda e^{-\lambda s_i} \mathbb{1}_{\{s_i > 0\}}$.
- (ii) S'obtient par récurrence sur n car $T_n = T_{n-1} + E_n$, ou à partir de (i) par intégration.
- (iii) Découle de (i) et (ii) en faisant le rapport de la densité de $\text{Loi}((T_1, \dots, T_{n+1}))$ par la densité de $\text{Loi}(T_{n+1})$.
- (iv) Nous avons $\{N_t = n\} = \{T_n \leq t < T_{n+1}\} = \{T_n \leq t\} \setminus \{T_{n+1} \leq t\}$, d'où $\mathbb{P}(N_t = n) = \mathbb{P}(T_n \leq t) - \mathbb{P}(T_{n+1} \leq t)$, tandis que (ii) donne, par intégration par parties, $\mathbb{P}(T_{n+1} \leq t) = -\frac{(\lambda t)^n}{n!} e^{-\lambda t} + \mathbb{P}(T_n \leq t)$.
- (v) Nous avons $\{N_t = n\} = \{T_n \leq t < T_{n+1}\}$, événement sur le vecteur aléatoire (T_1, \dots, T_{n+1}) , et il suffit alors d'utiliser la densité de $\text{Loi}((T_1, \dots, T_{n+1}))$ fournie par (i) et la formule $\mathbb{P}(N_t = n) = e^{-\lambda t} (\lambda t)^n / n!$ fournie par (iv). ■

Théorème 6.5.2. Caractérisation du processus de Poisson simple.

Soit $(N_t)_{t \in \mathbb{R}_+}$ un processus sur \mathbb{N} , issu de 0, et dont les trajectoires $t \mapsto N_t$ sont p.s. croissantes et continues à droite (donc avec limites à gauche). Alors les propriétés suivantes sont équivalentes, pour tout réel $\lambda > 0$.

- (i) *Comptage.* N est le processus de comptage de tops espacés par des durées i.i.d. de loi $\text{Exp}(\lambda)$. Autrement dit, les sauts de N valent +1 et les temps inter-sauts sont i.i.d. de loi $\text{Exp}(\lambda)$.
- (ii) *Structure des accroissements.* Pour tout entier $n \geq 1$ et tous réels $0 \leq t_0 \leq t_1 \leq \dots \leq t_n$, les accroissements $N_{t_n} - N_{t_{n-1}}, \dots, N_{t_1} - N_{t_0}$ sont indépendants et de lois de Poisson $\text{Poi}(\lambda(t_n - t_{n-1})), \dots, \text{Poi}(\lambda(t_1 - t_0))$.
- (iii) *Bernoulli infinitésimal.* Les accroissements de N sont indépendants, avec, uniformément en $t \in \mathbb{R}_+$,

$$\mathbb{P}(N_{t+\varepsilon} - N_t = 0) = 1 - \lambda\varepsilon + o_{\varepsilon \rightarrow 0^+}(\varepsilon) \quad \text{et} \quad \mathbb{P}(N_{t+\varepsilon} - N_t = 1) = \lambda\varepsilon + o_{\varepsilon \rightarrow 0^+}(\varepsilon).$$

Dans (ii), les accroissements du processus sont stationnaires : ils ne font intervenir que la durée ou différence de temps. Il s'agit d'un processus à accroissements indépendants et stationnaires (PAIS) de loi de Poisson.

La propriété de Markov faible du théorème 6.4.1 peut être déduite de (ii).

La propriété (iii) est une comparaison infinitésimale à un processus de Bernoulli de paramètre $\varepsilon\lambda$.

Démonstration.

(i) \Rightarrow (ii). Découle de la propriété de Markov (th. 6.4.1) et du fait que $N_t \sim \text{Poi}(\lambda t)$ (lemme 6.5.1 (iv)).

(ii) \Rightarrow (iii). Découle de $\mathbb{P}(N_{t+\varepsilon} - N_t = 0) = e^{-\lambda\varepsilon}$ et $\mathbb{P}(N_{t+\varepsilon} - N_t = 1) = e^{-\lambda\varepsilon} \lambda\varepsilon$.

(iii) \Rightarrow (ii). La propriété (iii) donne $\mathbb{P}(N_{t+\varepsilon} - N_t = m) = o_{\varepsilon \rightarrow 0^+}(\varepsilon)$ uniformément en $t \in \mathbb{R}_+$, pour tout $m \geq 2$. Par suite, en notant $p_k(t) = \mathbb{P}(N_t = k)$, on obtient pour tout $m \geq 1$, $p_m(t + \varepsilon) = \sum_{n=0}^m \mathbb{P}(N_{t+\varepsilon} - N_t = n) p_{m-n}(t)$, et donc

$$\frac{p_m(t + \varepsilon) - p_m(t)}{\varepsilon} = -\lambda p_m(t) + \lambda p_{m-1}(t) + o_{\varepsilon \rightarrow 0^+}(1)$$

uniformément en $t \in \mathbb{R}_+$. En exprimant cette propriété pour $t - \varepsilon$, on obtient

$$\frac{p_m(t) - p_m(t - \varepsilon)}{\varepsilon} = -\lambda p_m(t - \varepsilon) + \lambda p_{m-1}(t - \varepsilon) + o_{\varepsilon \rightarrow 0^+}(1).$$

La fonction $t \mapsto p_m(t)$ est donc dérivable et vérifie l'EDO $p'_m(t) = -\lambda p_m(t) + \lambda p_{m-1}(t)$. De même $p'_0(t) = -\lambda p_0(t)$. Les conditions initiales de ces EDO sont imposées par $N_0 = 0$. Par récurrence sur m , $p_m(t) = e^{-\lambda t} \frac{(\lambda t)^m}{m!}$, c'est-à-dire que $N_t \sim \text{Poi}(\lambda t)$. Comme $(N_{s+t} - N_s)_{t \in \mathbb{R}_+}$ vérifie également (iii), on obtient bien (ii).

(ii) \Rightarrow (i). La propriété (ii) spécifie la loi des trajectoires, et il existe donc un unique processus vérifiant (ii), qui est donc forcément le processus de Poisson simple issu de 0 et d'intensité λ construit par (i). Plus précisément, il est possible d'utiliser la même méthode que dans la preuve de la propriété de Markov forte (th. 6.4.2) : pour tout entier $n \geq 1$, en notant T_1, \dots, T_{n+1} les $n + 1$ premiers sauts de N , pour tous réels $0 =: t_0 < \dots < t_n$,

$$\{T_1 \leq t_1, \dots, T_n \leq t_n, T_{n+1} > t_n\} = \{N_{t_1} - N_{t_0} \geq 1, \dots, N_{t_n} - N_{t_{n-1}} \geq 1, N_{t_n} - N_{t_0} = n\}$$

qui permet de déduire la loi jointe des temps de sauts de la loi jointe des accroissements (et réciproquement!). L'hypothèse de trajectoires continues à droite intervient dans la preuve de la propriété de Markov forte. ■

Remarque 6.5.3. Processus ponctuels de Poisson.

La v.a. $N_{t+s} - N_s$ compte les tops dans l'intervalle de temps $(s, t + s]$. Plus généralement, si I et J sont deux intervalles disjoints quelconques de \mathbb{R}_+ , alors le nombre de tops dans I et J sont indépendants et de lois de Poisson $\text{Poi}(\lambda|I|)$ et $\text{Poi}(\lambda|J|)$. C'est le point de départ de la notion de processus ponctuel de Poisson.

Remarque 6.5.4. Martingales.

Le processus de Bernoulli $(B_n)_{n \in \mathbb{N}}$ de paramètre p est une sous-martingale pour sa filtration canonique, de compensateur $A_n = np$, de sorte que sa décomposition de Doob s'écrit $B_n = B_n - np + np = M_n + A_n$. De même, le processus de Poisson $(N_t)_{t \in \mathbb{R}_+}$ d'intensité λ est une sous-martingale pour sa filtration canonique^a, de compensateur $A_t = \lambda t$, de sorte que sa décomposition de Doob s'écrit $N_t = N_t - \lambda t + \lambda t = M_t + A_t$.

a. Il s'agit d'une filtration à temps continu, et de notre première martingale à temps continu!

La loi des grands nombres et le théorème limite central pour les suites de v.a. i.i.d. de loi de Bernoulli $\text{Bin}(p) = (1-p)\delta_0 + p\delta_1$ conduisent aux propriétés suivantes pour le processus de Bernoulli B de paramètre p : $\frac{B_n}{n} \rightarrow p$ p.s. et $\sqrt{n}(\frac{B_n}{n} - p) \rightarrow \mathcal{N}(0, p(1-p))$ en loi. Le théorème suivant fournit l'analogie pour le processus de Poisson simple.

Théorème 6.5.5. Loi des grands nombres et théorème limite central.

Si $(N_t)_{t \in \mathbb{R}_+}$ est un processus de Poisson simple issu de 0 et d'intensité λ , alors

$$\frac{N_t}{t} \xrightarrow[t \rightarrow \infty]{\text{p.s.}} \lambda \quad \text{et} \quad \sqrt{t} \left(\frac{N_t}{t} - \lambda \right) \xrightarrow[t \rightarrow \infty]{\text{loi}} \mathcal{N}(0, \lambda).$$

Démonstration. La convergence en loi vers $\mathcal{N}(0, \lambda)$ s'obtient en utilisant les fonctions caractéristiques et le fait que $N_t \sim \text{Poi}(\lambda t)$ (lemme 6.5.1). Pour la convergence p.s. vers λ , la croissance de $t \mapsto N_t$ donne $N_{\lfloor t \rfloor} \leq N_t \leq N_{\lfloor t \rfloor + 1}$, où $\lfloor t \rfloor$ est la partie entière de t . La décomposition $N_{\lfloor t \rfloor} = N_1 + (N_2 - N_1) + \dots + (N_{\lfloor t \rfloor} - N_{\lfloor t \rfloor - 1})$ et le théorème 6.5.2 montrent que $N_{\lfloor t \rfloor}$ est la somme de $\lfloor t \rfloor$ v.a. i.i.d. de loi $\text{Poi}(\lambda)$. Donc $t^{-1}N_{\lfloor t \rfloor}$ converge p.s. vers λ par la LGN pour des v.a. i.i.d. de loi $\text{Poi}(\lambda)$. Idem pour $t^{-1}N_{\lfloor t \rfloor + 1}$, et donc pour $t^{-1}N_t$, par encadrement. Alternativement, on peut observer que $N_{T_n} / T_n = n / T_n$ et utiliser la LGN sur la suite $(T_{n+1} - T_n)_{n \geq 1}$, comme le suggère la remarque suivante. ■

Remarque 6.5.6. Estimation de l'intensité λ .

Considérons un processus de Poisson simple $(N_t)_{t \in \mathbb{R}_+}$ issu de 0 et d'intensité λ , de temps de saut $(T_n)_{n \geq 1}$ et de durées inter-sauts $(E_n)_{n \geq 1}$. Une estimation sans biais de $1/\lambda$ peut être obtenue à partir d'une seule réalisation de la trajectoire du processus grâce aux convergences suivantes, fournies par la loi des grands nombres et le théorème limite central pour des variables aléatoires i.i.d. de loi $\text{Exp}(\lambda)$:

$$\frac{T_n}{n} = \frac{E_1 + \dots + E_n}{n} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \frac{1}{\lambda} \quad \text{et} \quad \sqrt{n} \left(\frac{T_n}{n} - \frac{1}{\lambda} \right) \xrightarrow[n \rightarrow \infty]{\text{loi}} \mathcal{N} \left(0, \frac{1}{\lambda^2} \right).$$

La loi de T_n/n est explicite car T_n suit une loi Gamma (le. 6.5.1). Il est donc possible d'utiliser cette loi pour la détermination d'intervalles de confiance non asymptotiques. D'autre part, il est possible d'estimer λ sans biais à partir d'une seule réalisation de N_t grâce à la convergence p.s. de N_t/t vers λ (th. 6.5.5). Ici encore, la loi de N_t/t est explicite, et fournit des intervalles de confiance non asymptotiques.

6.6 Processus de Poisson comme limite de processus de Bernoulli

Le lemme 6.2.2 suggère que les temps inter-sauts du processus de Bernoulli convergent, lorsque le temps est contracté, vers ceux d'un processus de Poisson simple. Le théorème suivant confirme cette intuition. D'autre part, il est bien connu que $\text{Bin}(n, p_n) \rightarrow \text{Poi}(\lambda)$ lorsque $n \rightarrow \infty$ avec $np_n \rightarrow \lambda$. Le théorème suivant généralise cette propriété aux lois marginales du processus de Bernoulli. Le tableau 6.1 dresse une comparaison analogique.

Théorème 6.6.1. Processus de Poisson et de Bernoulli.

Soit $\lambda > 0$ un réel et $(p_k)_{k \geq 1}$ une suite de réels dans $[0, 1]$ telle que $\lim_{k \rightarrow \infty} kp_k = \lambda$. Pour tout entier $k \geq 1$, soit $(B_n^k)_{n \in \mathbb{N}}$ un processus de Bernoulli issu de 0 et de paramètre p_k . Alors pour tout entier $n \geq 1$ et tous réels $0 \leq t_1 \leq \dots \leq t_n$, la suite de vecteurs aléatoires $((B_{\lfloor t_1 k \rfloor}^k, \dots, B_{\lfloor t_n k \rfloor}^k))_{k > 0}$ converge en loi vers la loi de $(N_{t_1}, \dots, N_{t_n})$, où $(N_t)_{t \in \mathbb{R}_+}$ est un processus de Poisson simple issu de 0 et d'intensité λ .

Démonstration. On a $\lim_{k \rightarrow \infty} \text{Bin}(r_k, p_k) = \text{Poi}(\mu)$ si $\lim_{k \rightarrow \infty} r_k p_k = \mu$. Pour $t \geq 0$ fixé, on pose $r_k = \lfloor tk \rfloor$. Comme $\lim_{k \rightarrow \infty} r_k/k = t$ et $\lim_{k \rightarrow \infty} kp_k = \lambda$, on a $\lim_{k \rightarrow \infty} r_k p_k = \lambda t$, et donc $\lim_{k \rightarrow \infty} \text{Bin}(\lfloor tk \rfloor, p_k) = \text{Poi}(\lambda t)$. Or le processus de Bernoulli est à accroissements indépendants et stationnaires, de lois binomiales, d'où le résultat. ■

Processus de Bernoulli (jeu de pile ou face)	Processus de Poisson simple
Espace d'état \mathbb{N} et loi initiale δ_0	
Processus $(B_n)_{n \in \mathbb{N}}$ avec $B_0 = 0$	Processus $(N_t)_{t \in \mathbb{R}_+}$ avec $N_0 = 0$
Temps discret $n \in \mathbb{N}$	Temps continu $t \in \mathbb{R}_+$
Paramètre $p \in [0, 1]$	Paramètre $\lambda \in \mathbb{R}_+$
$T_0 = 0$ et $T_m = E_1 + \dots + E_m$	
$(E_m)_{m \in \mathbb{N}^*}$ i.i.d. géométrique $\text{Geo}_{\mathbb{N}^*}(p)$	$(E_m)_{m \in \mathbb{N}^*}$ i.i.d. exponentielle $\text{Exp}(\lambda)$
$B_n = \inf\{m \in \mathbb{N} : T_{m+1} > n\}$	$N_t = \inf\{m \in \mathbb{N} : T_{m+1} > t\}$
$B_n = \sup\{m \in \mathbb{N} : T_m \leq n\}$	$N_t = \sup\{m \in \mathbb{N} : T_m \leq t\}$
Loi(T_m) = $\text{Geo}_{\mathbb{N}^*}(p)^{*m}$ = loi binomiale négative	Loi(T_m) = $\text{Exp}(\lambda)^{*m}$ = loi Gamma
Loi(B_n) = $\text{Bin}(n, p)$ = loi binomiale	Loi(N_t) = $\text{Poi}(\lambda t)$ = loi de Poisson
$\mathbb{E}(B_n) = np$	$\mathbb{E}(N_t) = \lambda t$
$B_n = B_{n_k} - B_{n_{k-1}} + \dots + B_{n_1} - B_{n_0}$	$N_t = N_{t_k} - N_{t_{k-1}} + \dots + N_{t_1} - N_{t_0}$
$0 = n_0 \leq n_1 \leq \dots \leq n_k = n$	$0 = t_0 \leq t_1 \leq \dots \leq t_k = t$
$(B_{n_k} - B_{n_{k-1}})_{k \in \mathbb{N}}$ indépendantes	$(N_{t_k} - N_{t_{k-1}})_{k \in \mathbb{N}}$ indépendantes
Loi($B_{n_k} - B_{n_{k-1}}$) = $\text{Bin}(n_k - n_{k-1}, p)$	Loi($N_{t_k} - N_{t_{k-1}}$) = $\text{Poi}(\lambda(t_k - t_{k-1}))$
$\lim_{k \rightarrow \infty} \text{Bin}(\lfloor tk \rfloor, p_k) = \text{Poi}(\lambda t)$ quand $\lim_{k \rightarrow \infty} kp_k = \lambda$	

TABLE 6.1 – Tableau analogique entre processus de Bernoulli et de Poisson.

6.7 Paradoxe de l'autobus

Soit $(N_t)_{t \in \mathbb{R}_+}$ un processus de Poisson simple issu de 0, d'intensité λ , et de temps de saut $(T_n)_{n \in \mathbb{N}}$, avec $T_0 = 0$. Pour tout réel $t \geq 0$, $T_{N_t} \leq t < T_{N_t+1}$ ce qui permet de considérer la dichotomie

$$[T_{N_t}, T_{N_t+1}] = [T_{N_t}, t] \cup [t, T_{N_t+1}], \text{ de longueurs } U_t = t - T_{N_t} \text{ et } V_t = T_{N_t+1} - t.$$

Or $\{U_t \geq u, V_t > v\} = \{N_{t+v} - N_{t-u} = 0\}$ pour tous $0 < u < t$ et $v > 0$, et $N_{t+v} - N_{t-u} \sim \text{Poi}(\lambda(u+v))$ (th. 6.5.2), d'où

$$\mathbb{P}(U_t \geq u, V_t > v) = \mathbb{P}(N_{t+v} - N_{t-u} = 0) = e^{-\lambda(v+u)} = e^{-\lambda v} e^{-\lambda u},$$

ce qui montre à la fois que U_t et V_t sont indépendantes, que $V_t \sim \text{Exp}(\lambda)$, et que $\mathbb{P}(U_t > u) = e^{-\lambda u}$ si $0 < u < t$, c'est-à-dire que U_t a la même loi³ que $\min(E, t)$ où $E \sim \text{Exp}(\lambda)$. La formule $T_{N_t+1} - T_{N_t} = U_t + V_t$ donne

$$\mathbb{E}(T_{N_t+1} - T_{N_t}) = \mathbb{E}(U_t) + \mathbb{E}(V_t) = \frac{1 - e^{-\lambda t}}{\lambda} + \frac{1}{\lambda} \quad \text{car} \quad \mathbb{E}(U_t) = \int_0^t \mathbb{P}(U_t > u) du.$$

En particulier, $\mathbb{E}(T_{N_t+1} - T_{N_t}) > 1/\lambda$ et $\lim_{t \rightarrow \infty} \mathbb{E}(T_{N_t+1} - T_{N_t}) = 2/\lambda$. En moyenne, l'intervalle $[T_{N_t}, T_{N_t+1}]$ délimité par les deux tops encadrant un temps déterministe fixé t est plus grand que la durée moyenne $1/\lambda$ séparant deux tops consécutifs. Le rapport croît avec t et tend même vers 2 lorsque $t \rightarrow \infty$. Cependant, $\lim_{t \rightarrow 0} \mathbb{E}(T_{N_t+1} - T_{N_t}) = 1/\lambda$.

Si les tops représentent l'arrivée d'un autobus à la station, l'écart moyen entre le dernier passage avant le temps t et le prochain passage après le temps t est plus grand que la durée moyenne des passages des autobus à la station! C'est le paradoxe de l'autobus. Il est possible d'interpréter ce paradoxe en faisant appel à la notion d'information, qui est à la base de la théorie des probabilités. L'introduction du temps déterministe t correspond à choisir implicitement un intervalle aléatoire, la loi de la longueur de cet intervalle est donc déformée par ce choix, qui constitue une information supplémentaire. Si I_0, I_1, \dots désignent les intervalles $[T_0, T_1], [T_1, T_2], \dots$, alors l'intervalle $[T_{N_t}, T_{N_t+1}]$ s'écrit I_{N_t} . Pour tout n , la loi de la longueur de I_n est exponentielle, mais la loi de la longueur de I_{N_t} n'est plus exponentielle car N_t est aléatoire! La quantité déterministe t s'est muée en quantité aléatoire N_t . L'intuition n'est correcte que pour des intervalles de longueur déterministe. Plus t est grand, plus le numéro N_t de l'intervalle contenant t est dispersé, de variance λt , et plus la longueur apparente de l'intervalle est grande⁴!

3. La loi de U_t est $\lambda e^{-\lambda u} \mathbb{1}_{[0, t]}(u) du + e^{-\lambda t} \delta_t$, composée d'une partie à densité sur l'intervalle $[0, t]$ et d'un atome en t . Ce type de loi composite est typique des phénomènes de seuillage et de saturation, très courants dans la nature. L'atome en t vient du fait que $\mathbb{P}(T_1 > t) = e^{-\lambda t}$.

4. De manière imagée, le doigt aléatoire désignant l'intervalle tremble d'autant plus que t est grand! Cela fait penser au phénomène de biais par la taille, classique en théorie de l'échantillonnage : estimer la taille moyenne des familles d'une population d'être humains en tirant au hasard des individus dans la population doit tenir compte de la sur ou sous représentation des familles en fonction de leur taille.

6.8 Poissonisation d'une chaîne de Markov à temps discret

Exemple 6.8.1. Processus de Poisson composé.

Soit $(Z_n)_{n \in \mathbb{N}}$ une marche aléatoire sur un semi-groupe E , avec $(\varepsilon_{n+1} = Z_{n+1} - Z_n)_{n \in \mathbb{N}}$ suite de v.a. i.i.d. sur E de loi ν , indépendante de Z_0 . Soit $(N_t)_{t \in \mathbb{R}_+}$ un PPS(0, λ), indépendant de Z . Le processus $(Z_{N_t} = \varepsilon_0 + \dots + \varepsilon_{N_t})_{t \in \mathbb{R}_+}$ est un processus de Poisson composé. Il vérifie

$$\mathbb{P}(Z_{N_t} = y \mid Z_{N_0} = x) = e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} \mathbb{P}(x + \varepsilon_1 + \dots + \varepsilon_n = y) = e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} \nu^{*n}(y - x)$$

par le théorème de Fubini–Tonelli. Lorsque $\nu(0) = 0$, les temps de saut de $(Z_{N_t})_{t \in \mathbb{R}_+}$ sont ceux du processus de Poisson sous-jacent $(N_t)_{t \in \mathbb{R}_+}$. Dans le cas particulier où $(Z_n)_{n \in \mathbb{N}}$ est un processus de Bernoulli de paramètre $p \in [0, 1]$, alors $\nu^{*n} = \text{Bin}(n, p)$ et on obtient que $(Z_{N_t})_{t \in \mathbb{R}_+}$ est un PPS(0, $p\lambda$).

Remarque 6.8.2. Formules de Wald.

Avec les notations de l'exemple 6.8.1 et $E = \mathbb{R}$, si T est une v.a. sur \mathbb{N} indépendante de $(\varepsilon_n)_{n \in \mathbb{N}}$, alors ^a

$$\mathbb{E}(\varepsilon_1 + \dots + \varepsilon_T) = \mathbb{E}(\mathbb{E}(\varepsilon_1 + \dots + \varepsilon_T \mid T)) = \mathbb{E}(T)\mathbb{E}(\varepsilon_1).$$

Comme $(\varepsilon_n)_{n \geq 1}$ et $(N_t)_{t \in \mathbb{R}_+}$ sont indépendants et que $N_t \sim \text{Poi}(\lambda t)$, on obtient

$$\mathbb{E}(Z_{N_t} \mid Z_{N_0} = x) = x + \mathbb{E}(\varepsilon_1 + \dots + \varepsilon_{N_t}) = x + \mathbb{E}(N_t)\mathbb{E}(\varepsilon_1) = x + \lambda t \mathbb{E}(\varepsilon_1).$$

La même méthode permet d'exprimer la fonction caractéristique de Z_{N_t} en fonction de celle de ν .

^a Le conditionnement est visuel et intuitif, mais l'indépendance permet de s'en affranchir grâce au théorème de Fubini–Tonelli après réduction au cas positif, plus précisément, $\mathbb{E}(\varepsilon_1 + \dots + \varepsilon_T) = \sum_n \mathbb{E}((\varepsilon_1 + \dots + \varepsilon_n) \mathbb{1}_{\{T=n\}}) = \sum_n n \mathbb{E}(\varepsilon_1) \mathbb{P}(T=n) = \mathbb{E}(T)\mathbb{E}(\varepsilon_1)$.

Exemple 6.8.3. Poissonisation d'une chaîne de Markov à temps discret.

Reprenons la notion de processus de Poisson composé en la débarrassant de la structure de groupe sur E . Soit $(Z_n)_{n \in \mathbb{N}}$ une chaîne de Markov d'espace d'état au plus dénombrable E et de noyau de transition \mathbf{P} . Soit $(N_t)_{t \in \mathbb{R}_+}$ un PPS(0, λ), indépendant de $(Z_n)_{n \in \mathbb{N}}$. Le processus à temps continu et à espace discret $(Z_{N_t})_{t \in \mathbb{R}_+}$ est appelée chaîne de Markov poissonisée. Pour tout réel $t \geq 0$,

$$Z_{N_t} = Z_n \quad \text{sur} \quad \{T_n \leq t < T_{n+1}\}.$$

L'indépendance de Z et N et le fait que $N_t \sim \text{Poi}(\lambda t)$ donnent, par le théorème de Fubini–Tonelli,

$$\mathbb{P}(Z_{N_t} = y \mid Z_{N_0} = x) = e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} \mathbf{P}^n(x, y).$$

Les interventions du mécanisme de transition \mathbf{P} sont espacées par des temps aléatoires exponentiels indépendants $(T_{n+1} - T_n)_{n \in \mathbb{N}}$. Or \mathbf{P} ne provoque un saut qu'au bout d'un nombre aléatoire géométrique de tentatives (section 5.5). Ainsi, sachant $\{Z_0 = x\}$, les temps de saut de $(Z_{N_t})_{t \in \mathbb{R}_+}$ sont des sommes géométrique de temps exponentiels indépendants. Le lemme 6.8.4 ci-dessous affirme que ces temps sont exponentiels!

Lemme 6.8.4. Somme géométrique d'exponentielles.

Soit $(X_n)_{n \geq 1}$ une suite de v.a. i.i.d. de loi exponentielle $\text{Exp}(\lambda)$. Soit Y une v.a. de loi géométrique $\text{Geo}_{\mathbb{N}^*}(p)$ avec $p \in [0, 1]$, indépendante de $(X_n)_{n \geq 1}$. Alors $X_1 + \dots + X_Y$ suit la loi exponentielle $\text{Exp}(p\lambda)$.

Démonstration. Pour tout $n \in \mathbb{N}^*$, $X_1 + \dots + X_n \sim \text{Gamma}(n, \lambda) = \text{Exp}(\lambda)^{*n}$. La densité de $X_1 + \dots + X_Y$ est

$$x \in \mathbb{R}_+ \mapsto \sum_{n=1}^{\infty} \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x} p(1-p)^{n-1} = p\lambda e^{-p\lambda x}.$$

C'est un mélange géométrique de lois gammas. Alternativement, avec la transformée de Laplace, pour tout $\theta > 0$,

$$\mathbb{E}(e^{-\theta(X_1 + \dots + X_Y)}) = \mathbb{E}(\mathbb{E}(e^{-\theta(X_1 + \dots + X_Y)} | Y)) = \mathbb{E}(\mathbb{E}(e^{-\theta X_1})^Y) = \sum_{n=1}^{\infty} p(1-p)^{n-1} \left(\frac{\lambda}{\lambda + \theta}\right)^n = \frac{p\lambda}{p\lambda + \theta}.$$

■

Plus généralement, la même méthode permet d'établir que si Y suit la loi binomiale négative $\text{Geo}_{\mathbb{N}^*}(p)^{*m}$, alors $X_1 + \dots + X_Y$ suit la loi Gamma $\text{Exp}(p\lambda)^{*m}$. Il y a stabilité de la famille des lois Gamma par mélange géométrique. Cela correspond au phénomène observé dans l'exemple 6.8.1 : le processus de Poisson composé d'intensité λ obtenu à partir d'un processus de Bernoulli de paramètre p est un processus de Poisson d'intensité λp .

6.9 Chaînes de Markov à temps continu : construction

Nous allons à présent dépasser la temporisation d'une chaîne de Markov à temps discret par un processus de Poisson indépendant, en considérant une temporisation qui dépend de la position spatiale de la chaîne.

Soit $(Z_n)_{n \in \mathbb{N}}$ une chaîne de Markov sur E au plus dénombrable de noyau de transition \mathbf{P} . On munit chaque état $z \in E$ d'un nombre réel déterministe $\lambda(z) > 0$ et d'une suite $(E_n^{(z)})_{n \geq 1}$ de v.a. i.i.d. de loi exponentielle $\text{Exp}(\lambda(z))$. On suppose de plus que les v.a. $(E_n^{(z)})_{z \in E, n \geq 1}$ sont indépendantes, et indépendantes de $(Z_n)_{n \in \mathbb{N}}$. Posons à présent

$$T_0 = 0, \quad \text{et} \quad T_n = E_1^{(Z_0)} + \dots + E_n^{(Z_{n-1})}, \quad n \geq 1.$$

Cette construction assure que conditionnellement à $\{Z_0 = z_0, \dots, Z_n = z_n\}$, la v.a. T_{n+1} est bien une somme de v.a. indépendantes de lois $\text{Exp}(\lambda(z_0)), \dots, \text{Exp}(\lambda(z_n))$. Presque sûrement, la suite $(T_n)_{n \in \mathbb{N}}$ est strictement croissante. Par conséquent, pour tout réel t vérifiant $0 \leq t < \sup\{T_m : m \in \mathbb{N}\}$, il existe presque sûrement un unique entier aléatoire N_t vérifiant $T_{N_t} \leq t < T_{N_t+1}$. Cela revient à dire que $N_t = n$ sur $\{T_n \leq t < T_{n+1}\}$, ou encore que

$$N_t = \sup\{m \in \mathbb{N} : T_m \leq t\}.$$

Par définition, le temps d'explosion est défini par

$$T_\infty = \sup\{T_m : m \in \mathbb{N}\} = \lim_{m \rightarrow \infty} T_m = \sum_{n=1}^{\infty} E_n^{(Z_{n-1})}.$$

C'est une v.a. qui prend ses valeurs dans $\mathbb{R}_+ \cup \{\infty\}$. Pour tout réel $t \in [0, T_\infty)$, on définit la v.a. X_t à valeurs dans E par

$$X_t = Z_{N_t}.$$

Cette variable est définie pour tout réel $t \geq 0$ ssi il n'y a pas explosion p.s. : $\mathbb{P}(T_\infty = \infty) = 1$.

Le processus $(X_t)_{t \in [0, T_\infty)}$ est une chaîne de Markov à temps continu sur l'espace discret E , associée à λ et \mathbf{P} .

Exemple 6.9.1. Cas particulier de la poissonisation.

Lorsque l'intensité λ est constante, alors $(N_t)_{t \in \mathbb{R}_+}$ est un processus de Poisson simple, $\mathbb{P}(T_\infty = \infty) = 1$, et $(X_t)_{t \in \mathbb{R}_+} = (Z_{N_t})_{t \in \mathbb{R}_+}$ est la poissonisation de la chaîne de Markov $(Z_n)_{n \in \mathbb{N}}$ de noyau de transition \mathbf{P} . D'autre part, lorsque $E = \mathbb{N}$, $\text{Loi}(Z_0) = \delta_0$, et $\mathbf{P}(z, \cdot) = \delta_{z+1}$ pour tout $z \in E$, alors $(X_t)_{t \in \mathbb{R}_+}$ est identique à $(N_t)_{t \in \mathbb{R}_+}$.

6.10 Chaînes de Markov à temps continu : non-explosion

Considérons la chaîne de Markov à temps continu $(X_t)_{t \in [0, T_\infty)}$ sur E construite comme précédemment à partir de λ et \mathbf{P} , de même que la suite $(T_n)_{n \in \mathbb{N}}$ associée, de sorte que $X_{T_n} = Z_n$ pour tout n , et en particulier $X_0 = Z_{T_0} = Z_0$. La trajectoire $t \mapsto X_t$ pour $0 \leq t < T_\infty$ est constante par morceaux, et ses temps de saut sont inclus dans la suite de v.a. $(T_n)_{n \geq 1}$ (ce sont tous des temps de saut lorsque \mathbf{P} est à diagonale nulle). Pour tout $z \in E$ et tout n ,

$$\text{Loi}(T_{n+1} - T_n | X_{T_n} = z) = \text{Exp}(\lambda(z)) \quad \text{et} \quad \text{Loi}(X_{T_{n+1}} | X_{T_n} = z) = \mathbf{P}(z, \cdot).$$

Les trajectoires $t \mapsto X_t$ ne sont définies que pour $t < T_\infty$. Le théorème 6.10.1 ci-dessous montre que lorsque \mathbf{P} est récurrent ou lorsque λ est bornée, alors $\mathbb{P}(T_\infty = \infty) = 1$ et les trajectoires sont donc définies pour tout $t \in \mathbb{R}_+$.

Théorème 6.10.1. Critère de non-explosion.

Le temps d'explosion T_∞ vérifie

$$\mathbb{P}(T_\infty = \infty) = \mathbb{P}\left(\sum_{n=0}^{\infty} \frac{1}{\lambda(Z_n)} = \infty\right).$$

En particulier, si λ est bornée ou si Loi(Z_0) ne charge que des états récurrents de \mathbf{P} , alors la chaîne de Markov à temps continu $(X_t)_{t \in [0, T_\infty)}$ n'explose pas : $T_\infty = \infty$ p.s., et X_t est bien définie pour tout $t \in \mathbb{R}_+$.

Pour une chaîne de Markov poissonisée, λ est constante donc bornée, donc $\mathbb{P}(T_\infty = \infty) = 1$, même si \mathbf{P} ne possède pas d'état récurrent. C'est le cas par exemple pour le processus de Poisson : $E = \mathbb{N}$ et $\mathbf{P}(z, \cdot) = \delta_{z+1}$.

Il n'y a jamais explosion si E est fini car alors λ est bornée.

Démonstration. Pour tout entier n , toute suite z_0, \dots, z_n dans E , et tout réel $\theta > 0$,

$$\mathbb{E}(e^{-\theta T_{n+1}} \mid Z_0 = z_0, \dots, Z_n = z_n) = \mathbb{E}(e^{-\theta(U_0 + \dots + U_n)}),$$

où U_0, \dots, U_n sont indépendantes avec, pour tout $0 \leq k \leq n$, $U_k \sim \text{Exp}(\lambda(z_k))$. Il en découle que pour tout réel $\theta > 0$,

$$\mathbb{E}(e^{-\theta T_{n+1}} \mid Z_0, \dots, Z_n) = \prod_{k=0}^n \left(1 + \frac{\theta}{\lambda(Z_k)}\right)^{-1}.$$

Cela donne la transformée de Laplace de T_{n+1} en prenant l'espérance, puis celle de T_∞ par convergence monotone :

$$\mathbb{E}(e^{-\theta T_\infty}) = \mathbb{E}\left(\prod_{k=0}^{\infty} \left(1 + \frac{\theta}{\lambda(Z_k)}\right)^{-1}\right).$$

Or pour toute suite $(a_k)_{k \in \mathbb{N}}$ dans \mathbb{R}_+ , $\prod_{k=0}^{\infty} (1 + \theta a_k)$ converge ssi $\sum_{k=0}^{\infty} \theta a_k$ converge. Comme $\theta > 0$, il en découle que

$$\mathbb{E}\left(e^{-\theta T_\infty} \mathbb{1}_{\{T_\infty < \infty\}}\right) = \mathbb{E}\left(\prod_{k=0}^{\infty} \left(1 + \frac{\theta}{\lambda(Z_k)}\right)^{-1} \mathbb{1}_{\{\sum_{k=0}^{\infty} (\lambda(Z_k))^{-1} < \infty\}}\right).$$

Les quantités $e^{-\theta T_\infty}$ et $\prod_{k=0}^{\infty} (1 + \theta / \lambda(Z_k))^{-1}$ sont monotones en θ . Lorsque $\theta \searrow 0$, par convergence monotone, nous obtenons $\mathbb{P}(T_\infty < \infty) = \mathbb{P}(\sum_{n=0}^{\infty} 1 / \lambda(Z_n) < \infty)$ comme attendu. En particulier, si λ est bornée, alors $\sum_{n=0}^{\infty} 1 / \lambda(Z_n)$ diverge p.s. D'autre part $\sum_{n=0}^{\infty} 1 / \lambda(Z_n)$ diverge sur l'événement $A = \{\sum_{n=0}^{\infty} \mathbb{1}_{\{Z_n = Z_0\}} = \infty\}$, donc si Loi(Z_0) ne charge que des états récurrents de \mathbf{P} , alors $\mathbb{P}(A) = 1$ et $\mathbb{P}(\sum_{n=0}^{\infty} 1 / \lambda(Z_k) = \infty) = \mathbb{P}(\sum_{n=0}^{\infty} 1 / \lambda(Z_k) = \infty \mid A) \mathbb{P}(A) = 1$. ■

Remarque 6.10.2. Non explosion et loi du zéro-un.

La loi du zéro-un pour les horloges exponentielles cumulées du lemme 6.3.1 est une conséquence directe du critère de non-explosion du lemme 6.10.1 appliqué au cas $E = \mathbb{N}$, $\mathbf{P}(z, \cdot) = \delta_{z+1}$, et Loi(Z_0) = δ_0 . Dans ce cas $(Z_n)_{n \in \mathbb{N}}$ est déterministe et $Z_n = n$ pour tout n . Par conséquent, $\sum_{n=0}^{\infty} 1 / \lambda(Z_n) = \sum_{n=0}^{\infty} 1 / \lambda(n)$ est déterministe. D'autre part, $T_\infty = \sum_{n=1}^{\infty} E_n^{(n-1)}$ et les v.a. $(E_n^{(n-1)})_{n \geq 1}$ sont indépendantes, et $E_n^{(n-1)} \sim \text{Exp}(\lambda(n-1))$.

6.11 Chaînes de Markov à temps continu : générateur

Considérons la chaîne de Markov à temps continu $(X_t)_{t \in [0, T_\infty)}$ sur E construite comme précédemment à partir de λ et \mathbf{P} . Supposons qu'il n'y a pas explosion : $\mathbb{P}(T_\infty = \infty) = 1$, de sorte que X_t est bien défini pour tout réel $t \in \mathbb{R}_+$. Le générateur ou Q-matrice $\mathbf{Q} : E \times E \rightarrow \mathbb{R}$ de cette chaîne à temps continu est défini pour tout $(x, y) \in E \times E$ par

$$\mathbf{Q}(x, y) = \begin{cases} \lambda(x) \mathbf{P}(x, y) & \text{si } x \neq y, \\ \lambda(x) (\mathbf{P}(x, x) - 1) & \text{si } x = y. \end{cases}$$

En utilisant le produit matriciel, on obtient $\mathbf{Q} = \mathbf{D}_\lambda (\mathbf{P} - \mathbf{I})$, où $\mathbf{D}_\lambda : E \times E \rightarrow \mathbb{R}_+$ est la matrice diagonale définie par

$$\mathbf{D}_\lambda(x, y) = \lambda(x) \mathbb{1}_{x=y}.$$

Notons que $\mathbf{P} - \mathbf{I}$ est le générateur de la chaîne à temps discret de noyau de transition \mathbf{P} .

Les propriétés caractéristiques d'un générateur sont les suivantes :

- (i) $\mathbf{Q}(x, x) \leq 0$ pour tout x dans E .
- (ii) $\mathbf{Q}(x, y) \geq 0$ pour tous x et y dans E avec $x \neq y$.

(iii) $\sum_{y \in E} \mathbf{Q}(x, y) = 0$ pour tout x dans E .

Le générateur fournit un nouvel algorithme de simulation des trajectoires en distinguant temps de saut et sauts. Soit $(S_n)_{n \in \mathbb{N}}$ la suite des temps de saut de $(X_t)_{t \in \mathbb{R}_+}$ à valeurs dans $\mathbb{R}_+ \cup \{\infty\}$ définie par

$$S_0 = 0 \quad \text{et} \quad S_{n+1} = \inf\{t > S_n : X_t \neq X_{S_n}\},$$

avec $\inf \emptyset := \infty$. La suite $(S_n)_{n \in \mathbb{N}}$ est strictement croissante, puis éventuellement constante et égale à ∞ .

Théorème 6.11.1. Algorithme de simulation des trajectoires basé sur le générateur.

Les propriétés suivantes ont lieu :

- (i) $S_n \nearrow T_\infty$ p.s.
- (ii) Pour tous $x \in E$ et $n \in \mathbb{N}$, conditionnellement à $\{S_n < \infty, X_{S_n} = x\}$,
 - (a) si $\mathbf{Q}(x, x) = 0$, alors $S_{n+1} = \infty$, et on dit que x est absorbant.
 - (b) si $\mathbf{Q}(x, x) \neq 0$, alors $S_{n+1} < \infty$ p.s., $S_{n+1} - S_n$ et $X_{S_{n+1}}$ sont indépendantes, et pour tout $y \in E, y \neq x$,

$$\text{Loi}(S_{n+1} - S_n \mid X_{S_n} = x) = \text{Exp}(-\mathbf{Q}(x, x)) \quad \text{et} \quad \mathbb{P}(X_{S_{n+1}} = y \mid X_{S_n} = x) = -\frac{\mathbf{Q}(x, y)}{\mathbf{Q}(x, x)}.$$

En particulier la loi (des trajectoires) de $(X_t)_{t \in \mathbb{R}_+}$ ne dépend de (λ, \mathbf{P}) qu'à travers \mathbf{Q} .

Les quantités $\mathbf{Q}(x, y)$ sont les taux de transition de la chaîne de Markov à temps continu $(X_t)_{t \in \mathbb{R}_+}$. Tout générateur permet de construire une chaîne de Markov à temps continu.

Exemple 6.11.2. Processus de Poisson composé.

Si λ est constante et égale à λ , alors le processus de comptage $(N_t)_{t \in \mathbb{R}_+}$ de la suite $(T_n)_{n \in \mathbb{N}}$ est un processus de Poisson simple issu de 0 et d'intensité λ , indépendant de $(Z_n)_{n \in \mathbb{N}}$, et le générateur de $(X_t)_{t \in \mathbb{R}_+}$ est

$$\mathbf{Q}(x, y) = \lambda(\mathbf{P} - \mathbf{I})(x, y) = \lambda \begin{cases} \mathbf{P}(x, y) & \text{si } x \neq y, \\ \mathbf{P}(x, x) - 1 & \text{si } x = y. \end{cases}$$

C'est le cas par exemple pour le processus de Poisson composé de l'exemple 6.8.1.

La simulation des trajectoires de $(X_t)_{t \in [0, T_\infty]}$ peut être menée comme pour les chaînes à temps discret, soit en utilisant λ et \mathbf{P} , soit en utilisant le générateur et la décomposition en sauts fournie par le théorème 6.11.1 (ii)(b) (approche événementielle). Lorsque le temps d'explosion T_∞ n'est pas infini p.s. alors ni la loi initiale ni le générateur ne précisent pas la manière dont le processus redémarre après explosion. Il existe une infinité de manières de procéder, et le prolongement le plus simple consiste à ajouter un point ∞ à l'espace d'état E . Ce point est absorbant pour le processus prolongé, et les trajectoires explosives sont choisies constantes et égales à ∞ après explosion. Ce prolongement particulier du processus après explosion est appelé prolongement minimal.

Preuve du théorème 6.11.1. Soient $(T_n)_{n \in \mathbb{N}}$ et $(Z_n)_{n \in \mathbb{N}}$ comme dans la construction de $(X_t)_{t \in \mathbb{R}_+}$ précédente. Comme $T_n \leq S_n$ pour tout entier n , il en découle que $S_n \nearrow T_\infty$ p.s. Soit $(U_n)_{n \in \mathbb{N}}$ les temps de saut de $(Z_n)_{n \in \mathbb{N}}$, définis par

$$U_0 = 0 \quad \text{et} \quad U_{n+1} = \inf\{m > U_n : Z_m \neq Z_{U_n}\}.$$

D'après la section 5.5, sachant $\{Z_{U_n} = x\}$, les variables aléatoires $U_{n+1} - U_n$ et $Z_{U_{n+1}}$ sont indépendantes et

$$U_{n+1} - U_n \sim \text{Geo}_{\mathbb{N}^*}(1 - \mathbf{P}(x, x)) \quad \text{et} \quad Z_{U_{n+1}} \sim \sum_{y \neq x} \frac{\mathbf{P}(x, y)}{1 - \mathbf{P}(x, x)} \delta_y.$$

Comme $X_{S_n} = Z_{U_n}$, il en découle tout d'abord que sachant $\{X_{S_n} = x\}$, les variables aléatoires $S_{n+1} - S_n$ et $X_{S_{n+1}}$ sont indépendantes. En fait $X_t = X_{S_n} = Z_{U_n}$ pour tout $t \in [S_n, S_{n+1})$.

Ensuite, comme $S_n = T_{U_n}$, nous avons $S_{n+1} - S_n = T_{U_{n+1}} - T_{U_n}$, et donc, pour tout entier $m \geq 1$,

$$\text{Loi}(S_{n+1} - S_n \mid X_{S_n} = x, U_{n+1} - U_n = m) = \text{Exp}(\lambda(x))^{*m} = \text{Gamma}(m, \lambda(x)).$$

Comme la loi de $U_{n+1} - U_n$ sachant $\{Z_{U_n} = x\} = \{X_{S_n} = x\}$ est géométrique $\text{Geo}_{\mathbb{N}^*}(1 - \mathbf{P}(x, x))$, le lemme 6.8.4 donne

$$\text{Loi}(S_{n+1} - S_n \mid X_{S_n} = x) = \text{Exp}(\lambda(x)(1 - \mathbf{P}(x, x))) = \text{Exp}(-\mathbf{Q}(x, x)).$$

Enfin, pour tout x tel que $-\mathbf{Q}(x, x) = \boldsymbol{\lambda}(x)(1 - \mathbf{P}(x, x)) > 0$, on a pour tous $y \in E$, $y \neq x$, et $n \in \mathbb{N}$,

$$\mathbb{P}(X_{S_{n+1}} = y \mid X_{S_n} = x) = \mathbb{P}(Z_{U_{n+1}} = y \mid Z_{U_n} = x) = \frac{\mathbf{P}(x, y)}{1 - \mathbf{P}(x, x)} = \frac{\boldsymbol{\lambda}(x)\mathbf{P}(x, y)}{\boldsymbol{\lambda}(x)(1 - \mathbf{P}(x, x))} = -\frac{\mathbf{Q}(x, y)}{\mathbf{Q}(x, x)}.$$

Si $\mathbf{Q}(x, x) = 0$, alors x est absorbant pour $(X_t)_{t \in \mathbb{R}_+}$, et $S_{n+1} = \infty$ sur l'événement $\{X_{S_n} = x\}$. ■

Remarque 6.11.3. Représentation canonique, chaîne incluse et chaîne harmonisée.

L'écriture sous la forme $\mathbf{Q} = \mathbf{D}_\lambda(\mathbf{P} - \mathbf{I})$ n'est pas unique. Plusieurs choix sont possibles. Il est commode de prendre pour $\boldsymbol{\lambda}$ l'opposé en signe de la diagonale de \mathbf{Q} , et pour \mathbf{P} le noyau donné par

$$\mathbf{P}(x, y) = \begin{cases} 0 & \text{si } x = y \text{ et } \mathbf{Q}(x, x) < 0, \\ 1 & \text{si } x = y \text{ et } \mathbf{Q}(x, x) = 0, \\ -\frac{\mathbf{Q}(x, y)}{\mathbf{Q}(x, x)} & \text{si } x \neq y \text{ et } \mathbf{Q}(x, x) < 0, \\ 0 & \text{si } x \neq y \text{ et } \mathbf{Q}(x, x) = 0. \end{cases}$$

Ce noyau \mathbf{P} est appelée noyau inclus dans \mathbf{Q} . La chaîne incluse saute à chaque étape vers un état différent, sauf bien sûr si l'état courant est absorbant. La valeur de $\boldsymbol{\lambda}(z)$ pour les états $z \in E$ tels que $\mathbf{Q}(z, z) = 0$ n'a pas d'impact sur la loi des trajectoires de $(X_t)_{t \in [0, T_\infty)}$. La diagonale de \mathbf{P} est constituée de 0 et de 1, et ces derniers correspondent aux états absorbants. Avec ce choix de $(\boldsymbol{\lambda}, \mathbf{P})$, la suite des temps exponentiels $(T_n)_{n \geq 1}$ liée à $\boldsymbol{\lambda}$ dans la construction de $(X_t)_{t \in [0, T_\infty)}$ coïncide avec la suite des temps de saut $(S_n)_{n \geq 1}$ de $(X_t)_{t \in [0, T_\infty)}$. Une autre paramétrisation de \mathbf{Q} sous la forme $\mathbf{Q} = \mathbf{D}_\lambda(\mathbf{P} - \mathbf{I})$ s'obtient en faisant en sorte que $\boldsymbol{\lambda}$ soit constant et égal à $\sup_x -\mathbf{Q}(x, x)$. Cela n'est possible que si ce supremum est fini c'est-à-dire que $\boldsymbol{\lambda}$ est bornée. Le noyau \mathbf{P} associé est alors appelé noyau harmonisé, et la chaîne associée est appelée chaîne harmonisée.

Remarque 6.11.4. Horloges exponentielles en compétition.

Pour tout $x \in E$, la ligne $(\mathbf{Q}(x, y))_{y \in E}$ a une interprétation en terme d'horloges exponentielles en compétition.

- Si $\mathbf{Q}(x, x) = 0$ alors x est absorbant et le processus devient constant.
- Si $\mathbf{Q}(x, x) \neq 0$ alors x n'est pas absorbant et le processus va sauter en un temps fini vers un état $y \neq x$ à déterminer. Associons à chaque $y \neq x$ une horloge qui sonne au bout d'un temps aléatoire de loi exponentielle de paramètre $\mathbf{Q}(x, y)$. Les horloges des différents états $y \neq x$ sont choisies indépendantes. Le saut se fera au bout d'un temps T vers l'état Y dont l'horloge aura sonné en premier. En effet, le lemme 6.11.5 entraîne que cet état aléatoire Y est presque sûrement bien défini, qu'il est indépendant de T , que $\text{Loi}(T) = \text{Exp}(-\mathbf{Q}(x, x))$, et que $\mathbb{P}(Y = y) = -\mathbf{Q}(x, y)/\mathbf{Q}(x, x)$ pour tout $y \neq x$.

L'interprétation en terme d'horloges exponentielles en compétition de la remarque 6.11.4 est commode en temps continu. Le lemme 6.11.5 n'a pas lieu lorsque les lois exponentielles sont remplacées par des lois géométriques, la probabilité d'égalité de deux v.a. géométriques indépendantes n'est pas nulle.

Lemme 6.11.5. Horloges exponentielles en compétition.

Si $(V_z)_{z \in E}$ est une famille au plus dénombrable de variables aléatoires exponentielles indépendantes telle que $\sum_{z \in E} \lambda_z < \infty$ où $V_z \sim \text{Exp}(\lambda_z)$, alors $V = \inf_{z \in E} V_z$ suit la loi exponentielle $\text{Exp}(\sum_{z \in E} \lambda_z)$. De plus, presque sûrement, l'infimum V est atteint pour un unique élément aléatoire Z de E , indépendant de V , et

$$\mathbb{P}(Z = z) = \mathbb{P}(V = V_z) = \frac{\lambda_z}{\sum_{z' \in E} \lambda_{z'}}.$$

Démonstration. Soit Z la v.a. sur $E \cup \{\infty\}$ valant z si $V_{z'} > V_z$ pour tout $z' \neq z$, et ∞ sinon. Déterminons la loi jointe de (V, Z) . Pour tout $z \in E$ et tout réel $t \geq 0$, comme $\{V \geq t, Z = z\} = \{V_z \geq t, V_{z'} > V_z \text{ pour tout } z' \neq z\}$, nous avons

$$\mathbb{P}(V \geq t \text{ et } Z = z) = \int_t^\infty \lambda_z e^{-s\lambda_z} \prod_{z' \neq z} \mathbb{P}(V_{z'} > s) ds = e^{-t \sum_{z' \in E} \lambda_{z'}} \frac{\lambda_z}{\sum_{z' \in E} \lambda_{z'}}.$$

Cela donne la loi de V et la loi de Z , en particulier le fait que $\mathbb{P}(Z \in E) = 1$, et l'indépendance de V et Z . ■

6.12 Chaînes de Markov à temps continu : semi-groupe et équations de Kolmogorov

Une chaîne de Markov à temps continu d'espace d'état au plus dénombrable E est un processus $(X_t)_{t \in [0, T_\infty)}$ construit comme précédemment à partir d'un champ de temporisation λ et d'un noyau de transition \mathbf{P} . Par le théorème 6.11.1, le générateur $\mathbf{Q} = \mathbf{D}_\lambda(\mathbf{P} - \mathbf{I})$ caractérise la loi des trajectoires de $(X_t)_{t \in [0, T_\infty)}$. Quitte à considérer le prolongement minimal, nous pouvons supposer qu'il n'y a pas explosion : $T_\infty = \infty$ p.s., et que le processus $(X_t)_{t \in \mathbb{R}_+}$ est bien défini sur $E \cup \{\infty\}$. Pour tout réel $t \geq 0$, on considère le noyau de transition

$$(x, y) \in E \times E \mapsto \mathbf{P}_t(x, y) := \mathbb{P}(X_t = y \mid X_0 = x) \in [0, 1].$$

En particulier $\mathbf{P}_0 = \mathbf{I}$. Pour toute $f : E \rightarrow \mathbb{R}$ bornée ou positive et tout $t \geq 0$, en notant $\nu = \text{Loi}(X_0)$ la loi initiale,

$$\mathbb{E}(f(X_t)) = \sum_{x \in E} \nu(x) \mathbf{P}_t(x, y) f(y) = \nu \mathbf{P}_t f.$$

En d'autres termes $X_t \sim \nu \mathbf{P}_t$. Il est possible d'établir les propriétés suivantes :

(i) Propriété de Markov. Pour tous $x, y \in E$ et $s, t \in \mathbb{R}_+$,

$$\mathbf{P}_{t+s}(x, y) = \sum_{z \in E} \mathbf{P}_s(x, z) \mathbf{P}_t(z, y) = \sum_{z \in E} \mathbf{P}_t(x, z) \mathbf{P}_s(z, y).$$

Matriciellement $\mathbf{P}_{t+s} = \mathbf{P}_s \mathbf{P}_t = \mathbf{P}_t \mathbf{P}_s$, et $\mathbf{P}_0 = \mathbf{I}$. On dit que $(\mathbf{P}_t)_{t \geq 0}$ est un semi-groupe de Markov.

(ii) Équations de Kolmogorov rétrograde et progressive. Pour tous $x, y \in E$, la fonction $t \in \mathbb{R}_+ \mapsto \mathbf{P}_t(x, y)$ est dérivable (notamment à droite en 0) et solution positive minimale des EDO

$$\frac{d}{dt} \mathbf{P}_t(x, y) = (\mathbf{Q} \mathbf{P}_t)(x, y) \quad \text{et} \quad \frac{d}{dt} \mathbf{P}_t(x, y) = (\mathbf{P}_t \mathbf{Q})(x, y),$$

En particulier $\mathbf{Q}(x, y) = \partial_{t=0^+} \mathbf{P}_t(x, y) := \lim_{t \rightarrow 0^+} \frac{\mathbf{P}_t(x, y) - \mathbf{P}_0(x, y)}{t}$.

Si E est fini, alors $\mathbf{P}_t = e^{t\mathbf{Q}}$ (exponentielle de matrice).

Exemple 6.12.1. Processus de Poisson simple.

Pour le processus de Poisson simple d'intensité $\lambda > 0$, on a $E = \mathbb{N}$ et pour tous x et y dans \mathbb{N}

$$\mathbf{Q}(x, y) = \begin{cases} -\lambda & \text{si } y = x \\ +\lambda & \text{si } y = x + 1, \\ 0 & \text{sinon} \end{cases} \quad \text{et} \quad \mathbf{P}_t(x, y) = e^{-t\lambda} \begin{cases} \frac{(t\lambda)^{y-x}}{(y-x)!} & \text{si } y \geq x \\ 0 & \text{sinon} \end{cases}.$$

Ce processus n'explose pas car la diagonale de \mathbf{Q} est constante, et donc bornée. Les temps inter-sauts sont des lois exponentielles i.i.d. de paramètre λ . On peut écrire $\mathbf{Q} = \lambda(\mathbf{P} - \mathbf{I})$ où $\mathbf{P}(x, x+1) = 1$. On reconnaît la loi de Poisson translatée sur les lignes de \mathbf{P}_t puisque $\text{Loi}(X_t \mid X_0 = x) = \delta_x * \text{Poi}(\lambda t)$ pour tous $x \in E$ et $t \geq 0$.

6.13 Chaînes de Markov à temps continu : synthèse

Trois descriptions de la loi d'une chaîne de Markov à temps continu $(X_t)_{t \in \mathbb{R}_+}$, liées par $\partial_{t=0^+} \mathbf{P}_t = \mathbf{Q} = \mathbf{D}_\lambda(\mathbf{P} - \mathbf{I})$:

— par la donnée de (λ, \mathbf{P}) .

— par la donnée de \mathbf{Q} .

— par la donnée de $(\mathbf{P}_t)_{t \geq 0}$.

Les deux premières descriptions fournissent un algorithme de simulation des trajectoires de la chaîne de Markov à temps continu, qui sont par construction constantes par morceaux, continues à droite avec limites à gauche.

L'étude de la chaîne à temps continu $(X_t)_{t \in \mathbb{R}_+}$ se ramène à celle de la chaîne à temps discret sous-jacente de noyau \mathbf{P} . La récurrence/transience des états est héritée de celle de \mathbf{P} . Une mesure μ est invariante lorsque $\mu \mathbf{P}_t = \mu$ pour tout $t \in \mathbb{R}_+$, ce qui est équivalent à $\mu \mathbf{Q} = 0$, et les théorèmes de convergence peuvent être développés en se ramenant à \mathbf{P} . Une étude détaillée se trouve dans [53]. La loi des grands nombres ou théorème ergodique et le théorème limite central prennent la forme suivante :

$$\frac{1}{t} \int_0^t f(X_s) ds \xrightarrow[t \rightarrow \infty]{\text{p.s.}} \int f d\mu \quad \text{et} \quad \sqrt{t} \left(\frac{1}{t} \int_0^t f(X_s) ds - \int f d\mu \right) \xrightarrow[t \rightarrow \infty]{\text{loi}} \mathcal{N}(0, \sigma_f^2).$$

Le tableau 6.2 dresse une comparaison analogique entre chaînes à temps discret et continu, en particulier :

5. À ne pas confondre avec l'invariance pour la chaîne incluse lorsque λ n'est pas constante.

- la notion de période est spécifique au cas discret⁶ (mais subsiste pour la chaîne incluse).
- la notion de temps d'explosion est spécifique au cas continu (et aux espaces d'états infinis).
- l'interprétation en terme d'horloges est commode pour le cas continu.
- la notion de générateur est commune au cas discret et continu.

Chaîne à temps discret	Chaîne à temps continu
Espace d'état au plus dénombrable E et loi initiale ν	Processus $(X_t)_{t \geq 0}$ avec $X_0 \sim \nu$
Suite $(X_n)_{n \in \mathbb{N}}$ avec $X_0 \sim \nu$	$\mathbf{P}_t(x, y) = \mathbb{P}(X_t = y \mid X_0 = x)$
$\mathbf{P}^n(x, y) = \mathbb{P}(X_n = y \mid X_0 = x)$	$\mathbf{P}_t(x, y) = \mathbb{P}(X_t = y \mid X_0 = x)$
Rien d'équivalent	Temps d'explosion $T_\infty \in \mathbb{R}_+ \cup \{\infty\}$
Semi-groupe de Markov $(\mathbf{P}^n)_{n \in \mathbb{N}}$	Semi-groupe de Markov $(\mathbf{P}_t)_{t \in \mathbb{R}_+}$
Noyau de transition \mathbf{P}	Rien d'équivalent
Générateur $\mathbf{L} = \mathbf{P} - \mathbf{I}$	Générateur $\mathbf{L} = \lim_{t \rightarrow 0^+} \frac{1}{t} (\mathbf{P}_t - \mathbf{I})$
Rien d'équivalent	Écriture temporisée $\mathbf{L} = \mathbf{D}_\lambda (\mathbf{P} - \mathbf{I})$
$\mathbb{E}(f(X_n)) = \nu \mathbf{P}^n f$	$\mathbb{E}(f(X_t)) = \nu \mathbf{P}_t f$
Notion de période	Rien d'équivalent
Sauts	Sauts
$S_{n+1} = \inf\{n > S_n : X_n \neq X_{S_n}\}$	$S_{n+1} = \inf\{t > S_n : X_t \neq X_{S_n}\}$
Loi $(S_{n+1} - S_n \mid X_{S_n} = x) = \text{Geo}_{\mathbb{N}^*}(-\mathbf{L}(x, x))$	Loi $(S_{n+1} - S_n \mid X_{S_n} = x) = \text{Exp}(-\mathbf{L}(x, x))$
Chaîne incluse $\mathbb{P}(X_{S_{n+1}} = y \mid X_{S_n} = x) = -\frac{\mathbf{L}(x, y)}{\mathbf{L}(x, x)}$	
x absorbant ssi $\mathbf{L}(x, x) = 0$	
μ invariante ssi $\mu \mathbf{L} = 0$	

TABLE 6.2 – Tableau analogue entre chaînes de Markov à temps discret et à temps continu. Le générateur de la chaîne à temps continu est noté ici \mathbf{L} plutôt que \mathbf{Q} pour renforcer l'analogie. Il ne doit pas être confondu avec le générateur $\mathbf{P} - \mathbf{I}$ de la chaîne à temps discret associée à une temporisation. Les chaînes de Markov à temps continu et espace discret sont des processus de sauts markoviens, au sens où l'aléa des trajectoires ne provient que des sauts (et de la condition initiale). De tels processus peuvent être définis sur des espaces quelconques, discrets ou non. Le mouvement brownien, qui fait l'objet du chapitre suivant, est markovien, mais n'est pas un processus à sauts.

6.14 Files d'attentes et quelques autres exemples

Exemple de processus explosif

Considérons le générateur \mathbf{Q} sur $E = \mathbb{N}$ défini par $\mathbf{Q}(x, x+1) = -\mathbf{Q}(x, x) = a_x$ pour tout $x \in E$, où $(a_x)_{x \in \mathbb{N}}$ est une suite de réels positifs telle que $\sum_{x=0}^{\infty} 1/a_x < \infty$. La représentation canonique $\mathbf{Q} = \mathbf{D}_\lambda (\mathbf{P} - \mathbf{I})$ de \mathbf{Q} est donnée par $\lambda(x) = -a_x$ et $\mathbf{P}(x, x+1) = 1$ pour tout $x \in E$. Par conséquent, toute chaîne de Markov $(Z_n)_{n \in \mathbb{N}}$ de noyau \mathbf{P} vérifie $Z_n = Z_0 + n$ pour tout $n \in \mathbb{N}$, et en particulier $\sum_{n=0}^{\infty} \frac{1}{\lambda(Z_n)} = \sum_{z=Z_0}^{\infty} \frac{1}{a_z} < \infty$. Ainsi $\mathbb{P}(T_\infty = \infty) = 0$. Explosion p.s.!

Files d'attente à temps continu

Considérons des atomes radioactifs dont les arrivées dans un détecteur sont espacées par des durées aléatoires i.i.d. de loi exponentielle $\text{Exp}(\lambda)$. Après arrivée, chaque atome survit pendant un temps aléatoire de loi $\text{Exp}(\mu)$, puis se désintègre. Les atomes sont tous indépendants. On note X_t le nombre d'atomes présents à l'instant t . Plaçons-nous sur l'événement $\{X_0 = n\}$, $n \geq 1$. Soient V_1, \dots, V_n les temps de survie des n atomes initiaux, A le temps d'arrivée d'un nouvel atome, et $T = \min(A, V_1, \dots, V_n) = \min(A, W)$ où $W = \min(V_1, \dots, V_n)$. Notons que $X_t = n$ pour tout $t \in [0, T)$. De plus, $X_T = n - 1$ sur l'événement $\{T = W\}$, et $X_T = n + 1$ sur l'événement $\{T = A\}$. Or A, V_1, \dots, V_n sont indépendantes. De plus, $A \sim \text{Exp}(\lambda)$ et $V_k \sim \text{Exp}(\mu)$ pour tout $1 \leq k \leq n$. Ainsi

$$T \sim \text{Exp}(\lambda + n\mu), \quad \mathbb{P}(T = A) = \frac{\lambda}{\lambda + n\mu}, \quad \text{et} \quad \mathbb{P}(T = W) = \frac{n\mu}{\lambda + n\mu}.$$

Si $n = 0$ alors l'étude est simplifiée car seul A est présent. Cette construction peut être renouvelée de manière indépendante après le temps T , grâce à l'absence de mémoire des lois exponentielles. Ce modèle correspond à une chaîne de Markov à temps continu sur $E = \mathbb{N}$ avec temporisation donnée par $\lambda(n) = \lambda + n\mu$ et noyau de transition

6. On pourrait penser que les phénomènes de période apparaissent également lors de la simulation de chaînes à temps continu, puisque l'informatique est le règne des structures discrètes. Cependant, les méthodes de simulation des chaînes à temps continu font appel à des simulations approximatives de variables exponentielles via des nombres réels dyadiques. Ces approximations brisent la spécificité des modèles et donnent naissance à des processus à temps discret qui ont toutes les chances d'être apériodiques.

donné par $\mathbf{P}(n, \cdot) = \frac{\lambda}{\lambda+n\mu} \delta_{n+1} + \frac{n\mu}{\lambda+n\mu} \delta_{n-1}$. Le générateur $\mathbf{Q} = \mathbf{D}_\lambda(\mathbf{P} - \mathbf{I})$ est donné par

$$\mathbf{Q}(n, m) = \begin{cases} n\mu & \text{si } m = n - 1 \\ \lambda & \text{si } m = n + 1 \\ -(\lambda + n\mu) & \text{si } n = m \\ 0 & \text{sinon} \end{cases}.$$

Le processus de Poisson simple correspond au cas où $\mu = 0$. Le noyau de vie ou de mort \mathbf{P} est irréductible et sa période vaut 2. Sa récurrence découle de la condition nécessaire et suffisante de récurrence des processus de vie ou de mort irréductibles sur \mathbb{N} (section 5.14.1). Comme \mathbf{P} est récurrent, le temps d'explosion est infini p.s. et X_t est bien défini pour tout réel $t \geq 0$. Le semi-groupe de Markov $(\mathbf{P}_t)_{t \in \mathbb{R}_+}$ possède l'expression explicite suivante :

$$\mathbf{P}_t(n, \cdot) = \text{Loi}(X_t | X_0 = n) = \text{Bin}(n, e^{-\mu t}) * \text{Poi}\left(\lambda \frac{1 - e^{-\mu t}}{\mu}\right) \xrightarrow{t \rightarrow \infty} \text{Poi}\left(\frac{\lambda}{\mu}\right).$$

En effet, on vérifie par calcul que $\partial_{t=0^+} \mathbf{P}_t(n, m) = \mathbf{Q}(n, m)$. Une vérification directe montre que la loi de Poisson $\text{Poi}(\lambda/\mu)$ est invariante pour \mathbf{Q} . Par conséquent, si $X_0 \sim \text{Poi}(\lambda/\mu)$, alors $X_t \sim \text{Poi}(\lambda/\mu)$ pour tout $t \in \mathbb{R}_+$.

Le processus $(X_t)_{t \in \mathbb{R}_+}$ peut être interprété de la manière suivante : des clients arrivent selon un processus de Poisson d'intensité λ , et ils sont servis immédiatement par un serveur dédié dont l'intensité de service est μ . Les temps de service ne dépendent pas du processus des arrivées. La variable aléatoire X_t représente le nombre de clients en cours de service à l'instant t . Ce processus est connu sous le nom de file d'attente M/M/∞. Les deux lettres M utilisées rappellent que les arrivées et les services suivent des lois exponentielles : ces processus sont Markoviens ou sans Mémoire. Cette terminologie due à Kendall est détaillée dans les premières pages de [56].

Cette interprétation en termes de file d'attente suggère la généralisation suivante : des clients arrivent selon un processus de Poisson d'intensité λ , mais seuls s serveurs sont disponibles. Lorsque tous les serveurs sont occupés, les clients en surnombre attendent qu'un serveur se libère. Le processus associé compte le nombre de clients en attente et en cours de service. Ce processus est appelé file d'attente M/M/s. L'interprétation en terme d'horloges exponentielles en compétition suggère de considérer le générateur défini, pour tous $n, m \in \mathbb{N}$, par

$$\mathbf{Q}(n, m) = \begin{cases} \min(s, n)\mu & \text{si } m = n - 1 \\ \lambda & \text{si } m = n + 1 \\ -(\lambda + \min(s, n)\mu) & \text{si } n = m \\ 0 & \text{sinon} \end{cases}.$$

Le cas $s = \infty$ correspond à la file d'attente M/M/∞. L'interprétation de \mathbf{Q} en terme d'horloges exponentielles en compétition permet également de simuler ces files d'attente de manière intuitive. De manière générale, les processus de vie ou de mort sur \mathbb{N} à temps continu sont des files d'attente.

Il est également possible de construire un générateur qui tient compte de la présence d'une salle d'attente de capacité finie, de phénomènes d'impatience, de phénomènes de compétition entre clients ou entre serveurs, de politiques de service, de variation des intensités au cours du temps, etc. Les files d'attente font partie des processus markoviens les plus utilisés en modélisation : lignes téléphoniques, aéroports, feux tricolores, caisses de supermarché, parkings, échanges de données sur les réseaux informatiques, allocation mémoire des ordinateurs, etc.

Versions à temps continu de chaînes de Markov usuelles

Les versions à temps continu de la marche aléatoire simple sur \mathbb{Z}^d , des processus de vie ou de mort sur \mathbb{N} , de la chaîne d'Ehrenfest, et de la chaîne de Wright – Fisher sont faciles à construire à partir de leur noyau de transition et d'une temporisation par un champ d'exponentielles. Idem pour la version à temps continu du processus de Galton – Watson, appelée processus de Yule.

Pour aller plus loin 6.14.1.

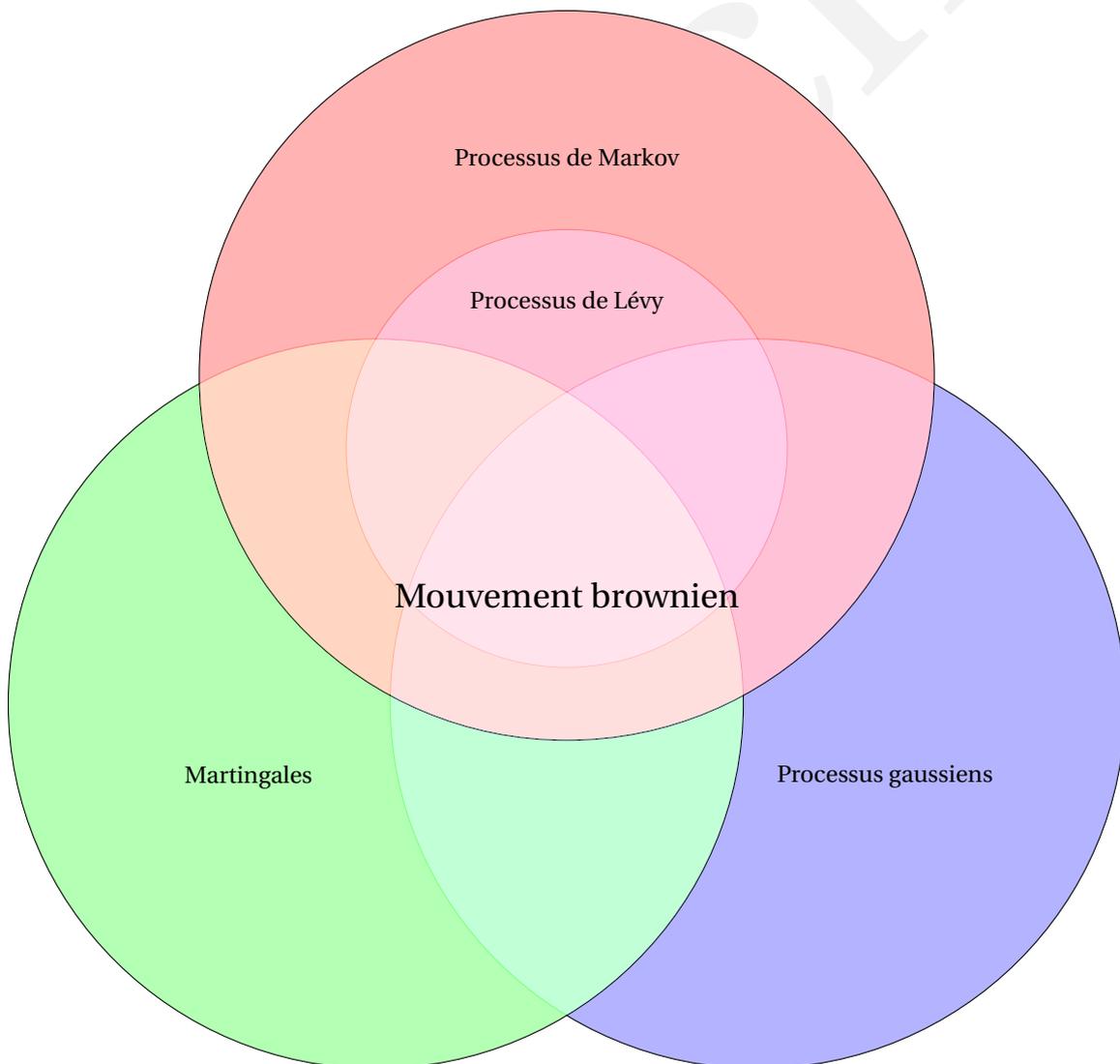
- Processus ponctuels de Poisson [56].
- Processus de Poisson composé avec dérive et ruine d'une compagnie d'assurance [11].
- Équations de Kolmogorov progressive et rétrograde pour les chaînes à temps continu [53, sec. 2.8].

Chapitre 7

Mouvement brownien

À propos de la loi normale ou gaussienne du théorème limite central, désigné sous le nom de loi des erreurs à l'époque : « Cette loi ne s'obtient pas par des déductions rigoureuses; plus d'une démonstration qu'on a voulu en donner est grossière, entre autres celle qui s'appuie sur l'affirmation que la probabilité des écarts est proportionnelle aux écarts. Tout le monde y croit cependant, me disait un jour M. Lippmann, car les expérimentateurs s'imaginent que c'est un théorème de mathématiques, et les mathématiciens que c'est un fait expérimental. » In *Le calcul des Probabilités* (1896) p. 149, Henri Poincaré (1854 – 1912). C'était avant l'œuvre de Paul Lévy (1886 – 1971) notamment. Comme le dit le physicien Oriol Bohigas i Guardiola (1925 – 2021), de nos jours, nous savons que c'est à la fois un fait expérimental et un théorème de mathématiques!

Tout comme le théorème limite central, le mouvement brownien, grâce à de multiples contributions, est à la fois un phénomène physique et un phénomène mathématique, voir figures 7.1, 7.2, et 7.3. Ce chapitre aborde quelques aspects du mouvement brownien mathématique.



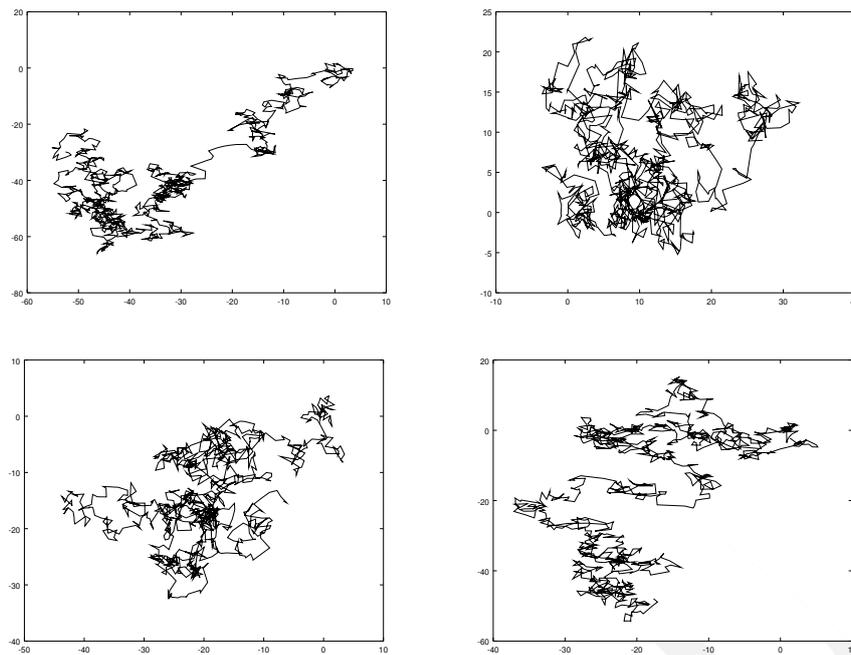


FIGURE 7.1 – Premiers pas de quatre trajectoires approchées du mouvement brownien plan issu de l'origine, avec une marche aléatoire à pas gaussiens, simulée avec le code `plot(cumsum(randn(2,1000)))`. La fonction `plot` effectue automatiquement une interpolation linéaire entre les points, coordonnées du vecteur passé en paramètre.

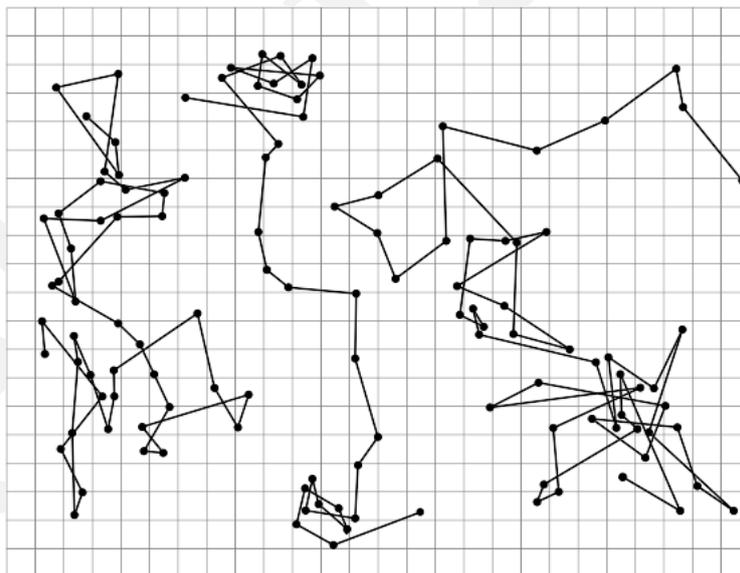


FIGURE 7.2 – Extrait du fameux livre [54] intitulé « Les Atomes » (1913) de Jean Perrin (1870–1942) : trois tracés de la trajectoire de grains de mastic de rayon $0.53 \mu\text{m}$, vus au microscope. Positions successives, toutes les 30 secondes, reliées par des segments linéaires (taille de maille de $3.1 \mu\text{m}$). Ces expériences, inspirées par les observations historiques de Robert Brown (1773–1858), ont permis de tester l'hypothèse atomiste ou moléculaire avancée par Ludwig Boltzmann (1844–1906), Albert Einstein (1879–1955), Marian Schmoluchovski (1872–1917), entre autres. « *Ainsi, la théorie moléculaire du mouvement brownien peut être regardée comme expérimentalement établie, et, du même coup, il devient assez difficile de nier la réalité objective des molécules.* ». Louis Bachelier (1870–1946) a identifié indépendamment un phénomène physique similaire des actifs financiers.

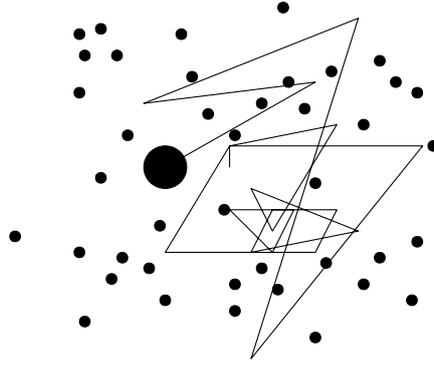


FIGURE 7.3 – Interprétation atomiste du mouvement brownien physique : une particule de poussière dans un liquide est sujette à un grand nombre de collisions avec les molécules du liquide, beaucoup plus petites et désordonnées par la chaleur. Dans la réalité, le ratio des diamètres est élevé, par exemple le rayon du grain de mastic observé par Perrin est de $0.53 \mu\text{m}$ tandis qu’une molécule d’eau a un rayon d’environ 0.17 nm , ce qui donne un ratio d’environ 3000. De plus, dans la réalité, la densité des molécules est élevée, la distance entre les molécules est d’environ 0.31 nm pour l’eau. Dans cette interprétation atomiste, le mouvement brownien physique est une sorte de marche aléatoire à temps continu et à espace continu, à trajectoires faites d’une succession de sauts à temps et positions aléatoires, vue à une échelle temps-espace qui la rend proche du mouvement brownien mathématique, sa limite d’échelle idéalisée. Les modèles sont des caricatures, les limites d’échelle des caricatures de caricatures.

Pour tous $t > 0$, $d \geq 1$, la densité de $\mathcal{N}(0, tI_d)$ sur \mathbb{R}^d est $x \in \mathbb{R}^d \mapsto p_t(x) = \frac{e^{-\frac{|x|^2}{2t}}}{(\sqrt{2\pi t})^d}$, où $|x|^2 = x_1^2 + \dots + x_d^2$, et

$$p_{t+s}(x) = (p_t * p_s)(x) = \int_{\mathbb{R}^d} p_t(x-z)p_s(z)dz, \quad s, t > 0.$$

Définition 7.0.1. Mouvement brownien ou processus de Wiener.

Un mouvement brownien (MB) de \mathbb{R}^d issu de B_0 , défini sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$, est un processus à temps continu $B = (B_t)_{t \in \mathbb{R}_+}$ à valeurs dans \mathbb{R}^d , c’est-à-dire une application $(t, \omega) \in \mathbb{R}_+ \times \Omega \mapsto B_t(\omega) \in \mathbb{R}^d$ telle que $\omega \mapsto B_t(\omega)$ est une variable aléatoire pour tout $t \in \mathbb{R}_+$, vérifiant les propriétés suivantes :

1. ses trajectoires $t \mapsto B_t$ sont presque sûrement continues, on dit que B est un processus continu.
2. ses accroissements sont indépendants, stationnaires, et gaussiens :
 - $B_{t_1} - B_{t_0}, \dots, B_{t_n} - B_{t_{n-1}}$ sont indépendantes pour tous $0 = t_0 < t_1 < \dots < t_n$, $n \geq 0$.
 - $B_t - B_s \sim \mathcal{N}(0, (t-s)I_d)$ pour tous $0 \leq s \leq t$.

- À ce stade, B n’est pas conçu comme une v.a. à valeurs dans $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^d)$, cela viendra plus tard.
- B est une sorte de vecteur gaussien sur $(\mathbb{R}^d)^{\mathbb{R}_+}$ (dimension infinie!)
- Il n’y a aucune condition sur B_0 , en particulier $B_t = B_0 + B_t - B_0$ peut ne pas être gaussienne!
- La stationnarité signifie que la loi de l’accroissement ne dépend que de l’écart temporel.
- Le théorème 3.3.4 donne $\partial_t \mathbb{E}(f(B_t) | B_0 = x) = \partial_t (p_t * f)(x) = \frac{1}{2} \Delta (p_t * f)(x)$ (équation de la chaleur).

Remarque 7.0.2. MB comme limite d’échelle temps-espace : principe d’invariance de Donsker.

Si X_1, X_2, \dots sont i.i.d. centrées de variance σ^2 , alors pour tous $0 = t_0 < t_1 < \dots < t_k$, $k \geq 1$, par le TLC,

$$(S_{t_1}^{(n)}, \dots, S_{t_k}^{(n)}) \xrightarrow[n \rightarrow \infty]{\text{loi}} \text{Loi}((B_{t_1}, \dots, B_{t_k})) \quad \text{où} \quad S_t^{(n)} = \frac{X_1 + \dots + X_{[nt]} + (nt - [nt])X_{[nt]+1}}{\sqrt{n\sigma^2}}.$$

Ceci exprime la convergence d’une marche aléatoire renormalisée en temps et en espace vers un MB :

$$(S_t^{(n)})_{t \in \mathbb{R}_+} \xrightarrow[n \rightarrow \infty]{} (B_t)_{t \in \mathbb{R}_+}$$

au sens des lois marginales de dimension finie. Avec un argument de tension, cela peut être renforcé en une convergence des lois sur les trajectoires, en concevant ces processus comme des v.a. sur $\mathcal{C}(\mathbb{R}_+, \mathbb{R})$ muni de la tribu cylindrique, qui coïncide avec la tribu borélienne de la topologie de la convergence uniforme sur les compacts. Ce phénomène limite central sur les trajectoires, appelé principe d’invariance de Donsker, exprime le fait que le MB est un objet limite gaussien universel, du moment que les accroissements de la

marche aléatoire sont centrés et de variance finie. En revanche, lorsque les accroissements de la marche aléatoire ont une variance infinie, on parle de vol de Lévy, la classe d'universalité n'est plus la même, le théorème limite est le TLC stable de Lévy, et la limite d'échelle est un processus de Lévy. La modélisation aléatoire « bain thermique » des particules du liquide remplace avantagement la modélisation déterministe par un grand nombre (Avogadro) de sphères dures satisfaisant aux équations de Newton.

```
# Code Python simulant des trajectoires de marches aléatoires à incréments iid gaussiens
import numpy as np ; import matplotlib.pyplot as plt
for i in range(1,11):
    plt.plot(np.cumsum(np.random.randn(1,1000)[0]), 'k-', linewidth=1)
plt.axis('off') ; plt.show()
```

```
# Code Julia simulant des trajectoires de marches aléatoires à incréments iid gaussiens
using Plots # using Pkg ; Pkg.add("Plots")
for i=1:10
    plot!(cumsum(randn(1000,1), dims = 1), lw=1, legend=false, grid=false, axes=([], false))
end
gui() # le graphique obtenu est celui figurant sur la couverture de ces notes !
```

— Lois finies dimensionnelles. Un processus de \mathbb{R}^d continu $(X_t)_{t \geq 0}$ issu de $x \in \mathbb{R}^d$ est un MB ssi pour tout $n \geq 1$, tous $0 < t_1 < \dots < t_n$, et tous $A_i \in \mathcal{B}_{\mathbb{R}^d}$, $1 \leq i \leq n$, on a

$$\mathbb{P}(X_{t_1} \in A_1, \dots, X_{t_n} \in A_n) = \int_{A_1 \times \dots \times A_n} p_{t_1}(x_1 - x) p_{t_2 - t_1}(x_2 - x_1) \dots p_{t_n - t_{n-1}}(x_n - x_{n-1}) dx_1 \dots dx_n.$$

- Reduction au cas centré. $B = (B_t)_{t \in \mathbb{R}_+}$ est un MB ssi $B - B_0 = (B_t - B_0)_{t \in \mathbb{R}_+}$ est un MB issu de 0.
- Reduction au cas unidimensionnel. Si $(X_t)_{t \geq 0}$ est d -dimensionnel et $X_t = (X_t^1, \dots, X_t^d)$ sont les coordonnées dans \mathbb{R}^d , alors X est un MB issu de 0 ssi les deux propriétés suivantes ont lieu :
 - (i) $(X_t^i)_{t \geq 0}$ est un MB sur \mathbb{R} issu de 0, pour tout $1 \leq i \leq d$.
 - (ii) $(X_t^1)_{t \geq 0}, \dots, (X_t^d)_{t \geq 0}$ sont des processus indépendants.

Lorsque X_0 n'est pas déterministe, les coordonnées de X peuvent être corrélées.

Définition 7.0.3. Processus gaussiens.

Un processus $X = (X_t)_{t \in \mathbb{R}_+}$ sur \mathbb{R}^d est un processus gaussien lorsque toutes ses lois fini-dimensionnelles sont gaussiennes : le vecteur aléatoire $(X_{t_1}, \dots, X_{t_n})$ est un gaussien pour tout $n \geq 1$ et tous $0 \leq t_1 < \dots < t_n$. Cela revient à dire que à la fois X_0 et les accroissements de X sont gaussiens au sens où le vecteur aléatoire $(X_{t_0}, X_{t_1} - X_{t_0}, \dots, X_{t_n} - X_{t_{n-1}})$ est gaussien pour tout $n \geq 1$ et tous $0 = t_0 < t_1 < \dots < t_n$.

Un MB est gaussien lorsque sa condition initiale est gaussienne et indépendante de ses incréments.

Remarque 7.0.4. Processus de Lévy.

Si B est un MB, alors pour tous $n \geq 1$ et $0 = t_0 < \dots < t_n$, les accroissements $B_{t_1} - B_{t_0}, \dots, B_{t_n} - B_{t_{n-1}}$ sont indépendants, et stationnaires au sens où leur loi ne dépend que des différences $t_1 - t_0, \dots, t_n - t_{n-1}$ entre les temps successifs. Aussi le MB est à accroissements indépendants et stationnaires, et ces processus sont appelés processus de Lévy, sous-classe des processus de Markov. Le processus de Poisson du chapitre 6 est aussi un processus de Lévy, pour lequel les accroissements sont de loi de Poisson, et les trajectoires constantes par morceaux, continues à droite avec limites à gauche.

7.1 Caractérisations, martingales, invariance d'échelle

Théorème 7.1.1. Caractérisation du MB par gaussianité et covariance.

Soit $X = (X_t)_{t \in \mathbb{R}_+}$ un processus sur \mathbb{R} , issu de 0, continu.

Alors X est un MB ssi X est gaussien, centré, de covariance donnée par $\mathbb{E}(X_t X_s) = s \wedge t$ pour tous $s, t \in \mathbb{R}_+$.

Démonstration. Soit $(X_t)_{t \geq 0}$ un MB issu de 0. Nous savons déjà qu'il est gaussien. De plus, pour tous $n \geq 1$ et $0 = t_0 < t_1 < \dots < t_n$, les v.a. $X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}}$ sont gaussiennes, centrées, indépendantes, de variance $t_1 - t_0, \dots, t_n - t_{n-1}$, donc le vecteur aléatoire $(X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}})$ est gaussien centré, donc le vecteur aléatoire $(X_{t_1}, \dots, X_{t_n})$ l'est aussi (image linéaire). De plus, pour tous $0 \leq s \leq t$,

$$\mathbb{E}(X_s X_t) = \mathbb{E}(X_s(X_t - X_s)) + \mathbb{E}(X_s^2) = 0 + s = s = s \wedge t.$$

Réciproquement, si $(X_t)_{t \in \mathbb{R}_+}$ est gaussien, centré, avec $\mathbb{E}(X_s X_t) = s \wedge t$ pour tous $s, t \geq 0$, alors pour tous $n \geq 1$ et $0 = t_0 < t_1 < \dots < t_n$, le vecteur aléatoire $(X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}})$ est gaussien, centré, de covariance

$$K_{i,j} = \mathbb{E}((X_{t_{i+1}} - X_{t_i})(X_{t_{j+1}} - X_{t_j})) = t_{i+1} \wedge t_{j+1} + t_i \wedge t_j - t_{i+1} \wedge t_j - t_i \wedge t_{j+1} = t_{i+1} + t_i - t_{i+1} \wedge t_j - t_i = (t_{i+1} - t_i) \mathbb{1}_{i=j}$$

lorsque $i \leq j$, elle est donc diagonale, égale à $\text{diag}(t_1 - t_0, t_2 - t_1, \dots, t_n - t_{n-1})$, ce qui implique l'indépendance des accroissements $X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}}$, leur gaussianité, et stationnarité, donc $(X_t)_{t \in \mathbb{R}_+}$ est un MB. ■

Corollaire 7.1.2. Invariance par rééchelonnement (scaling) espace-temps.

Si $B = (B_t)_{t \in \mathbb{R}_+}$ est un MB sur \mathbb{R} , issu de 0, alors $(\frac{1}{\sqrt{c}} B_{ct})_{t \geq 0}$ l'est aussi, pour tout réel $c > 0$.

Démonstration. Le processus $(\frac{1}{\sqrt{c}} B_{ct})_{t \geq 0}$ est continu, gaussien, de même covariance que le MB. ■

Théorème 7.1.3. Caractérisation martingale Fourier/Laplace du MB.

Soit $(X_t)_{t \in \mathbb{R}_+}$ un processus sur \mathbb{R}^d , issu de 0, continu, et sa filtration naturelle $(\mathcal{F}_t)_{t \in \mathbb{R}_+}$, $\mathcal{F}_t := \sigma(X_s : s \in [0, t])$. Alors les propriétés suivantes sont équivalentes :

- (i) X est un MB.
- (ii) $X_t - X_s$ est indépendante de \mathcal{F}_s et $X_t - X_s \sim \mathcal{N}(0, (t-s)I_d)$, pour tous $0 \leq s < t$.
- (iii) $\mathbb{E}(e^{i\lambda \cdot (X_t - X_s)} | \mathcal{F}_s) = \exp(-\frac{|\lambda|^2(t-s)}{2})$ pour tous $0 \leq s < t$ et $\lambda \in \mathbb{R}^d$.
- (iv) $(M_t^\lambda)_{t \in \mathbb{R}_+} = (e^{i\lambda \cdot X_t + \frac{|\lambda|^2 t}{2}})_{t \in \mathbb{R}_+}$ est une martingale pour $(\mathcal{F}_t)_{t \in \mathbb{R}_+}$, pour tout $\lambda \in \mathbb{R}^d$.
- (v) $(N_t^\lambda)_{t \in \mathbb{R}_+} = (e^{\lambda \cdot X_t - \frac{|\lambda|^2 t}{2}})_{t \in \mathbb{R}_+}$ est une martingale pour $(\mathcal{F}_t)_{t \in \mathbb{R}_+}$, pour tout $\lambda \in \mathbb{R}^d$.

- La notion de martingale s'adapte au temps continu : si $(M_t)_{t \in \mathbb{R}_+}$ est un processus sur $(\Omega, \mathcal{F}, \mathbb{P})$, et $(\mathcal{F}_t)_{t \in \mathbb{R}_+}$ une famille croissante de sous-tribus de \mathcal{F} (filtration à temps continu), alors M est une martingale lorsque M_t est \mathcal{F}_t -mesurable (adaptation), $M_t \in L^1$ (intégrabilité), et $\mathbb{E}(M_t | \mathcal{F}_s) = M_s$, pour tous $t \geq s \geq 0$.
- La notion de martingale s'adapte également immédiatement aux processus à valeurs complexes.
- Une v.a. X est indépendante d'une tribu \mathcal{F} ssi $\mathbb{E}(f(X) | \mathcal{F})$ est une v.a. constante pour une famille de fonctions test f qui caractérise la loi (exemples : transformée de Fourier ou de Laplace de la loi conditionnelle).

Démonstration. L'implication (i) \Rightarrow (ii) est immédiate. L'équivalence (ii) \Leftrightarrow (iii) provient des propriétés de l'espérance conditionnelle et de la fonction caractéristique conditionnelle. L'implication (iii) \Rightarrow (i) s'obtient en utilisant la propriété d'emboîtement des espérances conditionnelles pour obtenir la loi du vecteur des accroissements.

En multipliant les deux membres de (iii) par e^{iuZ} pour une v.a. Z \mathcal{F}_s -mesurable bornée, et en prenant l'espérance, il vient que $X_t - X_s$ est indépendante de \mathcal{F}_s et $X_t - X_s \sim \mathcal{N}(0, (t-s)I_d)$. Ceci démontre l'équivalence (iii) \Leftrightarrow (iv). La propriété (v) n'est que la version Laplace de (iv) (Fourier). ■

Définition 7.1.4. MB par rapport à une filtration.

Soit $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in \mathbb{R}_+}, \mathbb{P})$ un espace de probabilité filtré. On dit qu'un processus continu $(X_t)_{t \in \mathbb{R}_+}$ dans \mathbb{R}^d est un MB pour la filtration $(\mathcal{F}_t)_{t \in \mathbb{R}_+}$ lorsque X est $(\mathcal{F}_t)_{t \in \mathbb{R}_+}$ adapté et l'accroissement $X_t - X_s$ est in-

dépendant de \mathcal{F}_s pour tous $0 \leq s \leq t$, et suit la loi gaussienne $\mathcal{N}(0, (t-s)I_d)$. Cela revient à dire que $(\exp(i\lambda \cdot X_t + \frac{1}{2}|\lambda|^2 t))_{t \in \mathbb{R}_+}$ est une $(\mathcal{F}_t)_{t \in \mathbb{R}_+}$ -martingale pour tout $\lambda \in \mathbb{R}^d$.

- Si $(X_t)_{t \in \mathbb{R}_+}$ est un MB pour $(\mathcal{F}_t)_{t \in \mathbb{R}_+}$, alors X est un MB au sens de la définition 7.0.1. Réciproquement, si $(X_t)_{t \in \mathbb{R}_+}$ est un MB au sens de la définition 7.0.1, alors c'est un MB pour sa filtration naturelle.
- Être un MB est une affaire de loi des trajectoires : penser à un vecteur gaussien de dimension infinie. Il est établi par la suite que si B est un BM alors le processus défini par $X_0 := 0$ et $X_t := tB_{1/t}$ pour tout $t > 0$ est un MB, pour sa filtration naturelle, mais pour celle de B .

Théorème 7.1.5. Propriétés de martingale.

Soit $B = (B_t)_{t \in \mathbb{R}_+}$ un MB sur \mathbb{R}^d pour $(\mathcal{F}_t)_{t \in \mathbb{R}_+}$ et $B_t = (B_t^1, \dots, B_t^d)$ ses coordonnées. Pour tous $1 \leq j, k \leq d$,

- (i) $(B_t^j)_{t \in \mathbb{R}_+}$ est une martingale pour $(\mathcal{F}_t)_{t \in \mathbb{R}_+}$ lorsque $B_0 \in L^1$.
- (ii) $(B_t^j B_t^k - \mathbb{1}_{j=k} t)_{t \in \mathbb{R}_+}$ est une martingale pour $(\mathcal{F}_t)_{t \in \mathbb{R}_+}$ lorsque $B_0 \in L^2$.

Un fameux théorème de Paul Lévy, abordé en général dans un cours de calcul stochastique de M2, affirme que ces propriétés caractérisent le MB dans une classe assez large de processus continus.

Démonstration. Les mesurabilités et intégrabilités sont claires. Pour tous $1 \leq j \leq d, 0 \leq s \leq t$,

$$\mathbb{E}(B_t^j - B_s^j | \mathcal{F}_s) = \mathbb{E}(B_t^j - B_s^j) = 0 \quad \text{donc} \quad \mathbb{E}(B_t^j | \mathcal{F}_s) = \mathbb{E}(B_t^j - B_s^j + B_s^j | \mathcal{F}_s) = B_s^j.$$

De même, pour tous $1 \leq j, k \leq d$ et tous $0 \leq s \leq t$, en utilisant en plus l'indépendance des composantes,

$$\mathbb{E}((B_t^j - B_s^j)(B_t^k - B_s^k) | \mathcal{F}_s) = \mathbb{E}((B_t^j - B_s^j)(B_t^k - B_s^k)) = (t-s)\mathbb{1}_{j=k}$$

d'où

$$\mathbb{E}(B_t^j B_t^k - t\mathbb{1}_{j=k} | \mathcal{F}_s) = \mathbb{E}((B_t^j - B_s^j)(B_t^k - B_s^k) + B_t^j B_s^k + B_s^j B_t^k - B_s^j B_s^k - t\mathbb{1}_{j=k} | \mathcal{F}_s) = B_s^j B_s^k - s\mathbb{1}_{j=k}.$$

L'objet mathématique MB existe-t-il vraiment? Une construction explicite est donnée dans la section 7.6.

7.2 Variation des trajectoires et variation quadratique

Pour la régularité des trajectoires, la propriété de continuité s'oppose à celle d'accroissements indépendants.

Théorème 7.2.1. Variation et variation quadratique du MB.

Soit $B = (B_t)_{t \in \mathbb{R}_+}$ un MB sur \mathbb{R} issu de 0, $[u, v]$ un intervalle réel, $0 \leq u < v$, et δ une partition ou subdivision de $[u, v]$, $\delta : u = t_0 < \dots < t_n = v, n \geq 1$. Considérons les quantités suivantes :

$$r_1(\delta) := \sum_{i=0}^{n-1} |B_{t_{i+1}} - B_{t_i}| \quad \text{et} \quad r_2(\delta) := \sum_{i=0}^{n-1} |B_{t_{i+1}} - B_{t_i}|^2.$$

Alors les propriétés suivantes ont lieu :

- (i) $\lim_{|\delta| \rightarrow 0} r_2(\delta) = v - u$ dans L^2 et donc en \mathbb{P} , où $|\delta| := \sup_{0 \leq i \leq n-1} (t_{i+1} - t_i)$. Ainsi la variation quadratique de B sur un intervalle fini est donnée par la longueur de l'intervalle.
- (ii) $\sup_{\delta \in \mathcal{P}} r_1(\delta) = +\infty$ p.s., où \mathcal{P} est l'ensemble des partitions ou subdivisions de $[u, v]$. Ainsi les trajectoires de B ont une variation infinie sur tout intervalle, p.s.

- La propriété (i) révèle l'irrégularité des trajectoires du MB. En effet, pour $f : [u, v] \mapsto \mathbb{R}$ de régularité \mathcal{C}^1 , l'inégalité des accroissements finis donne $r_2(\delta) \leq \|f'\|_\infty^2 (\sum_{i=0}^{n-1} (t_{i+1} - t_i)) |\delta| = \|f'\|_\infty^2 (v - u) |\delta| \rightarrow 0$ si $|\delta| \rightarrow 0$.
- La propriété (ii) indique qu'on ne peut pas définir l'intégrale stochastique $\int_a^b \varphi_t dB_t(\omega)$ d'une fonction φ continue par rapport à la trajectoire $t \mapsto B_t(\omega)$ avec les sommes de Riemann. Cependant, comme l'ont remarqué Döblin et Itô, la propriété (i) est la clé d'une définition L^2 ou en \mathbb{P} de cette intégrale stochastique, abordée dans un cours de calcul stochastique, et qui permet de résoudre des EDS (EDO stochastiques).

Démonstration.

(i) Si $Z \sim \mathcal{N}(0, 1)$ alors $\mathbb{E}(Z^4) = 3$. Les propriétés des accroissements du MB donnent

$$\begin{aligned} \mathbb{E}((r_2(\delta))^2) &= \mathbb{E}\left(\left(\sum_i |B_{t_{i+1}} - B_{t_i}|^2\right)^2\right) \\ &= \sum_i \mathbb{E}(|B_{t_{i+1}} - B_{t_i}|^4) + 2 \sum_{i < j} \mathbb{E}(|B_{t_{i+1}} - B_{t_i}|^2 |B_{t_{j+1}} - B_{t_j}|^2) \\ &= 3 \sum_i (t_{i+1} - t_i)^2 + 2 \sum_{i < j} (t_{i+1} - t_i)(t_{j+1} - t_j) \\ &= 2 \sum_i (t_{i+1} - t_i)^2 + \left(\sum_i (t_{i+1} - t_i)\right)^2 = 2 \sum_i (t_{i+1} - t_i)^2 + (v - u)^2. \end{aligned}$$

De plus $\mathbb{E}(r_2(\delta)) = \sum_i (t_{i+1} - t_i) = v - u$. Donc

$$\mathbb{E}((r_2(\delta) - (v - u))^2) = 2 \sum_i (t_{i+1} - t_i)^2 \leq 2 \max_i (t_{i+1} - t_i)(v - u) \xrightarrow{|\delta| \rightarrow 0} 0.$$

(ii) La première propriété implique l'existence d'une suite $(\delta^k)_k$ de subdivisions de $[u, v]$ telle que

$$\lim_{k \rightarrow \infty} |\delta^k| = 0 \quad \text{et} \quad \lim_{k \rightarrow \infty} r_2(\delta^k) = \lim_{k \rightarrow \infty} \sum_i |B_{t_{i+1}^k} - B_{t_i^k}|^2 = v - u \quad \text{p.s.}$$

et donc, presque sûrement,

$$\sup_{\delta} r_1(\delta) \geq r_1(\delta^k) = \sum_i |B_{t_{i+1}^k} - B_{t_i^k}| \geq \frac{\sum_i |B_{t_{i+1}^k} - B_{t_i^k}|^2}{\max_i |B_{t_{i+1}^k} - B_{t_i^k}|} \xrightarrow{k \rightarrow \infty} +\infty,$$

où on a utilisé le fait que $\max_i |B_{t_{i+1}^k} - B_{t_i^k}| \rightarrow 0$ p.s. quand $k \rightarrow \infty$, car $|\delta^k| \rightarrow 0$ et p.s. les trajectoires de B sont continues donc uniformément continues sur tout intervalle fini donc compact comme $[u, v]$ (th. de Heine). ■

- La preuve de la propriété (i) reste valable au-delà du cadre gaussien, pourvu que les accroissements soient indépendants, stationnaires, centrés, de variance linéaire, et de moment d'ordre 4 fini.
- La preuve de la propriété (ii) révèle une propriété plus forte : de toute suite de subdivisions (δ^k) de $[u, v]$ telle que $|\delta^k| \rightarrow 0$, on peut extraire une sous-suite qui fait diverger presque sûrement r_1 vers $+\infty$.

25 → 26

7.3 Loi du zéro-un de Blumenthal et ses conséquences, loi de Bachelier – Lévy

Théorème 7.3.1. Propriétés des trajectoires du MB.

Soit $B = (B_t)_{t \in \mathbb{R}_+}$ un MB sur \mathbb{R} issu de 0, et $\mathcal{F}_t := \sigma(B_s : s \in [0, t])$. Alors :

- (i) Loi du zéro-un de Blumenthal : $\mathbb{P}(A) \in \{0, 1\}$ pour tout $A \in \mathcal{F}_{0^+}$, où $\mathcal{F}_{0^+} := \bigcap_{t > 0} \mathcal{F}_t$.
- (ii) Presque sûrement, $\inf_{s \in [0, \varepsilon]} B_s < 0$ et $\sup_{s \in [0, \varepsilon]} B_s > 0$ pour tout $\varepsilon > 0$.
- (iii) Pour tout $a \in \mathbb{R}$, presque sûrement a , $T_a = \inf\{t \in \mathbb{R}_+ : B_t = a\} < +\infty$.
- (iv) Presque sûrement b , $\lim_{t \rightarrow \infty} B_t = -\infty$ et $\overline{\lim}_{t \rightarrow \infty} B_t = +\infty$.
- (v) Presque sûrement, la fonction $t \in \mathbb{R}_+ \mapsto B_t$ n'est monotone sur aucun intervalle non singleton.

- a. En revanche T_a n'est pas borné, comme peut le montrer un argument de martingale arrêtée (exemple 4.4.4).
- b. Ceci n'implique pas que presque sûrement $\lim_{t \rightarrow \infty} |B_t| = +\infty$.

Dessin !

Démonstration.

- (i) L'idée est de montrer que \mathcal{F}_{0^+} est indépendante d'elle-même. Pour tout $A \in \mathcal{F}_{0^+}$, tout $k \geq 1$, toute $f : \mathbb{R}^k \rightarrow \mathbb{R}$ continue bornée, et tous $0 < t_1 < \dots < t_k$, les continuités de f et de B donnent, par convergence dominée,

$$\mathbb{E}(\mathbb{1}_A f(B_{t_1}, \dots, B_{t_k})) = \lim_{\varepsilon \rightarrow 0^+} \mathbb{E}(\mathbb{1}_A f(B_{t_1} - B_\varepsilon, \dots, B_{t_k} - B_\varepsilon)).$$

Or lorsque $0 < \varepsilon < t_1$, les v.a. $B_{t_1} - B_\varepsilon, \dots, B_{t_k} - B_\varepsilon$ sont indépendantes de \mathcal{F}_ε et donc de \mathcal{F}_{0^+} . D'où

$$\mathbb{E}(\mathbb{1}_A f(B_{t_1}, \dots, B_{t_k})) = \lim_{\varepsilon \rightarrow 0^+} \mathbb{P}(A) \mathbb{E}(f(B_{t_1} - B_\varepsilon, \dots, B_{t_k} - B_\varepsilon)) = \mathbb{P}(A) \mathbb{E}(f(B_{t_1}, \dots, B_{t_k})).$$

Ainsi \mathcal{F}_0^+ est indépendante de $\sigma(B_{t_1}, \dots, B_{t_k})$ pour tous $k \geq 1$ et $0 < t_1 < \dots < t_k$, et donc indépendante de la tribu $\sigma(B_t : t > 0)$ par classes monotones¹. Or $\sigma(B_t, t > 0) = \sigma(B_t, t \geq 0)$ car $B_0 = 0$, et $\mathcal{F}_0^+ \subset \sigma(B_t, t \geq 0)$.

(ii) Pour la propriété en sup, il suffit de montrer que $\mathbb{P}(A) = 1$ où $A = \bigcap_n \{\sup_{s \in [0, 1/n]} B_s > 0\}$. Observons que $A \in \bigcap_n \mathcal{F}_{1/n} = \bigcap_{t>0} \mathcal{F}_t = \mathcal{F}_0^+$. D'autre part,

$$\mathbb{P}\left(\sup_{s \in [0, 1/n]} B_s > 0\right) \underset{n \rightarrow \infty}{\searrow} \mathbb{P}(A) \quad \text{tandis que} \quad \mathbb{P}\left(\sup_{s \in [0, 1/n]} B_s > 0\right) \geq \mathbb{P}(B_{1/n} > 0) = \frac{1}{2},$$

donc $\mathbb{P}(A) \geq 1/2$, d'où $\mathbb{P}(A) = 1$ par (i). La propriété en inf s'obtient en utilisant $-B$ au lieu de B .

(iii) Le (ii) donne $\mathbb{P}(\sup_{s \in [0, 1]} B_s > \varepsilon) \underset{\varepsilon \rightarrow 0}{\nearrow} \mathbb{P}(\sup_{s \in [0, 1]} B_s > 0) = 1$. Par invariance par rééchelonnement (cor. 7.1.2),

$$\mathbb{P}\left(\sup_{s \in [0, 1]} B_s > \varepsilon\right) = \mathbb{P}\left(\sup_{s \in [0, \varepsilon^{-2}]} \varepsilon^{-1} B_{\varepsilon^2 s} > 1\right) = \mathbb{P}\left(\sup_{s \in [0, \varepsilon^{-2}]} B_s > 1\right) \underset{\varepsilon \rightarrow 0}{\nearrow} \mathbb{P}\left(\sup_{s \geq 0} B_s > 1\right).$$

Donc $\mathbb{P}(\sup_{s \geq 0} B_s > 1) = 1$. À nouveau par rééchelonnement, pour tout $R > 0$, et en remplaçant B par $-B$,

$$\mathbb{P}\left(\sup_{s \geq 0} B_s > R\right) = 1 \quad \text{et} \quad \mathbb{P}\left(\inf_{s \geq 0} B_s < -R\right) = 1.$$

Par conséquent, pour tout $R > 0$, la trajectoire de B sort de l'intervalle $[-R, R]$ p.s. D'autre part $B_0 = 0$.

Or comme B est continu, le théorème des valeurs intermédiaires donne $T_a < +\infty$ p.s. pour tout $a \in \mathbb{R}$.

(iv) Découle directement de la fin de la preuve de la propriété (iii). Notons qu'en combinant (iv) avec la continuité p.s. des trajectoires et le théorème des valeurs intermédiaires, nous obtenons que $\{t : B_t = 0\}$ est non-borné p.s., d'où $\inf\{t > 0 : B_t = 0\} < \infty$ p.s., et réciproquement grâce à la propriété de Markov forte!

(v) De la propriété (ii) sur les sup et inf, et la structure des accroissements, presque sûrement, pour tout $t \in \mathbb{Q} \cap \mathbb{R}_+$ et tout $\varepsilon > 0$, $\inf_{s \in [t, t+\varepsilon]} B_s < B_t$ et $\sup_{s \in [t, t+\varepsilon]} B_s > B_t$, d'où le résultat. ■

Corollaire 7.3.2. Loi du temps d'atteinte via transformée de Laplace.

Si $(B_t)_{t \in \mathbb{R}_+}$ est un MB sur \mathbb{R} issu de 0, alors pour tout $a > 0$, le temps d'atteinte $T_a = \inf\{t \geq 0 : B_t = a\}$ a pour transformée de Laplace $\lambda \geq 0 \mapsto \mathbb{E}(e^{-\lambda T_a}) = e^{-a\sqrt{2\lambda}}$ et pour densité $t \in \mathbb{R}_+ \mapsto \frac{a}{\sqrt{2\pi t^3}} e^{-\frac{a^2}{2t}}$.

— La loi de T_a est appelée loi de Lévy ou loi de Bachelier.

— La v.a. $T_a = \inf\{t \in \mathbb{R}_+ : B_t = a\}$ est bien mesurable car en notant $C \in \mathcal{F}$ un événement p.s. sur lequel B a des trajectoires continues, $C \cap \{T_a \leq t\} = C \cap \{\sup_{s \in [0, t] \cap \mathbb{Q}} B_s \geq a\}$, pour tous $a \in \mathbb{R}$ et $t \in \mathbb{R}_+$. Ceci montre également que T_a est un temps d'arrêt, à condition que $C \in \mathcal{F}_t$ pour tout $t \in \mathbb{R}_+$. Ceci est en particulier toujours le cas lorsque \mathcal{F}_0 contient tous les négligeables de \mathcal{F} , c'est-à-dire lorsque $(\mathcal{F}_t)_{t \in \mathbb{R}_+}$ est une tribu complète. Nous faisons toujours implicitement cette hypothèse lorsque T_a est utilisé comme temps d'arrêt.

Démonstration. Le th. 7.3.1 donne $T_a < \infty$ p.s. Pour tous $c, R > 0$, la version à temps continu du théorème d'arrêt de Doob² (th. 4.4.1) pour la martingale $(e^{cB_t - \frac{c^2}{2}t})_{t \in \mathbb{R}_+}$ et le temps d'arrêt $T_a \wedge R$ donne $\mathbb{E}(e^{cB_{T_a \wedge R} - \frac{c^2}{2}(T_a \wedge R)}) = 1$. Or comme $e^{cB_{T_a \wedge R} - \frac{c^2}{2}(T_a \wedge R)} \leq e^{ca}$, nous obtenons $\mathbb{E}(e^{cB_{T_a} - \frac{c^2}{2}T_a}) = 1$ par convergence dominée, quand $R \rightarrow \infty$. La continuité de B donne $B_{T_a} = a$ p.s. d'où la transformée de Laplace, puis la densité par inversion (difficile). ■

— La loi de T_a est également obtenue, et mieux comprise, dans le corollaire 7.5.3, via la propriété de Markov.

— La densité de la loi de T_a est unimodale, à queue lourde, sans espérance. Le comportement du MB ressemble à celui de la marche aléatoire simple, dont il est la limite d'échelle temps-espace, et bénéficie des symétries continues des limites d'échelle continues. La dépendance en a de la densité de la loi de T_a est liée à l'invariance par rééchelonnement du MB (corollaire 7.1.2), qui indique que T_a et $a^2 T_1$ ont même loi :

$$T_a = \inf\{t \in \mathbb{R}_+ : B_t = a\} = \inf\{t \in \mathbb{R}_+ : \frac{1}{a} B_t = 1\} = a^2 \inf\{t \in \mathbb{R}_+ : \frac{1}{a} B_{a^2 t} = 1\} \stackrel{\text{loi}}{=} a^2 \inf\{t \in \mathbb{R}_+ : B_t = 1\} = a^2 T_1.$$

La loi de T_1 a un mode et une médiane distincts. Ceux de T_a sont proportionnels à a^2 . Tout cela est relié à la relation entre échelles temps-espace du MB : l'écart-type de B_t est \sqrt{t} , de sorte que $\sqrt{t} = a$ donne $t = a^2$.

1. Soit H l'ensemble des v.a. bornées indépendantes de A , c'est-à-dire de \mathbf{I}_A . C est un espace vectoriel, contenant les constantes, stable par convergence monotone. Soit K l'ensemble des v.a. bornées de la forme $f(B_{t_1}, \dots, B_{t_k})$ pour un $k \geq 1$ et des $0 < t_1 < \dots < t_k$ arbitraires. Cet ensemble est stable par multiplication. Nous avons déjà démontré que $K \subset H$, donc par le théorème des classes monotones fonctionnel, H contient toutes les v.a. bornées mesurables pour la tribu engendrée par K , qui n'est rien d'autre que $\sigma(B_t : t > 0)$. Cet argument n'est plus valable si K est remplacé par l'ensemble des v.a. bornées de la forme $f(B_t)$ pour un $t > 0$ arbitraire, car cet ensemble n'est pas stable par produit. Du reste, si X, Y, Z sont des v.a. avec X indépendante de Y et de Z , alors X n'est pas forcément indépendante de (Y, Z) . Il n'est donc pas suffisant de démontrer que A est indépendant de $\sigma(B_t)$ pour tout $t > 0$, c'est la raison pour laquelle nous considérons des événements multi-temps.

2. Peut se démontrer de la même manière en approchant le temps d'arrêt par une suite de temps d'arrêts discrets sur les dyadiques.

7.4 Loi forte des grands nombres, invariance par inversion du temps

Tout comme pour le processus de Poisson, l'indépendance et la stationnarité des incréments du MB mène naturellement à une loi forte des grands nombres. Cependant, l'argument sandwich exploitant la monotonie des trajectoires du processus de Poisson dans la preuve du théorème 6.5.5 n'est pas valable pour le MB. La formule

$$\frac{B_t}{t} = \frac{B_1 - B_0 + \dots + B_{[t]} - B_{[t]-1}}{[t]} + \frac{B_t - B_{[t]}}{t},$$

notamment son dernier terme, suggère de chercher à contrôler uniformément B sur un intervalle de temps fini.

Théorème 7.4.1. Loi forte des grands nombres pour le MB.

Si $(B_t)_{t \in \mathbb{R}_+}$ est un MB sur \mathbb{R} issu de 0 alors $\lim_{t \rightarrow \infty} \frac{B_t}{t} = 0$ p.s. et dans L^p pour tout $p \in [1, \infty)$.

Le théorème limite central se résume à l'observation directe suivante : $\sqrt{t} \frac{B_t}{t} \sim \mathcal{N}(0, 1) \xrightarrow[t \rightarrow \infty]{\text{loi}} \mathcal{N}(0, 1)$.

La convergence p.s. reste valide pour un B_0 arbitraire, et la convergence L^p lorsque $B_0 \in L^p$.

Démonstration. Comme pour tout $t > 0$ et tout $p > 0$, $\mathbb{E}(|\frac{B_t}{t}|^p) = \frac{\mathbb{E}(|B_1|^p)}{t^{p/2}}$, et $B_1 \sim \mathcal{N}(0, 1)$, il vient $\frac{B_t}{t} \rightarrow 0$ dans L^p et donc en probabilité, quand $t \rightarrow \infty$. La convergence p.s. peut s'obtenir en contrôlant la vitesse de la convergence en probabilité, par un argument de tension. Plus précisément, soient a et b des réels tels que $0 < a < b$. Nous avons

$$\mathbb{E}\left(\sup_{a \leq t \leq b} \left(\frac{B_t}{t}\right)^2\right) \leq \frac{1}{a^2} \mathbb{E}\left(\sup_{a \leq t \leq b} B_t^2\right).$$

Par la version à temps continu de l'inégalité maximale de Doob (th. 4.7.2) pour la martingale $(B_{a+t})_{t>0}$ sur $[0, b-a]$,

$$\mathbb{E}\left(\sup_{a \leq t \leq b} \left(\frac{B_t}{t}\right)^2\right) \leq \frac{1}{a^2} 4\mathbb{E}(B_b^2) = \frac{4b}{a^2}, \quad \text{d'où} \quad \mathbb{E}\left(\sup_{2^n \leq t \leq 2^{n+1}} \left(\frac{B_t}{t}\right)^2\right) \leq \frac{8}{2^n}.$$

L'inégalité de Markov donne alors, pour tout $\varepsilon > 0$,

$$\mathbb{P}\left(\sup_{2^n \leq t \leq 2^{n+1}} \left|\frac{B_t}{t}\right| > \varepsilon\right) \leq \frac{1}{\varepsilon^2} \mathbb{E}\left(\sup_{2^n \leq t \leq 2^{n+1}} \left(\frac{B_t}{t}\right)^2\right) \leq \frac{8}{2^n \varepsilon^2}, \quad \text{d'où} \quad \sum_{n=0}^{\infty} \mathbb{P}\left(\sup_{2^n \leq t \leq 2^{n+1}} \left|\frac{B_t}{t}\right| > \varepsilon\right) < \infty.$$

Par le lemme de Borel–Cantelli, il existe $A_\varepsilon \in \mathcal{F}$ tel que $\mathbb{P}(A_\varepsilon) = 1$, et pour tout $\omega \in A_\varepsilon$, il existe n_ω tel que pour tout $n \geq n_\omega$, $\sup_{2^n \leq t \leq 2^{n+1}} \left|\frac{B_t(\omega)}{t}\right| \leq \varepsilon$, ce qui implique qu'il existe t_ω tel que pour tout $t \geq t_\omega$, $\left|\frac{B_t(\omega)}{t}\right| \leq \varepsilon$. Il reste à considérer l'événement $A = \bigcap_{r=1}^{\infty} A_{\frac{1}{r}}$, qui vérifie $\mathbb{P}(A) = 1$, et sur lequel $\lim_{t \rightarrow \infty} \frac{B_t}{t} = 0$. ■

Corollaire 7.4.2. Invariance par inversion du temps.

Si $B = (B_t)_{t \in \mathbb{R}_+}$ est un MB sur \mathbb{R} issu de 0, alors $X := (tB_{1/t})_{t \in \mathbb{R}_+}$, avec $X_0 := 0$, est aussi un MB.

Démonstration. Le processus X est gaussien, centré, et $\mathbb{E}(X_s X_t) = s \wedge t$ pour tous $s, t \geq 0$. Il reste à démontrer que X est continu. Par définition, X est continu sur $(0, \infty)$. La continuité p.s. en 0 découle de la loi forte des grands nombres du théorème 7.4.1 car $\lim_{t \rightarrow 0^+} X_t = \lim_{t \rightarrow 0^+} tB_{1/t} = \lim_{t \rightarrow \infty} (B_t/t) = 0$. p.s. ■

La preuve du corollaire 7.4.2 est réversible : si nous savons démontrer autrement que p.s. $t \mapsto tB_{1/t}$ est continue en $t = 0$, alors cela donne immédiatement en corollaire la loi forte des grands nombres du théorème 7.4.1. Ici aussi, il est clair que $tB_{1/t} \rightarrow 0$ dans L^p et donc en probabilité quand $t \rightarrow 0$, et le problème est d'établir la convergence p.s.

7.5 Propriété de Markov, principe de réflexion, temps d'atteinte

Si $(B_t)_{t \in \mathbb{R}_+}$ est un MB alors il est immédiat que pour tout $T > 0$ fixé, $(B_{T+t} - B_T)_{t \in \mathbb{R}_+}$ est un MB, issu de 0, et indépendant de \mathcal{F}_T . Il s'agit de la propriété de Markov faible. Elle s'étend aux temps d'arrêt T :

Théorème 7.5.1. Propriété de Markov forte du MB.

Soit $B = (B_t)_{t \in \mathbb{R}_+}$ un MB sur \mathbb{R}^d issu de 0, pour la filtration $(\mathcal{F}_t)_{t \in \mathbb{R}_+}$ sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. Soit T un temps d'arrêt, et $\mathcal{F}_T := \{A \in \mathcal{F} : \forall t \in \mathbb{R}_+, A \cap \{T \leq t\} \in \mathcal{F}_t\}$ sa tribu d'arrêt. Alors :

(i) Conditionnellement à $\{T < \infty\}$, $(B_{T+t} - B_T)_{t \in \mathbb{R}_+}$ est un MB issu de 0, indépendant de \mathcal{F}_T .

(ii) Pour tous $n \geq 1$, $f : (\mathbb{R}^d)^n \rightarrow \mathbb{R}$ mesurable et bornée, et $0 \leq t_1 \leq \dots \leq t_n$,

$$\mathbb{E}(f(B_{T+t_1}, \dots, B_{T+t_n}) \mathbb{1}_{\{T < \infty\}} \mid \mathcal{F}_T) = F(B_T) \mathbb{1}_{\{T < \infty\}} \quad \text{où} \quad F(x) = \mathbb{E}(f(x + B_{t_1}, \dots, x + B_{t_n})).$$

en particulier $\mathbb{E}(f(B_{T+t_1}, \dots, B_{T+t_n}) \mid T < \infty, B_T = x) = \mathbb{E}(f(x + B_{t_1}, \dots, x + B_{t_n}))$.

L'énoncé et la preuve sont semblables à ce qui a déjà été fait pour le processus de Poisson (théorème 6.4.2). La situation ici est cependant moins subtile car les trajectoires sont continues, et non plus seulement càdlàg.

Ceci n'est pas abordé lors du cours oral.

Démonstration.

(i) Supposons tout d'abord que $\mathbb{P}(T < \infty) = 1$. Posons $B^* := (B_{T+t} - B_T)_{t \in \mathbb{R}_+}$. Pour tout $n \in \mathbb{N}$, posons

$$T_n := \sum_{k=0}^{\infty} \frac{k+1}{2^n} \mathbb{1}_{T \in [k/2^n, (k+1)/2^n)},$$

à ne pas confondre avec le temps d'atteinte de n . Nous avons $T \leq T_n$ et T_n prend ses valeurs dans l'ensemble dyadique $D_n := \{k/2^n : k \in \mathbb{N}\}$. De plus T_n est un temps d'arrêt, et $T_n \searrow T$. Notons au passage que comme B a p.s. des trajectoires continues, nous avons $B_{T+t} = \lim_{n \rightarrow \infty} B_{T_n+t}$ p.s. et donc B_{T+t} est bien mesurable. Soit $A \in \mathcal{F}_T$, $m \geq 0$, et $0 = t_0 < \dots < t_m < \infty$. Par convergence dominée, pour toute $\varphi : (\mathbb{R}^d)^m \rightarrow \mathbb{R}$ continue et bornée, la continuité p.s. des trajectoires de B donne

$$\begin{aligned} \mathbb{E}(\mathbb{1}_A \varphi(B_{t_1}^*, \dots, B_{t_m}^*)) &= \mathbb{E}(\mathbb{1}_A \varphi(B_{T+t_1} - B_T, \dots, B_{T+t_m} - B_T)) \\ &= \lim_{n \rightarrow \infty} \mathbb{E}(\mathbb{1}_A \varphi(B_{T_n+t_1} - B_{T_n}, \dots, B_{T_n+t_m} - B_{T_n})). \end{aligned}$$

De plus, pour tout $n \geq 1$, $A \in \mathcal{F}_T \subset \mathcal{F}_{T_n}$ car $T \leq T_n$. En utilisant la propriété de Markov faible,

$$\begin{aligned} \mathbb{E}(\mathbb{1}_A \varphi(B_{T_n+t_1} - B_{T_n}, \dots, B_{T_n+t_m} - B_{T_n})) &= \sum_{r \in D_n} \mathbb{E}(\mathbb{1}_{A \cap \{T_n=r\}} \varphi(B_{r+t_1} - B_r, \dots, B_{r+t_m} - B_r)) \\ &= \sum_{r \in D_n} \mathbb{E}(\mathbb{1}_{A \cap \{T_n=r\}} \mathbb{E}(\varphi(B_{r+t_1} - B_r, \dots, B_{r+t_m} - B_r) \mid \mathcal{F}_r)) \\ &= \sum_{r \in D_n} \mathbb{P}(A \cap \{T_n = r\}) \mathbb{E}(\varphi(B_{t_1}, \dots, B_{t_m})) \\ &= \mathbb{P}(A) \mathbb{E}(\varphi(B_{t_1}, \dots, B_{t_m})). \end{aligned}$$

Ceci démontre que B^* est indépendant de \mathcal{F}_T . En prenant $A = \Omega$, ceci démontre également que B^* a les mêmes lois finies dimensionnelles que B , c'est donc un processus gaussien de même covariance que B . Comme B^* hérite par ailleurs de B le fait d'avoir p.s. des trajectoires continues, il s'agit bien d'un MB (l'argument doit être utilisé pour chacune des d composantes de B^*). Enfin, pour T pouvant prendre la valeur ∞ , le même argument fonctionne avec A remplacé par $A \cap \{T < \infty\}$.

(ii) Il suffit d'écrire, sur $\{T < \infty\}$, la décomposition

$$(B_{T+t_1}, \dots, B_{T+t_n}) = (B_{T+t_1} - B_T, \dots, B_{T+t_n} - B_T) + (B_T, \dots, B_T),$$

où $(B_{T+t_1} - B_T, \dots, B_{T+t_n} - B_T)$ est indépendant de \mathcal{F}_T et (B_T, \dots, B_T) est \mathcal{F}_T -mesurable, puis d'exploiter une propriété adéquate de l'espérance conditionnelle (aléa résiduel après moyennisation de l'aléa indépendant). Enfin, le « en particulier » vient du fait que $\sigma(\{T < \infty\}, B_T) \subset \mathcal{F}_T$. ■

Corollaire 7.5.2. Principe de réflexion.

Soit B un MB sur \mathbb{R} issu de 0. Pour tout $t \geq 0$, $S_t := \sup_{s \in [0, t]} B_s$ vérifie les propriétés suivantes :

- (i) $\mathbb{P}(T_a \leq t, B_t - a \leq -(a - b)) = \mathbb{P}(T_a \leq t, B_t - a \geq a - b)$ pour tous $a \geq 0$ et $b \in (-\infty, a]$.
- (ii) $\mathbb{P}(S_t \geq a, B_t \leq b) = \mathbb{P}(B_t \geq 2a - b)$ pour tous $a \geq 0$ et $b \in (-\infty, a]$.
- (iii) S_t et $|B_t|$ ont même loi.

- La propriété (i) exprime le fait que sur $\{T_a \leq t\}$, la probabilité que B_t soit en dessous du niveau $b = a - (a - b)$ est égale à celle d'être au-dessus du niveau $2a - b = a + (a - b)$. Nous avons $B_{T_a} = a$. Par la propriété de Markov forte, le processus après le temps d'arrêt T_a est encore un MB, issu de a , indépendant de T_a , ce qui entraîne que la loi des trajectoires après T_a est symétrique par rapport à a (réflexion), voir figure 7.4.
- Notons que $S_t = \sup_{s \in [0, t]} B_s$ est mesurable car en notant $C \in \mathcal{F}$ un événement sur lequel B a des trajectoires continues, $C \cap \{S_t \geq x\} = C \cap \{\sup_{s \in [0, t] \cap \mathbb{Q}} B_s \geq x\}$ pour tous $x \in \mathbb{R}$ et $t \in \mathbb{R}_+$. Ceci montre aussi que S_t est de plus adapté lorsque \mathcal{F}_0 contient les négligeables de \mathcal{F} , c'est-à-dire lorsque $(\mathcal{F}_t)_{t \in \mathbb{R}_+}$ est une tribu complète.

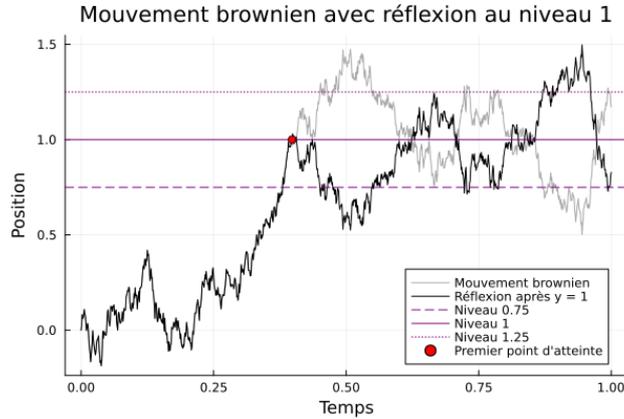


FIGURE 7.4 – Illustration du principe de réflexion du corollaire 7.5.2 avec $t = 1$, $a = 1$, et $b = 0.75$.

Démonstration.

- (i) Montrons que (i) et (ii) sont équivalents. Comme B a p.s. des trajectoires continues issues de 0, nous avons $\{T_a \leq t\} = \{S_t \geq a\}$ p.s.³. D'autre part, comme $a - b \geq 0$ et $a \geq 0$, nous avons

$$\{B_t \geq 2a - b\} = \{B_t - a \geq a - b\} \subset \{B_t - a \geq 0\} \subset \{S_t \geq a\}.$$

Et comme $b = a - (a - b)$, la propriété (i) se réécrit en (ii).

- (ii) L'idée est de se ramener au processus $B_t^* := B_{T_a+t} - B_{T_a} = B_{T_a+t} - a$, et d'exploiter la propriété de Markov forte. Avant tout, notons que $T_a = \inf\{t \geq 0 : B_t = a\} < \infty$ p.s. grâce au théorème 7.3.1. À présent, nous avons

$$\mathbb{P}(S_t \geq a, B_t \leq b) = \mathbb{P}(T_a \leq t, B_t \leq b) = \mathbb{P}(T_a \leq t, B_{t-T_a}^* \leq b - a),$$

où on a utilisé dans la dernière étape le fait que $B_{t-T_a}^* = B_{T_a+t-T_a} - B_{T_a} = B_t - a$, ce qui fait sens sur $\{T_a \leq t\}$. Par la propriété de Markov forte (théorème 7.5.1), B^* est indépendant de T_a et a la même loi que B . Comme B^* et $-B^*$ ont même loi, il vient que (T_a, B^*) a la même loi que $(T_a, -B^*)$. Aussi

$$\begin{aligned} \mathbb{P}(T_a \leq t, B_{t-T_a}^* \leq b - a) &= \mathbb{P}(T_a \leq t, -B_{t-T_a}^* \leq b - a) \\ &= \mathbb{P}(T_a \leq t, -(B_t - a) \leq b - a) \\ &= \mathbb{P}(T_a \leq t, B_t \geq 2a - b) \\ &= \mathbb{P}(B_t \geq 2a - b) \end{aligned}$$

où la dernière étape s'obtient grâce à l'argument de la preuve de la propriété (i).

- (iii) La propriété (ii) avec $b = a$, l'inégalité $B_t \leq S_t$, et le fait que B_t et $-B_t$ ont même loi, donnent

$$\begin{aligned} \mathbb{P}(S_t \geq a) &= \mathbb{P}(S_t \geq a, B_t \leq a) + \mathbb{P}(S_t \geq a, B_t \geq a) \\ &= \mathbb{P}(B_t \geq a) + \mathbb{P}(B_t \geq a) \\ &= \mathbb{P}(B_t \geq a) + \mathbb{P}(B_t \leq -a) \\ &= \mathbb{P}(|B_t| \geq a). \end{aligned}$$

■

La preuve du corollaire 7.5.2 ne fait intervenir ni la gaussianité ni la covariance du MB. En revanche, elle fait intervenir l'indépendance et la stationnarité des accroissements via la propriété de Markov forte, et la symétrie de leur loi. Le principe de réflexion est ainsi valable pour une large classe de processus, incluant la marche aléatoire simple sur \mathbb{Z} , analogue discret du MB. Cela est utile notamment pour le test statistique de Kolmogorov–Smirnov.

3. Par convention, $A \subset B$ p.s. lorsqu'il existe $\Omega' \in \mathcal{A}$ tel que $\mathbb{P}(\Omega') = 1$ et $A \cap \Omega' \subset B \cap \Omega'$.

Corollaire 7.5.3. Densités.

Soit $(B_t)_{t \in \mathbb{R}_+}$ un MB sur \mathbb{R} issu de 0. Alors :

- (i) Pour tout $t > 0$, la loi du couple $(\sup_{s \in [0, t]} B_s, B_t)$ a pour densité $(a, b) \in \mathbb{R}^2 \mapsto \frac{2(2a-b)}{\sqrt{2\pi t^3}} e^{-\frac{(2a-b)^2}{2t}} \mathbb{1}_{a \geq 0, b \leq a}$.
- (ii) Pour tout $a \in \mathbb{R}$, $T_a = \inf\{t \geq 0 : B_t = a\}$ a même loi que $\frac{a^2}{B_1^2}$ et a pour densité $t \in \mathbb{R}_+ \mapsto \frac{|a|}{\sqrt{2\pi t^3}} e^{-\frac{a^2}{2t}} \mathbb{1}_{t > 0}$.

Démonstration.

- (i) Conséquence directe du corollaire 7.5.2.
- (ii) Grâce au corollaire 7.5.2, $\mathbb{P}(T_a \leq t) = \mathbb{P}(S_t \geq a) = \mathbb{P}(|B_t| \geq a) = \mathbb{P}(B_t^2 \geq a^2) = \mathbb{P}(tB_1^2 \geq a^2) = \mathbb{P}(a^2 / B_1^2 \leq t)$. ■

Ainsi, à dilatation près, la loi de Lévy ou de Bachelier se trouve être finalement une loi inverse χ^2 , cas spécial de loi inverse gamma. En physique, elle est connue sous le nom de loi ou profil de van der Waals.

Le corollaire 7.3.2 aborde la loi de T_a via des martingales arrêtées au lieu de la propriété de Markov.

7.6 Construction avec le théorème de continuité de Kolmogorov, loi du logarithme itéré

L'existence mathématique du MB n'est pas évidente. Historiquement, Norbert Wiener semble être le premier à donner une construction rigoureuse, vers 1923, et pour cette raison, le MB est également appelé processus de Wiener. Une idée naturelle et intuitive pour construire le MB est de le réaliser comme une limite d'échelle d'une marche aléatoire à accroissements de variance finie, par exemple gaussiens, en exploitant le phénomène TLC, comme évoqué dans la remarque 7.0.2. Ceci est très utile pour obtenir par simulation une approximation numérique des trajectoires. La construction du MB proposée ci-dessous est basée sur une autre idée naturelle : le concevoir comme une superposition d'une infinité de v.a. gaussiennes centrées orthogonales, bref la limite d'une série aléatoire.

Théodéf 7.6.1. Pré-mouvement brownien, gaussien avec la bonne covariance, mais pas encore continu.

Considérons les ingrédients suivants :

- L'espace de Hilbert de dimension infinie $H = L^2(\mathbb{R}_+, dt)$ muni de $\langle f, g \rangle_H = \int_0^\infty f(t)g(t)dt$, $f, g \in H$.
- Une base hilbertienne (suite orthonormée à espace engendré dense) $(e_n)_{n \in \mathbb{N}}$ de H .
- Une suite $(X_n)_{n \in \mathbb{N}}$ de v.a. i.i.d. gaussiennes définies sur $(\Omega, \mathcal{A}, \mathbb{P})$ de loi $\mathcal{N}(0, 1)$.

Alors les propriétés suivantes ont lieu :

- (i) $\tilde{B}_f := \sum_{n \in \mathbb{N}} X_n \langle f, e_n \rangle_H$ est une série aléatoire qui converge dans $L^2(\Omega, \mathcal{A}, \mathbb{P})$, pour tout $f \in H$.
- (ii) \tilde{B}_f est une v.a. gaussienne réelle centrée de variance $\|f\|_H^2$, pour tout $f \in H$.
- (iii) $f \mapsto \tilde{B}_f$ est une isométrie linéaire $H \rightarrow L^2(\Omega, \mathcal{A}, \mathbb{P})$: pour tous $f, g \in H$ et $\alpha, \beta \in \mathbb{R}$,

$$\tilde{B}_{\alpha f + \beta g} = \alpha \tilde{B}_f + \beta \tilde{B}_g \quad \text{et} \quad \mathbb{E}(\tilde{B}_f \tilde{B}_g) = \langle f, g \rangle_H.$$

- (iv) Le pré-MB $(B_t)_{t \in \mathbb{R}_+}$ où $B_t := \tilde{B}_{\mathbb{1}_{[0, t]}}$ est gaussien de même moyenne et covariance que le MB :

$$\mathbb{E}(B_t) = 0 \quad \text{et} \quad \mathbb{E}(B_s B_t) = \langle \mathbb{1}_{[0, s]}, \mathbb{1}_{[0, t]} \rangle_H = \int_0^{s \wedge t} du = s \wedge t, \quad \text{pour tous } s, t \in \mathbb{R}_+.$$

- Le pré-MB n'a pas de raison d'être continu, il est construit comme une limite L^2 , et non pas uniforme.
- Les polynômes orthogonaux d'Hermite amortis $e_n = P_n(x) \sqrt{p(x)}$ sont un exemple de base. La construction du pré-MB sur l'intervalle fini $I \subset \mathbb{R}_+$, plutôt que sur \mathbb{R}_+ , peut se faire en utilisant l'espace de Hilbert $L^2(I, dt)$ et les bases de Fourier (construction de Wiener) ou de Haar (construction de Lévy) par exemple. Le passage d'une construction $B^{(0)} = (B_t^{(0)})_{t \in [0, 1]}$ sur l'intervalle de temps $[0, 1]$ à une construction sur \mathbb{R}_+ peut se faire naturellement en répétant à l'infini la même construction : $B_t := \sum_{k=0}^{\lfloor t \rfloor - 1} B_1^{(k)} + B_{t - \lfloor t \rfloor}^{(\lfloor t \rfloor)}$, pour tout $t \in \mathbb{R}_+$, où $(B^{(n)})_{n \geq 1}$ sont des copies indépendantes de $B^{(0)}$, ce qui produit un processus gaussien de covariance $s \wedge t$.

Démonstration.

- (i) Pour tout $f \in H$, la série $\sum_{n=0}^\infty X_n \langle f, e_n \rangle_H$ converge dans $L^2(\Omega, \mathcal{A}, \mathbb{P})$ car le critère de Cauchy est vérifié :

$$\mathbb{E} \left(\left(\sum_{n=p}^{p+q} X_n \langle f, e_n \rangle_H \right)^2 \right) = \sum_{n=p}^{p+q} \langle f, e_n \rangle_H^2 \xrightarrow{p, q \rightarrow \infty} 0.$$

- (ii) Comme la convergence dans (i) a lieu dans L^2 , les deux premiers moments convergent, en particulier $\mathbb{E}(\tilde{B}_f) = 0$ et $\mathbb{E}((\tilde{B}_f)^2) = \sum_{n \in \mathbb{N}} \langle f, e_n \rangle_H = \|f\|_H^2$ quand $f \in H$. D'autre part, par le lemme 7.6.2, \tilde{B}_f est une v.a. gaussienne.
- (iii) La linéarité découle du fait que le terme général de la série aléatoire est linéaire en f . Le caractère isométrique sur H provient de (ii). La formule s'obtient par polarisation, ou directement : pour tous $f, g \in H$,

$$\mathbb{E}(\tilde{B}_f \tilde{B}_g) = \sum_{n,m} \mathbb{E}(X_n X_m) \langle f, e_n \rangle_H \langle g, e_m \rangle_H = \sum_n \langle f, e_n \rangle_H \langle g, e_n \rangle_H = \langle \sum_n \langle f, e_n \rangle_H e_n, \sum_m \langle g, e_m \rangle_H e_m \rangle_H = \langle f, g \rangle_H.$$

- (iv) Pour tout $d \geq 1$ et tous $0 \leq t_1 \leq \dots \leq t_d$, le vecteur aléatoire $(B_{t_1}, \dots, B_{t_d})$ de \mathbb{R}^d est la limite dans L^2 d'une suite de vecteur aléatoires gaussiens de \mathbb{R}^d , par construction de ses composantes. Il est donc gaussien par le lemme⁴ 7.6.2. Il est par ailleurs centré avec la même matrice de covariance que pour le MB.

Lemme 7.6.2. Convergence gaussienne.

Si X_n est un vecteur gaussien de \mathbb{R}^d de moyenne m_n et de matrice de covariance K_n , $X \in L^2$, et $X_n \rightarrow X$ dans L^2 , alors $X \sim \mathcal{N}(m, K)$ où $m = \lim_{n \rightarrow \infty} m_n$ et $K = \lim_{n \rightarrow \infty} K_n$, et $X_n \rightarrow X$ dans L^p pour tout $p \geq 1$.

Il est possible de ne supposer qu'une convergence en probabilité, mais la preuve est alors plus subtile.

Preuve du lemme 7.6.2. La convergence dans L^2 donne $m_n \rightarrow m$ et $K_n \rightarrow K$. Ensuite, pour tout $t \in \mathbb{R}^d$,

$$\varphi_X(t) = \lim_{n \rightarrow \infty} \varphi_{X_n}(t) = \lim_{n \rightarrow \infty} e^{i \langle t, m_n \rangle - \frac{1}{2} \langle K_n t, t \rangle} = e^{i \langle t, m \rangle - \frac{1}{2} \langle K t, t \rangle} = \varphi_{\mathcal{N}(m, K)}(t)$$

d'où $X \sim \mathcal{N}(m, K)$. Enfin, pour tout $p \geq 1$, comme $\mathbb{E}(|X_n|^p)$ est continue en m_n et K_n , elle est bornée en n . Donc pour tout $p \geq 1$, $|X_n|^p$ est bornée dans disons L^2 , donc u.i., tandis que $X_n \rightarrow X$ en \mathbb{P} , d'où $X_n \rightarrow X$ dans L^p . ■

Concernant la régularité des trajectoires, nous savons déjà que la propriété de continuité s'oppose à celle d'indépendance des accroissements (théorème 7.2.1). Le théorème ci-dessous affirme que cela reste compatible lorsque les accroissements sont suffisamment intégrables (la queue de leur distribution n'est pas trop lourde).

Théorème 7.6.3. Critère de continuité ou de tension de Kolmogorov.

Soit $(X_t)_{t \in \mathbb{R}_+}$ un processus défini sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ et prenant ses valeurs dans un espace de Banach \mathbb{B} de norme $\|\cdot\|$, et tel que la propriété suivante de tension des accroissements a lieu :

$$\exists p \geq 1, \varepsilon > 0, c > 0, \quad \forall s, t \geq 0, \quad \mathbb{E}(\|X_t - X_s\|^p) \leq c |t - s|^{1+\varepsilon}.$$

Alors il existe sur le même espace de probabilité un processus continu $(X_t^*)_{t \in \mathbb{R}_+}$ à valeurs dans \mathbb{B} tel que :

- (i) X^* est une modification de X : pour tout $t \geq 0$, $\mathbb{P}(X_t^* = X_t) = 1$.
- (ii) X^* est à trajectoires γ -Hölder pour tout $\gamma \in (0, \varepsilon/p)$ sur tout intervalle fini : p.s., pour tout $t > 0$, pour tout $\gamma \in (0, \varepsilon/p)$, il existe une v.a. C_γ , telle que pour tous $u, v \in [0, t]$, $|X_u^* - X_v^*| \leq C_\gamma |u - v|^\gamma$.

Démonstration. La preuve fait appel à un chaînage⁵. Pour alléger les notations, nous allons nous contenter d'établir le résultat sur l'intervalle de temps $[0, 1]$. Montrons d'abord que X est Hölder sur les dyadiques $\mathcal{D} = \cup_{n \in \mathbb{N}} \mathcal{D}_n$ où $\mathcal{D}_n = \{k/2^n : k \in \{0, \dots, 2^n\}\} \subset \mathcal{D}_{n+1}$. Pour tous $n \geq 1$, $\varepsilon > 0$, et $\gamma > 0$, l'inégalité de Markov et (i) donnent

$$\begin{aligned} \mathbb{P}\left(\max_{1 \leq k \leq 2^n} \|X_{\frac{k}{2^n}} - X_{\frac{k-1}{2^n}}\| \geq 2^{-\gamma n}\right) &\leq \sum_{k=1}^{2^n} \mathbb{P}\left(\|X_{\frac{k}{2^n}} - X_{\frac{k-1}{2^n}}\| \geq 2^{-\gamma n}\right) \leq \sum_{k=1}^{2^n} 2^{\gamma p n} \mathbb{E}\left(\|X_{\frac{k}{2^n}} - X_{\frac{k-1}{2^n}}\|^p\right) \\ &\leq c 2^n 2^{-n(1+\varepsilon)+\gamma p n} = c 2^{-n(\varepsilon-\gamma p)}. \end{aligned}$$

Choisissons à présent $\gamma \in (0, \varepsilon/p)$, de sorte que $\sum_{n=1}^{\infty} \mathbb{P}\left(\max_{1 \leq k \leq 2^n} \|X_{\frac{k}{2^n}} - X_{\frac{k-1}{2^n}}\| \geq 2^{-\gamma n}\right) < \infty$. Le lemme de Borel-Cantelli fournit alors un événement $A = A_\gamma \in \mathcal{A}$ tel que $\mathbb{P}(A) = 1$ et pour tout $\omega \in A$, il existe N_ω tel que pour tout $n \geq N_\omega$, nous avons $\max_{1 \leq k \leq 2^n} \|X_{\frac{k}{2^n}} - X_{\frac{k-1}{2^n}}\| \leq 2^{-\gamma n}$. Donc il existe une v.a. $C = C_\gamma$ telle que pour tout $n \in \mathbb{N}$,

$$C < \infty \text{ p.s. et } \max_{1 \leq k \leq 2^n} \|X_{\frac{k}{2^n}} - X_{\frac{k-1}{2^n}}\| \leq C 2^{-\gamma n}.$$

4. Il est possible de ne faire appel au lemme que dans sa version $d = 1$, en considérant les accroissements de B , qui sont gaussiens par le lemme, mais aussi indépendants comme limite dans L^2 d'une suite de vecteurs aléatoires de \mathbb{R}^d à composantes indépendantes.

5. Inventé semble-t-il par Andrei Kolmogorov, et dont l'usage a contribué au succès de Michel Talagrand bien plus tard.

L'événement p.s. A dépend de γ , mais par monotonie, il peut être rendu valable pour tout $\gamma \in (0, \varepsilon/p)$. Démontrons que sur A , les trajectoires de X sont γ -Hölder sur \mathcal{D} , pour tout $\gamma \in (0, \varepsilon/p)$. Soient $s, t \in \mathcal{D}$, $s < t$. Comme $\mathcal{D} = \cup_n \mathcal{D}_n$ et $\mathcal{D}_n \subset \mathcal{D}_{n+1}$ pour tout n , nous avons $s, t \in \mathcal{D}_n$ à partir d'un certain rang sur n . Choisissons à présent n assez grand pour que $2^{-n} \leq |t - s|$. Notons également $s = k_1/2^n$ et $t = k_2/2^n$, $k_1 < k_2$, de sorte que $k_2 - k_1 = |t - s|2^n$. Nous pouvons alors exploiter le résultat précédent en utilisant un chaînage :

$$\|X_t - X_s\| \leq \sum_{k=k_1+1}^{k_2} \|X_{\frac{k}{2^n}} - X_{\frac{k-1}{2^n}}\| \leq (k_2 - k_1)C2^{-\gamma n} = |t - s|C2^{-n(\gamma-1)} \leq |t - s|C|t - s|^{\gamma-1} = C|t - s|^\gamma.$$

Donc sur l'événement A , les trajectoires de X sont γ -Hölder sur \mathcal{D} pour tout $\gamma \in (0, \varepsilon/p)$. En particulier, sur A , les trajectoires de X sont uniformément continues sur \mathcal{D} . Comme \mathcal{D} est dense sur $[0, 1]$ et comme \mathbb{B} est un espace métrique complet, nous pouvons donc considérer l'unique fonction continue⁶ $t \in [0, 1] \mapsto X_t^*(\omega)$ qui coïncide avec $t \mapsto X_t(\omega)$ sur \mathcal{D} , $\omega \in A$. Par continuité, cette fonction est γ -Hölder.

Montrons que X^* est une modification de X . Par construction, $X_t = X_t^*$ pour tout $t \in \mathcal{D}$. Soit $t \in [0, 1]$. Comme \mathcal{D} est dense dans \mathbb{R}_+ , il existe $(t_n)_{n \in \mathbb{N}}$ dans \mathcal{D} t.q. $t_n \rightarrow t$, ce qui, combiné à l'hypothèse de tension, donne $X_{t_n} \rightarrow X_t$ dans $L^p((\Omega, \mathcal{A}, \mathbb{P}), (\mathbb{B}, \|\cdot\|))$. Par le lemme de Borel–Cantelli, il existe donc une sous-suite $(t_{\varphi(n)})_{n \in \mathbb{N}}$ t.q. $X_{t_{\varphi(n)}} \rightarrow X_t$ p.s. Or par continuité de X^* , $X_{t_{\varphi(n)}}^* \rightarrow X_t^*$ p.s., et comme $X_{t_{\varphi(n)}} = X_{t_{\varphi(n)}}^*$ p.s., il vient que $X_t^* = X_t$ p.s. ■

Corollaire 7.6.4. Existence du MB.

Le MB sur \mathbb{R} existe, et donc le MB sur \mathbb{R}^d existe pour tout $d \geq 1$. De plus, p.s. les trajectoires du MB sur \mathbb{R} sont, sur chaque intervalle de temps fini, Hölder d'ordre γ pour tout $\gamma \in (0, 1/2)$, pas plus.

Démonstration. Le cas $d \geq 1$ se ramène au cas $d = 1$. Soit B le pré-MB du th. 7.6.1 avec $d = 1$. Pour tous $0 \leq s < t$,

$$\frac{B_t - B_s}{\sqrt{t - s}} \sim \mathcal{N}(0, 1), \quad \text{et donc}^7 \quad \mathbb{E}((B_t - B_s)^{2n}) = c_n(t - s)^n.$$

Le cas du moment d'ordre 4 ($n = 2$) permet d'utiliser le théorème 7.6.3, avec $p = 4$, $\varepsilon = 1$, $\gamma < \frac{\varepsilon}{p} = \frac{1}{4}$, ce qui donne un processus continu B^* tel que pour tout $t \geq 0$, $\mathbb{P}(B_t^* = B_t) = 1$. Le processus continu B^* est gaussien et de même covariance que B , c'est donc un MB sur \mathbb{R} issu de 0. De plus, en utilisant $p = 2n$ avec n arbitraire, nous obtenons la régularité γ -Hölder des trajectoires, pour tout $\gamma < \frac{n-1}{2n} \rightarrow \frac{1}{2}$. Enfin la loi du logarithme itéré (corollaire 7.6.6) donne la régularité optimale en excluant le cas $\gamma = \frac{1}{2}$. ■

Théorème 7.6.5. Loi du logarithme itéré (LLI).

Si $(B_t)_{t \in \mathbb{R}_+}$ est un MB sur \mathbb{R} alors

$$\mathbb{P}\left(\lim_{t \searrow 0} \frac{B_t}{\sqrt{2t \log(\log(1/t))}} = -1, \quad \overline{\lim}_{t \searrow 0} \frac{B_t}{\sqrt{2t \log(\log(1/t))}} = 1\right) = 1$$

et

$$\mathbb{P}\left(\lim_{t \rightarrow \infty} \frac{B_t}{\sqrt{2t \log(\log(t))}} = -1, \quad \overline{\lim}_{t \rightarrow \infty} \frac{B_t}{\sqrt{2t \log(\log(t))}} = 1\right) = 1.$$

La LLI met en lumière le rééchelonnement critique entre convergence p.s. (LGN) et en loi (TLC) (théorème 7.4.1). Elle est facile à deviner à partir de son analogue pour les sommes de v.a. i.i.d.

Ceci n'est pas abordé lors du cours oral.

Démonstration. La seconde propriété découle de la première en utilisant l'invariance par inversion du temps (corollaire 7.4.2). Démontrons la première propriété. On peut supposer sans perte de généralité que $B_0 = 0$. Comme B et $-B$ ont même loi, il suffit d'établir que $\mathbb{P}(\overline{\lim}_{t \searrow 0} \frac{B_t}{h(t)} = 1) = 1$ où $h(t) := \sqrt{2t \log(\log(1/t))}$ pour $t < 1/e$. Pour tous $\alpha > 0$ et $\beta > 0$, l'inégalité maximale de Doob (version continue du théorème 4.7.2)

6. Rappelons la propriété topologique générale suivante : si S et T sont des espaces métriques, $D \subset S$ est dense dans S , T est complet, et $f : D \rightarrow T$ est uniformément continue, alors il existe une unique $\tilde{f} : S \rightarrow T$ continue telle que $\tilde{f} = f$ sur D .

7. Nous avons $c_n = \mathbb{E}(Z^{2n}) = \frac{(2n)!}{2^n n!}$ où $Z \sim \mathcal{N}(0, 1)$ mais cette formule explicite pour c_n n'est pas utilisée ici.

utilisée pour la martingale $(\exp(\alpha B_t - \frac{\alpha^2}{2} t))_{t \in \mathbb{R}_+}$ donne, pour tout $t \in \mathbb{R}_+$,

$$\mathbb{P}\left(\max_{s \in [0, t]} \left(B_s - \frac{\alpha}{2} s\right) > \beta\right) = \mathbb{P}\left(\max_{s \in [0, t]} e^{\alpha B_s - \frac{\alpha^2}{2} s} \geq e^{\alpha \beta}\right) \leq e^{-\alpha \beta}.$$

Pour tous $\theta, \delta \in (0, 1)$ et $n \geq 1$, cette inégalité avec $t = \theta^n$, $\alpha = \frac{(1+\delta)h(\theta^n)}{\theta^n}$ et $\beta = \frac{h(\theta^n)}{2}$ donne

$$\mathbb{P}\left(\max_{s \in [0, \theta^n]} \left(B_s - \frac{(1+\delta)h(\theta^n)}{2\theta^n} s\right) > \frac{h(\theta^n)}{2}\right) \leq e^{-\alpha \beta} = \left(n \log \frac{1}{\theta}\right)^{-(1+\delta)} = O_{n \rightarrow \infty}(n^{-(1+\delta)}).$$

Par le lemme de Borel–Cantelli, on obtient que p.s. il existe N aléatoire tel que pour tout $n \geq N$,

$$\max_{s \in [0, \theta^n]} \left(B_s - \frac{(1+\delta)h(\theta^n)}{2\theta^n} s\right) \leq \frac{h(\theta^n)}{2}.$$

Ceci implique que pour tout $t \in [\theta^{n+1}, \theta^n]$, $B_t \leq \max_{s \in [0, \theta^n]} B_s \leq \frac{(2+\delta)h(\theta^n)}{2} \leq \frac{(2+\delta)h(t)}{2\sqrt{\theta}}$ d'où

$$\mathbb{P}\left(\overline{\lim}_{t \searrow 0} \frac{B_t}{h(t)} \leq \frac{2+\delta}{2\sqrt{\theta}}\right) = 1.$$

En faisant $\theta \rightarrow 1$ et $\delta \rightarrow 0$, on obtient bien que $\mathbb{P}(\overline{\lim}_{t \searrow 0} \frac{B_t}{h(t)} \leq 1) = 1$, et $\mathbb{P}(\underline{\lim}_{t \searrow 0} \frac{B_t}{h(t)} \geq -1) = 1$ puisque B et $-B$ ont même loi. Il ne reste plus qu'à établir que $\mathbb{P}(\overline{\lim}_{t \searrow 0} \frac{B_t}{h(t)} \geq 1) = 1$. Pour cela, considérons l'événement $A_n := \{B_{\theta^n} - B_{\theta^{n+1}} \geq (1 - \sqrt{\theta})h(\theta^n)\}$, et observons que si Φ est la fonction de répartition de $\mathcal{N}(0, 1)$, alors $1 - \Phi(x) \geq \frac{x}{x^2+1} \Phi'(x) \geq \frac{\Phi'(x)}{2x}$ pour tout $x \geq 1$. En particulier, cela donne, pour n assez grand,

$$\mathbb{P}(A_n) = 1 - \Phi(r_n) \geq \frac{1}{2r_n} \frac{e^{-\frac{r_n^2}{2}}}{\sqrt{2\pi}} = \frac{1}{2\sqrt{2\pi}r_n} \left(n \log \frac{1}{\theta}\right)^{-\frac{(1-\sqrt{\theta})^2}{1-\theta}} \quad \text{où } r_n := \frac{(1-\sqrt{\theta})h(\theta^n)}{\theta^{n/2}\sqrt{1-\theta}},$$

car $r_n^2 = \frac{2(1-\sqrt{\theta})^2}{1-\theta} \log(n \log \frac{1}{\theta})$. Or $\frac{(1-\sqrt{\theta})^2}{1-\theta} < 1$, donc $\sum_{n \geq 1} \mathbb{P}(A_n) = \infty$. À présent, l'indépendance des accroissements de B et le lemme de Borel–Cantelli^b donnent que p.s., pour une infinité de valeurs de n ,

$$B_{\theta^n} - B_{\theta^{n+1}} \geq (1 - \sqrt{\theta})h(\theta^n).$$

Comme $-B$ et B ont même loi, par la première partie de la preuve, p.s. il existe N t.q. pour tout $n \geq N$,

$$B_{\theta^{n+1}} > -2h(\theta^{n+1}) \geq -c\sqrt{\theta}h(\theta^n)$$

pour une constante $c > 0$ ne dépendant pas de θ . Par conséquent, p.s., pour une infinité de valeurs de n ,

$$B_{\theta^n} > (1 - (c+1)\sqrt{\theta})h(\theta^n).$$

Cela donne $\mathbb{P}(\overline{\lim}_{t \searrow 0} \frac{B_t}{h(t)} \geq 1 - (c+1)\sqrt{\theta}) = 1$. Il ne reste plus qu'à faire $\theta \rightarrow 0$. ■

a. L'encadrement de Mills $\frac{x}{x^2+1} \leq \frac{1-\Phi(x)}{\Phi'(x)} \leq \frac{1}{x}$ est valable pour tout $x > 0$.
b. Cette preuve utilise donc les deux sens du lemme de Borel–Cantelli!

Corollaire 7.6.6. Régularité Hölder des trajectoires du MB.

Si $(B_t)_{t \in \mathbb{R}_+}$ est un MB sur \mathbb{R} alors pour tout $s \geq 0$,

$$\mathbb{P}\left(\underline{\lim}_{t \searrow 0} \frac{B_{s+t} - B_s}{\sqrt{2t \log(\log(1/t))}} = -1, \quad \overline{\lim}_{t \searrow 0} \frac{B_{s+t} - B_s}{\sqrt{2t \log(\log(1/t))}} = 1\right) = 1.$$

En particulier, presque sûrement, la trajectoire $t \in \mathbb{R}_+ \mapsto B_t$ de B n'est pas $\frac{1}{2}$ -Hölder^a continue sur tout intervalle fini non réduit à un point, et en particulier n'est nulle part dérivable sur \mathbb{R}_+ .

a. $f : I \rightarrow \mathbb{R}$ est γ -Hölder continue quand il existe une constante $c > 0$ telle que pour tous $s, t \in I$, $|f(t) - f(s)| \leq c|t - s|^\gamma$.

Démonstration. Découle du théorème 7.6.5 et du fait que $(B_{s+t} - B_s)_{t \in \mathbb{R}_+}$ et $(B_t)_{t \in \mathbb{R}_+}$ ont même loi. ■

7.7 Mesure de Wiener et mouvement brownien canonique

Les MB ne sont pas uniques, mais ont tous la même loi, vus comme des v.a. sur un espace de trajectoires :

Théodéf 7.7.1. Mesure de Wiener, espace canonique, MB canonique.

Soit $W = \mathcal{C}(\mathbb{R}_+, \mathbb{R}^d)$, muni de la tribu cylindrique \mathcal{A}_W engendrée par les parties de la forme $\{f \in W : f(t_1) \in A_1, \dots, f(t_n) \in A_n\}$, $n \geq 1$, $A_1, \dots, A_n \in \mathcal{B}_{\mathbb{R}^d}$, $0 \leq t_1 < \dots < t_n$. Soit $\pi_t(w) := w_t$ pour tout $w \in W$ et tout $t \in \mathbb{R}_+$.

- (i) Il existe une unique loi de probabilité μ sur (W, \mathcal{A}_W) , appelée mesure de Wiener, telle que pour tout $n \geq 1$, tous $0 < t_1 < \dots < t_n$, tous $A_1, \dots, A_n \in \mathcal{B}_{\mathbb{R}^d}$, avec $t_0 := 0$ et $x_0 := 0$,

$$\mu(\{w : w_{t_1} \in A_1, \dots, w_{t_n} \in A_n\}) = \int_{A_1 \times \dots \times A_n} p_{t_1-t_0}(x_1-x_0) \cdots p_{t_n-t_{n-1}}(x_n-x_{n-1}) dx_1 \cdots dx_n$$

où p est le noyau de la chaleur défini pour tout $t > 0$ et $x \in \mathbb{R}^d$ par $p_t(x) = \frac{e^{-\frac{|x|^2}{2t}}}{(\sqrt{2\pi t})^d}$.

- (ii) Sous la loi μ sur (W, \mathcal{A}_W) , le processus des coordonnées ou des projections $\pi = (\pi_t)_{t \in \mathbb{R}_+}$ est un MB, appelé MB canonique, tandis que (W, \mathcal{A}_W, μ) est appelé espace canonique ou processus de Wiener.
- (iii) Pour tout MB $B = (B_t)_{t \in \mathbb{R}_+}$ sur \mathbb{R}^d issu de 0, et toute $\Phi : W \rightarrow \mathbb{R}$ cylindrique et bornée,

$$\mathbb{E}(\Phi(B)) = \int_W \Phi(w) \mu(dw) = \mathbb{E}(\Phi(\pi)).$$

- Une fonction cylindrique Φ est une fonction mesurable ne dépendant que d'un nombre fini de temps.
- Il est possible d'établir que la tribu cylindrique \mathcal{A}_W coïncide avec la tribu borélienne \mathcal{B}_W de la topologie de la convergence uniforme sur les compacts de \mathbb{R}_+ , grâce à la métrique $d(f, g) = \sum_n \max(1, \sup_{t \in [0, n]} |f(t) - g(t)|)$.
- Si $B = (B_t)_{t \in \mathbb{R}_+}$ est un MB sur \mathbb{R}^d , défini sur $(\Omega, \mathcal{A}, \mathbb{P})$, alors il existe $\Omega' \in \mathcal{A}$ tel que $\mathbb{P}(\Omega') = 1$ et $t \in \mathbb{R}_+ \mapsto B_t(\omega)$ est continue pour tout $\omega \in \Omega'$, et le processus B restreint à $(\Omega', \mathcal{A}' = \mathcal{A} \cap \Omega', \mathbb{P}|_{\mathcal{A}'})$ est une v.a. à valeurs dans (W, \mathcal{A}_W) , dont la loi est la mesure de Wiener.
- La mesure de Wiener est une loi gaussienne sur l'espace de dimension infinie $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^d) \subset (\mathbb{R}^d)^{\mathbb{R}_+}$, dont le support est inclus dans l'ensemble des fonctions qui sont γ -Hölder pour tout $\gamma \in (0, 1/2)$ et pas plus.

Démonstration.

- (i) La donnée des lois marginales sur les cylindres caractérise la loi (théorème d'extension de Kolmogorov).
(ii) Par construction!
(iii) Découle de ce qui précède. ■

7.8 Intégrale de Wiener

Toute combinaison linéaire finie des accroissements d'un MB est gaussienne. Or il est naturel de considérer un MB comme un vecteur gaussien de dimension infinie. On s'attend en particulier à ce que les combinaisons linéaires infinies sommables soient aussi gaussiennes. Cette intuition est réalisée mathématiquement par l'intégrale de Wiener du théorème 7.8.1 ci-dessous, qui donne un sens à la v.a. gaussienne

$$\int_0^\infty f(t) dB_t = \langle f, B \rangle$$

où l'intégrateur $(B_t)_{t \in \mathbb{R}_+}$ est un MB sur \mathbb{R}^d et l'intégrande f est dans $L^2_{\mathbb{R}^d}(\mathbb{R}_+, dt)$. Le premier est aléatoire gaussien et le second est déterministe de carré intégrable. C'est notre première intégrale stochastique.

Théodéf 7.8.1. Intégrale de Wiener.

Considérons les ingrédients suivants :

- Un MB $B = (B_t)_{t \in \mathbb{R}_+} = ((B_t^1, \dots, B_t^d))_{t \in \mathbb{R}_+}$ sur \mathbb{R}^d , issu de 0, défini sur $(\Omega, \mathcal{A}, \mathbb{P})$.
 - L'adhérence G du sous-espace de $L^2(\Omega, \mathcal{A}, \mathbb{P})$ engendré par les v.a. $\{B_t^i : t \geq 0, 1 \leq i \leq d\}$.
 - L'espace de Hilbert $H := L^2_{\mathbb{R}^d}(\mathbb{R}_+, dt)$ muni du produit scalaire $\langle f, g \rangle_H := \int_0^\infty f(t) \cdot g(t) dt$.
- Alors il existe une unique application linéaire continue $I : H \rightarrow G$ telle que

$$I(f) = a \cdot (B_t - B_s) \quad \text{lorsque } f = a \mathbb{1}_{(s, t]} \text{ avec } a \in \mathbb{R}^d \text{ et } 0 \leq s \leq t.$$

De plus, cette application vérifie les propriétés suivantes :

- (i) I est une isométrie : pour tous $f, g \in H$, $\langle I(f), I(g) \rangle_G = \mathbb{E}(I(f)I(g)) = \int_0^\infty f(t) \cdot g(t) dt = \langle f, g \rangle_H$.
- (ii) I est bijective.
- (iii) $I(f) \sim \mathcal{N}(0, \|f\|_H^2)$ pour tout $f \in H$, il s'agit de l'intégrale de Wiener de f , notée $\int_0^\infty f(t) dB_t$.

Espace gaussien : les éléments de G sont des v.a. gaussiennes d'après le lemme 7.6.2.

Démonstration. Le sous-espace $S = \{f \in H : f = \sum_{i=0}^{n-1} a_i \mathbb{1}_{(t_i, t_{i+1}]}, 0 = t_0 < t_1 < \dots < t_n, n \geq 0, a_i \in \mathbb{R}^d\}$ de H est dense. Si $f \in S$, alors $f = \sum_{\text{finie}} a_i \mathbb{1}_{(t_i, t_{i+1}]}$, et nous pouvons définir

$$I(f) := \sum_{\text{finie}} a_i \cdot (B_{t_{i+1}} - B_{t_i}).$$

Cette définition ne dépend pas de la décomposition choisie pour f , et l'application $f \mapsto I(f)$ est linéaire. De plus, grâce aux propriétés d'indépendance des accroissements et des incréments du MB,

$$\begin{aligned} \mathbb{E}((I(f))^2) &= \sum_{i,j} \mathbb{E}((a_i \cdot (B_{t_{i+1}} - B_{t_i})) (a_j \cdot (B_{t_{j+1}} - B_{t_j}))) \\ &= \sum_i \mathbb{E}((a_i \cdot (B_{t_{i+1}} - B_{t_i}))^2) \\ &= \sum_i \mathbb{E}\left(\left(\sum_j a_{i,j} (B_{t_{i+1}}^j - B_{t_i}^j)\right)^2\right) \\ &= \sum_i \sum_{j,k} a_{i,j} a_{i,k} \mathbb{E}((B_{t_{i+1}}^j - B_{t_i}^j)(B_{t_{i+1}}^k - B_{t_i}^k)) \\ &= \sum_i \sum_{j,k} a_{i,j} a_{i,k} (t_{i+1} - t_i) \mathbb{1}_{j=k} = \sum_i |a_i|^2 (t_{i+1} - t_i) = \int_0^\infty |f(t)|^2 dt. \end{aligned}$$

Ainsi I est une isométrie sur S , donc uniformément continue. Comme S est dense dans H , I admet un unique prolongement continu sur H . Alternativement, pour tout $f \in H$ et toute $(f_n)_{n \in \mathbb{N}}$ dans S t.q. $\|f_n - f\|_H \rightarrow 0$,

$$\|f_n - f_m\|_H = \|I(f_n) - I(f_m)\|_G \xrightarrow{m,n \rightarrow \infty} 0,$$

et comme G est complet, $I(f) = \lim_{n \rightarrow \infty} I(f_n)$ existe. Cette limite ne dépend pas de la suite choisie pour approcher f . De plus nous avons $\|f\|_H^2 = \mathbb{E}((I(f))^2)$, et, par polarisation et linéarité de I , pour tous $f, g \in H$,

$$\int_0^\infty f(t)g(t)dt = \frac{1}{4} \int_0^\infty ((f+g)^2 - (f-g)^2)(t)dt = \frac{1}{4} \mathbb{E}((I(f+g))^2 - (I(f-g))^2) = \mathbb{E}(I(f)I(g)).$$

L'application I ainsi définie est unique. Elle est injective car c'est une isométrie. Elle est également surjective car $F = I(H)$ est un sous-espace fermé de G , dense dans G (prendre $f = e_i \mathbb{1}_{(0,t]}$ donne $B_t^i \in F$). Le caractère gaussien centré des éléments de G , et donc de $I(f)$ pour tout $f \in H$, est garanti par le lemme 7.6.2. ■

L'ensemble $L^2(\Omega, \mathcal{A}, \mathbb{P}) \setminus G$ est énorme, la plupart des v.a. de carré intégrable ne sont pas gaussiennes!

Corollaire 7.8.2. Propriétés de l'intégrale de Wiener.

- (i) Pour tous $f, g \in H$, $I(f)$ et $I(g)$ sont indépendantes ssi $\langle f, g \rangle_H = \int_0^\infty f(t)g(t)dt = 0$.
- (ii) Pour tous $t \geq 0$, $1 \leq i \leq d$, et $f = (f^1, \dots, f^d) \in H$, $\mathbb{E}(B_t^i \int_0^\infty f(s)dB_s) = \int_0^t f^i(s)ds$.
- (iii) Pour tout $f \in H$, le processus $(I(f \mathbb{1}_{(0,t]}))_{t \in \mathbb{R}_+}$ est une martingale pour la filtration de $(B_t)_{t \in \mathbb{R}_+}$.
- (iv) Si $(h_n)_{n \in \mathbb{N}}$ est une base hilbertienne de H , alors $(I(h_n))_{n \in \mathbb{N}}$ est une suite de v.a. i.i.d. $\mathcal{N}(0, 1)$, base hilbertienne de G , et pour tout $t \in \mathbb{R}_+$, le développement en série suivant dans $L^2(\Omega, \mathcal{A}, \mathbb{P})$ est disponible :

$$B_t = \sum_{n=0}^{\infty} \underbrace{\left(\int_0^\infty h_n(s)dB_s \right)}_{\substack{I(h_n) \\ \text{v.a. gaussienne}}} \underbrace{\int_0^t h_n(s)ds}_{\text{vecteur déterministe}}$$

L'intégrale de Wiener produit des martingales à temps continu en masse à partir du MB! Mais comme elles sont définies par une approche L^2 , leur continuité (éventuellement à modification près) n'est pas évidente.

Démonstration.

- (i) Par le lemme 7.6.2 pour $d = 2$, le vecteur $(I(f), I(g))$ est gaussien, centré et de covariance 0 ssi $f \perp g$.
- (ii) Si $g = e_i \mathbb{1}_{(0,t]}$ alors $I(g) = B_t^i$ et $\mathbb{E}(B_t^i \int_0^\infty f(s) dB_s) = \mathbb{E}(I(g)I(f)) = \int_0^\infty g(s) \cdot f(s) ds = \int_0^t f^i(s) ds$.
- (iii) Pour tous $f \in H$, $t \geq 0$, et $s \in [0, t]$, $M_{s,t} = I(f \mathbb{1}_{(s,t]})$ est la limite L^2 des combinaisons linéaires $\sum_{\text{finie}} a_i (B_{v_i} - B_{u_i})$ avec $a_i \in \mathbb{R}^d$ et $u_i, v_i \in (s, t]$. Elle est donc intégrable, centrée, \mathcal{F}_t -mesurable, indépendante de \mathcal{F}_s . De plus $M_{0,t} = M_{0,s} + M_{s,t}$, d'où $\mathbb{E}(M_{0,t} | \mathcal{F}_s) = M_{0,s} + \mathbb{E}(M_{s,t}) = M_{0,s}$.
- (iv) Comme $(h_n)_{n \in \mathbb{N}}$ est une base hilbertienne de H , par isométrie, $(I(h_n))_{n \in \mathbb{N}}$ est bien une base hilbertienne de l'espace gaussien G . De plus, par (ii), $\langle B_t^i, I(h_n) \rangle_G = \int_0^t h_n^i(s) ds$, pour tout $1 \leq i \leq d$.

■

7.9 Vers les intégrales de Itô et de Stratonovich

Soit $(B_t)_{t \in \mathbb{R}_+}$ un MB sur \mathbb{R}^d . Examinons à présent le sens à donner à l'intégrale stochastique

$$\int_0^t B_s dB_s.$$

Ce n'est pas une intégrale de Wiener car l'intégrande est aléatoire.

En utilisant une approximation par des processus étagés, il vient, en réarrangeant les termes pour produire une sommation télescopique de sommes connues (on s'attend à une martingale, avec lemme 4.2.2 en tête),

$$\begin{aligned} \sum_i B_{t_i} (B_{t_{i+1}} - B_{t_i}) &= \sum_i (B_{t_{i+1}}^2 - B_{t_i}^2 - B_{t_{i+1}}^2 + B_{t_i} B_{t_{i+1}}) \\ &= \sum_i (B_{t_{i+1}}^2 - B_{t_i}^2 - B_{t_{i+1}} (B_{t_{i+1}} - B_{t_i})) \\ &= \sum_i (B_{t_{i+1}}^2 - B_{t_i}^2) - \sum_i B_{t_i} (B_{t_{i+1}} - B_{t_i}) - \sum_i (B_{t_{i+1}} - B_{t_i})^2, \end{aligned}$$

ce qui donne, si le passage à la limite en probabilité est licite,

$$\int_0^t B_s dB_s = B_t^2 - \int_0^t B_s dB_s - [B]_t,$$

où $[B]_t$ est la variation quadratique de B sur $[0, t]$ (théorème 7.2.1). Comme $[B]_t = t$ (théorème 7.2.1), il vient

$$\int_0^t B_s dB_s = \frac{B_t^2 - t}{2}.$$

Le processus $(\frac{1}{2}(B_t^2 - t))_{t \in \mathbb{R}_+}$ est une martingale centrée, et cela est dû au fait qu'on a considéré la valeur B_{t_i} de l'intégrande à l'extrémité gauche de l'intervalle $[t_i, t_{i+1}]$ dans la somme de Riemann. Il s'agit de l'intégrale stochastique d'Itô. Plus généralement, cela permet de donner un sens à $\int_0^t H_s dB_s$ où H est un processus de carré intégrable adapté et càdlàg, l'intégrale de Wiener étant alors le cas particulier où H est déterministe.

Examinons la notion d'intégrale stochastique de Stratonovich, qui correspond à considérer la valeur moyenne entre les extrémités gauche et droite sur chaque intervalle. Cela donne

$$\sum_i \frac{B_{t_{i+1}} + B_{t_i}}{2} (B_{t_{i+1}} - B_{t_i}) = \frac{1}{2} \sum_i (B_{t_{i+1}}^2 - B_{t_i}^2) = \frac{1}{2} B_t^2,$$

d'où, toujours en supposant que la convergence a lieu en probabilité, et cette fois-ci en notant \circ l'intégrale stochastique de Stratonovich pour la distinguer de celle d'Itô,

$$\int_0^t B_s \circ dB_s = \frac{1}{2} B_t^2.$$

Cette fois-ci, la règle de la chaîne du calcul différentiel est respectée, mais le résultat n'est plus une martingale⁸. La suite se trouve dans un cours de calcul stochastique! Un objectif important est de donner un sens rigoureux à la solution d'équations différentielles stochastiques (EDS = EDO stochastiques) $X_t = X_0 + \int_0^t \sigma(X_s) dB_s + \int_0^t b(X_s) ds$.

Le MB a la même physique que la marche aléatoire simple (réurrence/transience, martingale/Markov, etc.), et les symétries continues des limites d'échelles continues (invariance par rotation) que n'a pas la marche.

8. Par symétrie, il y a aussi une intégrale stochastique « anticipative » via $\sum_i B_{t_{i+1}} (B_{t_{i+1}} - B_{t_i})$, donnant $2(\text{Strato}_t - \frac{1}{2}t\delta_t) = \frac{1}{2}(B_t^2 + t)$.

Pour aller plus loin 7.9.1.

- Pont brownien et test statistique de Kolmogorov – Smirnov.
- Principe d'invariance de Donsker.
- Récurrence et transience du MB.
- Ensemble des zéros du mouvement brownien.
- Représentation de Kakutani de la solution du problème de Dirichlet au moyen du MB.
- Formule de Cameron – Martin
Application à la singularité des mesures de Wiener, et à la loi du temps d'atteinte du MB avec dérive.
- Brownien fractionnaire, indice de Hurst, p -variation, régularité et mémoire longue.
- Invariance conforme du MB, évolution de Schramm – Loewner (SLE).
- Décomposition en chaos de Wiener, calcul de Malliavin.
- Mesures gaussiennes sur les espaces de Hilbert et de Banach.
- Bruit blanc, champ libre gaussien.
- Intégrale d'Itô, formule d'Itô, caractérisation de Lévy du MB.

Annexe A

Variation totale sur espace au plus dénombrable

Si μ et ν sont des mesures de probabilité sur E au plus dénombrable, leur distance en variation totale est

$$d_{VT}(\mu, \nu) := \sup_{A \subset E} |\mu(A) - \nu(A)| \in [0, 1].$$

Par abus de notation $d_{VT}(X, Y) := d_{VT}(\text{Loi}(X), \text{Loi}(Y))$ et $d_{VT}(X, \mu) := d_{VT}(\text{Loi}(X), \mu)$.

Le nom vient du fait que $d_{VT}(\mu, \nu) = \|\mu - \nu\|_{VT}$ où $\|\eta\|_{VT} := \sup_{A \subset E} |\eta(A)|$ est la norme en variation totale de la mesure signée η de masse finie sur E . D'autre part, l'ensemble des mesures de probabilités sur E est un sous-ensemble fermé de $\ell^1(E)$, et par la propriété (iii) ci-dessous $2\|\cdot\|_{VT} = \|\cdot\|_{\ell^1(E, \mathbb{R})}$, en particulier $d_{VT}(\cdot, \cdot)$ est bien une distance sur cet ensemble, qui en fait un espace métrique complet. Cet ensemble est également convexe.

A.1 Formulations alternatives

Lemme A.1.1. Variation totale.

- (i) Le supremum dans $d_{VT}(\mu, \nu) := \sup_{A \subset E} |\mu(A) - \nu(A)|$ est atteint pour $A = A_* := \{x \in E : \mu(x) \geq \nu(x)\}$.
- (ii) $d_{VT}(\mu, \nu) = \frac{1}{2} \sup_{f: E \rightarrow [-1, 1]} |\mu f - \nu f|$ et le supremum est atteint pour $f = f_* := \mathbb{1}_{A_*} - \mathbb{1}_{A_*^c}$.
- (iii) $d_{VT}(\mu, \nu) = \frac{1}{2} \sum_{x \in E} |\mu(x) - \nu(x)| = \frac{1}{2} \|\mu - \nu\|_{\ell^1(E)}$.
- (iv) $d_{VT}(\mu, \nu) = 1 - \sum_{x \in E} (\mu(x) \wedge \nu(x))$ et en particulier $d_{VT}(\mu, \nu) = 1$ ssi μ et ν ont des supports disjoints.
- (v) $d_{VT}(\mu, \nu) = \min_{(X, Y)} \mathbb{P}(X \neq Y)$ où l'infimum porte sur les couples de v.a. sur $E \times E$ de lois marginales μ et ν .

Démonstration. L'équivalence entre (ii) et (iii) exprime la dualité $\ell^1(E) - \ell^\infty(E)$.

L'égalité entre les membres de droite des formules (ii) et (iii) provient de l'inégalité

$$|\mu f - \nu f| \leq \sum_{x \in E} |f(x)| |\mu(x) - \nu(x)| \leq \sup_{x \in E} |f(x)| \sum_{x \in E} |\mu(x) - \nu(x)|$$

qui est saturée pour $f = \mathbb{1}_{A_*} - \mathbb{1}_{A_*^c}$. Pour (ii) et (i), comme $|\mu(A) - \nu(A)| = \frac{1}{2} |\mu f_A - \nu f_A|$ où $f_A := \mathbb{1}_A - \mathbb{1}_{A^c}$, il vient

$$|\mu(A) - \nu(A)| \leq \frac{1}{2} \sup_{f: E \rightarrow [-1, 1]} |\mu f - \nu f| = \frac{1}{2} \sum_{x \in E} |\mu(x) - \nu(x)|$$

qui est saturée pour $A = A_*$ car, par définition de A_* ,

$$2|\mu(A_*) - \nu(A_*)| = \mu(A_*) - \nu(A_*) + \nu(A_*^c) - \mu(A_*^c) = \sum_{x \in A_*} |\mu(x) - \nu(x)| + \sum_{x \in A_*^c} |\mu(x) - \nu(x)|.$$

Pour (iv), il suffit d'observer que

$$\sum_{x \in E} (\mu(x) \wedge \nu(x)) = \frac{1}{2} \sum_{x \in E} (\mu(x) + \nu(x) - |\mu(x) - \nu(x)|) = 1 - d_{VT}(\mu, \nu).$$

Démontrons enfin (v). Soit (X, Y) un couple de v.a. sur $E \times E$ de lois marginales μ et ν . Alors (iv) donne

$$1 - d_{VT}(\mu, \nu) = \sum_{x \in E} (\mu(x) \wedge \nu(x)) \geq \sum_{x \in E} \mathbb{P}(X = x, Y = x) = \mathbb{P}(X = Y)$$

car $\mathbb{P}(X = x, Y = x) \leq \mu(x) \wedge \nu(x)$ pour tout $x \in E$. Il suffit donc de construire un couple (X, Y) réalisant l'égalité. Le cas indépendant de loi $\mu \otimes \nu$ ne convient pas forcément. Posons $p = 1 - d_{VT}(\mu, \nu) \in [0, 1]$ et distinguons trois cas.

- Cas où $p = 0$. On prend (X, Y) avec X et Y indépendantes de lois respectives μ et ν . Puisque $d_{VT}(\mu, \nu) = 1$, μ et ν ont des supports disjoints, d'où $\mathbb{P}(X = Y) = \sum_{x \in E} \mu(x)\nu(x) = 0$.
- Cas où $p = 1$. Alors $d_{VT}(\mu, \nu) = 0$, d'où $\mu = \nu$. On prend (X, Y) où $X \sim \mu$ et $Y = X$.
- Cas où $0 < p < 1$. Soit U, V, W des v.a. de lois $p^{-1}(\mu \wedge \nu)$, $(1-p)^{-1}(\mu - (\mu \wedge \nu))$, $(1-p)^{-1}(\nu - (\mu \wedge \nu))$. Notons que $p = \sum_{x \in E} (\mu(x) \wedge \nu(x))$ par (iv). Soit B une v.a. de loi de Bernoulli $p\delta_1 + (1-p)\delta_0$ indépendante du triplet (U, V, W) . Posons $(X, Y) = (U, U)$ si $B = 1$ et $(X, Y) = (V, W)$ si $B = 0$. On a alors $X \sim \mu$ et $Y \sim \nu$, et puisque les lois de V et W ont des supports disjoints, on a $\mathbb{P}(V = W) = 0$, et donc $\mathbb{P}(X = Y) = \mathbb{P}(B = 1) = p$. ■

A.2 Convergence étroite

Lemme A.2.1. Variation totale et convergence étroite ou en loi.

Soient μ, μ_1, μ_2, \dots des lois sur E au plus dénombrable. Les propriétés suivantes sont équivalentes :

- (i) $\lim_{n \rightarrow \infty} \int f d\mu_n = \int f d\mu$ pour toute fonction bornée $f : E \rightarrow \mathbb{R}$.
- (ii) $\lim_{n \rightarrow \infty} \mu_n(x) = \mu(x)$ pour tout $x \in E$.
- (iii) $\lim_{n \rightarrow \infty} d_{VT}(\mu_n, \mu) = 0$.
- (iv) si $X_n \sim \mu_n$ pour tout n alors $X_n \rightarrow \mu$ en loi.

En revanche, si $\lim_{n \rightarrow \infty} \mu_n(x)$ existe pour tout x alors la limite n'est pas forcément une mesure de probabilité lorsque E est infini, car la masse peut alors se disperser à l'infini. Si par exemple $E = \mathbb{N}$ alors $\frac{1}{n}(\delta_1 + \dots + \delta_n) \rightarrow 0$.

Démonstration. L'équivalence entre (i) et (iv) vient du fait que toute fonction $E \rightarrow \mathbb{R}$ est continue. Pour déduire (i) de (iii), il suffit d'utiliser l'expression variationnelle fonctionnelle de $d_{VT}(\cdot, \cdot)$. Pour déduire (ii) de (i) on peut prendre $f = \mathbb{1}_{\{x\}}$. Pour déduire (iii) de (ii) on observe que pour tout $A \subset E$,

$$\sum_{x \in E} |\mu_n(x) - \mu(x)| = \sum_{x \in A} |\mu_n(x) - \mu(x)| + \sum_{x \in A^c} |\mu_n(x) - \mu(x)|.$$

Grâce à (ii), si A est fini, alors pour tout $\varepsilon > 0$, il existe un entier $N = N(A, \varepsilon)$ tel que le premier terme du membre de droite est majoré par ε pour tout $n \geq N$. Le second terme du membre de droite peut se contrôler comme suit :

$$\sum_{x \in A^c} |\mu_n(x) - \mu(x)| \leq \sum_{x \in A^c} \mu_n(x) + \sum_{x \in A^c} \mu(x).$$

Puisqu'on a $\sum_{x \in A^c} \mu_n(x) = \sum_{x \in A} \mu(x) - \sum_{x \in A} \mu_n(x) + \sum_{x \in A^c} \mu(x)$, on obtient

$$\sum_{x \in A^c} |\mu_n(x) - \mu(x)| \leq \sum_{x \in A} |\mu_n(x) - \mu(x)| + 2 \sum_{x \in A^c} \mu(x).$$

Puisque μ est une mesure de probabilité, pour tout $\varepsilon' > 0$, on peut choisir A fini tel que $\mu(A^c) \leq \varepsilon'$. Ainsi, on obtient $\lim_{n \rightarrow \infty} \sum_{x \in E} |\mu_n(x) - \mu(x)| = 0$, qui n'est rien d'autre que (iii) d'après le lemme A.1.1. ■

Annexe B

Modèles emblématiques issus de l'informatique, de la physique, et de la biologie

Ils font appel pour la plupart aux martingales et aux chaînes de Markov.

Ils sont pratiquement tous étudiés dans [11], [5], ou [8].

- Marche aléatoire sur \mathbb{Z}^d
- Marche aléatoire sur $\mathbb{Z}/n\mathbb{Z}$
- Marche aléatoire sur l'hypercube discret $\{0, 1\}^n$
- Marche aléatoire sur le groupe symétrique et battage de cartes
- Jeu de pile ou face et processus de Bernoulli, processus de Poisson
- Processus de branchement de Galton – Watson et de Yule
- Modèle de Wright – Fisher et de Moran
- Urnes d'Ehrenfest
- Urnes de Pólya
- Processus des restaurants chinois et loi d'Ewens, coalescent de Kingman
- Processus de vie et de mort, files d'attente $M/M/k$
- Processus de renouvellement, processus semi-markoviens, fiabilité des systèmes
- Processus de taille de fenêtre TCP
- Processus d'Ornstein – Uhlenbeck
- Processus auto-régressif
- Bandit à deux bras
- Algorithme du recuit simulé
- Algorithme du gradient stochastique
- Algorithme d'approximation stochastique de Robbins – Monro
- Estimateur à noyau
- Voyageur de commerce randomisé
- Percolation de premier passage
- Cascades de Mandelbrot

Annexe C

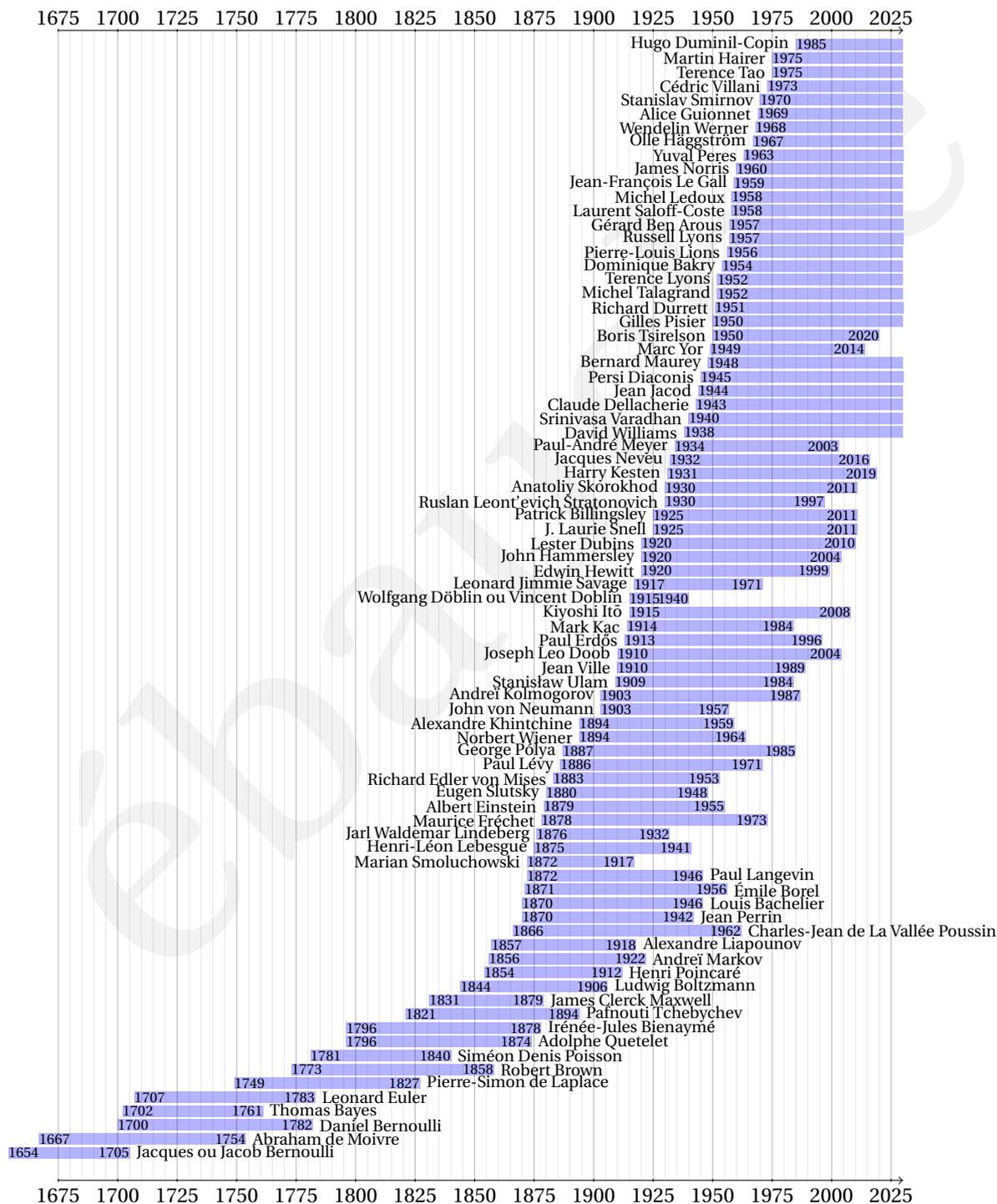
Lexique bilingue

Français	Anglais
si et seulement si (ssi)	if and only if (iff)
tribu	σ -field or σ -algebra (hence the notation \mathcal{F} and \mathcal{A} !)
indépendantes et identiquement distribuées (i.i.d.)	independent and identically distributed (i.i.d.)
pas de nom!	continuous mapping theorem
règle de dérivation en chaîne	chain rule
uniformément intégrable (u.i.)	uniformly integrable (u.i.)
temps d'arrêt	stopping time
théorème d'arrêt	stopping theorem
inégalité maximale	maximal inequality
décomposition de Doob	Doob decomposition
espérance conditionnelle	conditional expectation
propriété d'emboîtement de l'espérance conditionnelle	tower property of conditional expectation
processus arrêté	stopped process
martingale rétrograde ou renversée	backward or reversed martingale
montées et descentes	upcrossings and downcrossings
théorème limite central (TLC)	central limit theorem (CLT)
loi (forte) des grands nombres (LGN)	(strong) law of large numbers (LLN)
loi du logarithme itéré (LLI)	law of the iterated logarithm (LIL)
chaîne de Markov (CM)	Markov chain (MC)
mouvement brownien (MB)	Brownian motion (BM)
Khintchine, Liapounov, et Tchebychev	Khinchin, Lyapunov, and Chebyshev
marche paresseuse	lazy random walk
équation de bilan	balance equation
équation de bilan détaillé	detailed balance equation
file d'attente	queuing process

Annexe D

Chronologie

Quelques personnages historiques ou contemporains liés à ces notes ¹.



1. Merci à Julien Guillod pour le partage de son code Tikz.

Bibliographie

- [1] D. Aldous et J. A. Fill. Reversible markov chains and random walks on graphs. Monographie inachevée, 2002.
- [2] N. Bacaër. *Histoires de mathématiques et de populations*. Cassini, 2009.
- [3] P. Baldi, L. Mazliak, et P. Priouret. *Martingales et chaînes de Markov*. Hermann, 2008.
- [4] P. Barbe et M. Ledoux. *Probabilité*. EDP Sciences, 2007.
- [5] B. Bercu et D. Chafaï. *Modélisation stochastique et simulation : cours et applications*. Dunod, 2007.
- [6] N. Berglund. Processus aléatoires et applications - Algorithmes MCMC et vitesse de convergence. notes de cours arXiv:2412.07396v1, 2024.
- [7] P. Billingsley. *Probability and measure*. Wiley, 2012.
- [8] T. Bodineau. *Modélisation de phénomènes aléatoires*. École Polytechnique, 2023.
- [9] V. S. Borkar. *Probability theory*. Universitext. Springer, 1995. An advanced course.
- [10] P. Brémaud. *Markov chains—Gibbs fields, Monte Carlo simulation and queues*. Springer, 2020.
- [11] D. Chafaï et F. Malrieu. *Recueil de modèles aléatoires*. Springer, 2016.
- [12] D. Chafaï. Phénomènes de grande dimension. Cours de première année, École normale supérieure, 2024.
- [13] D. Chafaï et P.-A. Zitt. Probabilités - Préparation à l'agrégation interne. Notes de cours, 2020.
- [14] C. Coccozza-Thivent. *Processus stochastiques et fiabilité des systèmes*. Springer, 1997.
- [15] H. Cramér. Half a century with probability theory : some personal recollections. *Ann. Probability*, 4(4) :509–546, 1976.
- [16] D. Dacunha-Castelle et M. Duflo. *Probab. et statistiques. Tome 2*. Masson, 1983. ver. anglaise “Probab. and statistics. Vol. II.” Springer 1986.
- [17] C. Dellacherie et P.-A. Meyer. *Probabilités et potentiel. Chapitres I à IV; V à VIII; IX à XI; XII–XVI*. Hermann, Paris, 1975; 1980; 1983; 1987. Intro. générale; Th. des martingales; Th. discrète du potentiel; Th. du potentiel associée à une résolvante et Th. des processus de Markov.
- [18] J.-F. Delmas et B. Jourdain. *Modèles aléatoires*. Springer, 2006.
- [19] P. Diaconis et J. Fulman. *The mathematics of shuffling cards*. American Mathematical Society, 2023.
- [20] J. L. Doob. *Stochastic processes*. Wiley, 1953.
- [21] J. L. Doob. *Classical potential theory and its probabilistic counterpart*. Springer, 1984.
- [22] P. G. Doyle et J. L. Snell. *Random walks and electric networks*. Mathematical Association of America, 1984.
- [23] L. E. Dubins et L. J. Savage. *Inequalities for stochastic processes (how to gamble if you must)*. Dover, 1976. Corrected repub. 1965 ed.
- [24] M. Duflo. *Random iterative models*. Springer, 1997.
- [25] R. Durrett. *Probability—theory and examples*. Cambridge University Press, 2019.
- [26] P. Ehrenfest et T. Ehrenfest. *The conceptual foundations of the statistical approach in mechanics*. Dover, 1990 (initialement 1912).
- [27] M. Fathi. Processus discrets et théorèmes limites. Notes de cours de Master 1, 2020.
- [28] W. Feller. *An introduction to probability theory and its applications. Vol. I et II*. Wiley, 1968; 1971.
- [29] P. A. Ferrari et A. Galves. Construction of stochastic processes, coupling and regeneration. Notes de cours, monographie inachevée, 2000.
- [30] G. R. Grimmett et G. F. Lawler. Harry Kesten (1931–2019) : a personal and scientific tribute. *Notices Amer. Math. Soc.*, 67(6) :822–831, 2020.
- [31] G. R. Grimmett et D. R. Stirzaker. *Probability and random processes*. Oxford University Press, fourth edition, 2020.
- [32] A. Gut. *Probability : a graduate course*. Springer, New York, 2013.
- [33] O. Häggström. *Finite Markov chains and algorithmic applications*. Cambridge University Press, 2002.
- [34] P. Hall et C. C. Heyde. *Martingale limit theory and its application*. Academic Press, 1980.
- [35] C. C. Heyde et E. Seneta, editors. *Statisticians of the centuries*. Springer, 2001.
- [36] J. Jacod et P. Protter. *Probability essentials*. Springer, 2003.
- [37] J. Jacod et A. N. Shiryaev. *Limit theorems for stochastic processes*, volume 288. Springer, 2003.
- [38] M. Kac. *Enigmas of chance*. Harper & Row, 1985.
- [39] O. Kallenberg. *Foundations of modern probability*. Springer, 2021.
- [40] J.-M. Kantor et L. Graham. *Au nom de l'infini. Une histoire vraie de mysticisme religieux et de création mathématique*. Belin, Paris, 2010.
- [41] F. P. Kelly. *Reversibility and stochastic networks*. Cambridge University Press, 2011.
- [42] A. Klenke. *Probability theory—a comprehensive course*. Springer, 2020.
- [43] J.-F. Le Gall. *Intégration, Probabilités et Processus Aléatoires*. École normale supérieure, 2006. Version anglaise : « Measure theory, probability, and stochastic processes » (2022).
- [44] D. A. Levin et Y. Peres. *Markov chains and mixing times*. American Mathematical Society, 2017.
- [45] P. Lévy. *Processus stochastiques et mouvement brownien*. Gauthier-Villars, 1965. Suivi d'une note de M. Loève, Deuxième édition.
- [46] R. Lyons et Y. Peres. *Probability on trees and networks*. Cambridge University Press, 2016.
- [47] P. Lévy. *Quelques aspects de la vie d'un mathématicien*. Librairie Armand Blanchard, 1970.
- [48] A. P. Maitra et W. D. Sudderth. *Discrete gambling and stochastic games*. Springer, 1996.

- [49] S. Meyn et R. L. Tweedie. *Markov chains and stochastic stability*. Cambridge University Press, 2009.
- [50] M. Mitzenmacher et E. Upfal. *Probab. and computing. Randomization and probab. techniques in algorithms and data analysis*. CUP, 2017.
- [51] R. Motwani et P. Raghavan. *Randomized algorithms. Theory of computation handbook*. CRC, 1999.
- [52] J. Neveu. *Bases mathématiques du calcul des probabilités; Martingales à temps discret*. Masson, 1964; 1972.
Version anglaise : « Mathematical foundations of the calculus of probability » (1965); « Discrete-parameter martingales » (1975).
- [53] J. R. Norris. *Markov chains*. Cambridge University Press, 1997.
- [54] J. Perrin. *Les Atomes (1913)*. CNRS Éditions, 2014.
- [55] D. Revuz. *Markov chains*. North-Holland, 1984.
- [56] P. Robert. *Réseaux et files d'attente : méthodes probabilistes*. Springer, 2000.
- [57] E. Seneta. *Non-negative matrices and Markov chains*. Springer, 2006.
- [58] J. Sjöstrand et M. Zworski. Elementary linear algebra for advanced spectral problems. *Ann. Inst. Fourier*, 57(7) :2095–2141, 2007.
- [59] J. L. Snell. A conversation with Joe Doob. *Statist. Sci.*, 12(4) :301–311, 1997.
- [60] S. M. Ulam. *Adventures of a mathematician*. Scribner, 1976.
- [61] D. Williams. *Probability with martingales*. Cambridge University Press, 1991.

ébauche