

Modèles stochastiques

Djalil Chafaï (Paris-Est Marne-la-Vallée, puis Paris-Dauphine)
Florent Malrieu (Rennes, puis Tours)

Notes de cours 2013-2014
Master 2 Mathématiques
(7+4) séances de 3 heures



Table des matières

1	Jeu de pile ou face et collectionneur de coupons	7
1.1	Jeu de pile ou face	7
1.2	Approximation de la loi binomiale par la loi normale	8
1.3	Distance en variation totale	9
1.4	Approximation de la loi binomiale par la loi de Poisson	11
1.5	Collectionneur de coupons	12
1.6	Notes et commentaires	16
2	Modèle de Fisher-Wright	17
2.1	Modèle de Fisher-Wright et fixation	18
2.2	Modèle de Fisher-Wright avec mutations	21
2.3	Modèle de Fisher-Wright avec sélection	22
2.4	Modèle de Cannings	23
2.5	Modèle de Moran	25
2.6	Notes et commentaires	26
3	Processus des restaurants chinois	27
3.1	Lois d'Ewens	28
3.2	Nombre de tables	30
3.3	Tables extrêmes	32
3.4	Compléments de combinatoire	33
3.5	Notes et commentaires	34
4	Processus de branchement de Galton-Watson	37
4.1	Extinction et phénomène de seuil	38
4.2	Étude des trois cas possibles	41
4.3	Taille de l'arbre en régimes critique et sous-critique	44
4.4	Immigration	46
4.5	Notes et commentaires	47
5	Marches aléatoires	49
5.1	Marche aléatoire simple sur \mathbb{Z}	49
5.2	Marche aléatoire simple symétrique sur \mathbb{Z}^d	52
5.3	Une marche aléatoire sur le groupe symétrique Σ_r	54
5.4	Notes et commentaires	57
6	Matrices aléatoires et loi du demi-cercle de Wigner	59
6.1	Théorème de Wigner	60
6.2	Convergence des moments et convergence étroite	61
6.3	Moments de la loi du demi-cercle	63

6.4	Convergence des moments et théorème de Wigner faible	64
6.5	Retour aux matrices de covariance empiriques	65
6.6	Notes et commentaires	66
7	Processus TCP window-size	69
7.1	Processus TCP window-size en informatique	69
7.2	Intensité des sauts variable selon la position	73
7.3	Branchement et croissance-fragmentation	74
7.4	Notes et commentaires	76
8	Processus ponctuels de Poisson	77
8.1	Processus ponctuels de Poisson	78
8.2	Processus ponctuels de Poisson sur \mathbb{R}_+	83
8.3	Processus auto-excités	83
9	Généalogies et coalescence	85
9.1	Modèle généalogique à temps continu	87
9.2	Longueur de l'arbre généalogique	88
9.3	Mutations	90
9.4	Coalescent de Kingman	92
9.5	Notes et commentaires	93
10	Simulation de lois discrètes	95
10.1	Algorithme basique pour les lois discrètes	95
10.2	Permutations aléatoires	97
10.3	Partitions aléatoires	98
10.4	Algorithme de Metropolis-Hastings	99
10.5	Algorithme de Propp-Wilson	101
10.6	Notes et commentaires	103
11	Problème du voyageur de commerce	105
11.1	Concentration pour le cas uniforme	106
11.2	Évaluation de la moyenne du cas uniforme	108
11.3	Preuve du cas uniforme	111
11.4	Courbes qui remplissent le plan	111
11.5	Notes et commentaires	112
	Bibliographie	113

Avant propos

Le but de ce cours est de contribuer à la construction d'une culture probabiliste classique mais éclectique, basée sur des modèles, des problèmes, des outils en action. Chacun des onze chapitres est consacré à un modèle stochastique particulier. Les sept premiers chapitres correspondent aux sept séances de cours de trois heures. Les quatre derniers chapitres peuvent servir de support pour une évaluation par projet. Certaines séances débutent par quinze minutes d'exposé d'un étudiant sur un passage peu ou pas détaillé lors des séances précédentes. Le choix des modèles abordés peut varier d'une année sur l'autre, ce qui rend ce cours plutôt flexible et renouvelable. Ces notes de cours 2014-2015 ont bénéficié des remarques et questionnements des étudiants de plusieurs promotions des Master de mathématiques des universités Paris-Est Marne-la-Vallée, Rennes, et Paris-Dauphine.

Djalil Chafaï
Marne-la-Vallée et Vincennes
Printemps 2014

Séance 1

Jeu de pile ou face et collectionneur de coupons

Mots-clés. Combinatoire ; lois discrètes ; théorèmes limites.

Le jeu de pile ou face et le collectionneur de coupons constituent deux modèles probabilistes fondamentaux qu'il est bon de connaître et savoir reconnaître.

1.1 Jeu de pile ou face

Ce jeu consiste en des lancers successifs d'une pièce de monnaie qui donnent à chaque fois soit pile (succès, codé 1) soit face (échec, codé 0). On modélise cela par une suite $(X_n)_{n \geq 1}$ de variables aléatoires i.i.d. de loi de Bernoulli :

$$\mathbb{P}(X_n = 1) = 1 - \mathbb{P}(X_n = 0) = p \in [0, 1].$$

Le nombre de succès dans les n premiers lancers $S_n = X_1 + \dots + X_n$ suit la loi binomiale $\text{Bin}(n, p)$ de taille n et de paramètre p , donnée pour tout $k = 0, 1, \dots, n$ par

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

On a

$$\mathbb{E}(S_n) = np \quad \text{et} \quad \text{Var}(S_n) = np(1-p).$$

Si $p > 0$, alors le nombre de lancers pour obtenir un succès $T = \inf\{n \geq 1 : X_n = 1\}$ suit la loi géométrique $\text{Geo}(p)$ sur \mathbb{N}^* de paramètre p donnée pour tout $k \in \mathbb{N}^*$ par

$$\mathbb{P}(T = k) = (1-p)^{k-1} p.$$

On a $T \equiv \infty$ si $p = 0$ et $\mathbb{P}(T < \infty) = 1$ sinon. On a

$$\mathbb{E}(T) = \frac{1}{p} \quad \text{et} \quad \text{Var}(T) = \frac{1-p}{p^2}.$$

Le nombre d'échecs avant le premier succès $T' = \inf\{n \geq 0 : X_{n+1} = 1\} = T - 1$ suit la loi géométrique $\text{Geo}_{\mathbb{N}}(p)$ sur \mathbb{N} et de paramètre p donnée pour tout $k \in \mathbb{N}$ par

$$\mathbb{P}(T' = k) = \mathbb{P}(T - 1 = k) = (1-p)^k p$$

et on a

$$\mathbb{E}(T') = \mathbb{E}(T) - 1 = \frac{1-p}{p} \quad \text{et} \quad \text{Var}(T') = \text{Var}(T) = \frac{1-p}{p^2}.$$

Pour tout $r \in \mathbb{N}^*$, le nombre de lancers T_r nécessaires pour obtenir r succès est défini par récurrence par $T_1 = T$ et $T_{r+1} = \inf\{n > T_r : X_n = 1\}$. Les variables aléatoires $T_1, T_2 - T_1, T_3 - T_2, \dots$ sont i.i.d. de loi géométrique $\text{Geo}(p)$. La variable aléatoire T_r suit la loi de Pascal ou loi binomiale-négative $\text{Geo}(p)^{*r}$. On a pour tout $k \geq r$,

$$\mathbb{P}(T_r = k) = \sum_{\substack{k_1 \geq 1, \dots, k_r \geq 1 \\ k_1 + \dots + k_r = k}} (1-p)^{k_1-1} p \cdots (1-p)^{k_r-1} p = (1-p)^{k-r} p^r \binom{k-1}{r-1}$$

et

$$\mathbb{E}(T_r) = r\mathbb{E}(T) = \frac{r}{p} \quad \text{et} \quad \text{Var}(T_r) = r\text{Var}(T) = r \frac{1-p}{p^2}.$$

Le processus de Bernoulli $(S_n)_{n \geq 0}$ a des trajectoires constantes par morceaux, avec des sauts d'amplitude $+1$, et les temps de saut sont donnés par $(T_r)_{r \geq 1}$ (temps inter-sauts i.i.d. géométriques). Il constitue le processus de comptage de tops espacés par des durées indépendantes de même loi géométrique (analogue discret du processus de Poisson). Comme S_n est une somme de variables indépendantes, $(S_n)_{n \geq 0}$ est une chaîne de Markov sur \mathbb{N} de noyau $\mathbf{P}(x, y) = p\mathbb{1}_{y=x+1} + (1-p)\mathbb{1}_{y=x}$, et $(S_n - np)_{n \geq 0}$ est une martingale.

La loi forte des grands nombres et le théorème central limite s'écrivent ici

$$\frac{S_n}{n} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} p \quad \text{et} \quad \frac{\sqrt{n}}{\sqrt{p(1-p)}} \left(\frac{S_n}{n} - p \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

Cela donne un intervalle de confiance asymptotique pour p appelé intervalle de Wald (assez mauvais en pratique). Il est également possible de confectionner des intervalles de confiance pour p non asymptotiques, comme celui de Clopper-Pearson par exemple, basé sur la correspondance beta-binomiale : si U_1, \dots, U_n sont i.i.d. de loi uniforme sur $[0, 1]$ et si $U_{(1)} \leq \dots \leq U_{(n)}$ est leur réordonnement croissant alors $U_{(k)}$ suit la loi $\text{Beta}(k, n-k+1)$ sur $[0, 1]$ de densité $t \in [0, 1] \mapsto \left(\int_0^1 s^{k-1} (1-s)^{n-k} ds \right)^{-1} t^{k-1} (1-t)^{n-k}$, et

$$\mathbb{P}(S_n \geq k) = \mathbb{P}(\mathbb{1}_{\{U_1 \leq p\}} + \dots + \mathbb{1}_{\{U_n \leq p\}} \geq k) = \mathbb{P}(U_{(k)} \leq p).$$

Remarque 1.1.1 (Motifs répétés et lois du zéro-un). *Lorsque $0 < p < 1$, le second lemme de Borel-Cantelli (cas indépendant) entraîne que toute suite finie de 0 et de 1 apparaît presque sûrement une infinité de fois dans la suite X_1, X_2, \dots . L'indépendance est capitale.*

Remarque 1.1.2 (Lien avec la loi uniforme sur $[0, 1]$). *Une variable aléatoire U suit la loi uniforme sur $[0, 1]$ si et seulement si les bits de son écriture en base 2 sont des variables aléatoires de Bernoulli indépendantes et de paramètre $1/2$.*

1.2 Approximation de la loi binomiale par la loi normale

Soit $(X_n)_{n \geq 1}$ des v.a.r. i.i.d. de moyenne m et variance $0 < \sigma^2 < \infty$. On pose $S_n = X_1 + \dots + X_n$. Le théorème central limite indique que pour tout $t \in \mathbb{R}$,

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{S_n - nm}{\sqrt{ns}} \leq t \right) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

Cela suggère que la loi de S_n est proche de $\mathcal{N}(nm, n\sigma^2)$ lorsque n est grand. Le théorème de Berry-Esseen précise cette proximité en affirmant que pour tout $t \in \mathbb{R}$ et $n \geq 1$,

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P} \left(\frac{S_n - nm}{\sqrt{n}\sigma} \leq t \right) - \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \right| \leq \frac{\tau^3}{\sqrt{n}\sigma^3}$$

où $\tau^3 = \mathbb{E}(|X_1 - \mathbb{E}(X_1)|^3)$. Lorsque X_1, X_2, \dots sont des variables de Bernoulli de paramètre $p \in]0, 1[$, on trouve $m = p$, $\sigma^2 = p(1-p)$ et $\tau^3 = p(1-p)(1-2p(1-p))$ ce qui donne

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P} \left(S_n \leq \sqrt{np(1-p)}t + np \right) - \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \right| \leq \frac{1-2p(1-p)}{\sqrt{p(1-p)}\sqrt{n}}.$$

Cette approximation de la loi binomiale par la loi normale est d'autant meilleure que $(1-2p(1-p))/\sqrt{np(1-p)}$ est petit. À n fixé, cette borne est minimale pour $p = 1/2$ mais explose quand p se rapproche de 0 ou de 1.

1.3 Distance en variation totale

Dans toute cette section, E est un ensemble au plus dénombrable muni de la topologie et de la tribu discrètes $\mathcal{P}(E)$. L'ensemble des lois sur E est un espace métrique complet pour la *distance en variation totale*

$$d_{\text{VT}}(\mu, \nu) := \sup_{A \subset E} |\mu(A) - \nu(A)|.$$

Notons que $d_{\text{VT}}(\mu, \nu) = \|\mu - \nu\|_{\text{VT}}$ où $\|\eta\|_{\text{VT}} := \sup_{A \subset E} |\eta(A)|$ pour toute mesure signée η de masse finie sur E . On a $0 \leq d_{\text{VT}}(\mu, \nu) \leq 1$ et $d_{\text{VT}}(\mu, \nu) = 1$ si μ et ν ont des supports disjoints.

Théorème 1.3.1 (Expressions alternatives). *Si μ et ν sont des lois sur E alors*¹

$$d_{\text{VT}}(\mu, \nu) = \frac{1}{2} \sup_{f: E \rightarrow [-1,1]} \left| \int f d\mu - \int f d\nu \right| = \frac{1}{2} \sum_{x \in E} |\mu(x) - \nu(x)|.$$

De plus, le supremum dans la définition de $d_{\text{VT}}(\cdot, \cdot)$ est atteint pour l'ensemble

$$A_* = \{x \in E : \mu(x) \geq \nu(x)\}$$

tandis que dans l'expression variationnelle fonctionnelle de $d_{\text{VT}}(\cdot, \cdot)$ il est atteint pour

$$f = \mathbb{1}_{A_*} - \mathbb{1}_{A_*^c}.$$

Démonstration. La seconde égalité provient de l'inégalité

$$\left| \int f d\mu - \int f d\nu \right| \leq \sum_{x \in E} |f(x)| |\mu(x) - \nu(x)| \leq \sup_{x \in E} |f(x)| \sum_{x \in E} |\mu(x) - \nu(x)|$$

qui est saturée pour $f = \mathbb{1}_{A_*} - \mathbb{1}_{A_*^c}$. Pour la première égalité, on écrit

$$|\mu(A) - \nu(A)| = \frac{1}{2} \left| \int f_A d\mu - \int f_A d\nu \right|$$

où $f = \mathbb{1}_A - \mathbb{1}_{A^c}$, ce qui donne

$$|\mu(A) - \nu(A)| \leq \frac{1}{2} \sup_{f: E \rightarrow [-1,1]} \left| \int f d\mu - \int f d\nu \right| = \frac{1}{2} \sum_{x \in E} |\mu(x) - \nu(x)|$$

1. En particulier on a $2\|\cdot\|_{\text{VT}} = \|\cdot\|_{\ell^1(E, \mathbb{R})}$.

qui est saturée pour $A = A_*$ car

$$2|\mu(A_*) - \nu(A_*)| = \sum_{x \in A_*} |\mu(x) - \nu(x)| + \sum_{x \in A_*^c} |\mu(x) - \nu(x)|.$$

□

Théorème 1.3.2 (Convergence en loi). *Si (X_n) est une suite de variables aléatoires sur E et si μ_n désigne la loi de X_n , alors pour toute loi μ sur E , les propriétés suivantes sont équivalentes :*

1. $\lim_{n \rightarrow \infty} \int f d\mu_n = \int f d\mu$ pour toute fonction bornée $f : E \rightarrow \mathbb{R}$;
2. $\lim_{n \rightarrow \infty} \mu_n(x) = \mu(x)$ pour tout $x \in E$;
3. $\lim_{n \rightarrow \infty} d_{VT}(\mu_n, \mu) = 0$.

Lorsqu'elles ont lieu on dit que (X_n) converge en loi vers μ quand $n \rightarrow \infty$.

Comme E est muni de la topologie discrète, toutes les fonctions $E \rightarrow \mathbb{R}$ sont continues.

Démonstration. Pour déduire 1. de 3. il suffit d'utiliser l'expression variationnelle fonctionnelle de $d_{VT}(\cdot, \cdot)$. Pour déduire 2. de 1. on peut prendre $f = \mathbb{1}_{\{x\}}$. Pour déduire 3. de 2. 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 2., 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 $n \geq N$. Par ailleurs, on peut contrôler le second terme du membre de droite de la manière suivante :

$$\sum_{x \in A^c} |\mu_n(x) - \mu(x)| \leq \sum_{x \in A^c} \mu_n(x) + \sum_{x \in A^c} \mu(x).$$

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 $\mu \in \mathcal{P}$, 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 3. d'après le théorème 1.3.1. □

Remarque 1.3.3 (Dispersion à l'infini). *Si (μ_n) sont des lois et $\mu(x) := \lim_{n \rightarrow \infty} \mu_n(x)$ alors μ n'est pas forcément une loi, sauf si E est fini. En effet, lorsque E est infini, il peut se produire un phénomène de dispersion de la masse à l'infini. Contre exemple : $E = \mathbb{N}$ et μ_n affecte la masse $1/n$ aux singletons $\{1\}, \dots, \{n\}$, ce qui donne μ identiquement nulle.*

Théorème 1.3.4 (Autre expression et cas extrémal). *Si μ et ν sont des lois sur E alors*

$$d_{VT}(\mu, \nu) = 1 - \sum_{x \in E} (\mu(x) \wedge \nu(x)).$$

En particulier, $d_{VT}(\mu, \nu) = 1$ si et seulement si μ et ν ont des support disjoints.

Démonstration. Il suffit d'écrire

$$\sum_{x \in E} (\mu(x) \wedge \nu(x)) = \frac{1}{2} \sum_{x \in E} (\mu(x) + \nu(x) - |\mu(x) - \nu(x)|) = 1 - d_{\text{VT}}(\mu, \nu).$$

□

Théorème 1.3.5 (Couplage). *Si μ et ν sont des lois sur E alors*

$$d_{\text{VT}}(\mu, \nu) = \inf_{(X, Y)} \mathbb{P}(X \neq Y)$$

où l'infimum porte sur les couples de v.a. sur $E \times E$ de lois marginales μ et ν . De plus, il existe un couple de ce type pour lequel l'égalité est atteinte (i.e. l'infimum est un minimum).

Démonstration. Soit (X, Y) un couple de variables aléatoires sur $E \times E$ de lois marginales μ et ν . Comme $\mathbb{P}(X = x, Y = x) \leq \mu(x) \wedge \nu(x)$ pour tout $x \in E$ on a

$$1 - d_{\text{VT}}(\mu, \nu) = \sum_{x \in E} (\mu(x) \wedge \nu(x)) \geq \sum_{x \in E} \mathbb{P}(X = x, Y = x) = \mathbb{P}(X = Y).$$

Il suffit donc de construire un couple (X, Y) pour lequel l'égalité est atteinte. Posons

$$p = 1 - d_{\text{VT}}(\mu, \nu) \in [0, 1].$$

Cas où $p = 0$. On a alors $d_{\text{VT}}(\mu, \nu) = 1$ et μ et ν ont des supports disjoints. Cela donne $\mathbb{P}(X = Y) = \sum_{x \in E} \mu(x)\nu(x) = 0$. On prend (X, Y) avec $X \sim \mu$ et $Y \sim \nu$ indépendantes.

Cas où $p = 1$. On a alors $d_{\text{VT}}(\mu, \nu) = 0$ et donc $\mu = \nu$. On prend (X, X) où $X \sim \mu$.

Cas où $0 < p < 1$. Soit (U, V, W) un triplet de variables aléatoires de lois respectives

$$p^{-1}(\mu \wedge \nu), \quad (1 - p)^{-1}(\mu - (\mu \wedge \nu)), \quad (1 - p)^{-1}(\nu - (\mu \wedge \nu)).$$

Notons que $p = \sum_{x \in E} (\mu(x) \wedge \nu(x))$. Soit B une variable aléatoire de loi de Bernoulli $\text{Bin}(1, p)$ indépendante de (U, V, W) . Définissons $(X, Y) = (U, U)$ 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 $\mathbb{P}(X = Y) = \mathbb{P}(B = 1) = p$. □

1.4 Approximation de la loi binomiale par la loi de Poisson

Si S_n suit la loi binomiale $\text{Bin}(n, p)$ alors pour tout $k \in \mathbb{N}$, on a,

$$\mathbb{P}(S_n = k) - e^{-np} \frac{(np)^k}{k!} = \left(\frac{n}{n(1-p)} \cdots \frac{n-k+1}{n(1-p)} (1-p)^n - e^{-np} \right) \frac{(np)^k}{k!}.$$

Ceci montre que si p dépend de n avec $\lim_{n \rightarrow \infty} np = \lambda$ alors la loi de S_n tend vers $\text{Poi}(\lambda)$. La distance en variation permet de quantifier cette convergence en loi : l'inégalité de poissonisation de Le Cam du théorème 1.4.1 ci-dessous donne (utile si np^2 est petit)

$$\sum_{k=0}^{\infty} \left| \mathbb{P}(S_n = k) - e^{-np} \frac{(np)^k}{k!} \right| \leq 2np^2.$$

Théorème 1.4.1 (Inégalité de poissonisation de Le Cam). *Soient X_1, \dots, X_n des v.a. indépendantes de lois de Bernoulli $\text{Bin}(1, p_1), \dots, \text{Bin}(1, p_n)$. Soit μ_n la loi de $S_n = X_1 + \dots + X_n$ et soit $\nu_n = \text{Poi}(p_1 + \dots + p_n)$ la loi de Poisson de même moyenne que S_n . Alors on a*

$$d_{\text{VT}}(\mu_n, \nu_n) \leq p_1^2 + \dots + p_n^2.$$

Démonstration. On commence par établir par récurrence sur n que si $\alpha_1, \dots, \alpha_n$ et β_1, \dots, β_n sont des lois de probabilité sur \mathbb{N} alors on a l'inégalité sous-additive

$$d_{VT}(\alpha_1 * \dots * \alpha_n, \beta_1 * \dots * \beta_n) \leq d_{VT}(\alpha_1, \beta_1) + \dots + d_{VT}(\alpha_n, \beta_n).$$

Ensuite on établit que $d_{VT}(\text{Bin}(1, p), \text{Poi}(p)) \leq p^2$. Rappel : $\text{Poi}(a) * \text{Poi}(b) = \text{Poi}(a + b)$. \square

Notons que $p_1^2 + \dots + p_n^2 \leq (p_1 + \dots + p_n) \max_{1 \leq k \leq n} p_k$. Cela permet de retrouver la loi des petits nombres : si $(X_{n,k})_{1 \leq k \leq n}$ est un tableau triangulaire de v.a.r. indépendantes de lois de Bernoulli avec $X_{n,k} \sim \text{Bin}(1, p_{n,k})$ pour tous $n \geq k \geq 1$, et si

$$\lim_{n \rightarrow \infty} p_{n,1} + \dots + p_{n,n} = \lambda \quad \text{et} \quad \lim_{n \rightarrow \infty} \max_{1 \leq k \leq n} p_{n,k} = 0,$$

alors $X_{n,1} + \dots + X_{n,n}$ converge en loi vers $\text{Poi}(\lambda)$ quand $n \rightarrow \infty$.

1.5 Collectionneur de coupons

Le collectionneur de coupons constitue un modèle stochastique fondamental important, à ranger dans la même boîte à outils que le jeu de pile ou face, auquel il est intimement relié. Un grand nombre de situations concrètes sont modélisables par le collectionneur de coupons ou une de ses variantes. Nous nous limitons ici à la variante la plus simple.

Il faut jouer un nombre de fois (aléatoire) géométrique à pile ou face pour voir apparaître les deux côtés de la pièce. Si on remplace la pièce de monnaie par un dé à $r \geq 2$ faces, combien de fois faut-il lancer le dé pour voir apparaître les r faces différentes ? On modélise cela, pour un entier fixé $r \geq 2$, en considérant la variable aléatoire

$$T = \min\{n \geq 1 : \{X_1, \dots, X_n\} = \{1, \dots, r\}\} = \min\{n \geq 1 : \text{card}\{X_1, \dots, X_n\} = r\}$$

où $(X_n)_{n \geq 1}$ est une suite de variables aléatoires i.i.d. de loi uniforme sur $\{1, \dots, r\}$. La variable aléatoire T est le temps de complétion de la collection. Le nom *collectionneur de coupons* provient des coupons à collectionner présents dans certains paquets de céréales.

Théorème 1.5.1 (Expression combinatoire de la loi). *On a $T \geq r$ et pour tout $n \geq r$,*

$$\mathbb{P}(T = n) = \frac{r!}{r^n} \left\{ \begin{matrix} n-1 \\ r-1 \end{matrix} \right\}$$

où la notation en accolades désigne le nombre de Stirling de seconde espèce².

Démonstration. On a $X_T \notin \{X_1, \dots, X_{T-1}\}$ car le coupon qui termine la collection n'a forcément jamais été vu auparavant. Si on fixe $n \geq r$, l'événement $\{T = n\}$ correspond à choisir le type du dernier coupon puis à répartir les $n - 1$ coupons restants sur les $r - 1$ types restants. Le résultat désiré en découle car la loi des types est uniforme. \square

Le théorème 1.5.1 n'est malgré tout pas très parlant. Le résultat intuitif suivant va beaucoup nous aider à étudier T , et montre en particulier que $\mathbb{P}(T < \infty) = 1$.

Lemme 1.5.2 (Décomposition). *On a $T = G_1 + \dots + G_r$ où G_1, \dots, G_r sont des variables aléatoires indépendantes avec $G_i \sim \text{Geo}(\pi_i)$ où $\pi_i := (r - i + 1)/r$. En particulier, on a*

$$\mathbb{E}(T) = r(\log(r) + \gamma) + o_{r \rightarrow \infty}(r) \quad \text{et} \quad \text{Var}(T) = \frac{\pi^2}{6} r^2 + o_{r \rightarrow \infty}(r^2),$$

où $\gamma = \lim_{n \rightarrow \infty} (\sum_{i=1}^n 1/i - \log(n)) \approx 0.577$ est la constante d'Euler.

2. Nombre de manières de partitionner un ensemble à $n - 1$ éléments en $r - 1$ sous-ensembles non vides.

Démonstration. On pose $G_1 \equiv 1$ et pour tout $1 < i \leq r$,

$$G_i = \min\{n \geq 1 : X_{G_{i-1}+n} \notin \{X_1, \dots, X_{G_{i-1}}\}\}.$$

On a $\text{card}(\{X_1, \dots, X_{G_i}\}) = i$ pour tout $1 \leq i \leq n$. Les variables aléatoires $G_1, G_1 + G_2, \dots, G_1 + \dots + G_r$ sont les temps d'apparition des r premiers succès dans un jeu de pile ou face spécial dans lequel la probabilité de gagner change après chaque succès : cette probabilité vaut successivement $\pi_1 = 1, \pi_2 = (r-1)/r, \pi_3 = (r-2)/r, \dots, \pi_r = 1/r$. Il est de plus en plus difficile d'obtenir un coupon d'un nouveau type au fil de la collection.

La linéarité de l'espérance donne

$$\mathbb{E}(T) = \sum_{i=1}^r \mathbb{E}(G_i) = \sum_{i=1}^r \frac{1}{\pi_i} = \sum_{i=1}^r \frac{r}{r-i+1} = r \sum_{i=1}^r \frac{1}{i} = r(\log(r) + \gamma + o_{r \rightarrow \infty}(1)).$$

D'autre part, l'indépendance des variables aléatoires G_1, \dots, G_r (exercice !) donne

$$\text{Var}(T) = \sum_{i=1}^r \text{Var}(G_i) = \sum_{i=1}^r \frac{1-\pi_i}{\pi_i^2} = r \sum_{i=1}^{r-1} \frac{r-i}{i^2} = \frac{\pi^2}{6} r^2 - r(\log(r) + \gamma) + o_{r \rightarrow \infty}(r^2).$$

□

Théorème 1.5.3 (Queue de distribution). *Pour tout $n \geq 1$,*

$$\mathbb{P}(T > n) = \sum_{k=1}^r (-1)^{k-1} \binom{r}{k} \left(1 - \frac{k}{r}\right)^n.$$

Démonstration. On a

$$\{T > n\} = E_{n,1} \cup \dots \cup E_{n,r} \quad \text{où} \quad E_{n,i} := \{X_1 \neq i, \dots, X_n \neq i\}.$$

Si i_1, \dots, i_k sont des éléments distincts de $\{1, \dots, r\}$ alors, en notant $R = \{1, \dots, r\} \setminus \{i_1, \dots, i_k\}$,

$$\mathbb{P}(E_{n,i_1} \cap \dots \cap E_{n,i_k}) = \mathbb{P}(X_1 \in R) \cdots \mathbb{P}(X_n \in R) = \left(\frac{r-k}{r}\right)^n = \left(1 - \frac{k}{r}\right)^n.$$

Le résultat désiré découle alors du principe d'inclusion-exclusion. □

Il est délicat de déduire le comportement de la queue de T en fonction de n et r à partir du théorème 1.5.3 en raison de l'alternance des signes.

Théorème 1.5.4 (Déviation). *Pour tout entier $m \geq 0$,*

$$\mathbb{P}(T > rm + r \log(r)) = \mathbb{P}\left(\frac{T - r \log(r)}{r} > m\right) \leq e^{-m}.$$

Démonstration. Pour tout entier $n \geq 1$, on peut écrire

$$\mathbb{P}(T > n) = \mathbb{P}(\cup_{i=1}^r E_{n,i}) \leq \sum_{i=1}^r \mathbb{P}(E_{n,i}) \quad \text{où} \quad E_{n,i} = \{X_1 \neq i, \dots, X_n \neq i\}.$$

Comme $\mathbb{P}(E_{n,i}) = (1 - 1/r)^n \leq e^{-n/r}$, le choix $n = rm + r \log(r)$ mène au résultat. □

Pour $\alpha = 0.05$ et r fixé, on peut choisir m assez grand pour que $e^{-m} \leq \alpha$, par exemple $m = -\log(\alpha)$, ce qui donne l'intervalle de prédiction $[r, r \log(r) + rm]$ de niveau α pour T .

Théorème 1.5.5 (Comportement asymptotique).

$$\frac{T}{r \log(r)} \xrightarrow[r \rightarrow \infty]{\mathbb{P}} 1.$$

Démonstration. Pour tout $t > 0$, l'inégalité de Markov et le lemme 1.5.2 donnent

$$\begin{aligned} \mathbb{P}\left(\left|\frac{T}{r \log(r)} - 1\right| > t\right) &\leq \frac{\mathbb{E}((T - r \log(r))^2)}{t^2 r^2 \log(r)^2} \\ &= \frac{\text{Var}(T) + (\mathbb{E}(T) - r \log(r))^2}{t^2 r^2 \log(r)^2} \\ &= \mathcal{O}_{r \rightarrow \infty}\left(\frac{1}{\log(r)^2}\right). \end{aligned}$$

Dans le jargon probabiliste, cette approche est une « méthode de second moment ». \square

La borne établie dans la preuve du théorème précédent n'est pas sommable en r et ne permet donc pas de démontrer une convergence presque sûre en utilisant le lemme de Borel-Cantelli³. D'autre part, la borne établie permet d'obtenir un intervalle de prédiction : pour $\alpha = 0.05$, r fixé, et t bien choisi, on a $\mathbb{P}(|T - r \log(r)|/r > t) = \mathcal{O}(1/t^2) = 1 - \alpha$. L'intervalle de prédiction est de largeur $2rt$, et se dégrade quand t croît (α diminue).

Le théorème suivant affirme que les fluctuations asymptotiques dans la convergence précédente suivent une loi de Gumbel de fonction de répartition $t \in \mathbb{R} \mapsto e^{-e^{-t}}$. Par le lemme de Slutsky et le comportement des deux premiers moments, $(T - \mathbb{E}(T))/\sqrt{\text{Var}(T)}$ converge quand $r \rightarrow \infty$ vers une loi de Gumbel translatée de γ . Cela suggère que la loi de Gumbel est de moyenne γ , ce qui est tout à fait vrai, bien que son mode soit à zéro.

Théorème 1.5.6 (Fluctuations asymptotiques). *On a*

$$\frac{T - r \log(r)}{r} = \log(r) \left(\frac{T}{r \log(r)} - 1 \right) \xrightarrow[r \rightarrow \infty]{\text{loi}} \text{Gumbel}.$$

Démonstration. Il suffit d'établir que pour tout $t \in \mathbb{R}$ on a

$$\lim_{r \rightarrow \infty} \mathbb{P}(T > r \log(r) + tr) = S(t) := 1 - e^{-e^{-t}}.$$

Fixons donc $t \in \mathbb{R}$ et supposons que r est assez grand pour que $r \log(r) + tr > r$. Introduisons l'entier $n_{t,r} = \lceil r \log(r) + tr \rceil$ (entier supérieur ou égal). Le théorème 1.5.3 donne

$$\mathbb{P}(T > r \log(r) + tr) = \sum_{k=1}^r (-1)^{k-1} \binom{r}{k} \left(1 - \frac{k}{r}\right)^{n_{t,r}}.$$

Comme $\binom{r}{k} \leq r^k/k!$ et $1 - u \leq e^{-u}$ pour tout $u \geq 0$, on a

$$\binom{r}{k} \left(1 - \frac{k}{r}\right)^{n_{t,r}} \leq \frac{e^{-tk}}{k!}.$$

Enfin, par convergence dominée, on obtient

$$\lim_{r \rightarrow \infty} \sum_{k=1}^r (-1)^{k-1} \binom{r}{k} \left(1 - \frac{k}{r}\right)^{n_{t,r}} = \sum_{k=1}^{\infty} (-1)^{k-1} \frac{e^{-tk}}{k!} = S(t).$$

3. Et il faut définir un espace probabilisé unique pour tout r pour envisager une convergence presque sûre !

Ceci achève la preuve du théorème. Alternativement, il est possible d'établir le théorème en utilisant l'approximation exponentielle des lois géométriques. Posons $H_k := \frac{1}{r}G_{r-k+1}$. Pour tout $k \geq 1$ fixé, la suite $(H_k)_{r \geq k}$ converge en loi vers la loi exponentielle de paramètre k . On dispose de plus de la version quantitative suivante sur les fonctions caractéristiques : pour tout $t \in \mathbb{R}$ et tout $k \geq 1$, il existe une constante $C > 0$ telle que pour r assez grand,

$$|\mathbb{E}(e^{itH_k}) - \mathbb{E}(e^{itE_k})| \leq \frac{C}{rk}$$

où E_k est une v.a.r. de loi exponentielle de paramètre k . En utilisant l'inégalité

$$|a_1 \cdots a_r - b_1 \cdots b_r| \leq |a_1 - b_1| + \cdots + |a_r - b_r|$$

valable pour tous $a_1, \dots, a_r, b_1, \dots, b_r \in \{z \in \mathbb{C} : |z| \leq 1\}$, il vient, en prenant E_1, \dots, E_k indépendantes de lois exponentielles de paramètres respectifs $1, \dots, k$,

$$\left| \mathbb{E}(e^{it\frac{T}{r}}) - \mathbb{E}(e^{it(E_1 + \dots + E_r)}) \right| \leq C \frac{\log(r)}{r}.$$

Comme dans le théorème 9.2.1, on observe que $E_1 + \dots + E_r$ a même loi que $\max(F_1, \dots, F_r)$ où F_1, \dots, F_r sont des v.a.r. i.i.d. de loi exponentielle de paramètre 1, puis on utilise le fait que $(\max(F_1, \dots, F_r) - \log(r))_{r \geq 1}$ converge en loi vers la loi de Gumbel. \square

Le théorème 1.5.6 fournit un intervalle de prédiction pour T : pour tous réels $b \geq a$,

$$\lim_{r \rightarrow \infty} \mathbb{P}(T \in [r \log(r) - ra, r \log(r) + rb]) = e^{-e^{-b}} - e^{-e^{-a}}.$$

La loi de Gumbel est très concentrée, et sa fonction de répartition fait apparaître une montée abrupte de 0 à 1. Cela donne un phénomène de seuil pour T . La quantité $\mathbb{P}(T > n)$ passe abruptement de ≈ 1 à ≈ 0 autour de $n = r \log(r)$ si $r \gg 1$:

Théorème 1.5.7 (Convergence abrupte autour de $n = r \log(r)$). *Pour tout réel $c > 0$,*

$$\mathbb{P}(T > r \log(r) + rc) \leq e^{-c}.$$

Si $(c_r)_{r \geq 1}$ dans \mathbb{R}_+ vérifie $\lim_{r \rightarrow \infty} c_r = \infty$ et $r \log(r) - rc_r > 0$ pour tout $r \geq 1$, alors

$$\lim_{r \rightarrow \infty} \mathbb{P}(T > r \log(r) - rc_r) = 1.$$

Démonstration. Avec $n = r(\log(r) + c)$ on obtient

$$\mathbb{P}(T > n) \leq \sum_{i=1}^r \mathbb{P}(E_{n,i}) = r(1 - 1/r)^n \leq re^{-n/r} = e^{-c}.$$

D'autre part, $(T - r \log(r))/r$ converge en loi (Gumbel) quand $r \rightarrow \infty$, donc

$$\mathbb{P}(T > r \log(r) - rc_r) = \mathbb{P}\left(\frac{T - r \log(r)}{r} > -c_r\right) = 1.$$

\square

1.6 Notes et commentaires

Notre étude du jeu de pile ou face fait l'impasse sur un certain nombre de propriétés remarquables, comme la loi de l'arc-sinus ou le principe d'invariance gaussien ou poissonnien, abordées en détail dans le livre de Feller [Fel68, Fel71]. Ce livre contient également une preuve du théorème de Berry-Esseen, qui fait appel à l'analyse de Fourier. La meilleure constante n'est pas connue à l'heure où nous écrivons ces lignes. L'article [Ste94] sur l'approximation par la loi de Poisson est bien éclairant. On trouvera également dans le livre [BHJ92] une version renforcée de l'inégalité de Le Cam, qui donne

$$d_{\text{VT}}(\text{Bin}(1, p_1) * \cdots * \text{Bin}(1, p_n), \text{Poi}(p_1 + \cdots + p_n)) \leq (1 - e^{-(p_1 + \cdots + p_n)}) \frac{p_1^2 + \cdots + p_n^2}{p_1 + \cdots + p_n}. \quad (1.1)$$

À titre de comparaison, une version inhomogène du théorème de Berry-Esseen affirme que si $(X_n)_{n \geq 1}$ sont des v.a.r. indépendantes alors en notant $S_n = X_1 + \cdots + X_n$, $\sigma_k^2 = \mathbb{E}(|X_k - \mathbb{E}(X_k)|^2)$, $\tau_k^3 = \mathbb{E}(|X_k - \mathbb{E}(X_k)|^3)$, on a

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P} \left(\frac{S_n - \mathbb{E}(S_n)}{\sqrt{\text{Var}(S_n)}} \leq x \right) - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt \right|^2 \leq \frac{(\tau_1^3 + \cdots + \tau_n^3)^2}{(\sigma_1^2 + \cdots + \sigma_n^2)^3}. \quad (1.2)$$

D'autre part, on peut voir $d_{\text{VT}}(\cdot, \cdot)$ comme une distance de couplage de Fréchet-Wasserstein. En effet, comme $\mathbb{P}(X \neq Y) = \mathbb{E}(d(X, Y))$ pour la distance atomique $d(x, y) = \delta_{x \neq y}$, on a, en notant $\Pi(\mu, \nu)$ l'ensemble des lois sur $E \times E$ de marginales μ et ν ,

$$d_{\text{VT}}(\mu, \nu) = \min_{\pi \in \Pi(\mu, \nu)} \int_{E \times E} d(x, y) d\pi(x, y).$$

Le problème du collectionneur de coupons est notamment abordé dans [Fel68, Fel71], [MR95], [Hol01], [DP13]. Newman et Shepp ont montré dans [NS60] que si on impose que chaque type soit observé m fois, alors le temps de complétion de la collection vaut en moyenne $r \log(r) + (m-1)r \log(\log(r)) + \mathcal{O}(r)$. Le théorème 1.5.6 révélant une fluctuation asymptotique de loi de Gumbel a été obtenu par Erdős et Rényi. Une analyse du cas non uniforme associé à une probabilité discrète (p_1, \dots, p_r) est menée dans [Hol01]. Lorsque r n'est pas connu, on dispose de l'estimateur $\hat{r}_n := \text{card}\{X_1, \dots, X_n\}$. Si e_1, \dots, e_r est la base canonique de \mathbb{R}^r alors le vecteur aléatoire $C_n := (C_{n,1}, \dots, C_{n,r}) := e_{X_1} + \cdots + e_{X_n}$ de \mathbb{N}^r suit la loi multinomiale de taille n et de paramètre (p_1, \dots, p_r) , et on a $\hat{r}_n = \sum_{i=1}^r \mathbb{1}_{\{C_{n,i} > 0\}} = r - \sum_{i=1}^r \mathbb{1}_{\{C_{n,i} = 0\}}$. En particulier,

$$\mathbb{E}(\hat{r}_n) = r - \sum_{i=1}^r \mathbb{P}(C_{n,i} = 0) = r - \sum_{i=1}^r (1 - p_i)^n.$$

Comme les r types sont ordonnés, on dispose de l'estimateur alternatif $\max\{X_1, \dots, X_n\}$, qui estime le bord droit du support (c'est l'analogie discret de la loi uniforme sur $[0, \theta]$).

Le collectionneur de coupons est un cas particulier du problème du recouvrement abordé dans [AF01], [Dem05], et [LPW09, Chapitre 11] : si $(X_n)_{n \geq 1}$ est une suite de variables aléatoires (pas nécessairement indépendantes ou de même loi) prenant leurs valeurs dans un même ensemble fini E , alors le *temps de recouvrement*⁴ de E est

$$T := \inf\{n \geq 1 : \{X_1, \dots, X_n\} = E\}.$$

4. En anglais : *covering time*.

Séance 2

Modèle de Fisher-Wright

Mots-clés. chaînes de Markov finies ; martingales ; absorption ; invariance.

La théorie de Mendel assure la préservation de la variabilité des génotypes.

Théorème 2.0.1 (Loi de Hardy-Weinberg pour la théorie de Mendel). *Considérons les fréquences alléliques dans une population pour un gène pouvant s'exprimer sous la forme de deux allèles A et B dans une population hermaphrodite¹ diploïde² idéale :*

1. Les générations sont séparées (elles ne coexistent pas)
2. La population est de taille infinie
3. Les individus s'y unissent aléatoirement (pas de choix du conjoint)
4. Il n'y a pas de migration, pas de mutation, pas de sélection

Si $p \in [0, 1]$ est la proportion d'allèles A dans la population initiale et $1 - p$ la proportion d'allèles B alors c'est encore le cas à chaque génération suivante. De plus, si r, s, t sont les proportions respectives des génotypes AA, AB et BB dans la population initiale avec $r + s + t = 1$ alors les fréquences de AA, AB , et BB pour toutes les générations suivantes sont égales à $(r + s/2)^2, 2(r + s/2)(t + s/2)$ et $(t + s/2)^2$.

Notons que $p = r + s/2$ (contributions de AA et AB normalisées à 1).

Démonstration. Un individu AA est issu

- à coup sûr d'un couple AA et AA qui a une probabilité r^2 de se former,
- avec probabilité $1/2$ d'un couple AA et AB qui a une probabilité $2rs$ de se former,
- avec probabilité $1/4$ d'un couple AB et AB qui a une probabilité s^2 de se former.

La fréquence du génotype AA à la génération 1 est donc donnée par

$$r^2 + rs + \frac{s^2}{4} = \left(r + \frac{s}{2}\right)^2.$$

Par symétrie, la fréquence du génotype BB est $(t + s/2)^2$. La dernière est obtenue par différence. En réappliquant ces formules on voit que les proportions sont constantes dès la première génération. La fréquence de l'allèle A dans la population initiale est $2r + s$ (normalisée à 2 car chaque individu a deux allèles). À la génération suivante, elle vaut

$$2\left(r + \frac{s}{2}\right)^2 + 2\left(r + \frac{s}{2}\right)\left(t + \frac{s}{2}\right) = 2r + s.$$

Elle est donc bien constante au cours du temps. □

1. Hermaphrodite : chaque individu peut être mâle ou femelle.
2. Diploïde : qui possède un double assortiment de chromosomes semblables.

Le résultat est généralisable à un locus³ avec plusieurs allèles A_1, \dots, A_k . Les relations de dominance entre allèles n'ont pas d'effet sur l'évolution des fréquences alléliques. Il n'y a pas d'évolution. Il faut violer les hypothèses du théorème pour expliquer l'apparition d'hétérogénéité darwinienne dans les populations au cours des générations.

2.1 Modèle de Fisher-Wright et fixation

Il s'agit d'un modèle simple de l'évolution de la fréquence d'un gène à plusieurs allèles au fil des générations dans une population. On suppose que la population est de **taille finie constante**, hermaphrodite, que les générations sont séparées, qu'il n'y a pas de migration, mutation, sélection, et que les unions sont indépendantes du caractère étudié.

On néglige dans un premier temps les phénomènes de mutation et sélection. On note $N \in \mathbb{N}^*$ la taille de la population et $n = 0, 1, 2, \dots$ les générations successives. Sur le locus étudié, on peut trouver deux allèles différents notés A et B . La variable aléatoire X_n compte le nombre d'allèles A à la génération n . La population à la génération $n + 1$ est déduite de celle de la génération n par un tirage avec remise de N individus de la génération n où la probabilité d'obtenir A est X_n/N . On peut réaliser $(X_n)_{n \geq 0}$ comme une suite récurrente aléatoire $X_{n+1} = \sum_{k=1}^N \mathbb{1}_{\{U_{n,k} \leq X_n/N\}}$ où $(U_{n,k})_{n \geq 1, 1 \leq k \leq N}$ sont des v.a.r. i.i.d. de loi uniforme sur $[0, 1]$, indépendantes de X_0 . Pour tous x_0, \dots, x_n , on a

$$\mathcal{L}(X_{n+1} | X_0 = x_0, \dots, X_n = x_n) = \mathcal{L}(X_{n+1} | X_n = x_n) = \text{Bin}(N, \psi_{x_n}) \quad \text{où} \quad \psi_x := \frac{x}{N}.$$

Le processus $(X_n)_{n \geq 0}$ est une chaîne de Markov homogène d'espace d'états $\{0, 1, \dots, N\}$ et de matrice de transition \mathbf{P} donnée pour tous $x, y \in \{0, 1, \dots, N\}$ par

$$\mathbf{P}(x, y) := \mathbb{P}(X_{n+1} = y | X_n = x) = \binom{N}{y} \psi_x^y (1 - \psi_x)^{N-y}.$$

Les états 0 et N sont absorbants (deux classes de récurrence singleton) tandis que tous les autres états mènent à $\{0, N\}$ et sont donc transitoires. Lorsque la population ne contient plus qu'un allèle, on dit qu'il est *fixé* dans la population. Le *temps de fixation* est défini par

$$T := \inf\{n \geq 0 : X_n \in \{0, N\}\}.$$

L'espace d'états étant fini, le temps d'atteinte T de $\{0, N\}$ est fini p.s. et intégrable :

$$\mathbb{P}(T < \infty) = 1 \quad \text{et} \quad \mathbb{E}(T) < \infty.$$

De plus $(X_n)_{n \geq 0}$ converge p.s. vers la v.a. X_T qui suit une loi de Bernoulli portée par 0 et N . On a $\mathbb{P}(X_T = 0) + \mathbb{P}(X_T = N) = 1$. L'événement $\{X_T = N\}$ signifie que l'allèle A est fixé tandis que l'événement $\{X_T = 0\}$ signifie que l'allèle B est fixé.

Remarque 2.1.1 (Couplage avec un jeu de pile ou face). *Il est possible d'obtenir une minoration de T par une variable aléatoire géométrique par couplage. On commence par observer que pour tout x et tout n on a $\mathbb{P}(X_{n+1} \in \{0, N\} | X_n = x) = \mathbf{P}(x, \{0, N\}) \geq p_*$ où*

$$p_* = \min_{0 \leq x \leq N} \mathbf{P}(x, \{0, N\}) = \min_{0 \leq x \leq N} (1 - \psi_x)^N + \psi_x^N \geq 2^{-N+1}$$

(atteint en $\psi_x = 1/2$). Ainsi, la chaîne atteint un état absorbant 0 ou N avant qu'un lanceur de pièce n'obtienne son premier pile avec une pièce qui fait pile avec probabilité 2^{-N+1} .

3. Locus : localisation précise d'un gène particulier sur un chromosome.

En toute rigueur, le raisonnement nécessite de pouvoir construire le couple de processus (jeu de pile ou face et chaîne de Fisher-Wright) sur le même espace de probabilités. Pour y parvenir, représentons la chaîne de Markov comme une suite récurrente aléatoire : pour tout $x \in \{0, \dots, N\}$, soit $(I_{x,y})_{0 \leq y \leq N}$ une partition de l'intervalle $[0, 1]$, avec $I_{x,y}$ de longueur $|I_{x,y}| = \mathbf{P}(x, y)$, et soit $(U_n)_{n \geq 1}$ une suite de v.a.r. i.i.d. de loi uniforme sur $[0, 1]$, indépendante de X_0 . On réalise $(X_n)_{n \geq 0}$ en utilisant la récurrence $X_{n+1} = y$ si $U_{n+1} \in I_{X_n, y}$. Il est bon de faire un dessin figurant une tour de $N + 1$ copies de l'intervalle $[0, 1]$ avec ses $N + 1$ partitions. Dans cette construction, on peut toujours ordonner les intervalles à loisir, et supposer que $I_{x,0}$ et $I_{x,N}$ sont les deux premiers en partant de la gauche. Par construction $p_* = \min_{0 \leq x \leq N} |I_{x,0} \cup I_{x,N}|$ et

$$T := \inf\{n \geq 0 : X_n \in \{0, N\}\} \leq T_* := \inf\{n \geq 1 : \mathbf{1}_{\{U_n \leq p_*\}} = 1\}$$

et T_* suit la loi géométrique sur \mathbb{N}^* de paramètre p_* . En particulier, T est intégrable et donc fini p.s.. Cette méthode de couplage fournit plus généralement, pour des chaînes de Markov, un minorant géométrique des temps d'atteinte d'ensembles.

Remarque 2.1.2 (Modèle à plus de deux allèles). Le modèle de Fisher-Wright à $\ell \geq 2$ allèles A_1, \dots, A_ℓ fait intervenir la loi multinomiale. Plus précisément, X_n prend ses valeurs dans $\{x \in \mathbb{N}^\ell : x_1 + \dots + x_\ell = N\}$ et $\mathcal{L}(X_{n+1} | X_0, \dots, X_n) = \mathcal{L}(X_{n+1} | X_n) = \text{Mul}(N, \frac{X_n}{N})$.

On adopte les notations usuelles $\mathbb{P}_x = \mathbb{P}(\cdot | X_0 = x)$ et $\mathbb{E}_x = \mathbb{E}(\cdot | X_0 = x)$.

Théorème 2.1.3 (Probabilité de fixation). Pour tout $x \in \{0, 1, \dots, N\}$,

$$\mathbb{P}_x(X_T = N) = 1 - \mathbb{P}_x(X_T = 0) = \frac{x}{N}.$$

Démonstration. Plaçons-nous dans le cas où $X_0 \sim \delta_x$. On a $\mathbb{P}_x(T < \infty) = 1$, et comme T est un temps de fixation, on sait que $(X_n)_{n \geq 0}$ converge p.s. vers X_T . Comme l'espace d'états est borné, le théorème de convergence dominée s'applique et la convergence a lieu dans L^1 , d'où $\lim_{n \rightarrow \infty} \mathbb{E}_x(X_n) = \mathbb{E}(X_T) = 0\mathbb{P}_x(X_T = 0) + N\mathbb{P}_x(X_T = N) = N\mathbb{P}_x(X_T = N)$. Or grâce à la loi conditionnelle binomiale $\mathcal{L}(X_n | \mathcal{F}_{n-1}) = \mathcal{L}(X_n | X_{n-1}) = \text{Bin}(N, \frac{X_{n-1}}{N})$, on a $\mathbb{E}(X_n) = \mathbb{E}(\mathbb{E}(X_n | \mathcal{F}_{n-1})) = \mathbb{E}(N \frac{X_{n-1}}{N}) = \mathbb{E}(X_{n-1}) = \dots = \mathbb{E}(X_0) = x$, d'où le résultat. \square

Les formules pour la probabilité de fixation peuvent s'obtenir en utilisant le système linéaire vérifié par le vecteur des probabilités d'atteinte pour la chaîne de Markov $(X_n)_{n \geq 0}$.

Remarque 2.1.4 (Martingales). En reprenant la preuve du théorème précédent, on a $\mathbb{E}(X_n | \mathcal{F}_{n-1}) = \mathbb{E}(X_n | X_{n-1}) = N \frac{X_{n-1}}{N} = X_{n-1}$. Ainsi, en plus d'être une chaîne de Markov, la suite $(X_n)_{n \geq 0}$ est une martingale pour sa filtration naturelle $(\mathcal{F}_n)_{n \geq 0}$ où $\mathcal{F}_n = \sigma(X_0, \dots, X_n)$. En particulier, on a bien sûr la conservation $\mathbb{E}(X_0) = \mathbb{E}(X_n)$ pour tout $n \geq 1$. Le théorème de Doob indique que $(X_{n \wedge T})_{n \geq 0}$ est aussi une martingale et donc $\mathbb{E}(X_0) = \mathbb{E}(X_{n \wedge T})$, qui converge vers $\mathbb{E}(X_T)$ par convergence dominée. Comme T est un temps de fixation, on a ici $X_n = X_{n \wedge T}$. Par ailleurs, nous sommes aussi dans le cadre d'application du théorème de convergence p.s. et dans L^p des martingales et bornées dans L^p , $p > 1$.

La vitesse de fixation peut être quantifiée par l'hétérozygotie, c'est-à-dire la probabilité que deux gènes choisis aléatoirement et sans remise dans la population totale à la génération n soient représentés par des allèles différents. Elle est donnée par

$$H_n = \frac{2X_n(N - X_n)}{N(N - 1)} = \frac{2}{N - 1} \text{Var}(X_{n+1} | X_n).$$

Notons que $H_n = 0$ si $n \geq T$ et donc $(H_n)_{n \geq 0}$ converge p.s. vers 0 quand $n \rightarrow \infty$.

Théorème 2.1.5 (Décroissance exponentielle de l'hétérozygotie). *Pour tout $n \geq 1$,*

$$\mathbb{E}(H_n) = \lambda^n h_0 \quad \text{où} \quad h_0 = \mathbb{E}(H_0), \quad \lambda = 1 - \frac{1}{N}$$

et

$$\text{Var}(X_n) = \mathbb{E}(X_0)(N - \mathbb{E}(X_0))(1 - \lambda^n) + \lambda^n \text{Var}(X_0).$$

Démonstration. Il suffit d'établir que pour tout $n \geq 1$,

$$\mathbb{E}(X_n(N - X_n)) = \left(1 - \frac{1}{N}\right)^n \mathbb{E}(X_0(N - X_0)).$$

On a

$$\mathbb{E}(X_n(N - X_n)) = N\mathbb{E}(X_n) - \mathbb{E}(X_n^2) = N\mathbb{E}(X_{n-1}) - \mathbb{E}(\mathbb{E}(X_n^2|X_{n-1})).$$

On écrit alors

$$\mathbb{E}(X_n^2|X_{n-1}) = \text{Var}(X_n|X_{n-1}) + (\mathbb{E}(X_n|X_{n-1}))^2 = X_{n-1} \left(1 - \frac{X_{n-1}}{N}\right) + X_{n-1}^2.$$

En regroupant les termes, on obtient bien

$$\mathbb{E}(X_n(N - X_n)) = \left(1 - \frac{1}{N}\right) \mathbb{E}(X_{n-1}(N - X_{n-1})).$$

Ce qui fournit la formule pour $\mathbb{E}(H_n)$. Pour la variance, on utilise la formule

$$\text{Var}(X_n) = \mathbb{E}(\text{Var}(X_n|X_{n-1})) + \text{Var}(\mathbb{E}(X_n|X_{n-1})).$$

□

Il est naturel de vouloir étudier $\mathbb{E}(T)$. Si l'on note $m(x) = \mathbb{E}_x(T)$ alors $m(0) = m(N) = 0$ et, grâce à la propriété de Markov (faible), on obtient une équation linéaire :

$$m(x) = \sum_{y=0}^n \mathbb{E}_x(T \mathbf{1}_{\{X_1=y\}}) = \sum_{y=0}^n (1 + \mathbb{E}_y(T)) \mathbf{P}(x, y) = 1 + \sum_{y=0}^N m(y) \mathbf{P}(x, y),$$

dont m est l'unique solution positive minimale. Bien qu'on puisse calculer m numériquement, il n'existe pas d'expression simple et explicite de m . Il est toutefois possible de trouver un équivalent de m lorsque la taille de la population N tend vers l'infini. Renormalisons le temps et l'espace en considérant le processus $Y^{(N)} = \left(\frac{1}{N} X_{\lfloor Nt \rfloor}\right)_{t \in \mathbb{R}_+}$ à valeurs dans $[0, 1]$. Heuristiquement, quand N est grand, la quantité $m(x)/N$ est proche d'une fonction régulière $m_*(p)$ de $p = x/N$ qui vérifie une équation

$$m_*(p) \approx 1 + \int_0^1 f(p, p') m_*(p') dp'$$

où la fonction $p' \mapsto f(p, p')$ est une densité de moyenne p et de variance $p(1-p)$. Comme $m_*(p') = m_*(p) + (p' - p)m_*'(p) + \frac{1}{2}(p' - p)^2 m_*''(p) + o((p' - p)^2)$ on obtient

$$m_*(p) \approx 1 + m_*(p) + m_*'(p) \int_0^1 (p' - p) f(p, p') dp' + m_*''(p) \frac{1}{2} \int_0^1 (p' - p)^2 f(p, p') dp'.$$

En utilisant les deux premiers moments de $f(p, \cdot)$ on obtient

$$0 \approx 1 + \frac{p(1-p)}{2} m_*''(p).$$

Comme $m_*(0) = m_*(1) = 0$ on trouve (c'est l'entropie !)

$$m_*(p) \approx -2(p \log(p) + (1-p) \log(1-p)).$$

Ainsi, pour une petite proportion de A au départ, disons $p \approx 1/N$, le temps moyen de fixation est de l'ordre de $\log(N)$, tandis que pour une grande proportion de A au départ, disons $p \approx 1/2$, le temps moyen de fixation est de l'ordre de N , bien plus grand que $\log(N)$. On peut établir rigoureusement que $Y^{(N)}$ converge en loi quand $N \rightarrow \infty$ vers un processus de diffusion Y sur $[0, 1]$, solution de l'équation différentielle stochastique $dY_t = \sqrt{Y_t(1-Y_t)}dB_t$, pour lequel la formule précédente du temps de fixation est exacte (via la formule d'Itô).

2.2 Modèle de Fisher-Wright avec mutations

Supposons à présent qu'avant le tirage avec remise qui permet de passer d'une génération à l'autre, chaque individu puisse muter indépendamment des autres : l'allèle A mute en allèle B avec probabilité $u \in [0, 1]$ et B mute en A avec probabilité $v \in [0, 1]$. Le passage d'une génération à la suivante se fait donc en quelque sorte en appliquant d'abord un opérateur de mutation, puis un opérateur d'héritage (procéder dans l'ordre inverse conduirait à un processus différent). Ce mécanisme définit un nouveau processus $(X_n)_{n \geq 0}$ à valeurs dans $\{0, 1, \dots, N\}$ qui vérifie $\mathcal{L}(X_{n+1} | X_0, \dots, X_n) = \text{Bin}(N, \psi_{X_n})$ où

$$\psi_x = \frac{x(1-u) + (N-x)v}{N}.$$

Le processus X est une chaîne de Markov d'espace d'états $\{0, 1, \dots, N\}$.

- si $u = v = 0$, alors on retrouve le modèle sans mutations étudié précédemment. Dans ce cas, les états 0 et N sont absorbants tandis que $\{1, \dots, N-1\}$ est une classe transitoire, l'ensemble des lois invariantes est $\{(1-p)\delta_0 + p\delta_N : p \in [0, 1]\}$, et partant de x , la chaîne converge en loi vers $(1-p_x)\delta_0 + p_x\delta_N$ où $p_x = x/N$;
- si $u = 0$ et $0 < v \leq 1$ alors l'état N est absorbant tandis que $\{0, \dots, N-1\}$ est une classe transitoire, et la chaîne converge en loi vers l'unique loi invariante δ_N ;
- si $0 < u \leq 1$ et $v = 0$ alors l'état 0 est absorbant tandis que $\{1, \dots, N\}$ est une classe transitoire, et la chaîne converge en loi vers l'unique loi invariante δ_0 ;
- si $u = v = 1$, alors $\{0, N\}$ est une classe de récurrence de période 2 tandis que $\{1, \dots, N-1\}$ est une classe transitoire. La chaîne est presque sûrement absorbée par la classe $\{0, N\}$ et oscille ensuite périodiquement entre les états 0 et N ;
- si $0 < u \leq 1$ et $0 < v < 1$ ou si $0 < u < 1$ et $0 < v \leq 1$ alors la chaîne est irréductible. Comme l'espace d'états est fini, elle est récurrente positive et possède une unique loi invariante μ , et cette loi charge tous les états. La loi des grands nombres pour les chaînes de Markov indique que quelle que soit la loi initiale, presque sûrement, pour tout $x \in \{0, 1, \dots, N\}$, on a

$$\lim_{n \rightarrow \infty} \frac{\text{card}\{0 \leq k \leq n : X_k = x\}}{n+1} = \mu(x).$$

Comme de plus la matrice de transition possède un coefficient diagonal non nul, la chaîne est apériodique, et converge donc en loi vers μ , quelle que soit la loi initiale, c'est-à-dire que pour tout $x \in \{0, 1, \dots, N\}$,

$$\lim_{n \rightarrow \infty} \mathbb{P}(X_n = x) = \mu(x).$$

La loi μ est calculable numériquement, mais ne possède pas de formulation simple. Cependant, il est possible de déterminer explicitement sa moyenne m par exemple. En effet, comme μ est invariante,

$$m = \sum_{x=0}^N y\mu(y) = \sum_{y=0}^N y \sum_{x=0}^N \mu(x)\mathbf{P}(x, y) = \sum_{x=0}^N \mu(x) \sum_{y=0}^N y\mathbf{P}(x, y).$$

Or comme $\mathbf{P}(x, \cdot) = \text{Bin}(N, \frac{(1-u)x+v(N-x)}{N})$, on a $\sum_{y=0}^N y\mathbf{P}(x, y) = (1-u)x + v(N-x)$. Cette formule affine en x donne une équation fermée pour m :

$$m = \sum_{x=0}^N ((1-u)x + v(N-x))\mu(x) = (1-u)m + v(N-m),$$

d'où enfin

$$m = \frac{Nv}{u+v}.$$

On peut montrer de même que la variance σ^2 de μ est de la forme :

$$\sigma^2 = \frac{N^2 uv}{(u+v)^2(2N(u+v)+1)} + o_{N \rightarrow \infty}(N).$$

On peut établir que $Y^{(N)} = (\frac{1}{N}X_{[Nt]})_{t \in \mathbb{R}_+}$ converge en loi quand $(N, Nu, Nv) \rightarrow (\infty, \alpha, \beta)$ vers un processus de diffusion Y sur $[0, 1]$, solution de l'équation différentielle stochastique $dY_t = \sqrt{Y_t(1-Y_t)}dB_t - \alpha Y_t dt + \beta(1-Y_t)dt$, dont la loi invariante est une loi Beta($2\beta, 2\alpha$), dont la moyenne est $\frac{\beta}{\alpha+\beta}$ et la variance $\frac{\alpha\beta}{(\alpha+\beta)^2(2(\alpha+\beta)+1)}$.

2.3 Modèle de Fisher-Wright avec sélection

Sur le plan biologique, il est possible que les allèles A et B ne procurent pas à l'individu qui en est doté des capacités identiques : viabilité, potentiel attractif, etc. Cette asymétrie est responsable de la sélection naturelle. Considérons pour simplifier le modèle sans mutations. Il est possible de prendre en compte un avantage sélectif de l'allèle A en effectuant la transformation suivante de la génération n avant la fabrication de la génération $n+1$: chaque individu de type A est remplacé par $(1+s)$ individus de type A , où $s > 0$ est un paramètre fixé (penser à une urne). Cela revient à remplacer $\psi_x = x/N$ par

$$\psi_x = \frac{x(1+s)}{x(1+s) + N-x},$$

de sorte que $\mathcal{L}(X_{n+1} | X_0, \dots, X_n) = \mathcal{L}(X_{n+1} | X_n) = \text{Bin}(N, \psi_{X_n})$. Le processus $(X_n)_{n \geq 0}$ est une chaîne de Markov. Son espérance ne vérifie pas une équation de récurrence linéaire car ψ_x n'est pas linéaire en x (contrairement au cas sans sélection $s=0$). Les états 0 et N sont absorbants, tandis que $\{1, \dots, N-1\}$ est une classe transitoire. Comme ψ_x n'est plus linéaire en x , on ne peut plus calculer la probabilité de fixation avec la méthode utilisée dans le cas sans sélection ($s=0$). Soit $\pi(x) = \mathbb{P}_x(X_T = N)$ la probabilité de fixation de l'allèle A . En notant \mathbf{P} la matrice de transition de X , on a $\pi(x) = \sum_{y=0}^N \mathbf{P}(x, y)\pi(y)$. Il est possible de trouver un équivalent de π lorsque la taille de la population N tend vers l'infini et s est proche de 0. Renormalisons le temps et l'espace en considérant le processus $Y^{(N)} = (\frac{1}{N}X_{[Nt]})_{t \in \mathbb{R}_+}$ à valeurs dans $[0, 1]$. Heuristiquement, quand N est grand, la quantité $\pi(x)$ est proche d'une fonction régulière $\pi_*(p)$ de $p = x/N$ qui vérifie une équation

$$\pi_*(p) \approx \int_0^1 f(p, p')\pi_*(p') dp'$$

où la fonction $p' \mapsto f(p, p')$ est une densité de moyenne $\psi_*(p)$ et de variance $\psi_*(p)(1 - \psi_*(p))$ où $\psi_*(p) = p(1 + s)(1 - ps + o(s))$ (provient d'un développement limité de $\psi_*(p)$ en $s = 0$). En utilisant le développement limité $\pi_*(p') = \pi_*(p) + (p' - p)\pi'_*(p) + \frac{1}{2}(p' - p)^2\pi''_*(p) + o((p' - p)^2)$ on obtient $0 \approx p(1 - p)s\pi'_*(p) + \frac{1}{2}p(1 - p)\pi''_*(p)$. Comme $\pi_*(0) = 0$ et $\pi_*(1) = 1$, on obtient

$$\pi_*(p) = \frac{1 - e^{-2sp}}{1 - e^{-2s}}.$$

Ainsi, le passage de $s = 0$ à $s > 0$ modifie considérablement les probabilités de fixation. Il est possible de rendre rigoureuse cette analyse heuristique en établissant que $Y^{(N)}$ converge en loi quand $N \rightarrow \infty$ et $Ns \rightarrow \alpha$ vers un processus de diffusion Y sur $[0, 1]$, solution de l'équation différentielle stochastique $dY_t = \sqrt{Y_t(1 - Y_t)}dB_t + \alpha Y_t(1 - Y_t)dt$ pour lequel la formule précédente de la probabilité de fixation est exacte.

2.4 Modèle de Cannings

Dans le modèle de Fisher-Wright, on passe de la génération n à la génération $n + 1$ en effectuant un tirage avec remise de N individus dans la génération n . Cela revient à attribuer un certain nombre d'enfants à chaque individu de la génération n , les enfants héritant de l'allèle du père. Si y_1, \dots, y_N désignent le nombre d'enfants de chaque individu de la génération n (numérotés de 1 à N) alors on a $y_1 + \dots + y_N = N$ et (y_1, \dots, y_N) suit la loi multinomiale symétrique de taille N et de paramètre $(1/N, \dots, 1/N)$. Cela correspond à lancer N fois un dé à N faces équilibrées. Ce mécanisme qui associe chaque individu de la génération $n + 1$ à un individu de la génération n peut être utilisé pour étudier la transmission de gènes quel que soit le nombre d'allèles considéré : les enfants héritent de l'allèle du père. Le modèle de Cannings va au delà de cette structure multinomiale symétrique en permettant une dépendance entre les individus d'une même génération : on suppose seulement que le vecteur (y_1, \dots, y_N) vérifie $y_1 + \dots + y_N = N$ et possède une loi échangeable, c'est-à-dire que pour toute permutation $\tau \in \mathcal{S}_N$,

$$(y_1, \dots, y_N) \stackrel{\text{loi}}{=} (y_{\tau(1)}, \dots, y_{\tau(N)}).$$

En particulier, les y_1, \dots, y_N ont même loi. Comme $y_1 + \dots + y_N = N$ on en déduit que $\mathbb{E}(y_1) = 1$ et on note $\sigma^2 = \text{Var}(y_1)$. On en déduit également $\text{Cov}(y_1, y_2) = -\sigma^2/(N - 1)$ car

$$0 = \text{Var}(y_1 + \dots + y_N) = N\text{Var}(y_1) + N(N - 1)\text{Cov}(y_1, y_2).$$

Dans le cas à deux allèles, ce mécanisme de transition donne lieu, tout comme pour le modèle de Fisher-Wright, à une chaîne de Markov $(X_n)_{n \geq 0}$ d'espace d'états $\{0, 1, \dots, N\}$ donnée pour tout $x \in \{0, 1, \dots, N\}$ par

$$\mathcal{L}(X_{n+1} | X_n = x) = \mathcal{L}(y_1 + \dots + y_x).$$

Les états 0 et N sont absorbants tandis que tous les autres y mènent et sont donc transitoires. Notons que $\mathbb{E}(X_{n+1} | X_n = x) = x$ et donc $(X_n)_{n \geq 0}$ est une martingale. De plus,

$$\begin{aligned} \text{Var}(X_{n+1} | X_n = x) &= \text{Var}(y_1 + \dots + y_x) \\ &= x\sigma^2 + x(x - 1)\text{Cov}(y_1, y_2) \\ &= x\sigma^2 - \frac{x(x - 1)\sigma^2}{N - 1} \\ &= \sigma^2 \frac{x(N - x)}{N - 1}. \end{aligned}$$

Remarque 2.4.1 (Loi multinomiale et ses marges). Soit $n, d \in \mathbb{N}^*$ et $(p_1, \dots, p_d) \in [0, 1]^d$ avec $p_1 + \dots + p_d = 1$. La loi multinomiale de taille n et de paramètre (p_1, \dots, p_d) est la loi sur l'ensemble fini $\{(n_1, \dots, n_d) \in \mathbb{N}^d : n_1 + \dots + n_d = n\}$ donnée par

$$(p_1 \delta_{e_1} + \dots + p_d \delta_{e_d})^{*n} = \sum_{\substack{(n_1, \dots, n_d) \in \mathbb{N}^d \\ n_1 + \dots + n_d = n}} \frac{n!}{n_1! \dots n_d!} p_1^{n_1} \dots p_d^{n_d} \delta_{(n_1, \dots, n_d)}.$$

Elle correspond aux résultats de n lancers d'un dé à d faces avec probabilités d'apparition p_1, \dots, p_d . On dit que la loi est symétrique lorsque $p_1 = \dots = p_d = 1/d$. Si $(X_1, \dots, X_d) \sim \text{Mul}(n, (p_1, \dots, p_d))$ et si I_1, \dots, I_r est une partition de $\{1, \dots, d\}$ alors

$$(X_{I_1}, \dots, X_{I_r}) \sim \text{Mul}(n, (p_{I_1}, \dots, p_{I_r}))$$

où $X_I = \sum_{i \in I} X_i$ et $p_I = \sum_{i \in I} p_i$. Si $I \subset \{1, \dots, d\}$ alors $X_I \sim \text{Bin}(n, p_I)$. En particulier, les lois marginales d'une loi multinomiale sont binomiales. On identifie $\text{Bin}(n, p)$ à $\text{Mul}(n, (p, 1-p))$. Si $(X_1, \dots, X_d) \sim \text{Mul}(n, (p_1, \dots, p_d))$ alors $X_d = n - X_{\{1, \dots, d-1\}}$.

Théorème 2.4.2 (Spectre de la matrice de transition). Les valeurs propres de la matrice de transition du modèle de Cannings sont données par $\lambda_0 = 1$ et pour tout $j \in \{1, \dots, N\}$,

$$\lambda_j = \mathbb{E}(y_1 y_2 \dots y_j).$$

En particulier, pour le modèle de Fisher-Wright, $\lambda_0 = 1$ et pour tout $j \in \{1, \dots, N\}$,

$$\lambda_j = \frac{N(N-1) \dots (N-j+1)}{N^j} = \left(1 - \frac{1}{N}\right) \dots \left(1 - \frac{j-1}{N}\right).$$

Démonstration. Considérons la matrice de Vandermonde $V \in \mathcal{M}_{N+1, N+1}(\mathbb{R})$ suivante :

$$V = \begin{pmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ 1 & 1 & 1^2 & 1^3 & \dots & 1^N \\ 1 & 2 & 2^2 & 2^3 & \dots & 2^N \\ 1 & 3 & 3^2 & 3^3 & \dots & 3^N \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & N & N^2 & N^3 & \dots & N^N \end{pmatrix}.$$

Soit $\mathbf{P} \in \mathcal{M}_{N+1, N+1}(\mathbb{R})$ la matrice de transition de Cannings. Pour tous $i, j \in \{0, 1, \dots, N\}$,

$$(\mathbf{P}V)_{i,j} = \sum_{k=0}^N \mathbf{P}_{i,k} V_{k,j} = \sum_{k=0}^N \mathbf{P}_{i,k} k^j = \mathbb{E}(X_1^j | X_0 = i).$$

Par ailleurs, comme $\mathcal{L}(X_1 | X_0 = i) = \mathcal{L}(y_1 + \dots + y_i)$, il existe des réels $b_{i,k}$ tels que

$$\mathbb{E}(X_1^j | X_0 = i) = \mathbb{E}((y_1 + \dots + y_i)^j) = i^{[j]} \mathbb{E}(y_1 y_2 \dots y_j) + \sum_{k=0}^{j-1} b_{i,k} i^k,$$

où $i^{[j]} = i(i-1)(i-2) \dots (i-j+1)$. Pour le voir, on peut par exemple procéder par récurrence sur j en observant que la loi conditionnelle $\mathcal{L}(y_1, \dots, y_i | y_k)$ est échangeable. Il existe donc une matrice $T \in \mathcal{M}_{N+1, N+1}(\mathbb{R})$ triangulaire supérieure telle que $T_{0,0} = 1$, $T_{j,j} = \mathbb{E}(y_1 \dots y_j)$ pour tout $j \in \{1, \dots, N\}$, et pour tous $i, j \in \{0, 1, \dots, N\}$,

$$\mathbb{E}(X_1^j | X_0 = i) = \sum_{k=0}^j T_{k,j} i^k = (VT)_{i,j}.$$

On a donc $\mathbf{P}V = VT$. Comme V est inversible, les matrices \mathbf{P} et T ont même spectre. \square

On a $\mathbb{E}(y_1) = 1$ et donc 1 est valeur propre de multiplicité 2, ce qui correspond au fait que la chaîne possède deux classes de récurrence $\{0\}$ et $\{N\}$. La deuxième plus grande valeur propre est $\lambda_2 = \mathbb{E}(y_1 y_2) = \text{Cov}(y_1, y_2) + \mathbb{E}(y_1)^2 = 1 - \sigma^2/(N - 1)$.

2.5 Modèle de Moran

Tout comme le modèle de Fisher-Wright, le modèle de Moran est un cas particulier du modèle de Cannings. Il correspond à la loi de (y_1, \dots, y_N) suivante : la transition de la génération n à la génération $n + 1$ se fait en tirant avec remise deux individus dans la génération n , le premier a deux fils, le second aucun, et tous les autres individus ont exactement un seul descendant. Dans le cas à deux allèles, la matrice de transition est alors donnée pour tous $x, y \in \{0, 1, \dots, N\}$ par

$$\mathbf{P}(x, y) = \frac{1}{N^2} \begin{cases} x(N - x) & \text{si } |x - y| = 1; \\ x^2 + (N - x)^2 & \text{si } y = x; \\ 0 & \text{sinon.} \end{cases}$$

Tout comme pour le modèle de Fisher-Wright, on peut poser $T = \inf\{n \in \mathbb{N} : X_n \in \{0, N\}\}$ et s'intéresser au temps moyen de fixation $m(x) = \mathbb{E}_x(T)$. On a $m(0) = m(N) = 0$ et $m(x) = 1 + \sum_{y=0}^N \mathbf{P}(x, y)m(y)$ pour tout $x \in \{1, \dots, N - 1\}$. Contrairement au modèle de Fisher-Wright, on obtient cette fois une formule de récurrence à trois termes explicite :

$$m(x + 1) - 2m(x) + m(x - 1) = -\frac{N^2}{x(N - x)}.$$

On en déduit que

$$m(x) = N \left(\sum_{y=1}^x \frac{N - x}{N - y} + \sum_{y=x+1}^{N-1} \frac{x}{y} \right).$$

Soit $0 < p < 1$. Pour tout $N \geq 1$, posons $x_N = \lfloor pN \rfloor$. On a alors

$$\begin{aligned} \frac{m(x_N)}{N^2} &= \frac{1}{N} \sum_{y=1}^{x_N} \frac{N - x_N}{N - y} + \frac{1}{N} \sum_{y=x_N+1}^{N-1} \frac{x_N}{y} \\ &= \frac{1 - N^{-1}x_N}{N} \sum_{y=1}^{x_N} \frac{1}{1 - N^{-1}y} + \frac{N^{-1}x_N}{N} \sum_{y=x_N+1}^{N-1} \frac{1}{N^{-1}y} \\ &\xrightarrow{N \rightarrow \infty} (1 - p) \int_0^p \frac{1}{1 - t} dt + p \int_p^1 \frac{1}{t} dt = -((1 - p) \log(1 - p) + p \log(p)). \end{aligned}$$

Ainsi, dans une population de taille N avec N grand, le temps moyen d'absorption partant de $x = pN$ avec $0 < p < 1$ est de l'ordre de

$$-N^2(p \log p + (1 - p) \log(1 - p)).$$

Pour le modèle de Fisher-Wright, l'espérance du temps d'absorption est de l'ordre de N tandis qu'il est de l'ordre de N^2 pour le modèle de Moran. Cela vient tout simplement du fait que lors d'une transition du modèle de Moran, un seul allèle est modifié tandis que tous sont concernés à chaque transition du modèle de Fisher-Wright.

2.6 Notes et commentaires

Les mécanismes de la génétique ont été considérablement discutés depuis leur découverte par Mendel en plein dix-neuvième siècle. La loi de Hardy et de Weinberg a été découverte de manière indépendante par ces deux auteurs au tout début du vingtième siècle. Le modèle de Fisher et de Wright date des années 1930. Le modèle de Moran date des années 1950 tandis que le modèle de Cannings date des années 1970. Ces modèles et leurs extensions sont étudiés par exemple dans les livres [DJ06], [Dur08], [Ewe04], [Tav04], et inspirent encore aujourd'hui des recherches mêlant biologie et mathématiques. L'approximation des chaînes de Markov par des diffusions fait l'objet du chapitre ??.

Séance 3

Processus des restaurants chinois

Mots-clés. Partitions aléatoires ; groupe symétrique ; combinatoire.

Pour tout entier $n \geq 1$, on note \mathcal{S}_n l'ensemble des permutations de $\{1, \dots, n\}$ (groupe symétrique). Le processus des restaurants chinois est une chaîne de Markov inhomogène $(\sigma_n)_{n \geq 1}$, d'espace d'états $\cup_{n \geq 1} \mathcal{S}_n$, où σ_n est à valeurs dans \mathcal{S}_n pour tout $n \geq 1$, de valeur initiale $\sigma_1 = (1)$, et de noyau de transition donné pour tous $(\sigma, \sigma') \in \mathcal{S}_n \times \mathcal{S}_{n+1}$ par

$$\mathbb{P}(\sigma_{n+1} = \sigma' \mid \sigma_n = \sigma) = \begin{cases} \frac{1}{\theta + n} & \text{si } \sigma' \text{ s'obtient en insérant } n+1 \text{ dans l'un des cycles de } \sigma; \\ \frac{\theta}{\theta + n} & \text{si } \sigma' \text{ s'obtient en ajoutant le cycle } (n+1) \text{ à } \sigma; \\ 0 & \text{sinon;} \end{cases}$$

où $\theta \geq 0$ est un paramètre fixé qui ne dépend pas de n . Il s'agit bien d'un noyau de transition car la somme des longueurs des cycles de σ est n . Le nom de ce processus provient de l'interprétation des cycles de σ_n comme les tables circulaires occupées par les n clients d'un restaurant chinois, ces restaurants où les gens s'attablent sans se connaître. Au temps initial 1, le restaurant ne compte qu'une seule table occupée par un seul client numéroté 1. Par récurrence sur n , à l'instant $n+1$, et conditionnellement à tout le passé $\sigma_1, \dots, \sigma_n$, un nouveau client, numéroté $n+1$, pénètre dans le restaurant, et décide soit de rejoindre l'une des tables déjà occupées (cycles de σ_n) avec une probabilité proportionnelle à la taille de la table (longueur du cycle), soit de s'asseoir à une table vide (créer un nouveau cycle de longueur 1 contenant $n+1$). Cette interprétation gastronomique suppose que les convives ne changent jamais de table, que les tables peuvent avoir un nombre arbitrairement grand de convives, et que le restaurant peut comporter un nombre arbitrairement grand de tables.

Associons à la permutation aléatoire σ_n la partition aléatoire π_n de $\{1, \dots, n\}$ donnée par le support des cycles. Chaque bloc de π_n regroupe les clients d'une table du restaurant. On a $\pi_1 = \{1\}$, et pour tout $n \geq 1$, π_n est à valeurs dans l'ensemble Π_n des partitions de $\{1, \dots, n\}$. En général, l'image d'une chaîne de Markov par une fonction n'est pas une chaîne de Markov¹. Cependant, il se trouve ici que $(\pi_n)_{n \geq 1}$ est une chaîne de Markov inho-

1. Notons qu'un célèbre critère dû à Dynkin fournit une condition suffisante sur la fonction.

mogène sur $\cup_{n \geq 1} \Pi_n$, de noyau de transition donné pour tous $(\pi, \pi') \in \Pi_n \times \Pi_{n+1}$ par

$$\mathbb{P}(\pi_{n+1} = \pi' \mid \pi_n = \pi) = \begin{cases} \frac{|b|}{\theta + n} & \text{si } \pi' \text{ s'obtient en insérant } n+1 \text{ dans le bloc } b \text{ de } \pi; \\ \frac{\theta}{\theta + n} & \text{si } \pi' \text{ s'obtient en ajoutant le bloc singleton } \{n+1\} \text{ à } \pi; \\ 0 & \text{sinon;} \end{cases}$$

où $|b|$ désigne le cardinal de b . On a $\sum_{b \in \pi_n} |b| = n$ où $b \in \pi_n$ signifie que b est un bloc de π_n .

Remarque 3.0.1 (Cas extrêmes). Lorsque $\theta = 0$ on a $\pi_n = \{\{1, \dots, n\}\}$ tandis que si $\theta = \infty$ alors $\pi_n = \{\{1\}, \dots, \{n\}\}$. Les probabilités sont monotones en θ . Plus θ est grand, plus les clients ont tendance à s'asseoir à une table vide plutôt que de rejoindre une table occupée.

Remarque 3.0.2 (Remonter le temps). Voici un exemple de début de trajectoire $(\pi_n)_{n \geq 1}$:

$$\begin{aligned} \pi_1 &= \{\{1\}\} \\ \pi_2 &= \{\{1\}, \{2\}\} \\ \pi_3 &= \{\{1, 3\}, \{2\}\} \\ \pi_4 &= \{\{1, 3\}, \{2\}, \{4\}\} \\ &\vdots \quad \quad \quad \vdots \end{aligned}$$

On remarque immédiatement qu'il est possible de remonter le temps. Par exemple, à partir de π_4 , on en déduit π_3 en repérant le bloc contenant 4, puis on en déduit π_2 en repérant le bloc contenant 3, puis π_1 en repérant le bloc contenant 2. Plus généralement, pour tout $n \geq 1$ et tout $\pi' \in \Pi_{n+1}$, il existe un unique $\pi \in \Pi_n$ tel que $\mathbb{P}(\pi_{n+1} = \pi' \mid \pi_n = \pi) > 0$. Il s'agit là d'une propriété commune à tous les processus qui transportent intégralement leur passé (ici l'information sur π_n est intégralement contenue dans π_{n+1} sans dégradation).

3.1 Lois d'Ewens

On note $|\pi|$ le nombre de blocs de la partition $\pi \in \Pi_n$ et $|\sigma|$ le nombre de cycles de la permutation $\sigma \in \mathcal{S}_n$. En particulier, on a $|\pi_n| = |\sigma_n|$ pour tout $n \geq 1$.

Théorème 3.1.1 (Loi d'Ewens sur \mathcal{S}_n). Pour tout $\sigma \in \mathcal{S}_n$, on a

$$\mathbb{P}(\sigma_n = \sigma) = \frac{\theta^{|\sigma|}}{\theta(\theta+1) \cdots (\theta+n-1)}.$$

Démonstration. La formule est vraie pour $n = 1$. Procédons par récurrence sur n et supposons la vraie pour n . Soit $\sigma' \in \mathcal{S}_{n+1}$. On observe tout d'abord qu'il existe un unique $\sigma \in \mathcal{S}_n$ tel que $\mathbb{P}(\sigma_{n+1} = \sigma' \mid \sigma_n = \sigma) > 0$. Si $|\sigma'| = |\sigma| + 1$ alors $\mathbb{P}(\sigma_{n+1} = \sigma' \mid \sigma_n = \sigma) = \theta/(\theta+n)$ et si $|\sigma'| = |\sigma|$ alors $\mathbb{P}(\sigma_{n+1} = \sigma' \mid \sigma_n = \sigma) = 1/(\theta+n)$. Dans les deux cas, on a bien

$$\mathbb{P}(\sigma_{n+1} = \sigma') = \mathbb{P}(\sigma_n = \sigma) \mathbb{P}(\sigma_{n+1} = \sigma' \mid \sigma_n = \sigma) = \frac{\theta^{|\sigma'|}}{\theta(\theta+1) \cdots (\theta+n)}.$$

□

Remarque 3.1.2 (Loi d'Ewens et loi uniforme sur \mathcal{S}_n). Le cas particulier $\theta = 1$ correspond à la loi uniforme sur \mathcal{S}_n étudiée dans le chapitre 10. En effet, lorsque $\theta = 1$, et conditionnellement à l'événement $\{\sigma_n = \sigma\}$, on peut construire σ_{n+1} à partir de σ en choisissant tout d'abord uniformément un élément aléatoire U de $\{1, \dots, n+1\}$, puis en rejoignant le cycle de cet élément dans σ si $U \leq n$ ou en créant un nouveau cycle $(n+1)$ si $U = n+1$, ce qui revient à calculer le produit $\sigma\tau$ où τ est la transposition aléatoire $(n+1, U)$. Ainsi, si $\theta = 1$ alors $(\sigma_n)_{n \geq 1}$ a même loi que $((1, U_1) \cdots (n, U_n))_{n \geq 1}$ où $(U_n)_{n \geq 1}$ est une suite de variables aléatoires indépendantes avec U_n de loi uniforme sur $\{1, \dots, n\}$ pour tout $n \geq 1$.

Théorème 3.1.3 (Loi d'Ewens sur Π_n). Pour tout $\pi \in \Pi_n$, on a

$$\mathbb{P}(\pi_n = \pi) = \frac{\theta^{|\pi|}}{\theta(\theta+1) \cdots (\theta+n-1)} \prod_{b \in \pi} (|b| - 1)!.$$

Démonstration. Le théorème 3.1.1 donne

$$\mathbb{P}(\pi_n = \pi) = \sum_{\sigma \in E_\pi} \mathbb{P}(\sigma_n = \sigma) = \frac{\theta^{|\pi|}}{\theta(\theta+1) \cdots (\theta+n-1)} \sum_{\sigma \in E_\pi} 1$$

où E_π est l'ensemble des $\sigma \in \mathcal{S}_n$ dont la partition de la décomposition en cycles est π . Il y a $(k-1)!$ cycles de longueur k d'un ensemble à k éléments donc $\sum_{\sigma \in E_\pi} 1 = \prod_{b \in \pi} (|b| - 1)!$. \square

Remarque 3.1.4 (Loi d'Ewens et loi uniforme sur Π_n). Pour $n \geq 3$, quelque soit θ , la loi d'Ewens sur Π_n n'est jamais la loi uniforme étudiée dans le chapitre 10. En effet, notons

$$\pi = \{\{1\}, \{2\}, \dots, \{n\}\}, \quad \pi' = \{\{1, 2\}, \{3\}, \dots, \{n\}\}, \quad \text{et} \quad \pi'' = \{\{1, 2, \dots, n\}\}.$$

Alors $\mathbb{P}(\pi_n = \pi) = \mathbb{P}(\pi_n = \pi')$ uniquement pour $\theta = 1$ et dans ce cas $\mathbb{P}(\pi_n = \pi) \neq \mathbb{P}(\pi_n = \pi'')$.

Pour tout temps $n \geq 1$ et tout $1 \leq k \leq n$, soit $A_{n,k}$ le nombre de tables de taille k au temps n , c'est-à-dire le nombre de blocs de taille k dans la partition aléatoire π_n . On a

$$n = A_{n,1} + 2A_{n,2} + \cdots + nA_{n,n} \quad \text{et} \quad |\pi_n| = A_{n,1} + \cdots + A_{n,n}.$$

Théorème 3.1.5 (Loi d'Ewens sur les blocs). Pour tout $n \in \mathbb{N}^*$ et tout $(a_1, \dots, a_n) \in \mathbb{N}$ vérifiant $a_1 + 2a_2 + \cdots + na_n = n$, on a

$$\mathbb{P}(A_{n,1} = a_1, \dots, A_{n,n} = a_n) = \frac{n!}{\theta(\theta+1) \cdots (\theta+n-1)} \prod_{j=1}^n \frac{1}{a_j!} \left(\frac{\theta}{j}\right)^{a_j}.$$

Démonstration. Le théorème 3.1.3 donne

$$\mathbb{P}(A_{n,1} = a_1, \dots, A_{n,n} = a_n) = \frac{\theta^{a_1 + \cdots + a_n}}{\theta(\theta+1) \cdots (\theta+n-1)} \prod_{j=1}^n ((j-1)!)^{a_j} \sum_{\pi \in \Pi_n(a)} 1$$

où $\Pi_n(a)$ désigne l'ensemble des $\pi \in \Pi_n$ comportant a_k blocs de taille k pour tout $1 \leq k \leq n$. Or l'ensemble $\Pi_n(a)$ a pour cardinal $n! / \prod_{j=1}^n (j!)^{a_j} a_j!$, d'où la formule. \square

Remarque 3.1.6. Le processus des restaurants chinois permet l'apparition de petites tables et leur maintien au fil du temps car elles sont choisies au prorata de leur taille !

3.2 Nombre de tables

Au temps $n \geq 1$, la salle du restaurant se compose de $|\pi_n|$ tables occupées.

Théorème 3.2.1 (Nombre de tables). *Pour tout $n \geq 1$ on a*

$$\mathbb{E}(|\pi_n|) = \sum_{k=0}^{n-1} \frac{\theta}{\theta + k} = \theta \log(n) + \mathcal{O}_{n \rightarrow \infty}(1) \sim_{n \rightarrow \infty} \theta \log(n)$$

et

$$\text{Var}(|\pi_n|) = \sum_{k=1}^{n-1} \frac{\theta k}{(\theta + k)^2} = \theta \log(n) + \mathcal{O}_{n \rightarrow \infty}(1) \sim_{n \rightarrow \infty} \theta \log(n).$$

Démonstration. Comme $(\pi_n)_{n \geq 1}$ est markovienne, c'est une suite récurrente aléatoire générée à partir de $\pi_1 = \{\{1\}\}$ par $\pi_{n+1} = f(\pi_n, \varepsilon_{n+1})$ où $(\varepsilon_n)_{n \geq 1}$ sont i.i.d. de loi uniforme sur $[0, 1]$. La structure de la fonction f entraîne alors qu'il existe une suite $(\xi_k)_{k \geq 1}$ de variables aléatoires indépendantes de lois de Bernoulli, vérifiant pour tout $k \geq 1$

$$\mathbb{P}(\xi_k = 1) = 1 - \mathbb{P}(\xi_k = 0) = \frac{\theta}{\theta + k - 1},$$

et telle que la variable aléatoire $|\pi_n|$ a même loi que $\xi_1 + \dots + \xi_n$ pour tout $n \geq 1$. L'événement $\{\xi_n = 1\}$ signifie que le client n décide de créer sa propre table, tandis que l'événement $\{\xi_n = 0\}$ signifie qu'il décide de rejoindre une table existante. En particulier, $\xi_1 = 1$ p.s. Cela donne pour $\mathbb{E}(|\pi_n|)$ et $\text{Var}(|\pi_n|)$ les formules en sommes. Une comparaison série-intégrale² montre que ces sommes sont des $\theta \log(n) + \mathcal{O}_{n \rightarrow \infty}(1) \sim_{n \rightarrow \infty} \theta \log(n)$. \square

Remarque 3.2.2 (Un jeu de pile ou face inhomogène). *Par construction, $(|\pi_n|)_{n \geq 1}$ est p.s. croissante et ne fait que des sauts de +1. Elle constitue le processus de comptage partant de 1 de tops espacés par des durées aléatoires indépendantes. Cependant, ces durées ne sont pas de même loi, et ne sont pas de loi géométrique. Il s'agit plutôt d'un jeu de pile ou face où la probabilité de gagner change à chaque lancer. Bien que $(\pi_n)_{n \geq 1}$ soit croissante, elle est cependant de moins en moins croissante en quelque sorte puisque la quantité*

$$p_n = \mathbb{P}(|\pi_{n+1}| = |\pi_n| + 1) = \mathbb{P}(\xi_{n+1} = 1) = \theta(\theta + n)^{-1}$$

décroit quand n croît. Malgré tout, la moyenne et la variance de $|\pi_n|$ sont équivalentes à $\theta \log(n)$ lorsque n tend vers ∞ . La situation diffère du cas du collectionneur de coupons du chapitre 1, pour lequel la probabilité de gagner change après chaque succès.

Théorème 3.2.3 (Comportement asymptotique du nombre de tables en \mathbb{P}). *On a*

$$\frac{|\pi_n|}{\log(n)} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \theta.$$

Démonstration. Par l'inégalité de Markov et le théorème 3.2.1, pour tous $\varepsilon > 0$ et $n \geq 2$,

$$\mathbb{P}\left(\left|\frac{|\pi_n|}{\log(n)} - \theta\right| > \varepsilon\right) \leq \frac{\text{Var}(|\pi_n|) + (\mathbb{E}(|\pi_n|) - \theta \log(n))^2}{(\varepsilon \log(n))^2} = \mathcal{O}\left(\frac{1}{\log(n)}\right) = o(1).$$

\square

On peut raffiner ce résultat en établissant une convergence presque sûre.

2. Si $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ est continue et décroissante alors $\int_0^{n+1} f(t) dt \leq \sum_{k=0}^n f(k) \leq f(0) + \int_0^n f(t) dt$.

Théorème 3.2.4 (Comportement asymptotique presque sûr du nombre de tables). *On a*

$$\frac{|\pi_n|}{\log(n)} \xrightarrow[n \rightarrow \infty]{p.s.} \theta.$$

Démonstration. Soient $(\xi_n)_{n \geq 1}$ comme dans la preuve du théorème 3.2.1 de sorte que

$$\left(\frac{|\pi_n|}{\log(n)} \right)_{n \geq 2} \stackrel{\text{loi}}{=} \left(\frac{1}{\log(n)} \sum_{k=1}^n \xi_k \right)_{n \geq 2}.$$

Notons que $\xi_1 = 1$ p.s. et donc $\xi_1 - \mathbb{E}(\xi_1) = 0$. Pour tout $n \geq 2$, on a la décomposition

$$\frac{1}{\log(n)} \sum_{k=1}^n \xi_k = \frac{1}{\log(n)} \sum_{k=2}^n (\xi_k - \mathbb{E}(\xi_k)) + \frac{1}{\log(n)} \sum_{k=1}^n \mathbb{E}(\xi_k).$$

Le second terme du membre de droite converge vers θ quand $n \rightarrow \infty$ (théorème 3.2.1). Il suffit donc d'établir que le premier terme du membre de droite converge p.s. vers 0. Or si

$$S_n := \sum_{k=2}^n Y_k \quad \text{avec} \quad Y_k := \frac{\xi_k - \mathbb{E}(\xi_k)}{\log(k)}$$

alors $(S_n)_{n \geq 2}$ est une martingale de carré intégrable bornée dans L^2 car (théorème 3.2.1)

$$\mathbb{E}(S_n^2) = \sum_{k=2}^n \frac{\text{Var}(\xi_k)}{(\log(k))^2} \leq \sum_{k=2}^{\infty} \frac{\theta}{(\theta + k - 1)(\log(k))^2} < \infty$$

(il est utile de rappeler ici que la série de Bertrand $\sum_{k=2}^{\infty} \frac{1}{k(\log(k))^\beta}$ converge ssi $\beta > 1$). Étant bornée dans L^2 , la martingale $(S_n)_{n \geq 2}$ converge p.s. (et dans L^2) vers une v.a.r. finie p.s. (et dans L^2). Le lemme de Kronecker³ assure alors la convergence p.s. vers 0 quand $n \rightarrow \infty$ de

$$\frac{1}{\log(n)} \sum_{k=2}^n \log(k) Y_k = \frac{1}{\log(n)} \sum_{k=2}^n (\xi_k - \mathbb{E}(\xi_k)).$$

□

Théorème 3.2.5 (Comportement asymptotique en loi du nombre de tables). *On a*

$$\frac{|\pi_n| - \mathbb{E}(|\pi_n|)}{\sqrt{\text{Var}(|\pi_n|)}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

De plus $\mathbb{E}(|\pi_n|)$ et $\text{Var}(|\pi_n|)$ peuvent être remplacés par $\theta \log(n)$.

Démonstration. Tout d'abord, on observe que $\mathbb{E}(|\pi_n|)$ et $\text{Var}(|\pi_n|)$ peuvent être remplacés par $\theta \log(n)$ en utilisant le lemme de Slutsky et le théorème 3.2.1.

Le résultat découle du théorème central limite quantitatif de Berry-Esseen page 16 utilisé avec $X_k = \xi_k$ de loi de Bernoulli de moyenne $\theta/(\theta + k - 1)$. On obtient même une vitesse de convergence uniforme pour les fonctions de répartition ! Une preuve alternative peut être obtenue en utilisant l'inégalité d'approximation poissonnienne de Le Cam renforcée page 16 avec $p_k = \theta/(\theta + k - 1)$. En effet, cela donne $d_V(\mu_n, \nu_n) = o_{n \rightarrow \infty}(1)$, et par conséquent, pour toute fonction continue et bornée $f : \mathbb{R} \rightarrow \mathbb{R}$, on a, avec $\sigma_n^2 = \text{Var}(|\pi_n|)$,

$$\left| \mathbb{E} \left(f \left(\frac{|\pi_n| - \lambda_n}{\sigma_n} \right) \right) - \mathbb{E} \left(f \left(\frac{P_n - \lambda_n}{\sigma_n} \right) \right) \right| \leq 2 \|f\|_\infty d_V(\mu_n, \nu_n) \xrightarrow[n \rightarrow \infty]{} 0.$$

3. Si la série réelle $\sum_n x_n$ converge alors $\lim_{n \rightarrow \infty} b_n^{-1} \sum_{k=1}^n b_k x_k = 0$ lorsque $0 < b_n \nearrow \infty$.

où P_n est une v.a.r. de loi de Poisson de moyenne $\lambda_n = p_1 + \dots + p_n$. D'autre part, on a

$$\frac{P_n - \lambda_n}{\sigma_n} = \frac{P_n - \lambda_n}{\sqrt{\lambda_n}} \frac{\sqrt{\lambda_n}}{\sigma_n} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1)$$

par le théorème central limite, le théorème 3.2.1 ($\sqrt{\lambda_n}/\sigma_n \rightarrow 1$), et le lemme de Slutsky. \square

Remarque 3.2.6. La preuve précédente montre que la poissonnisation est utile même si la moyenne diverge : $\lambda_n = p_1 + \dots + p_n \rightarrow \infty$. D'autre part, on a $p_1^2 + \dots + p_n^2 \not\rightarrow 0$ ce qui fait que la simple inégalité de Le Cam $d_V(\mu_n, \nu_n) = \mathcal{O}(p_1^2 + \dots + p_n^2)$ ne suffit pas.

3.3 Tables extrêmes

À l'instant n le restaurant compte $A_{n,1}$ tables réduites à un seul client. La combinatoire permet d'obtenir la loi de $A_{n,1}$. Contentons-nous ici des deux premiers moments.

Théorème 3.3.1 (Clients solitaires). Pour tout temps $n \geq 1$, le nombre $A_{n,1}$ de tables réduites à un seul client vérifie

$$\mathbb{E}(A_{n,1}) = \frac{n\theta}{n + \theta - 1} \quad \text{et} \quad \text{Var}(A_{n,1}) = \frac{n(n-1)(n-2+2\theta)\theta}{(n+\theta-2)(n+\theta-1)^2}.$$

En particulier,

$$\lim_{n \rightarrow \infty} \mathbb{E}(A_{n,1}) = \theta \quad \text{et} \quad \lim_{n \rightarrow \infty} \text{Var}(A_{n,1}) = \theta.$$

Notons que pour $\theta = 1$, la permutation aléatoire σ_n suit la loi uniforme sur \mathcal{S}_n , et possède $|\sigma_n| = |\pi_n| \sim \log(n)$ cycles, et en moyenne $\mathbb{E}(A_{n,1}) \sim 1$ point fixe quand $n \rightarrow \infty$.

Démonstration. Il s'agit de décrire la loi de la première composante $A_{n,1}$ d'un vecteur aléatoire A_n de \mathbb{N}^n qui suit la loi d'Ewens. Il est cependant plus commode de voir $A_{n,1}$ comme le nombre de blocs de taille 1 dans π_n . On a $A_{1,1} = 1$ et $0 \leq A_{n,1} \leq n$ pour tout $n \in \mathbb{N}^*$. De plus, pour tous entiers $1 \leq a_1, \dots, a_n \leq n$ et $0 \leq a_{n+1} \leq n+1$ on a

$$\mathbb{P}(A_{n+1,1} = a_{n+1} \mid A_{1,1} = a_1, \dots, A_{n,1} = a_n) = \mathbb{P}(A_{n+1,1} = a_{n+1} \mid A_{n,1} = a_n)$$

qui vaut

$$\begin{cases} \frac{\theta}{\theta + n} & \text{si } a_{n+1} = a_n + 1 \text{ (s'attabler seul à une table vide);} \\ \frac{a_n}{\theta + n} & \text{si } a_{n+1} = a_n - 1 \text{ (rejoindre la table d'un solitaire);} \\ \frac{n - a_n}{\theta + n} & \text{si } a_{n+1} = a_n \text{ (rejoindre une table comptant 2 clients ou +);} \\ 0 & \text{sinon.} \end{cases}$$

En particulier, $(A_{n,1})_{n \geq 1}$ est une chaîne de Markov inhomogène d'espace d'état \mathbb{N} . On obtient la remarquable formule affine (une martingale se cache dans la chaîne de Markov !)

$$\mathbb{E}(A_{n+1,1} \mid A_{n,1} = a) = \frac{a(\theta + n - 1) + \theta}{\theta + n}.$$

Cela donne $(n + \theta)m_{n+1} = (\theta + n - 1)m_n + \theta$ où $m_n := \mathbb{E}(A_{n,1})$. La formule annoncée pour m_n s'en déduit. Pour la variance, on procède de la même manière avec la formule

$$\text{Var}(X) = \mathbb{E}(\text{Var}(X | Y)) + \text{Var}(\mathbb{E}(X | Y)).$$

Les calculs sont cependant un peu plus lourds que pour l'espérance. □

L'entier $A_{n,n}$ représente le nombre de tables comportant n clients, autrement dit le nombre de blocs de taille n dans π_n . Lorsqu'une telle table existe, elle regroupe tous les clients, et on dit donc qu'il s'agit d'une table unique. On a $A_{n,n} = 1$ si et seulement si $|\pi_n| = 1$. Le théorème suivant précise les choses, et montre en particulier qu'asymptotiquement, le modèle des restaurants chinois ne fait pas apparaître de table unique.

Théorème 3.3.2 (Table unique). *On a $A_{1,1} = 1$ et $A_{n,n} \in \{0, 1\}$ pour tout $n \geq 1$. Pour tout $n \geq 2$, la probabilité qu'il n'y ait qu'une seule table vaut*

$$\mathbb{P}(A_{n,n} = 1) = \prod_{k=1}^{n-1} \frac{k}{\theta + k}.$$

Enfin, la suite $(A_{n,n})_{n \geq 1}$ décroît presque-sûrement vers 0.

Démonstration. Les premières assertions du théorème découlent directement de la définition de π_n . L'expression de la probabilité $\mathbb{P}(A_{n,n} = 1)$ s'obtient en observant que $\mathbb{P}(A_{n,n} = 1) = \mathbb{P}(\xi_2 = 0, \dots, \xi_n = 0) = \mathbb{P}(\xi_2 = 0) \cdots \mathbb{P}(\xi_n = 0)$. Pour la convergence presque-sûre, on remarque tout d'abord que, puisque $\theta > 0$, le produit infini $\prod_{k=1}^{\infty} (1 + \theta k^{-1})$ diverge car la série harmonique diverge, et par conséquent $\lim_{n \rightarrow \infty} \mathbb{P}(A_{n,n} = 1) = 0$. D'autre part, la suite d'événements $(E_n)_{n \geq 1}$ définie par $E_n = \{A_{n,n} = 0\}$ est croissante, et donc

$$\mathbb{P}(\lim_{n \rightarrow \infty} A_{n,n} = 0) = \mathbb{P}(\cup_{n=1}^{\infty} E_n) = \lim_{n \rightarrow \infty} \mathbb{P}(E_n) = 1 - \lim_{n \rightarrow \infty} \mathbb{P}(A_{n,n} = 1) = 1.$$

En fait, $A_{n,n} = \mathbf{1}_{\{n < T\}}$ où $T := \inf\{n \geq 1 : \xi_n = 1\}$ est l'instant de premier succès dans un jeu de pile ou face dont la probabilité de gagner change à chaque lancer, et décroît vers 0 au fil du temps. La décroissance est cependant suffisamment lente pour assurer un succès certain. En effet, on a $\{T < \infty\} = \cup_{n=1}^{\infty} E_n$ et donc $\mathbb{P}(T < \infty) = 1$. Alternativement, on peut utiliser le lemme de Borel-Cantelli pour les événements indépendants $F_n = \{\xi_n = 1\}$ qui vérifient $\sum_{n \geq 1} \mathbb{P}(F_n = 1) = \sum_{n \geq 1} \frac{\theta}{\theta + n} = \infty$ et $\lim F_n = \{\sum_{n \geq 1} \mathbf{1}_{\{\xi_n = 1\}} = \infty\} \subset \{T < \infty\}$. □

3.4 Compléments de combinatoire

D'après le chapitre 10, le cardinal de Π_n est le n -ème nombre de Bell π_n . On a

$$\pi_n = \sum_{k=1}^n \left\{ \begin{matrix} n \\ k \end{matrix} \right\}$$

où la notation entre accolades désigne le nombre de Stirling de seconde espèce, qui compte le nombre de partitions à k blocs de $\{1, \dots, n\}$. On dispose de la formule de récurrence

$$\left\{ \begin{matrix} n \\ k \end{matrix} \right\} = \left\{ \begin{matrix} n-1 \\ k-1 \end{matrix} \right\} + k \left\{ \begin{matrix} n-1 \\ k \end{matrix} \right\} \quad \text{avec conditions au bord} \quad \left\{ \begin{matrix} n \\ 1 \end{matrix} \right\} = 1 \quad \text{et} \quad \left\{ \begin{matrix} n \\ n \end{matrix} \right\} = 1$$

car pour choisir une partition de $\{1, \dots, n\}$ ayant k blocs il faut et il suffit soit de choisir une partition de $\{1, \dots, n-1\}$ ayant $k-1$ blocs et de la compléter avec le bloc singleton $\{n\}$,

soit d'ajouter l'élément n à l'un des k blocs d'une partition de $\{1, \dots, n-1\}$ ayant k blocs. Si X est une variable aléatoire de loi de Poisson de paramètre λ alors

$$\mathbb{E}(X^n) = \sum_{k=1}^n \left\{ \begin{matrix} n \\ k \end{matrix} \right\} \lambda^k \quad \text{en particulier} \quad \mathbb{E}(X^n) = \pi_n \quad \text{si } \lambda = 1.$$

On dispose également de la formule explicite suivante :

$$\left\{ \begin{matrix} n \\ k \end{matrix} \right\} = \frac{1}{k!} \sum_{j=1}^k (-1)^{k-j} \binom{k}{j} j^n$$

qui peut s'obtenir grâce au principe d'inclusion-exclusion en remarquant que le nombre de Stirling de seconde espèce est égal au nombre de surjections de $\{1, \dots, n\}$ dans $\{1, \dots, k\}$ divisé par $k!$.

Une partition de $\{1, \dots, n\}$ prescrit le support des cycles mais ne précise pas la structure interne de chaque cycle. Il peut donc y avoir plusieurs permutations pour chaque partition. Le nombre de permutations de $\{1, \dots, n\}$ qui possèdent exactement k cycles est donné par le nombre de Stirling de première espèce non signé, noté

$$\left[\begin{matrix} n \\ k \end{matrix} \right],$$

égal au coefficient de x^k dans le polynôme de Pochhammer $x_{(n)} = x(x+1) \cdots (x+n-1)$.

3.5 Notes et commentaires

L'essentiel de ce chapitre est tiré de l'article didactique [CDM12]. Warren Ewens est un professeur de biologie mathématique né en 1937 en Australie. C'est à la fin des années 1960 qu'il découvre la loi qui porte aujourd'hui son nom, en étudiant un problème d'échantillonnage en génétique des populations, lié au modèle de Fisher et Wright, abordé dans le chapitre 2. Le paramètre θ apparaît comme un taux de mutation des allèles, comme expliqué dans le chapitre 9. Une synthèse sur le sujet se trouve dans son livre [Ewe04], ainsi que dans celui de Kingman [Kin80] et de Durrett [Dur08]. De nombreux aspects statistiques sont abordés dans le cours de Tavaré [Tav04]. Le travail d'Ewens a engendré un nombre considérable de travaux en biologie quantitative et en probabilités. La loi d'Ewens apparaît dans une large gamme de structures aléatoires discrètes dites logarithmiques, allant de la combinatoire à la théorie des nombres. On pourra consulter à ce sujet le livre de Arratia, Barbour, et Tavaré [ABT03]. Il semble que le processus des restaurants chinois doive son nom à Pitman et Dubin. Il apparaît sous ce nom dans un cours d'Aldous [Ald85]. On peut le relier aux urnes de Pólya ainsi qu'aux processus de Dirichlet. De nos jours, le processus des restaurants chinois et la loi d'Ewens font désormais partie du folklore d'une théorie plus générale de la (fragmentation et de la) coalescence. On pourra à ce sujet consulter les livres de Kingman [Kin93], de Bertoin [Ber06], de Pitman [Pit06], ainsi que de Berestycki [Ber09].

La loi d'Ewens possède d'autres représentations remarquables. En voici deux :

1. Si Z_1, \dots, Z_n sont indépendantes de lois de Poisson de moyennes $\theta/1, \dots, \theta/n$ alors

$$\mathcal{L}(Z_1, \dots, Z_n \mid Z_1 + 2Z_2 + \cdots + nZ_n = n) \sim \text{Ewens}(\theta);$$

2. Soit $(P_r)_{r \geq 1}$ la loi sur \mathbb{N}^* aléatoire générée en posant (partition de $[0, 1]$)

$$P_1 = W_1, P_2 = (1 - W_1)W_2, P_3 = (1 - W_1)(1 - W_2)W_3, \dots$$

où $(W_r)_{r \geq 1}$ sont i.i.d. de loi Beta $w \mapsto \theta(1 - w)^{\theta-1} \mathbf{1}_{[0,1]}(w)$. Conditionnellement à $(P_r)_{r \geq 1}$, soient X_1, \dots, X_n des v.a. i.i.d. sur \mathbb{N}^* de loi $(P_r)_{r \geq 1}$. La suite X_1, \dots, X_n fait apparaître au plus n entiers différents, et leurs effectifs suit la loi Ewens(θ).

Signalons qu'il est possible d'établir que pour tous $0 \leq k \leq n$,

$$\mathbb{P}(|\pi_n| = k) = \binom{n}{k} \frac{\theta^k}{\theta(\theta + 1) \cdots (\theta + n - 1)}$$

et d'obtenir une formule du même genre pour la loi du couple $(A_n, |\pi_n|)$, cf. [JKB97].

Séance 4

Processus de branchement de Galton-Watson

Mots-clés. Arbres ; martingales ; chaînes de Markov à espace d'états dénombrable.

Le processus de Galton-Watson décrit l'évolution d'une population asexuée au fil de générations qui ne se recoupent pas. Soit $P = p_0\delta_0 + p_1\delta_1 + \dots$ une loi sur \mathbb{N} appelée *loi de reproduction*. On passe de la génération n à la génération $n + 1$ comme suit : chaque individu de la génération n fait, indépendamment de tout le reste, un nombre aléatoire d'enfants de loi P puis meurt. Si Z_n désigne le nombre d'individus de la génération $n \in \mathbb{N}$, alors Z_0 représente la taille de la population initiale et Z_{n+1} vérifie l'équation de récurrence

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

où $(X_{n,k})_{n \geq 1, k \geq 1}$ sont i.i.d. de loi P , indépendantes de Z_0 . On adopte la convention $\sum_{\emptyset} = 0$ de sorte que si $Z_n = 0$ alors $Z_{n+1} = 0$. Pour tous $n \in \mathbb{N}$ et $z_0, \dots, z_n \in \mathbb{N}$, on a

$$\mathcal{L}(Z_{n+1} | Z_0 = z_0, \dots, Z_n = z_n) = \mathcal{L}(Z_{n+1} | Z_n = z_n) = P^{*z_n}.$$

La suite $(Z_n)_{n \geq 0}$ est donc une chaîne de Markov d'espace d'état \mathbb{N} et de noyau de transition

$$\mathbf{P}(z, \cdot) = P^{*z}.$$

L'état 0 est absorbant (extinction de la population). Le temps d'extinction est

$$T := \inf\{n \geq 0 : Z_n = 0\} \in \mathbb{N} \cup \{\infty\}.$$

Ce modèle rudimentaire de dynamique de population ne tient pas compte du sexe. Il permet par exemple de modéliser de manière grossière le nombre de noms de familles en occident (car il suffit de se focaliser sur la descendance mâle).

Remarque 4.0.1 (Arbre de Galton-Watson). *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 . La variable Z_n est le nombre de noeuds de profondeur n . Un arbre r -naire correspond à $P = \delta_r$ mais on prendra garde au fait qu'il ne s'agit pas d'un graphe r -régulier car la racine (Abraham) possède r voisins tandis que tous les autres sommets (descendants d'Abraham) possèdent $r + 1$ voisins.*

On effectue à présent les observations suivantes :

- Le comportement de $(Z_n)_{n \geq 0}$ se ramène au cas $Z_0 = 1$ car¹ 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 $P = \delta_z$ et donc $Z_{n+1} = zZ_n = \dots = z^{n+1}Z_0$;
 - Si $p_0 = 0$ et $p_1 < 1$ alors $\mathbb{P}(Z_n \nearrow \infty \mid Z_0 = 1) = 1$ ($Z_n \nearrow$, comparer à pile ou face) ;
 - Si $p_0 + p_1 = 1$ et $p_0 \neq 0$, alors $\mathbb{P}(Z_n \searrow 0 \mid Z_0 = 1) = 1$ ($Z_n \searrow$, T est géométrique).
- Ces observations nous conduisent à supposer que dans toute la suite :

$$Z_0 = 1 \quad \text{et} \quad 0 < p_0 \leq p_0 + p_1 < 1 \quad (\text{implique que } p_z < 1 \text{ pour tout } z \in \mathbb{N}).$$

Notons qu'on a alors $Z_1 \sim P$ car $Z_0 = 1$. Dans toute la suite, on note $m = p_1 + 2p_2 + \dots$ la moyenne de P lorsqu'elle existe, et $\sigma^2 \in \overline{\mathbb{R}}_+$ la variance de P lorsque m existe.

4.1 Extinction et phénomène de seuil

Nous commençons par les deux premiers moments de Z_n .

Théorème 4.1.1 (Moyenne et variance). *Si m existe alors $\mathbb{E}(Z_n) = m^n$ pour tout $n \geq 0$ et si $\sigma^2 < \infty$ alors $\text{Var}(Z_n) = \sigma^2(m^n(m^n - 1))/(m^2 - m)$ prolongé en $n\sigma^2$ si $m = 1$.*

Démonstration. La formule de l'espérance découle de $\mathbb{E}(Z_{n+1} \mid Z_n) = mZ_n$ et $Z_0 = 1$. Pour la variance, $\mathbb{E}(Z_{n+1}^2 \mid Z_n) = Z_n(\sigma^2 + m^2) + Z_n(Z_n - 1)m^2$ et donc

$$\mathbb{E}(Z_{n+1}^2) = \mathbb{E}(Z_n)(\sigma^2 + m^2) + (\mathbb{E}(Z_n^2) - \mathbb{E}(Z_n))m^2.$$

Un calcul conduit alors à la récurrence linéaire $\text{Var}(Z_{n+1}) = \text{Var}(Z_n)m^2 + m^n\sigma^2$. □

Le comportement asymptotique de la taille moyenne $\mathbb{E}(Z_n) = m^n$ ne dépend que du paramètre m et révèle un phénomène de seuil autour de la valeur critique $m = 1$. On adopte tout naturellement la terminologie suivante polarisée par le paramètre m :

- cas sous-critique ($m < 1$) : $\mathbb{E}(Z_n) \searrow 0$
- cas critique ($m = 1$) : $\mathbb{E}(Z_n) = 1$
- cas sur-critique ($m > 1$) : $\mathbb{E}(Z_n) \nearrow \infty$

Les formules pour $\mathbb{E}(Z_n)$ et $\text{Var}(Z_n)$ peuvent être obtenues en utilisant une fonction génératrice. La fonction génératrice $g : [0, 1] \mapsto [0, 1]$ de la loi P est la série entière réelle

$$g(s) = \mathbb{E}(s^{X_{1,1}}) = \sum_{z=0}^{\infty} p_z s^z.$$

Si P a un moment d'ordre r alors $\lim_{s \rightarrow 1} g^{(r)}(s) = \mathbb{E}(X_{1,1}(X_{1,1} - 1) \cdots (X_{1,1} - r + 1))$ (moment factoriel d'ordre r de P). En particulier, $g'(1^-) = m$ et $g''(1^-) = \sigma^2 + m^2 - m$.

Théorème 4.1.2 (Génératrice). *Pour tout $n \geq 1$ la fonction génératrice g_n de Z_n vérifie*

$$g_n = g^{\circ n}.$$

Démonstration. On a $g_1 = g$ car $Z_0 = 1$. Pour tout $n \geq 1$ et tout $s \in [0, 1]$,

$$g_{n+1}(s) = \mathbb{E}(s^{Z_{n+1}}) = \mathbb{E}(\mathbb{E}(s^{Z_{n+1}} \mid Z_n)) = \mathbb{E}((\mathbb{E}(s^{X_{1,1}}))^{Z_n}) = \mathbb{E}(g(s)^{Z_n}) = g_n(g(s)).$$

□

1. C'est la propriété de branchement : $Z^{(x+y)} \stackrel{d}{=} Z^{(x)} + Z^{(y)}$ où $Z^{(x)}$ et $Z^{(y)}$ sont indépendants.

Remarque 4.1.3 (Liens entre Galton-Watson et Fisher-Wright). *Considérons le cas où la loi de reproduction P est une loi de Poisson $\text{Poi}(\lambda)$. Pour tout entier $N \geq 1$ fixé,*

$$\mathcal{L}((X_{n+1,1}, \dots, X_{n+1,N}) \mid X_{n+1,1} + \dots + X_{n+1,N} = N) = \text{Mul}(N, (1/N, \dots, 1/N)).$$

Ainsi, le processus de Fisher-Wright du chapitre 2 est lié au processus de Galton-Watson de loi de reproduction Poisson conditionné à être de taille constante (exercice 9.0.1). Le processus de Galton-Watson décrit l'évolution de la taille de la population, et ne dit rien à l'échelle individuelle. Réciproquement, si $(Y_1, \dots, Y_N) \sim \text{Mul}(N, (1/N, \dots, 1/N))$ alors pour tout $k \geq 1$ fixé, (Y_1, \dots, Y_k) converge en loi quand $N \rightarrow \infty$ vers la loi produit $\text{Poi}(1)^{\otimes k}$, ce qui indique que les individus dans un processus de Fisher-Wright de grande taille ont tendance à faire des enfants de manière poissonnienne et indépendante les uns des autres.

Théorème 4.1.4 (Dichotomie). *Presque sûrement, soit le processus s'éteint c'est-à-dire $\lim_{n \rightarrow \infty} Z_n = 0$ soit il tend vers l'infini c'est-à-dire $\lim_{n \rightarrow \infty} Z_n = \infty$. Autrement dit, p.s.*

$$\lim_{n \rightarrow \infty} Z_n = \begin{cases} \infty & \text{si } T = \infty; \\ 0 & \text{si } T < \infty. \end{cases}$$

Démonstration. 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}^* . \square

Le théorème 4.1.4 suggère de considérer la probabilité d'extinction $\mathbb{P}(T < \infty)$.

Théorème 4.1.5 (Probabilité d'extinction). *La probabilité d'extinction $\mathbb{P}(T < \infty)$ est point fixe de g . 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, $s \in]0, 1[$ (attractif) et 1 (répulsif) et $\mathbb{P}(T < \infty) = s$.*

Démonstration. Comme Z_n est entière, on a $\{T < \infty\} = \{\lim_{n \rightarrow \infty} Z_n = 0\} = \bigcup_{n \geq 0} \{Z_n = 0\}$. La suite $(\{Z_n = 0\})_{n \geq 0}$ est croissante et la probabilité d'extinction vérifie donc

$$\mathbb{P}(T < \infty) = \mathbb{P}(\lim_{n \rightarrow \infty} Z_n = 0) = \mathbb{P}(\bigcup_{n \geq 0} \{Z_n = 0\}) = \lim_{n \rightarrow \infty} \mathbb{P}(Z_n = 0).$$

Or du théorème 4.1.2 on tire $\mathbb{P}(Z_n = 0) = g_n(0) = g^{\circ n}(0)$ et donc

$$\mathbb{P}(T < \infty) = \lim_{n \rightarrow \infty} g_n(0) = \lim_{n \rightarrow \infty} g^{\circ n}(0).$$

La probabilité d'extinction $\mathbb{P}(T < \infty) = \lim_{n \rightarrow \infty} g^{\circ n}(0)$ est un point fixe de g (i.e. $g(s) = s$) car $g : [0, 1] \rightarrow [0, 1]$ est continue. La fonction g est convexe car

$$g''(s) = \sum_{z=0}^{\infty} (z+2)(z+1)s^z p_{z+2}.$$

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 $s \in]0, 1[$. Reste à observer que si $m > 1$ alors $g'(1^-) = m > 1$ et donc 1 n'est pas un point fixe attractif. La figure 4.1 illustre des cas particuliers tirés des exemples 4.1.6 et 4.1.7. \square

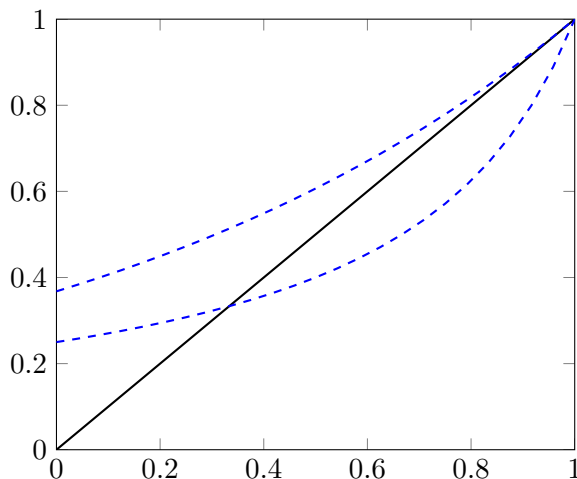


Figure 4.1 – En pointillés, la fonction génératrice de la loi géométrique de paramètre 1/4 (deux points fixes : 1/3, attractif, et 1, répulsif) ainsi que celle de la loi de Poisson de paramètre 1 (unique point fixe en 1). La ligne continue est la première bissectrice.

Exemple 4.1.6 (Reproduction poissonnienne). Si $P = \text{Poi}(\lambda)$ alors $m = \lambda$ et $g(s) = e^{\lambda(s-1)}$ et donc si $\lambda \leq 1$ alors la population s'éteint p.s. tandis que si $\lambda > 1$ alors la probabilité d'extinction s est l'unique solution sur $]0, 1[$ de l'équation $\lambda = \log(s)/(s - 1)$.

Exemple 4.1.7 (Reproduction géométrique). Si $P = \text{Geo}_{\mathbb{N}}(p) = \sum_{n \geq 0} q^n p \delta_n$ alors $m = q/p$ et $g(s) = p/(1 - sq)$ 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 $qs^2 - s + p = 0$. Un petit calcul fort agréable donne $s = p/q$.

On a donc la classification suivante d'après le théorème 4.1.5 :

- 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 $p \in]0, 1[$.

Pour étudier plus finement la suite $(Z_n)_{n \geq 0}$, on introduit la suite $(Y_n)_{n \geq 0}$ définie par

$$Y_n = \frac{Z_n}{\mathbb{E}(Z_n)} = \frac{Z_n}{m^n}.$$

Son utilité vient du fait qu'elle constitue une martingale pour la filtration naturelle :

$$\mathbb{E}(Y_{n+1} | \sigma(Y_0, \dots, Y_n)) = \frac{\mathbb{E}(X_{n+1,1} + \dots + X_{n+1,Z_n} | Z_n)}{m^{n+1}} = \frac{mZ_n}{m^{n+1}} = Y_n.$$

On a $\mathbb{E}(Y_n) = 1$. Comme $(Y_n)_{n \geq 0}$ est une martingale positive, elle converge p.s. vers une v.a.r. positive Y_∞ . Ainsi, sur l'événement $\{Y_\infty > 0\}$, on a $Z_n \sim_{n \rightarrow \infty} m^n Y_\infty$ p.s..

Comme $Y_n = 0$ si $n \geq T$, on obtient $\{T < \infty\} \subset \{Y_\infty = 0\}$ p.s.. D'après le théorème 4.1.5, dans le cas sous-critique ($m < 1$) et critique ($m = 1$), on a $\mathbb{P}(T < \infty) = 1$, d'où $\mathbb{P}(Y_\infty > 0) = 0$. Néanmoins, toujours si $m \leq 1$ alors $\sup_{n \geq 0} \frac{Z_n}{m^n}$ n'est pas intégrable, car sinon, par convergence dominée, on aurait $0 = \mathbb{E}(Y_T) = \mathbb{E}(\lim_{n \rightarrow \infty} Y_n) = \lim_{n \rightarrow \infty} \mathbb{E}(Y_n) = 1$.

4.2 Étude des trois cas possibles

Cas sous-critique

Si $m < 1$ alors le théorème 4.1.5 donne $\mathbb{P}(T < \infty) = 1$ et $Z_n = Z_{n \wedge T} \rightarrow Z_T = 0$ p.s. tandis que le théorème 4.1.1 donne $\mathbb{E}(Z_n) = m^n \rightarrow 0$ et $\text{Var}(Z_n) \rightarrow 0$ (si $\sigma^2 < \infty$).

Théorème 4.2.1 (Cas sous-critique (théorème de Yaglom)). *Si $m < 1$ alors $\mathcal{L}(Z_n | Z_n > 0)$ converge lorsque $n \rightarrow \infty$ vers une loi μ sur \mathbb{N}^* dont la fonction génératrice h vérifie*

$$(h - 1) \circ g = m(h - 1).$$

Démonstration. Soit h_n la fonction génératrice de $\mu_n := \mathcal{L}(Z_n | Z_n > 0)$. La suite $(h_n(1))_{n \geq 1}$ est constante et égale à 1 et converge donc vers $h(1) := 1$. Pour tout $s \in [0, 1[$, on a

$$1 - h_n(s) = 1 - \frac{\mathbb{E}(s^{Z_n} \mathbf{1}_{\{Z_n > 0\}})}{\mathbb{P}(Z_n > 0)} = 1 - \frac{\mathbb{E}(s^{Z_n}) - 1}{\mathbb{P}(Z_n > 0)} = \frac{1 - g_n(s)}{1 - g_n(0)}$$

car $\mathbb{P}(Z_n > 0) = 1 - \mathbb{P}(Z_n = 0) = 1 - g_n(0)$. On rappelle que $g_{n+1} = g \circ g_n$, et que $g_n(s) \rightarrow 1$ et $g_n(0) \rightarrow 1$ quand $n \rightarrow \infty$ car 1 est l'unique point fixe de g car $m < 1$ (théorème 4.1.5). Comme les fonctions $s \mapsto r(s) := (1 - g(s))/(1 - s)$ et g_n sont croissantes sur $[0, 1[$, on a

$$\frac{1 - h_{n+1}(s)}{1 - h_n(s)} = \frac{1 - g_{n+1}(s)}{1 - g_n(s)} \frac{1 - g_n(0)}{1 - g_{n+1}(0)} = \frac{r(g_n(s))}{r(g_n(0))} \geq 1.$$

Par conséquent, la suite positive $(1 - h_n(s))_{n \geq 1}$ est croissante, et comme elle est majorée par 1, elle converge vers une limite $1 - h(s)$. On a

$$1 - h_n(g(s)) = \frac{1 - g_{n+1}(0)}{1 - g_n(0)} \frac{1 - g_{n+1}(s)}{1 - g_{n+1}(0)}.$$

Comme $g_n(0) \nearrow \mathbb{P}(T < \infty) = 1$ on en déduit que

$$\frac{1 - g_{n+1}(0)}{1 - g_n(0)} = \frac{g(1) - g(g_n(0))}{g(1) - g_n(0)} \rightarrow g'(1) = m.$$

Donc $(1 - h) \circ g = m(1 - h)$ sur $[0, 1[$ et donc sur $[0, 1]$. Reste à établir que h est la fonction génératrice d'une loi μ sur \mathbb{N}^* et que $(\mu_n)_{n \geq 1}$ converge (étroitement) vers μ . Comme $(h_n)_{n \geq 1}$ converge ponctuellement sur $[0, 1[$ vers h , il est bien classique d'établir que $\mu_n(z)$ converge vers une limite $\mu(z) \in [0, 1]$ pour tout $z \in \mathbb{N}^*$, dont la fonction génératrice est égale à h sur $[0, 1[$. Le fait que la sous-probabilité μ est une loi de probabilité peut se déduire de l'équation fonctionnelle vérifiée par h en considérant la limite en 1. \square

Cas critique

Si $m = 1$ alors le théorème 4.1.5 donne $\mathbb{P}(T < \infty) = 1$ et $Z_n = Z_{n \wedge T} \rightarrow Z_T = 0$ p.s. tandis que le théorème 4.1.1 donne $\mathbb{E}(Z_n) = 1$ et $\text{Var}(Z_n) \rightarrow \infty$ (si $\sigma^2 < \infty$). La «largeur» de l'arbre $\sup_{n \geq 0} Z_n$ n'est pas intégrable car cela entraînerait par convergence dominée que $1 = \lim_{n \rightarrow \infty} \mathbb{E}(Z_n) = \mathbb{E}(\lim_{n \rightarrow \infty} Z_n) = 0$ ce qui est absurde. Les arbres critiques sont donc souvent très larges mais leur profondeur est finie p.s. (une folie des grandeurs passagère est rattrapée irrémédiablement par un manque de fertilité). Le cas est critique ! Notons par ailleurs que si $m = 1$ alors $(Z_n)_{n \geq 0} = (Y_n)_{n \geq 0}$ et on retrouve au passage que $Y_\infty = 0$ p.s.

Théorème 4.2.2 (Cas critique). *Si $m = 1$ et $\sigma^2 < \infty$ alors*

1. $\lim_{n \rightarrow \infty} n\mathbb{P}(Z_n > 0) = 2\sigma^{-2}$
2. $\lim_{n \rightarrow \infty} \frac{1}{n}\mathbb{E}(Z_n | Z_n > 0) = \frac{1}{2}\sigma^2$
3. $\mathcal{L}(n^{-1}Z_n | Z_n > 0)$ tend quand $n \rightarrow \infty$ vers la loi exponentielle de paramètre $2\sigma^{-2}$.

Le 2. n'est pas une conséquence directe du 3. car la convergence en loi n'entraîne pas la convergence des moments, en particulier n'entraîne pas la convergence des espérances. Un contre-exemple est fourni par $(Z_n)_{n \geq 1}$ qui converge p.s. vers 0 donc en loi vers δ_0 mais $1 = \mathbb{E}(Z_n)$ ne converge pas vers la moyenne de δ_0 qui est 0. Le théorème affirme cependant qu'il y a convergence de l'espérance de $n^{-1}Z_n$ vers l'espérance $2\sigma^{-2}$ de la limite en loi de $n^{-1}Z_n$. La liaison entre convergence en loi et convergence des moments est abordée dans le chapitre 6 pour la preuve du théorème de Wigner.

Démonstration. On a

$$n\mathbb{P}(Z_n > 0) = n(1 - g_n(0)) = \left(\frac{1}{n} \left(\frac{1}{1 - g_n(0)} - 1 \right) + \frac{1}{n} \right)^{-1}.$$

La formule de Taylor avec reste intégral en 1 à l'ordre 2 donne

$$g(s) = s + \frac{1}{2}(1-s)^2\sigma^2 + (1-s)^2\alpha(s)$$

avec α bornée sur $[0, 1]$ et $\lim_{s \rightarrow 1} \alpha(s) = 0$. Il en découle que

$$\frac{1}{1-g(s)} - \frac{1}{1-s} = \frac{g(s) - s}{(1-g(s))(1-s)} = \frac{\frac{1}{2}\sigma^2 + \alpha(s)}{1 - \frac{1}{2}(1-s)\sigma^2 - (1-s)\alpha(s)} = \frac{\sigma^2}{2} + \beta(s)$$

où β est comme α . Comme $(g_k)_{k \geq 1}$ converge uniformément vers 1, il vient

$$\lim_{n \rightarrow \infty} \frac{1}{n} \left(\frac{1}{1-g_n(s)} - \frac{1}{1-s} \right) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \left(\frac{1}{1-g_k(s)} - \frac{1}{1-g_k(s)} \right) = \frac{\sigma^2}{2}$$

uniformément pour tout $s \in [0, 1[$. De la même manière, comme $Z_n \mathbb{1}_{\{Z_n > 0\}} = Z_n$, on a

$$\mathbb{E}(n^{-1}Z_n | Z_n > 0) = \frac{\mathbb{E}(Z_n \mathbb{1}_{\{Z_n > 0\}})}{n\mathbb{P}(Z_n > 0)} = \frac{\mathbb{E}(Z_n)}{n(1-g_n(0))} \rightarrow \frac{\sigma^2}{2}.$$

Toujours de la même manière, on a pour tout $t \in \mathbb{R}_+^*$, en notant $s_n = e^{-t/n}$,

$$\begin{aligned} \mathbb{E}(e^{-t\frac{1}{n}Z_n} | Z_n > 0) &= \frac{\mathbb{E}(s_n^{Z_n} \mathbb{1}_{\{Z_n > 0\}})}{\mathbb{P}(Z_n > 0)} \\ &= \frac{g_n(s_n) - g_n(0)}{1 - g_n(0)} \\ &= 1 - \frac{1 - g_n(s_n)}{1 - g_n(0)} \\ &= 1 - \frac{1}{n(1-g_n(0))} \left(\frac{1}{n} \left(\frac{1}{1-g_n(s_n)} - \frac{1}{1-s_n} \right) + \frac{1}{n(1-s_n)} \right)^{-1} \\ &\rightarrow 1 - \frac{\sigma^2}{2} \left(\frac{\sigma^2}{2} + \frac{1}{t} \right)^{-1} = \frac{1}{1 + \frac{\sigma^2}{2}t} = \frac{2}{\sigma^2} \int_0^\infty e^{-tx} e^{-\frac{\sigma^2}{2}x} dx. \end{aligned}$$

□

Cas sur-critique

Si $m > 1$ alors le théorème 4.1.5 donne $0 < \mathbb{P}(T < \infty) < 1$, sur $\{T < \infty\}$ on a $Z_n \rightarrow 0$ tandis que sur $\{T = \infty\}$ on a $Z_n \rightarrow \infty$ (théorème 4.1.4), et d'autre part, le théorème 4.1.1 donne $\mathbb{E}(Z_n) = m^n \rightarrow \infty$ et $\text{Var}(Z_n) \rightarrow \infty$. Enfin, si $m > 1$, on sait que sur $\{T < \infty\}$ on a $Z_n \rightarrow 0$ tandis que sur $\{Y_\infty > 0\}$ on a $Z_n \sim_{n \rightarrow \infty} m^n Y_\infty$. On a aussi $\{T < \infty\} \subset \{Y_\infty = 0\}$ p.s. et le théorème suivant, fort agréable, affirme que cette inclusion p.s. est une égalité p.s.

Théorème 4.2.3 (Cas sur-critique – Dichotomie). *On a $\{Y_\infty = 0\} = \{T < \infty\}$ p.s. et donc p.s. soit $Y_\infty = 0$ et $\lim_{n \rightarrow \infty} Z_n = 0$ soit $Y_\infty > 0$ et $Z_n \sim_{n \rightarrow \infty} m^n Y_\infty \nearrow +\infty$.*

Démonstration. Comme $\mathbb{E}(Y_\infty) = 1$ et $Y_\infty \geq 0$ on en déduit que $\mathbb{P}(Y_\infty > 0) > 0$. Sur $\{Y_\infty > 0\}$ on a $Z_n \sim_{n \rightarrow \infty} m^n Y_\infty \rightarrow \infty$ et on retrouve $\mathbb{P}(T < \infty) < 1$. Reste à établir que l'inclusion p.s. $\{Y_\infty > 0\} \subset \{T = \infty\}$ est une égalité p.s. La réponse est positive car on a

$$\mathbb{P}(Y_\infty = 0) = \mathbb{P}(T < \infty).$$

Pour le voir, pour tout $z \in \mathbb{N}^*$, en utilisant les propriétés de Markov et de branchement,

$$\mathbb{P}(Y_\infty = 0 \mid Z_1 = z) = \mathbb{P}(\lim_{n \rightarrow \infty} Y_n = 0 \mid Z_1 = z) = \mathbb{P}(\lim_{n \rightarrow \infty} Y_n = 0)^z = \mathbb{P}(Y_\infty = 0)^z,$$

et comme $Z_1 \sim P$, on a $\mathbb{P}(Y_\infty = 0) = g(\mathbb{P}(Y_\infty = 0))$ donc $\mathbb{P}(Y_\infty = 0)$ est point fixe de g . Comme $\mathbb{P}(Y_\infty = 0) < 1$ on obtient $\mathbb{P}(Y_\infty = 0) = \mathbb{P}(T < \infty)$ grâce au théorème 4.1.5. \square

Le lemme de Fatou donne $\mathbb{E}(Y_\infty) \leq \liminf_{n \rightarrow \infty} \mathbb{E}(Y_n) = 1$, ce qui assure que Y_∞ est intégrable, mais la convergence de $(Y_n)_{n \geq 0}$ vers Y_∞ n'a pas forcément lieu dans L^1 . Enfin, le lemme de Scheffé nous dit que la convergence a lieu dans L^1 ssi $\mathbb{E}(Y_\infty) = 1$.

Théorème 4.2.4 (Cas sur-critique – Loi de Y_∞). *Si $m > 1$ et $\sigma^2 < \infty$ alors la martingale $(Y_n)_{n \geq 0}$ converge p.s. et dans L^2 vers une v.a.r. $Y_\infty \geq 0$ de moyenne $\mathbb{E}(Y_\infty) = 1$ et de variance $\text{Var}(Y_\infty) = \sigma^2/(m^2 - m)$. De plus, la transformée de Laplace $t \in \mathbb{R}_+ \mapsto L_\infty(t) = \mathbb{E}(e^{-tY_\infty})$ de Y_∞ est caractérisée par les propriétés suivantes :*

$$L'_\infty(0) = -1 \quad \text{et} \quad L_\infty(mt) = g(L_\infty(t)) \quad \text{pour tout } t \in \mathbb{R}_+.$$

Démonstration. La martingale $(Y_n)_{n \geq 0}$ est bornée dans L^2 car par le théorème 4.1.1 :

$$\mathbb{E}(Y_n^2) = \frac{\text{Var}(Z_n) + \mathbb{E}(Z_n)^2}{m^{2n}} = \frac{\sigma^2}{m^2 - m} - \frac{\sigma^2}{m^n(m^2 - m)} + 1$$

qui converge car $m > 1$. Par conséquent, $(Y_n)_{n \geq 0}$ converge p.s. et dans L^2 vers une v.a.r. Y_∞ dont les deux premiers moments sont la limite de ceux de Y_n . Il est possible de procéder directement sans faire appel à un théorème de martingales, grâce au fait que la convergence L^2 est ici assez rapide. En effet, par le théorème 4.1.1 on obtient, pour tout $n, k \geq 0$,

$$\mathbb{E}((Y_{n+k} - Y_n)^2) = \frac{\sigma^2}{m^n} \frac{1 - m^{-k}}{m^2 - m}.$$

Comme $m > 1$, ceci montre que $(Y_n)_{n \geq 0}$ est une suite de Cauchy dans L^2 . Comme L^2 est complet, elle converge vers une v.a.r. $Y_\infty \in L^2$. La série $\sum_{n=0}^{\infty} \mathbb{E}((Y_n - Y_\infty)^2)$ converge également grâce à la borne géométrique en m^{-n} sur $\mathbb{E}((Y_\infty - Y_n)^2)$ obtenue en faisant $k \rightarrow \infty$. Par convergence monotone, on obtient $\mathbb{E}(\sum_{n=0}^{\infty} (Y_n - Y_\infty)^2) < \infty$ et donc $(Y_n)_{n \geq 0}$ converge p.s. vers Y_∞ . Les deux premiers moments de Y_∞ s'obtiennent facilement.

Pour la transformée de Laplace, on a $L'_\infty(0) = -\mathbb{E}(Y_\infty) = -1$. De plus, la transformée de Laplace $t \in \mathbb{R}_+ \mapsto L_n(t) = \mathbb{E}(e^{-tY_n})$ de Y_n vérifie pour tout $n \geq 0$

$$L_{n+1}(mt) = g_{n+1}(e^{-\frac{mt}{m^{n+1}}}) = g(g_n(e^{-\frac{t}{m^n}})) = g(L_n(t)),$$

d'où $L_\infty(mt) = g(L_\infty(t))$ quand $n \rightarrow \infty$ car $L_\infty = \lim_{n \rightarrow \infty} L_n$ et g est continue. \square

Notons que si $s = \mathbb{P}(T < \infty)$ alors $(M_n)_{n \geq 0} = (s^{Z_n})_{n \geq 0}$ est une martingale positive et bornée, car $\mathbb{E}(M_{n+1} | \sigma(M_0, \dots, M_n)) = \mathbb{E}(s^{Z_{n+1}} | Z_n) = g(s)^{Z_n} = s^{Z_n}$, car $g(s) = s$.

4.3 Taille de l'arbre en régimes critique et sous-critique

Du théorème 4.1.1 on déduit que la population totale à l'instant n a pour moyenne

$$\mathbb{E}(Z_0 + \dots + Z_n) = \frac{1 - m^{n+1}}{1 - m} \mathbb{1}_{m \neq 1} + (n + 1) \mathbb{1}_{m=1}.$$

On peut obtenir une formule similaire pour la variance (exercice !). On se place dans cette section en régime critique ou sous-critique $m \leq 1$. Nous savons que la population s'éteint presque sûrement, et la population totale (ou taille de l'arbre) est donc donnée par

$$N = \lim_{n \rightarrow \infty} Z_0 + \dots + Z_n \quad (= Z_0 + \dots + Z_{T-1}),$$

d'où, par convergence monotone,

$$\mathbb{E}(N) = \begin{cases} \frac{1}{1 - m} & \text{si } m < 1, \\ \infty & \text{si } m = 1. \end{cases}$$

Dans le cas critique, la population totale N est finie p.s. mais n'est pas intégrable.

Nous allons caractériser la loi de N en fonction de la loi de reproduction P . Tout d'abord, nous pouvons exploiter la propriété de branchement : comme l'individu racine donne naissance à $X_{1,1}$ arbres de Galton-Watson i.i.d., la loi de N vérifie

$$N \stackrel{\text{loi}}{=} 1 + \sum_{i=1}^{X_{1,1}} N^{(i)}$$

où les v.a. $(N^{(i)})_{i \geq 1}$ sont i.i.d. de même loi que N . Cette égalité en loi permet de retrouver la formule $\mathbb{E}(N) = (1 - m)^{-1} \mathbb{1}_{m < 1} + \infty \mathbb{1}_{m=1}$, et fournit plus généralement le lemme suivant.

Lemme 4.3.1 (Fonction génératrice de N). *Si g et f sont les fonctions génératrices respectives de $X_{1,1}$ et N , alors la fonction f est solution de l'équation*

$$f(s) = s(g \circ f)(s) \quad \text{pour } s \in [0, 1].$$

Exemple 4.3.2 (Reproduction géométrique). *Si $P = \text{Geo}_{\mathbb{N}}(p) = \sum_{n \geq 0} q^n p \delta_n$ avec $p \geq 1/2$, alors $f(s)$ est solution de l'équation $qf(s)^2 - f(s) + sp = 0$. On a donc*

$$f(s) = \frac{1 - \sqrt{1 - 4pqs}}{2q} \quad \text{pour } s \in [0, 1].$$

Cette équation sur la fonction génératrice n'est pas toujours facile à utiliser. On peut aussi exprimer la loi de N à partir des convolutions successives de P

Théorème 4.3.3 (Loi de la taille de l'arbre). *Pour tout $n \geq 1$,*

$$\mathbb{P}(N = n) = \frac{\mathbf{P}(n, n-1)}{n} = \frac{P^{*n}(\{n-1\})}{n}.$$

Démonstration. Bien qu'il soit possible d'obtenir cette relation par un calcul direct utilisant les propriétés de l'arbre de Galton-Watson, nous présentons une méthode féconde exploitant une analogie entre arbres et marches aléatoires. L'idée consiste à associer de manière bijective un arbre fini à une portion de trajectoire de marche aléatoire. Étant donné un arbre fini de taille n , on numérote ses sommets en partant de la racine (notée 1) et de gauche à droite pour des individus d'une même génération. On note ensuite $X^{(i)}$ le nombre d'enfants de l'individu i et $U_i = X^{(i)} - 1$ pour $1 \leq i \leq n$. On pose enfin

$$S_0 = 0 \quad \text{et} \quad S_{i+1} = S_i + U_{i+1} \quad \text{pour} \quad 0 \leq i \leq n - 1.$$

La figure 4.2 donne un exemple d'association. Le mode de numérotation étant fixé, l'arbre se reconstruit aisément à partir de la trajectoire de S .

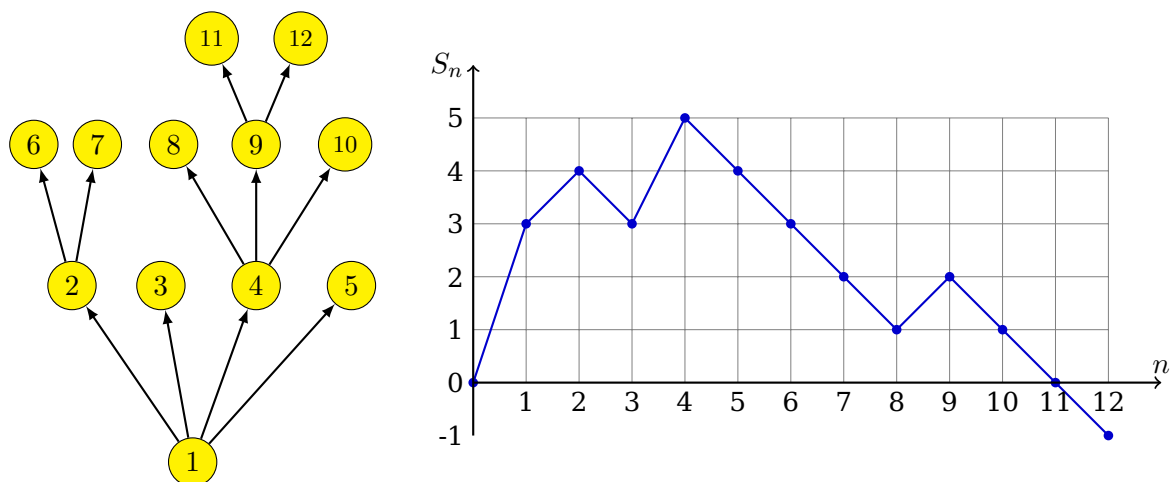


Figure 4.2 – Arbre et marche aléatoire associée.

La suite $(S_i)_{0 \leq i \leq n}$ a la loi d'une marche aléatoire d'accroissement $(U_i)_{i \geq 1}$ issue de 0 arrêtée au temps d'atteinte T_{-1} de -1 . Comme les accroissements sont de moyenne négative $m - 1$ et que les sauts négatifs de la marche sont d'amplitude 1, ce temps est fini presque sûrement. Cette bijection fournit le résultat suivant.

$$N \stackrel{\text{loi}}{=} T_{-1}.$$

Enfin, la loi de T_{-1} est fournie par le théorème 4.3.4. □

Théorème 4.3.4 (Principe de rotation). *Soit $(S_n)_{n \geq 0}$ une marche aléatoire issue de 0 d'incrémentations $(U_n)_{n \geq 1}$ i.i.d. à valeurs entières, et $T_{-1} := \inf\{n \geq 1 : S_n = -1\}$. Pour tout $n \geq 1$,*

$$\mathbb{P}(T_{-1} = n) = \frac{1}{n} \mathbb{P}(S_n = -1).$$

Démonstration. Soit $n \geq 1$ et U_1, \dots, U_n tels que $S_n = -1$. Pour tout entier k , notons $\sigma(k)$ l'entier compris entre 1 et n égal à k modulo n . Pour tout $l = 1, 2, \dots, n$, on définit $S^{(l)}$ par

$$S_0^{(l)} = 0 \quad \text{et} \quad S_{k+1}^{(l)} = S_k^{(l)} + U_{\sigma(k+l)}.$$

La figure 4.3 fournit un exemple. Remarquons que pour tout l , $S_n^{(l)} = S_n = -1$. Cependant, il existe une et une seule trajectoire $(S_i^{(l)})_{0 \leq i \leq n}$ parmi les n possibles qui reste positive ou nulle jusqu'à l'instant $n - 1$. Il s'agit de celle qui correspond à l'indice k où S atteint son minimum global pour la première fois. On a donc bien le résultat attendu. □

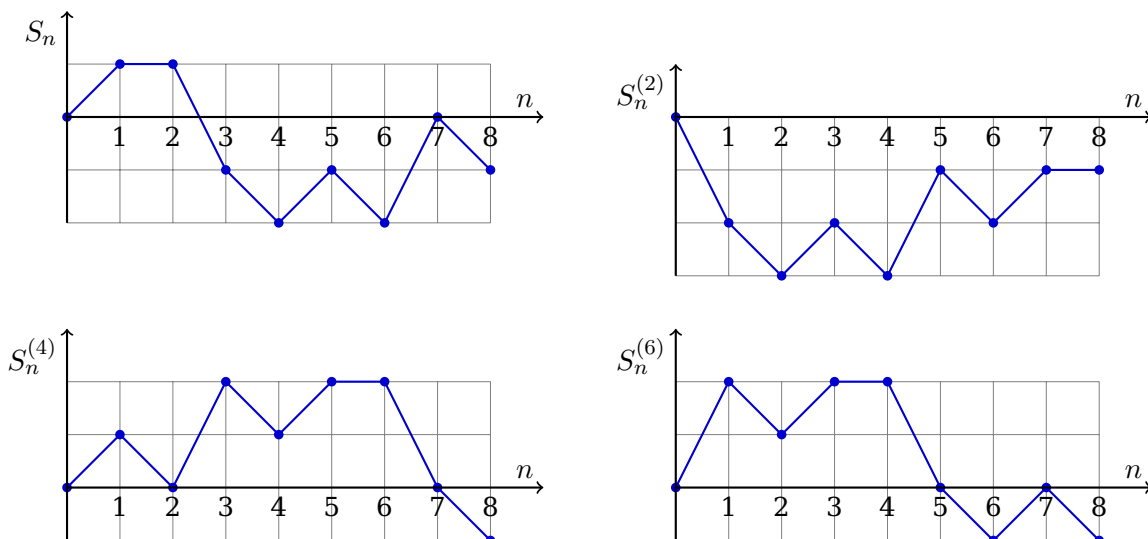


Figure 4.3 – Une marche parmi n reste positive avant d’atteindre -1.

Exemple 4.3.5 (Reproduction poissonnienne). Si la loi de reproduction P est la loi de Poisson Poi(λ) avec λ < 1, alors la loi de U_1 est la loi Poi(λ) * δ_{-1} et, pour tout n ≥ 1,

$$\mathbb{P}(N = n) = e^{-\lambda n} \frac{(\lambda n)^{n-1}}{n!}$$

puisque S_n suit la loi Poi(nλ) * δ_{-n}. D’autre part, la série génératrice f de N est solution de l’équation fonctionnelle f(s) = se^{λ(f(s)-1)} qui n’est pas facile à résoudre...

4.4 Immigration

Le processus de Galton-Watson avec immigration (Z_n)_{n ≥ 0} issu de Z_0 est défini par

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

pour tout n ≥ 0, où (X_{n,k})_{n ≥ 1, k ≥ 1} sont i.i.d. de loi P sur ℕ, (I_n)_{n ≥ 1} i.i.d. de loi P_+ sur ℕ, toutes ces variables formant avec Z_0 une famille indépendante. On suppose que P et P_+ possèdent une moyenne m et m_+ et une variance σ^2 < ∞ et σ_+^2 < ∞, et que Z_0 = 1. Soit F_n la tribu engendrée par Z_0, (X_{i,j})_{1 ≤ i ≤ n, j ≥ 1}, I_1, ..., I_n. Pour tout n ≥ 0,

$$\mathbb{E}(Z_{n+1} | \mathcal{F}_n) = mZ_n + m_+$$

d’où on tire (récurrence linéaire) que pour tout n ≥ 1,

$$\mathbb{E}(Z_n) = \left(m^n + \frac{m^n - 1}{m - 1} m_+ \right) \mathbf{1}_{m \neq 1} + (1 + nm_+) \mathbf{1}_{m=1}.$$

Pour la variance, on a

$$\mathbb{E}(Z_{n+1}^2 | \mathcal{F}_n) = Z_n(\sigma^2 + m^2) + Z_n(Z_n - 1)m^2 + 2Z_nmm_+ + \sigma_+^2 + m_+^2$$

ce qui donne

$$\text{Var}(Z_{n+1}) = \text{Var}(Z_n)m^2 + \sigma^2\mathbb{E}(Z_n) + \sigma_+^2.$$

Théorème 4.4.1 (Cas sur-critique). Si $m > 1$ alors $(Y_n)_{n \geq 0} = (Z_n/m^n)_{n \geq 0}$ converge p.s. et dans L^2 vers une v.a.r. $Y_\infty \geq 0$ qui vérifie² $\mathbb{E}(Y_\infty) = 1 + m_+/(m - 1)$.

Démonstration. La suite $(Y_n)_{n \geq 0}$ est une $(\mathcal{F}_n)_{n \geq 0}$ -sous-martingale car

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

On a $\mathbb{E}(Y_0) = 1$ et $\mathbb{E}(Y_n) = \mathbb{E}(Y_{n-1}) + m^{-n}m_+$ pour $n \geq 1$ d'où

$$\mathbb{E}(Y_n) = 1 + (m^{-1} + \dots + m^{-n})m_+.$$

Comme $m > 1$ et $Y_n \geq 0$ on obtient $\sup_n \mathbb{E}(|Y_n|) = \sup_n \mathbb{E}(Y_n) < \infty$, et un théorème de Doob sur les sous-martingales bornées dans L^1 entraîne que $(Y_n)_{n \geq 0}$ converge p.s. vers une v.a.r. $Y_\infty \geq 0$ intégrable (la convergence a lieu dans L^1 si la suite est uniformément intégrable). Nous devons établir que la convergence a lieu dans L^2 . Comme $m > 1$, on a

$$\mathbb{E}(Y_n^2) = m^{-2n}\text{Var}(Z_n) + m^{-2n}\mathbb{E}(Z_n)^2$$

ce qui donne (après calculs) $\ell := \lim_{n \rightarrow \infty} \mathbb{E}(Y_n^2) < \infty$ car $m > 1$, et pour tous $n, k \geq 1$,

$$\begin{aligned} \mathbb{E}((Y_{n+k} - Y_n)^2) &= \mathbb{E}(Y_{n+k}^2) + \mathbb{E}(Y_n^2) - 2\mathbb{E}(Y_{n+k}Y_n) \\ &= \mathbb{E}(Y_{n+k}^2) + \mathbb{E}(Y_n^2) - 2\mathbb{E}(\mathbb{E}(Y_{n+k} | \mathcal{F}_n)Y_n) \\ &= \mathbb{E}(Y_{n+k}^2) + \mathbb{E}(Y_n^2) - 2\mathbb{E}((Y_n + (m^{-n-k} + \dots + m^{-n-1})m_+)Y_n) \\ &= \mathbb{E}(Y_{n+k}^2) - \mathbb{E}(Y_n^2) - 2(m^{-n-k} + \dots + m^{-n-1})m_+\mathbb{E}(Y_n) \\ &= \ell + o_{n \rightarrow \infty}(1) - \ell - o_{n \rightarrow \infty}(1) + o_{n \rightarrow \infty}(1). \end{aligned}$$

Par conséquent, $(Y_n)_{n \geq 0}$ est de Cauchy dans L^2 et converge donc dans L^2 . □

4.5 Notes et commentaires

C'est vers 1875 que Francis Galton et Henry William Watson écrivent leur article sur l'évolution du nombre de noms de familles aristocratiques anglaises. Ils ne connaissaient sans doute pas les travaux antérieurs de Irénée-Jules Bienaymé sur le même sujet.

L'étude du processus de Galton-Watson peut être considérablement raffinée. Le modèle lui-même peut être enrichi et modifié afin de tenir compte de diverses situations d'intérêt : existence de sexes différents, survivance des individus à plusieurs générations, etc. On trouvera de nombreux développements dans les livres [DJ06], [AN04], [HJV07], [Har02].

Une étude du processus avec émigration se trouve dans [Rug01].

Le théorème 4.2.1 de Yaglom constitue un bon point de départ pour l'étude plus générale des distributions quasi-stationnaires des processus de populations, présentée dans [MV12].

Le théorème 4.3.4 est obtenu dans [DM47]. L'expression pour la taille d'un arbre de Galton-Watson est établie dans [Dwa69] avant que le lien entre les deux situations soit fait. On pourra consulter l'habilitation à diriger des recherches [Mar04] de Jean-François Marckert ainsi que l'article de synthèse autour du théorème du ballottage [ABR08].

2. Il est également possible de calculer $\text{Var}(Y_\infty)$ (exercice !).

Séance 5

Marches aléatoires

Mots-clés. Marches aléatoires; chaînes de Markov; martingales; temps de sortie; loi multinomiale; groupe symétrique; collectionneur de coupons.

5.1 Marche aléatoire simple sur \mathbb{Z}

Il s'agit de la suite $(X_n)_{n \geq 0}$ définie par la relation récurrente (ou auto-régressive)

$$X_{n+1} = X_n + \varepsilon_{n+1} = X_0 + \varepsilon_1 + \cdots + \varepsilon_{n+1}$$

pour tout $n \geq 0$, où $(\varepsilon_n)_{n \geq 1}$ est une suite de v.a.r. i.i.d. de loi de Rademacher, indépendante de X_0 . On pose $p = \mathbb{P}(\varepsilon_n = 1) = 1 - \mathbb{P}(\varepsilon_n = -1) \in]0, 1[$. La suite $(X_n)_{n \geq 0}$ est une chaîne de Markov d'espace d'état \mathbb{N} et de noyau de transition $\mathbf{P}(x, \cdot) = p\delta_{x+1} + (1-p)\delta_{x-1}$ (la matrice \mathbf{P} est tridiagonale). Une variable aléatoire B suit la loi de Bernoulli $(1-p)\delta_0 + p\delta_1$ si et seulement si $2B - 1$ suit la loi de Rademacher $(1-p)\delta_{-1} + p\delta_1$. Ainsi, pour tout $n \geq 0$,

$$\frac{X_n - X_0 + n}{2} \sim \text{Bin}(n, p).$$

Théorème 5.1.1 (Récurrence). *La marche aléatoire simple sur \mathbb{Z} est récurrente nulle si $p = 1/2$ (m.a.s. symétrique) et transitoire si $p \neq 1/2$ (m.a.s. asymétrique).*

Démonstration. On rappelle que x est récurrent si et seulement si $\sum_n \mathbf{P}^n(x, x) = \infty$. La chaîne est de période 2 et donc $\mathbf{P}^{2n+1}(x, x) = 0$ pour tout $n \geq 0$. Comme la chaîne est irréductible, tous les états ont même nature, et on peut donc se ramener à l'état 0. La formule binomiale donne $\mathbf{P}^{2n}(0, 0) = \mathbb{P}(X_{2n} = 0 | X_0 = 0) = \binom{2n}{n} p^n (1-p)^n$. La formule de Stirling $n! \sim \sqrt{2\pi n} (n/e)^n$ donne, en notant $\rho = 4p(1-p)$,

$$\mathbf{P}^{2n}(0, 0) \sim \frac{\rho^n}{\sqrt{\pi n}}.$$

À présent, si $p = 1/2$ alors $\rho = 1$ et la chaîne est récurrente tandis que si $p \neq 1/2$ alors $\rho < 1$ et la chaîne est transitoire. Lorsque $p = 1/2$ la mesure de comptage est symétrique donc invariante, et comme elle n'est pas finie, la chaîne est récurrente nulle.

On peut alternativement traiter le cas $p \neq 1/2$ avec un peu plus d'intuition probabiliste. En effet, la loi forte des grands nombres affirme que p.s. $X_n = X_0 + n(2p - 1 + o_{n \rightarrow \infty}(1))$, donc $X_n \rightarrow +\infty$ p.s. si $p > 1/2$ et $X_n \rightarrow -\infty$ p.s. si $p < 1/2$, ce qui interdit toute récurrence. \square

Le théorème suivant permet d'étudier le problème de la ruine d'un joueur qui gagne 1 Euro avec probabilité p et perd 1 Euro avec probabilité $1 - p$. La fortune initiale est x et le joueur quitte le jeu lorsqu'il possède $a < x$ Euros (ruine) ou $b > x$ Euros (gain).

Théorème 5.1.2 (Sortie de boîte¹). Soient $a, b \in \mathbb{Z}$ avec $a < b$ et les temps d'arrêt

$$\tau_a = \inf\{n \geq 0 : X_n = a\}, \quad \tau_b = \inf\{n \geq 0 : X_n = b\}, \quad \text{et} \quad \tau = \min(\tau_a, \tau_b).$$

Alors pour tout $a \leq x \leq b$ on a $\mathbb{E}_x(\tau) < \infty$ et en posant $\rho = \frac{1-p}{p}$ on a

$$\mathbb{P}_x(X_\tau = a) = \begin{cases} \frac{\rho^b - \rho^x}{\rho^b - \rho^a} & \text{si } p \neq \frac{1}{2}, \\ \frac{b-x}{b-a} & \text{si } p = \frac{1}{2}. \end{cases} \quad \text{et} \quad \mathbb{E}_x(\tau) = \begin{cases} \frac{x-a}{1-2p} - \frac{(b-a)\rho^x - \rho^a}{1-2p} & \text{si } p \neq \frac{1}{2}, \\ (b-x)(x-a) & \text{si } p = \frac{1}{2}. \end{cases}$$

Si $p = 1/2$ alors la chaîne est récurrente et visite presque sûrement chaque état une infinité de fois et donc $\mathbb{P}_x(\tau_a < \infty) = 1$ et $\mathbb{P}_x(\tau_b < \infty) = 1$ pour tout $a \leq x \leq b$. En revanche, si $p \neq 1/2$ alors la chaîne est transitoire et les temps d'atteinte de a ou de b ne sont plus finis p.s. (dépend de p et du point de départ x). On le voit bien dans les formules du théorème 5.1.2 en faisant tendre a ou b vers l'infini. Le temps de sortie τ de $[a, b]$ est identique en loi au temps d'absorption T par $\{a, b\}$ de la chaîne $Y = (Y_n)_{n \geq 0}$ d'espace d'état fini $\{a, \dots, b\}$ de mêmes transitions que $(X_n)_{n \geq 0}$ mais avec absorption en a et b . Comme pour Y , les états a et b sont récurrents et tous les autres (en nombre fini) transitoires, et comme p.s. la chaîne Y ne visite qu'un nombre fini de fois chaque état transitoire, on en déduit que $\mathbb{P}_x(\tau < \infty) = \mathbb{P}_x(T < \infty) = 1$ pour tout $a \leq x \leq b$.

Démonstration. Montrons que $\mathbb{E}_x(\tau) < \infty$ pour tout $a \leq x \leq b$. Il suffit d'obtenir une majoration géométrique pour la queue de la loi de τ (il serait possible de procéder par couplage, comme pour le modèle de Fisher-Wright, cf. remarque 2.1.1). Pour tout $a \leq x \leq b$, il existe un chemin ℓ_x de longueur $|\ell_x| \leq (b-a)$ qui mène de x à a ou b . On a

$$\mathbb{P}_x(\tau > (b-a)) \leq \mathbb{P}(X_{1:|\ell_x|} \neq \ell_x) = 1 - \mathbb{P}(X_{1:|\ell_x|} = \ell_x) \leq 1 - \min(p, 1-p)^{|\ell_x|}.$$

Si $\eta = \max_{a < x < b} (1 - \min(p, 1-p)^{|\ell_x|}) < 1$ alors on obtient, par récurrence, pour tout $k \geq 1$, en utilisant l'indépendance conditionnelle du passé et du futur sachant le présent,

$$\begin{aligned} \mathbb{P}_x(\tau > k(b-a)) &= \sum_{a < y < b} \mathbb{P}_x(\tau > k(b-a), X_{(k-1)(b-a)} = y, \tau > (k-1)(b-a)) \\ &= \sum_{a < y < b} \mathbb{P}_y(\tau > (b-a)) \mathbb{P}_x(X_{(k-1)(b-a)} = y, \tau > (k-1)(b-a)) \\ &\leq \eta \mathbb{P}_x(\tau > (k-1)(b-a)) \leq \eta \eta^{k-1} = \eta^k. \end{aligned}$$

Comme $\eta < 1$ on obtient que $\mathbb{E}_x(\tau) < \infty$ (tous les moments sont finis en fait) et en particulier $\mathbb{P}_x(\tau < \infty) = 1$. Calculons $r(x) = \mathbb{P}_x(X_\tau = a)$. On a pour tout $a < x < b$

$$\begin{aligned} r(x) &= \mathbb{P}_x(X_\tau = a | X_1 = x+1)p + \mathbb{P}_x(X_\tau = a | X_1 = x-1)(1-p) \\ &= pr(x+1) + (1-p)r(x-1). \end{aligned}$$

L'ensemble des solutions de cette récurrence linéaire d'ordre deux est un espace vectoriel qui contient la solution constante 1. Si $p \neq 1/2$ alors ρ^x est aussi solution, linéairement

1. Quoi de plus naturel pour un ivrogne (modélisé traditionnellement par la marche aléatoire).

indépendante de 1, et donc les solutions sont de la forme $A + B\rho^x$ avec A et B constantes. Les conditions aux bords $r(a) = 1$, $r(b) = 0$ fixent A et B , ce qui donne l'unique solution

$$r(x) = \frac{\rho^b - \rho^x}{\rho^b - \rho^a}.$$

Si $p = 1/2$ alors $\rho = 1$ et les deux solutions fondamentales précédentes sont confondues. Cependant, on observe que dans ce cas, x est également solution, linéairement indépendante de 1, et donc les solutions sont de la forme $A + Bx$ où A et B sont des constantes. Les conditions aux bords $r(a) = 1$ et $r(b) = 0$ fixent A et B , ce qui donne l'unique solution

$$r(x) = \frac{b - x}{b - a}.$$

Calculons $R(x) = \mathbb{E}_x(\tau)$. En conditionnant selon X_1 on obtient pour tout $a < x < b$ la récurrence linéaire (la méthode est valable pour toute chaîne de Markov, idem pour $r(x)$)

$$R(x) = pR(x + 1) + (1 - p)R(x - 1) + 1.$$

La présence du second membre 1 fait rechercher des solutions particulières. Si $p \neq 1/2$ alors $x/(1 - 2p)$ est solution particulière, et les solutions de l'équation sont de la forme $R(x) = x/(1 - 2p) + A + B\rho^x$. Les conditions aux bords $R(a) = 0$ et $R(b) = 0$ donnent enfin

$$R(x) = \frac{x - a}{1 - 2p} - \frac{(b - a)\rho^b - \rho^x}{1 - 2p(\rho^b - \rho^a)}.$$

Si $p = 1/2$ alors $-x^2$ est solution particulière, et les solutions sont de la forme $-x^2 + A + Bx$. Les conditions aux bords $R(a) = R(b) = 0$ donnent enfin

$$R(x) = (b - x)(x - a).$$

La même approche permet de calculer la fonction génératrice $\mathbb{E}_x(s^\tau | X_\tau = a)$. □

Remarque 5.1.3 (Les théorèmes limites à la rescousse). *Voici un autre argument pour établir que $\mathbb{P}_x(\tau < \infty) = 1$. Posons $m = 2p - 1$ et $\sigma^2 = 4p(1 - p)$. Si $m \neq 0$ alors par la loi forte des grands nombres, p.s. $(X_n)_{n \geq 1}$ tend vers $+\infty$ si $m > 0$ et vers $-\infty$ si $m < 0$, et donc $\mathbb{P}_x(\tau < \infty) = 1$. Si $m = 0$ alors pour tout $n \geq 1$, en posant $I_n = \frac{1}{\sqrt{n}}]a, b[$, on a*

$$\mathbb{P}_x(\tau = \infty) \leq \mathbb{P}(a < X_n < b) = \mathbb{P}\left(\frac{X_n}{\sqrt{n}} \in I_n\right).$$

Or $(n^{-1/2}X_n)_{n \geq 1}$ converge en loi vers $\mathcal{N}(0, \sigma^2)$ par le théorème central limite. Mais I_n dépend de n . Cependant, comme $(I_n)_{n \geq 1}$ est décroissante,

$$\limsup_{n \rightarrow \infty} \mathbb{P}\left(\frac{X_n}{\sqrt{n}} \in I_n\right) \leq \inf_{m \geq 1} \frac{1}{\sqrt{2\pi\sigma^2}} \int_{I_m} e^{-\frac{t^2}{2\sigma^2}} dt = 0.$$

Remarque 5.1.4 (Martingales). *Si $(Z_n)_{n \geq 0}$ est une chaîne de Markov sur E fini de noyau $\mathbf{P} = \mathbf{L} + \mathbf{I}$ alors $(M_n)_{n \geq 0}$ où $M_n = f(Z_n) - f(Z_0) - \sum_{k=0}^{n-1} (\mathbf{L}f)(Z_k)$ est une martingale pour toute fonction $f : E \rightarrow \mathbb{R}$. Il s'agit d'une formule d'Itô discrète. Le vecteur des probabilités d'atteinte d'un ensemble clos est harmonique : $\mathbf{L}f = 0$ et $(f(Z_n) - f(Z_0))_{n \geq 0}$ est alors une martingale. Il est possible de retrouver les formules du théorème 5.1.2 en utilisant des martingales bien choisies et le théorème d'arrêt de Doob-Wald. Par exemple, la martingale $(X_n - nm)_{n \geq 0}$ donne $\mathbb{E}_0(X_\tau) = m\mathbb{E}_0(\tau)$ tandis que la martingale exponentielle $(\rho^{X_n})_{n \geq 0}$ donne $\mathbb{E}_0(\rho^{X_\tau}) = 1$. Cette méthode à base de martingales permet également d'obtenir les formules pour le temps de sortie pour un processus de diffusion sur \mathbb{R} , et dans cas, les martingales s'obtiennent par exemple via la formule d'Itô pour $f(X_t)$ avec f bien choisie.*

Théorème 5.1.5 (Nombres de Catalan). Si $\tau = \inf\{n \geq 1 : X_n = 0\}$ alors pour tout $n \geq 0$,

$$\mathbb{P}_0(\tau = 2n + 2) = \frac{2}{n+1} \binom{2n}{n} p^{n+1} (1-p)^{n+1}.$$

On reconnaît le n^e nombre de Catalan $\frac{1}{n+1} \binom{2n}{n}$. Ces nombres comptent, outre les chemins de la marche aléatoire simple, les mots de Dyck, les parenthésages, les triangulations d'un polygone, les partitions non croisées, les chemins sous-diagonaux dans le carré, les arbres planaires, etc. Les moments pairs de la loi du demi-cercle sont les nombres de Catalan.

Preuve du théorème 5.1.5. Sachant $\{X_0 = 0\}$, l'événement $\{\tau = 2n + 2\}$ correspond à une trajectoire de longueur $2n + 2$ partant de 0 et revenant à zéro en restant strictement positive ou strictement négative. Ces deux cas sont équiprobables, d'où le facteur 2 dans le résultat. Dans les deux cas, il y a eu forcément $n + 1$ incréments $+1$ et $n + 1$ incréments -1 , d'où

$$\mathbb{P}_0(\tau = 2n + 2) = 2C_n p^{n+1} (1-p)^{n+1}.$$

où C_n est le nombre de chemins de longueur $2n + 2$ partant de zéro et revenant à zéro, et restant strictement positifs. Le premier incrément est forcément $+1$ et le dernier forcément -1 et C_n est égal au nombre de chemins de longueur $2n$ partant de zéro et revenant à zéro et restant positifs. Il y a n incréments $+1$ et n incréments -1 . Considérons les chemins partant de zéro et revenant à zéro et contenant n incréments $+1$ et n incréments -1 . Il y en a $\binom{2n}{n}$. Si un chemin de ce type n'est pas positif alors juste après la première position négative, modifions tous les incréments en permutant le signe des $+1$ et des -1 . On obtient de la sorte un chemin avec $n - 1$ incréments $+1$ et $n + 1$ incréments -1 , et il s'avère que tous les chemins partant de zéro avec $n - 1$ incréments $+1$ et $n + 1$ incréments -1 s'obtiennent de la sorte, et il y en a $\binom{2n}{n-1}$. Ainsi, $C_n = \binom{2n}{n} - \binom{2n}{n-1} = \frac{1}{n+1} \binom{2n}{n}$ (formule de Désiré André). \square

5.2 Marche aléatoire simple symétrique sur \mathbb{Z}^d

Pour tout $d \geq 1$, la marche aléatoire simple symétrique sur \mathbb{Z}^d est définie par

$$X_{n+1} = X_n + \varepsilon_{n+1} = X_0 + \varepsilon_1 + \dots + \varepsilon_{n+1}$$

où $(\varepsilon_n)_{n \geq 1}$ est une suite de v.a.r. i.i.d. de loi uniforme sur $\{\pm e_1, \dots, \pm e_d\}$, indépendante de X_0 . La suite $(X_n)_{n \geq 0}$ est une chaîne de Markov d'espace d'état \mathbb{Z}^d et de noyau de transition $\mathbf{P}(x, y) = \frac{1}{2d} \mathbb{1}_{|x-y|_1=1}$. On parle également de marche aléatoire symétrique aux plus proches voisins, au sens de la distance $|\cdot|_1$. En concevant les incréments $\varepsilon_1, \dots, \varepsilon_n$ comme issus de n jets d'un dé équilibré à $2d$ faces, on voit que $\mathcal{L}(X_n - X_0)$ est l'image de

$$\left(\frac{1}{2d} \delta_{e_1} + \frac{1}{2d} \delta_{-e_1} + \dots + \frac{1}{2d} \delta_{-e_d} + \frac{1}{2d} \delta_{e_d} \right)^{*n} = \text{Mul} \left(n, \left(\frac{1}{2d}, \dots, \frac{1}{2d} \right) \right)$$

par l'application

$$(k_1, \dots, k_{2d}) \in \mathbb{N}^{2d} \mapsto (k_1 - k_2, \dots, k_{2d-1} - k_{2d}) \in \mathbb{Z}^d.$$

Notons au passage que le temps (aléatoire) au bout duquel la marche aléatoire a emprunté les $2d$ directions qui s'offrent à elle suit la loi du collectionneur de coupon étudiée dans le chapitre 1.

Théorème 5.2.1 (Récurrence de la marche aléatoire simple symétrique). La marche aléatoire simple symétrique sur \mathbb{Z}^d est récurrente nulle si $d \leq 2$ et transitoire si $d \geq 3$.

Démonstration. La chaîne est irréductible et de période 2 pour tout $d \geq 1$ et il suffit donc d'étudier $\sum_n \mathbf{P}^{2n}(0, 0)$. La formule multinomiale donne

$$\mathbf{P}^{2n}(0, 0) = \frac{1}{(2d)^{2n}} \sum_{\substack{r_1 + \dots + r_{2d} = n \\ r_1 = r_2, \dots, r_{d-1} = r_d}} \binom{2n}{r_1, \dots, r_{2d}} = \frac{1}{(2d)^{2n}} \sum_{r_1 + \dots + r_d = n} \frac{(2n)!}{r_1!^2 \dots r_d!^2}.$$

Le cas $d = 1$ est déjà traité. Restent les cas $d = 2$ et $d \geq 3$.

Cas $d = 2$ (méthode directe). La chaîne est récurrente car, par la formule Vandermonde,

$$\mathbf{P}^{2n}(0, 0) = \frac{1}{4^{2n}} \sum_{n_1 + n_2 = n} \frac{(2n)!}{n_1!^2 n_2!^2} = \frac{(2n)!}{4^{2n} n!^2} \sum_{k=0}^n \binom{n}{k}^2 = \frac{(2n)!}{4^{2n} n!^2} \frac{(2n)!}{n!^2} \sim_{n \rightarrow \infty} \frac{1}{\pi n}$$

qui n'est pas sommable. La récurrence nulle vient du fait que la mesure de comptage, qui n'est pas finie car \mathbb{Z}^2 est infini, est une mesure symétrique et donc invariante.

Cas $d = 2$ (méthode astucieuse). Soient U_n et V_n les projections de X_n sur les premières et seconde bissectrices (pentes ± 1). On vérifie que $(U_n)_{n \geq 0}$ et $(V_n)_{n \geq 0}$ sont des marches aléatoires simples symétriques sur $\frac{1}{\sqrt{2}}\mathbb{Z}$, indépendantes ! De plus, $X_n = 0$ si et seulement si $U_n = V_n = 0$, ce qui donne $\mathbf{P}^{2n}(0, 0) \sim \frac{1}{\pi n}$, et la chaîne est donc récurrente. Attention : les composantes de $(X_n)_{n \geq 0}$ (abscisse et ordonnée) ne sont pas des marches aléatoires simples sur \mathbb{Z} , mais constituent des chaînes de Markov de noyau $\mathbf{Q}(x, y) = \frac{1}{4} \mathbb{1}_{|x-y|=1} + \frac{1}{2} \mathbb{1}_{x=y}$ (il est possible d'établir qu'elles sont récurrentes, mais cela ne conduit à rien de bien utile).

Cas $d = 3$ (méthode directe). La formule de la multinomiale donne ici

$$\mathbf{P}^{2n}(0, 0) = \frac{1}{6^{2n}} \sum_{r_1 + r_2 + r_3 = n} \frac{(2n)!}{r_1!^2 r_2!^2 r_3!^2} = \frac{(2n)!}{2^{2n} 3^n n!^2} \sum_{r_1 + r_2 + r_3 = n} \binom{n}{r_1 \ r_2 \ r_3}^2 \left(\frac{1}{3}\right)^n$$

Si $n = 3m$ alors une petite étude montre que $\binom{n}{r_1 \ r_2 \ r_3} \leq \binom{n}{m \ m \ m}$ et donc, grâce à la formule du trinôme de taille n et de paramètre $(1/3, 1/3, 1/3)$, on obtient, pour $n = 3m$,

$$\mathbf{P}^{2n}(0, 0) \leq \frac{(2n)!}{2^{2n} 3^n n!^2} \binom{n}{m \ m \ m} \sim \frac{1}{2} \left(\frac{3}{\pi n}\right)^{3/2} = \frac{c}{n^{3/2}}.$$

Ainsi, $\sum_m \mathbf{P}^{6m}(0, 0) < \infty$. Comme $(1/6)^{2k} \mathbf{P}^{6m}(0, 0) \leq \mathbf{P}^{6m+2k}(0, 0)$ pour $k = 1, 2$, il vient $\sum_n \mathbf{P}^{2n}(0, 0) = \sum_{k=0,1,2} \sum_m \mathbf{P}^{6m+2k}(0, 0) < \infty$ et donc la chaîne est transitoire.

Cas $d \geq 3$ (méthode directe). La formule de la multinomiale donne

$$\mathbf{P}^{2n}(0, 0) = \frac{1}{(2d)^{2n}} \sum_{n_1 + \dots + n_d = n} \frac{(2n)!}{n_1!^2 \dots n_d!^2} = \frac{(2n)!}{n!^2 (4d)^n} \sum_{n_1 + \dots + n_d = n} \left(\frac{n!}{n_1!^2 \dots n_d!^2}\right)^2 \left(\frac{1}{d}\right)^n.$$

Si $m_n = m_{n,d}$ désigne la valeur maximale du coefficient multinomial $n!/(n_1! \dots n_d!)$ sur $\{(n_1, \dots, n_d) \in \mathbb{N}^d : n_1 + \dots + n_d = n\}$ alors grâce à la formule du multinôme et de Stirling,

$$\mathbf{P}^{2n}(0, 0) \leq \frac{(2n)! m_n}{n!^2 (4d)^n} \sum_{n_1 + \dots + n_d = n} \frac{n!}{n_1! \dots n_d!} \left(\frac{1}{d}\right)^n = \frac{(2n)! m_n}{n!^2 (4d)^n} \sim \frac{m_n}{d^n \sqrt{\pi n}}.$$

Le lemme 5.2.2 donne $m_{n+1}/m_n = (n+1)/(q+1) \leq d$ et donc $(m_n/d^n)_{n \geq 1}$ décroît, d'où

$$\sum_n \frac{m_n}{d^n \sqrt{\pi n}} = \sum_q \sum_{n=qd}^{(q+1)d-1} \frac{m_n}{d^n \sqrt{\pi n}} \leq \sum_q \frac{1}{\sqrt{q}} \sum_{n=qd}^{(q+1)d-1} \frac{m_n}{d^n} \leq \sum_q \frac{d m_{qd}}{d^{qd} \sqrt{q}}.$$

À présent, le lemme 5.2.2 et la formule de Stirling donnent, lorsque $q \rightarrow \infty$,

$$\frac{m_{qd}}{d^{qd} \sqrt{q}} = \frac{(dq)!}{d^{dq} q!^d \sqrt{q}} \sim \frac{\sqrt{d}}{(2\pi)^{(d-1)/2} q^{d/2}}.$$

Or $d \geq 3$, donc $\sum_q q^{-d/2} < \infty$, d'où $\sum_n \mathbf{P}^{2n}(0, 0) < \infty$, et donc la chaîne est transitoire.

Cas $d > 3$ (méthode astucieuse à partir du cas $d = 3$). Soit $X' = (X'_n)_{n \geq 0}$ la chaîne projetée sur \mathbb{Z}^3 c'est-à-dire constituée par les trois premières coordonnées de $X = (X_n)_{n \geq 0}$. Il s'agit d'une chaîne de Markov sur \mathbb{Z}^3 de noyau $\mathbf{Q} = (1 - \frac{d-3}{d})\mathbf{P} + \frac{d-3}{d}\mathbf{I}$ où \mathbf{P} est le noyau de la marche aléatoire simple symétrique X'' sur \mathbb{Z}^3 . Un retour à zéro de X s'accompagne toujours d'un retour à zéro de X' , et donc si X' est transitoire alors forcément X l'est aussi. Or les trajectoires de X' sont les trajectoires de X'' avec des temporisations géométriques à chaque site, et en particulier, les fréquences asymptotiques de passage sont les mêmes, ce qui fait que la transience de X'' implique celle de X' , et donc X est transitoire. \square

Lemme 5.2.2 (Coefficient multinomial central). *Pour tout $d \geq 1$ et tout $n \geq 1$,*

$$m_{n,d} := \max_{\substack{(n_1, \dots, n_d) \in \mathbb{N}^d \\ n_1 + \dots + n_d = n}} \frac{n!}{n_1! \cdots n_d!} = \frac{n!}{q^{d-r} (q+1)!^r}$$

où $n = qd + r$ et $0 \leq r < d$ (division euclidienne). En particulier, $m_{n,d} = n!/q!^d$ si $n = qd$.

Démonstration. Si $n_1 + \dots + n_d = n$ et si $n_1 < q$ alors on a nécessairement $n_k \geq q+1$ pour un $1 \leq k \leq d$, et donc, par exemple dans le cas où $k = 2$ pour alléger,

$$\frac{n!}{n_1! \cdots n_d!} < \frac{n_2}{n_1 + 1} \frac{n!}{n_1! \cdots n_d!} = \frac{n!}{(n_1 + 1)!(n_2 - 1)!n_3! \cdots n_d!}.$$

Ce type de raisonnement par monotonie permet d'établir que le maximum $m_{n,d}$ est atteint lorsque $q \leq n_k \leq q+1$ pour tout $1 \leq k \leq n$, par exemple $n_1 = \dots = n_r = q+1$ et $n_{r+1} = \dots = n_d = q$, ce qui donne le résultat annoncé. \square

5.3 Une marche aléatoire sur le groupe symétrique Σ_r

On considère $r \geq 2$ cartes à jouer empilées en un paquet vertical, et numérotées de 1 à r . On dit que la carte du dessus est en position 1, etc, et que celle du dessous est en position r . Une configuration du paquet correspond donc à un élément $\sigma \in \Sigma_r$, de sorte que $\sigma(k)$ désigne la position de la carte numéro k . On étudie une manière particulière de mélanger le paquet de cartes. Plus précisément, on considère la k -insertion qui consiste à prendre la carte du sommet du paquet et à l'insérer entre la k^e et $k+1^e$ positions. La 1-insertion n'a aucun effet. On convient que la r -insertion place la carte du sommet sous le paquet. Une k -insertion fait passer de la configuration σ à la configuration $(k, k-1, \dots, 1)\sigma$ où

$$(k, k-1, \dots, 1)$$

désigne la permutation correspondant au cycle $k \rightarrow k-1 \rightarrow \dots \rightarrow 1 \rightarrow k$. On choisit d'effectuer des k -insertions successives en utilisant une suite i.i.d. uniforme sur $\{1, \dots, r\}$ pour choisir k . En notant X_n la configuration du paquet à l'instant n , on obtient une suite aléatoire $(X_n)_{n \geq 0}$ de Σ_r vérifiant pour tout $n \geq 0$,

$$X_{n+1} = \varepsilon_{n+1} X_n = \varepsilon_{n+1} \cdots \varepsilon_1 X_0$$

où $(\varepsilon_n)_{n \geq 1}$ sont i.i.d. de loi uniforme sur $C := \{(k, k-1, \dots, 1) : 1 \leq k \leq r\} \subset \Sigma_r$. La suite $(X_n)_{n \geq 0}$ est une marche aléatoire à gauche sur le groupe (non abélien) Σ_r , d'incrémentes C . C'est aussi une chaîne de Markov d'espace d'état fini Σ_r (de cardinal $r!$) et de noyau

$$\mathbf{P}(\sigma, \sigma') = \frac{\mathbb{1}_C(\sigma'\sigma^{-1})}{|C|} = \frac{\mathbb{1}_C(\sigma'\sigma^{-1})}{r}.$$

Théorème 5.3.1 (Convergence en loi). *La chaîne $(X_n)_{n \geq 0}$ est récurrente irréductible apériodique. Son unique loi invariante est la loi uniforme sur Σ_r :*

$$\mu = \sum_{\sigma \in \Sigma_r} \frac{1}{|\Sigma_r|} \delta_\sigma = \frac{1}{r!} \sum_{\sigma \in \Sigma_r} \delta_\sigma.$$

La chaîne $(X_n)_{n \geq 0}$ converge en loi vers μ quelque soit la loi initiale $\mathcal{L}(X_0)$.

Démonstration. Les transpositions spéciales $\mathcal{T} := \{(r, r-1), \dots, (2, 1), (1, r)\}$ engendrent Σ_r . Comme les k -insertions $(r, r-1, \dots, 1)$ et $(2, 1)$ correspondant à $k = r$ et à $k = 2$ engendrent \mathcal{T} , on en déduit qu'elles engendrent Σ_r . Donc C engendre Σ_r et la chaîne est irréductible. Comme l'identité (1) est également une k -insertion ($k = 1$), la diagonale de la matrice de transition de la chaîne est > 0 et donc la chaîne est apériodique. Comme Σ_r est un groupe, il y a exactement r états qui conduisent à chaque état, donc les colonnes de la matrice de transition ont exactement r entrées non nulles, toutes égales à $1/r$. Ainsi, la transposée de la matrice de transition est également une matrice de transition² et donc la loi uniforme est invariante. Or toute chaîne finie récurrente apériodique possède une unique loi invariante vers laquelle elle converge en loi quelque soit sa loi initiale. \square

On souhaite à présent étudier la vitesse de convergence de la chaîne vers sa mesure invariante, partant d'une configuration initiale X_0 fixée. Par invariance par translation, on peut supposer que $X_0 = (1)$, c'est-à-dire que **toutes les cartes sont dans l'ordre au départ**. Au temps 0 la carte r est en position r (tout en bas du paquet), et subit une remontée pas à pas au fil du temps jusqu'au sommet. Cette remontée est de plus en plus rapide. Pour tout $1 \leq k \leq r-1$, on note T_k le temps que cette carte passe en position $r-k$. Pour tout $k \geq 1$, on a, avec la convention $T_1 + \dots + T_{k-1} = 0$ si $k = 1$,

$$\begin{aligned} T_k &= \min\{n \geq 1 : X_{T_1 + \dots + T_{k-1} + n} = r - k\} \\ &= \min\{n \geq 1 : X_n(r) = r - k\} - (T_1 + \dots + T_{k-1}). \end{aligned}$$

Au temps $T_1 + \dots + T_{r-1}$ la carte r est en position 1 (sommet du paquet). La variable aléatoire

$$T := 1 + T_1 + \dots + T_{r-1}$$

suit la loi du collectionneur de coupons de r coupons de probabilité d'apparition uniforme étudiée dans le chapitre 1. Les v.a. T_1, \dots, T_{r-1} sont indépendantes avec $T_k \sim \text{Geo}((r-k)/r)$ pour tout $1 \leq k \leq r-1$. Notons que $T \geq r$.

Théorème 5.3.2 (Bon mélange après remontée). *Pour tout $n \geq 0$, les variables aléatoires X_{T+n} et T sont indépendantes. De plus, X_{T+n} suit la loi uniforme μ sur Σ_r .*

On dit alors que T est un temps fort de stationnarité (d'après Aldous et Diaconis).

2. Elle est doublement stochastique (ou bistochastique). L'ensemble des matrices doublement stochastiques $n \times n$ est un polytope (intersection de demi-espaces) convexe et compact à $(n-1)^2$ degrés de liberté. Ses points extrémaux sont les matrices de permutations (Birkhoff et von Neumann). Encore le groupe symétrique !

Démonstration. La loi uniforme sur Σ_r peut s'obtenir en tirant uniformément sans remise les images de $1, \dots, r$. D'autre part, la loi uniforme sur Σ_r est invariante par translation.

Au temps T_1 , la carte r se trouve pour la première fois en position $r - 1$, car la carte numéro 1 a été glissée sous le paquet. Au temps $T_1 + T_2$, la carte r se trouve pour la première fois en position $r - 2$ car la carte numéro 2 a été glissée sous ou sur la carte 1. Les numéros des deux cartes sous le paquet sont $1, 2$ ou $2, 1$ avec probabilité $1/2$. Par récurrence, au temps $T_1 + \dots + T_{r-1} = T - 1$, la carte r se trouve au sommet du paquet pour la première fois et les $r - 1$ cartes qui sont sous elle ont des numéros répartis uniformément sans remise dans $\{1, \dots, r - 1\}$. Au temps T , la carte r est placée aléatoirement et uniformément dans le paquet à une position entre 1 et r et donc X_T suit la loi uniforme. Comme la probabilité $\mathbb{P}(X_T = \sigma, T = k)$ ne dépend pas de σ , on en déduit que T et X_T sont indépendantes. Comme μ est invariante, on obtient $X_{T+n} \sim \mu$ pour tout $n \geq 0$. \square

Pour bien mélanger le paquet de cartes, on pourrait s'arrêter au temps T . Malheureusement, on ne connaît pas T en pratique ! Alternativement, on pourrait chercher à déterminer une valeur de n déterministe, aussi petite que possible, telle que $d_{VT}(\mathcal{L}(X_n), \mu)$ est proche de zéro, où $d_{VT}(\cdot, \cdot)$ est la distance en variation totale (section 1.3). Il s'avère que pour r assez grand, la quantité $d_{VT}(\mathcal{L}(X_n), \mu)$ passe de 1 à 0 de manière abrupte autour de $n = r \log(r)$. Ce phénomène de convergence abrupte est quantifié par le théorème suivant.

Théorème 5.3.3 (Convergence abrupte autour de $n = r \log(r)$). *Pour tout réel $c > 0$,*

$$d_{VT}(\mathcal{L}(X_{r \log(r) + cr}), \mu) \leq e^{-c}.$$

Si $(c_r)_{r \geq 1}$ dans \mathbb{R}_+ vérifie $\lim_{r \rightarrow \infty} c_r = \infty$ et $r \log(r) - rc_r > 0$ pour tout $r \geq 1$, alors

$$\lim_{r \rightarrow \infty} d_{VT}(\mathcal{L}(X_{r \log(r) - rc_r}), \mu) = 1.$$

Démonstration. Grâce au théorème 5.3.2, on a, pour tout $A \subset \Sigma_r$,

$$\begin{aligned} \mathbb{P}(X_n \in A) &= \sum_{k=0}^n \mathbb{P}(X_n \in A, T = k) + \mathbb{P}(X_n = \sigma, T > n) \\ &\leq \sum_{k=0}^n \mathbb{P}(X_n \in A, T = k) + \mathbb{P}(T > n) \\ &= \sum_{k=0}^n \mathbb{P}(X_{T+n-k} \in A, T = k) + \mathbb{P}(T > n) \\ &= \sum_{k=0}^n \mu(A) \mathbb{P}(T = k) + \mathbb{P}(T > n) \\ &= \mu(A) \mathbb{P}(T \leq n) + \mathbb{P}(T > n) \\ &\leq \mu(A) + \mathbb{P}(T > n). \end{aligned}$$

d'où $\mathbb{P}(X_n \in A) - \mu(A) \leq \mathbb{P}(T > n)$. Appliqué à A et A^c , cela donne, pour tout $n \geq 0$,

$$d_{VT}(\mathcal{L}(X_n), \mu) \leq \mathbb{P}(T > n).$$

Le résultat voulu découle à présent du théorème 1.5.7 sur le collectionneur de coupons. \square

5.4 Notes et commentaires

La formule de Désiré André se trouve dans [And].

Norris [Nor98], Cottrell et al [CDGC80], Feller [Fel68], Doyle et Snell [DS84]. Pour la convergence abrupte des battages de cartes, les articles de Diaconis et al [AD86, DMP95, DFP92, BD92] et le livre [LPW09].

Cette manière de mélanger assez peu usuelle a le mérite de permettre une analyse mathématique simple et rapide qui mène au phénomène de convergence abrupte à l'équilibre. Ce phénomène est commun à toutes les manières de mélanger les cartes qui ont été étudiées. On peut penser par exemple au *riffle shuffle* : on coupe le paquet en deux et on fusionne les deux sous-paquets en intercalant leurs cartes.

Au delà de nos exemples, on peut définir une marche aléatoire sur un groupe G comme la suite $(X_n)_{n \geq 0}$ de v.a.r. sur G donnée par la récursion $X_{n+1} = \varepsilon_{n+1} X_n = \varepsilon_{n+1} \cdots \varepsilon_1 X_0$ où $(\varepsilon_n)_{n \geq 1}$ sont i.i.d. de loi uniforme η sur $C \subset G$ fixé. Comme G n'est pas forcément commutatif, cette marche à gauche n'est pas forcément égale à la marche à droite $X_{n+1} = X_n \varepsilon_{n+1} = X_0 \varepsilon_1 \cdots \varepsilon_{n+1}$. On a $X_n \sim \eta^{*n} * \nu$ pour la marche à gauche et $X_{n+1} \sim \nu * \eta^{*n}$ pour la marche à droite, où $X_0 \sim \nu$. Si G est compact, par exemple fini, et si C engendre G , alors on peut espérer que $(X_n)_{n \geq 0}$ convergera en loi vers la mesure de Haar normalisée de G , qui est la loi uniforme μ sur G lorsque G est fini.

Le collectionneur de coupons intervient également dans l'analyse de la marche aléatoire sur l'hypercube (section ??) qui constitue un exemple remarquable de marche aléatoire sur un graphe régulier. Il est également possible de définir des marches aléatoires sur des graphes orientés ou non plus généraux, en considérant, pour chaque sommet, une loi sur ses voisins. Ces marches constituent des chaînes de Markov. Réciproquement, toute chaîne de Markov d'espace d'état au plus dénombrable peut être vue comme une marche aléatoire sur son graphe squelette. L'algorithme PageRank de Google est basé sur une marche aléatoire sur le graphe orienté des pages web. Un autre exemple classique est celui de la marche aléatoire sur le graphe de Cayley d'un groupe finiment engendré.

Il est possible d'étudier des marches aléatoires en milieu aléatoire, en paysage aléatoire, en auto interaction (évitement, renforcement, excitation, ...). Le sujet est d'une grande richesse. Par ailleurs, le théorème central limite permet de concevoir le mouvement brownien comme un analogue en temps et espace continu de la marche aléatoire simple symétrique.

Séance 6

Matrices aléatoires et loi du demi-cercle de Wigner

Mots-clés. Matrices aléatoires ; matrices de covariance ; spectre ; convergence étroite ; moments ; fonctions caractéristiques ; combinatoire ; nombres de Catalan ; théorème de Wigner ; théorème de Marchenko-Pastur ; vecteurs gaussiens.

Nous commençons par une motivation statistique. Soit X_1, \dots, X_n des vecteurs colonne aléatoires i.i.d. de \mathbb{R}^d de moyenne 0 et de matrice de covariance Σ . La matrice Σ est symétrique $d \times d$ et ses valeurs propres sont positives ou nulles. On a $\Sigma_{ij} = \mathbb{E}(X_{ki}X_{kj})$ pour tous $1 \leq i, j \leq d$ et tout $1 \leq k \leq n$ ou encore

$$\Sigma = \mathbb{E}(X_1 X_1^\top) = \dots = \mathbb{E}(X_n X_n^\top).$$

La *matrice de covariance empirique* $\widehat{\Sigma}_n$ est la matrice symétrique $d \times d$ définie par

$$\widehat{\Sigma}_n = \frac{1}{n} \sum_{k=1}^n X_k X_k^\top = \frac{1}{n} (X_1 \cdots X_n) (X_1 \cdots X_n)^\top.$$

On a $\mathbb{E}(\widehat{\Sigma}_n) = \Sigma$, et par la loi des grands nombres appliquée à chacune des $d \times d$ entrées,

$$\widehat{\Sigma}_n \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \Sigma.$$

On souhaite étudier le comportement de la matrice aléatoire $\widehat{\Sigma}_n$ lorsque la dimension des données $d = d_n$ dépend de n et tend vers ∞ . Simplifions en posant $\Sigma = I$. Cela nous conduit au modèle suivant : on se donne une famille $(Y_{ij})_{i,j \geq 1}$ de v.a.r. i.i.d. de moyenne 0 et de variance 1, et pour tout entier $n \geq 1$, on considère la matrice aléatoire $d_n \times d_n$ symétrique

$$\frac{1}{n} Y Y^\top$$

où $Y = (Y_{ij})_{1 \leq i \leq d_n, 1 \leq j \leq n}$. Si $n \mapsto d_n$ est constante et égale à d alors $\frac{1}{n} Y Y^\top$ converge p.s. vers I . Il est naturel de chercher à comprendre le comportement de la matrice $\frac{1}{n} Y Y^\top$ lorsque d_n dépend de n , par exemple en étudiant son spectre. L'analyse de la matrice symétrique $Y Y^\top$ est rendue difficile par le fait que ses coefficients sont dépendants. Cela suggère d'analyser en première approche des matrices aléatoires symétriques dont les entrées sont indépendantes, ce qui conduit au théorème de Wigner, objet principal de ce chapitre. Nous reviendrons brièvement aux matrices de covariance empiriques en fin de chapitre.

6.1 Théorème de Wigner

Soit M une matrice symétrique aléatoire obtenue en extrayant le carré $n \times n$ situé dans le coin supérieur gauche d'un tableau symétrique infini de variables aléatoires i.i.d. de moyenne m et de variance finie et non nulle σ^2 . Soit $\lambda_{n,1}, \dots, \lambda_{n,n}$ les valeurs propres de la matrice symétrique $\frac{1}{\sqrt{n}}M$, ordonnées de sorte que $\lambda_{n,1} \geq \dots \geq \lambda_{n,n}$. On s'intéresse à leur mesure de comptage, appelée *mesure spectrale empirique*, définie par

$$\mu_n := \frac{1}{n} \sum_{k=1}^n \delta_{\lambda_{n,k}}.$$

Il s'agit d'une mesure de probabilité aléatoire, de même nature que la mesure empirique dans le théorème de Glivenko-Cantelli¹, mis à part que les atomes $\lambda_{n,1}, \dots, \lambda_{n,n}$ sont ici des variables aléatoires dépendantes². Pour tout borélien B de \mathbb{R} on a

$$\mu_n(B) = \frac{\text{card}\{1 \leq k \leq n : \lambda_{n,k} \in B\}}{n}.$$

La fonction de répartition F_n de μ_n est donnée pour tout $x \in \mathbb{R}$ par

$$F_n(x) := \mu_n(]-\infty, x]) = \frac{\text{card}\{1 \leq k \leq n : \lambda_{n,k} \leq x\}}{n}.$$

Le théorème de Wigner constitue une sorte d'analogue de la loi des grands nombres.

Théorème 6.1.1 (Théorème de Wigner). *Presque sûrement, la suite de mesures de probabilités $(\mu_n)_{n \geq 1}$ converge étroitement³ vers la loi du demi-cercle μ_σ^{DC} de densité*

$$x \mapsto \frac{1}{2\pi\sigma^2} \sqrt{4\sigma^2 - x^2} \mathbb{1}_{[-2\sigma, 2\sigma]}(x).$$

Le nom « loi du demi-cercle » utilisé pour désigner μ_σ^{DC} vient de fait que la densité est donnée par un demi-cercle (en vérité une demi ellipse en raison de la constante de normalisation, égale à la moitié de l'aire du cercle). Comme μ_σ^{DC} n'a pas d'atomes, le théorème de Wigner affirme que presque sûrement, pour tout intervalle $I \subset \mathbb{R}$,

$$\lim_{n \rightarrow \infty} \mu_n(I) = \mu_\sigma^{\text{DC}}(I) = \frac{1}{2\pi\sigma^2} \int_{I \cap [-2\sigma, 2\sigma]} \sqrt{4\sigma^2 - x^2} dx.$$

Ainsi, la proportion de valeurs propres de M n'appartenant pas à l'intervalle $[-2\sigma\sqrt{n}, 2\sigma\sqrt{n}]$ tend vers zéro quand $n \rightarrow \infty$. Le théorème de Wigner met en lumière un phénomène d'universalité, en ce sens que la loi limite μ_σ^{DC} ne dépend de la loi des coefficients de la matrice M qu'à travers leur variance σ^2 . Le fait que la moyenne m n'intervient pas dans la loi limite peut être compris en observant que le spectre d'une matrice symétrique est peu sensible aux perturbations de faible rang, et le rang de $M - \mathbb{E}(M)$ est au plus 1.

Le comportement de la moyenne de μ_n peut être compris en observant que

$$\int x d\mu_n(x) = \frac{1}{n} \sum_{k=1}^n \lambda_{n,k} = \frac{1}{n} \text{Tr} \left(\frac{1}{\sqrt{n}} M \right) = \frac{1}{n^{3/2}} \sum_{1 \leq i \leq n} M_{i,i} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} 0 = \int x d\mu_\sigma^{\text{DC}}(x)$$

1. Si $(Y_n)_{n \geq 1}$ sont i.i.d. de loi η alors presque sûrement, la mesure empirique $\frac{1}{n} \sum_{k=1}^n \delta_{Y_k}$ converge en loi vers η quand $n \rightarrow \infty$, et la convergence des fonctions de répartition est uniforme. Il s'agit d'une conséquence de la loi forte des grands nombres, et du fait que la topologie de la convergence en loi sur \mathbb{R} est séparable.

2. Ce sont des fonctions non linéaires des coefficients de M (racines du polynôme caractéristique!).

3. Il s'agit de la convergence faible pour les fonctions continues et bornées, ou alternativement de la convergence ponctuelle des fonctions caractéristiques, ou encore de la convergence ponctuelle des fonctions de répartition en tout point de continuité de la limite. Pour les v.a., cela correspond à la convergence en loi.

où la convergence presque sûre provient de la loi forte des grands nombres. La normalisation en $1/\sqrt{n}$ de M peut être comprise à son tour en observant tout d'abord que

$$\int x^2 d\mu_n(x) = \frac{1}{n} \sum_{k=1}^n \lambda_{n,k}^2 = \frac{1}{n} \operatorname{Tr} \left(\left(\frac{1}{\sqrt{n}} M \right)^2 \right) = \frac{1}{n^2} \sum_{1 \leq i, j \leq n} M_{ij}^2 \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \sigma^2 + m^2$$

où la convergence presque sûre provient ici encore de la loi forte des grands nombres. D'autre part, grâce à l'inégalité de Markov, presque sûrement, pour tout $r > 0$,

$$\mu_n([-r, r]^c) \leq \frac{1}{r^2} \int x^2 d\mu_n(x) \xrightarrow[n \rightarrow \infty]{} \frac{\sigma^2 + m^2}{r^2}.$$

Par conséquent, presque sûrement, pour tout $\varepsilon > 0$ il existe un compact $K_\varepsilon = [-r_\varepsilon, r_\varepsilon]$ tel que $\mu_n(K_\varepsilon) \leq \varepsilon$ pour tout $n \geq 1$. Ainsi, presque sûrement, la suite $(\mu_n)_{n \geq 1}$ est *tendue*, et le théorème de Prohorov affirme alors que cette suite est relativement compacte pour la convergence étroite, en ce sens que si elle possède une unique valeur d'adhérence pour la convergence étroite alors elle converge étroitement. Malgré cette observation réconfortante, nous allons utiliser une autre approche pour démontrer le théorème de Wigner. Pour simplifier, on se propose de démontrer la version faible suivante du théorème de Wigner, qui concerne la suite de lois $(\mathbb{E}\mu_n)_{n \geq 1}$. La réduction du théorème de Wigner au théorème de Wigner faible se fait en utilisant des techniques de centrage (inégalité de Weyl), de troncature (inégalité de Hoffman-Wielandt), et de concentration de la mesure (inégalité de Azuma-Hoeffding), que nous ne développons pas ici par souci de simplicité.

Théorème 6.1.2 (Théorème de Wigner faible). *Si les M_{ij} sont centrées ($m = 0$), réduites ($\sigma^2 = 1$), et à support compact, alors pour toute fonction continue et bornée $f : \mathbb{R} \rightarrow \mathbb{R}$,*

$$\lim_{n \rightarrow \infty} \mathbb{E} \int f d\mu_n = \int f d\mu_1^{\text{DC}}.$$

La preuve du théorème 6.1.2 consiste à se ramener à des fonctions test f polynomiales (théorème 6.2.1 ci-dessous), puis à établir la convergence des moments (théorème 6.4.1 ci-dessous), en s'inspirant de la méthode utilisée ci-dessus pour les moments d'ordre 1 et 2.

6.2 Convergence des moments et convergence étroite

Soit \mathcal{P} l'ensemble des lois de probabilité μ sur \mathbb{R} telles que $\mathbb{R}[X] \subset L^1(\mu)$, muni de la relation d'équivalence $\mu_1 \sim \mu_2$ ssi $\int P d\mu_1 = \int P d\mu_2$ pour tout $P \in \mathbb{R}[X]$, c'est-à-dire que μ_1 et μ_2 ont les mêmes moments. On dit que $\mu \in \mathcal{P}$ est **caractérisée par ses moments** lorsque sa classe d'équivalence est un singleton.

Théorème 6.2.1 (Convergence des moments et convergence étroite). *Si μ, μ_1, μ_2, \dots sont des éléments de \mathcal{P} vérifiant, pour tout $P \in \mathbb{R}[X]$,*

$$\lim_{n \rightarrow \infty} \int P d\mu_n = \int P d\mu$$

et si μ est caractérisée par ses moments alors pour toute $f : \mathbb{R} \rightarrow \mathbb{R}$ continue bornée,

$$\lim_{n \rightarrow \infty} \int f d\mu_n = \int f d\mu.$$

4. La loi $\mathbb{E}\mu_n$ est définie par $\int f d\mathbb{E}\mu_n = \mathbb{E} \int f d\mu_n$ pour f mesurable et positive ou mesurable et bornée.

Démonstration. Par hypothèse, pour tout $P \in \mathbb{R}[X]$,

$$C_P := \sup_{n \geq 1} \int P d\mu_n < \infty.$$

Par conséquent, par l'inégalité de Markov, pour tout réel $R > 0$,

$$\mu_n([-R, R]^c) \leq \frac{C_{X^2}}{R^2}$$

et donc $(\mu_n)_{n \geq 1}$ est tendue. Grâce au théorème de Prohorov, il suffit donc d'établir que si $(\mu_{n_k})_{k \geq 1}$ converge étroitement vers ν alors $\nu = \mu$. Fixons $P \in \mathbb{R}[X]$ et un réel $R > 0$. Soit $\varphi_R : \mathbb{R} \rightarrow [0, 1]$ continue telle que $\mathbb{1}_{[-R, R]} \leq \varphi_R \leq \mathbb{1}_{[-R-1, R+1]}$. On a la décomposition

$$\int P d\mu_{n_k} = \int \varphi_R P d\mu_{n_k} + \int (1 - \varphi_R) P d\mu_{n_k}.$$

Comme $(\mu_{n_k})_{k \geq 1}$ converge étroitement vers ν on a

$$\lim_{k \rightarrow \infty} \int \varphi_R P d\mu_{n_k} = \int \varphi_R P d\nu.$$

De plus, par les inégalités de Cauchy-Schwarz et de Markov, on a

$$\left| \int (1 - \varphi_R) P d\mu_{n_k} \right|^2 \leq \mu_{n_k}([-R, R]^c) \int P^2 d\mu_{n_k} \leq \frac{C_{X^2} C_{P^2}}{R^2}.$$

D'un autre côté, on sait que

$$\lim_{k \rightarrow \infty} \int P d\mu_{n_k} = \int P d\mu$$

et on obtient donc

$$\lim_{R \rightarrow \infty} \int \varphi_R P d\nu = \int P d\mu.$$

En utilisant cela pour P^2 , qui est ≥ 0 , on obtient d'abord par convergence monotone que $P \in L^2(\nu) \subset L^1(\nu)$, puis par convergence dominée que

$$\int P d\nu = \int P d\mu.$$

Comme P est arbitraire et μ est caractérisée par ses moments, on obtient $\mu = \nu$. \square

Un théorème de Weierstrass affirme que pour tout compact $K \subset \mathbb{R}$, les polynômes $\mathbb{R}[X]$ restreints à K sont denses dans $\mathcal{C}(K, \mathbb{R}, \|\cdot\|_\infty)$. Ainsi, toute loi de probabilités à support compact est caractérisée par ses moments *parmi l'ensemble des lois à support compact*. Le théorème suivant permet d'aller plus loin, car il implique que toute loi à support compact est caractérisée par ses moments parmi l'ensemble des lois dont tous les moments sont finis. C'est le cas en particulier de la loi du demi-cercle μ_σ^{DC} .

Théorème 6.2.2 (Analyticité de la transformée de Fourier et moments). *Soit $\mu \in \mathcal{P}$. On pose $\varphi(t) = \int e^{itx} d\mu(x)$ et $\kappa_n = \int x^n d\mu(x)$. Les propositions suivantes sont équivalentes :*

1. φ est analytique sur un voisinage de 0
2. φ est analytique sur \mathbb{R}
3. $\overline{\lim}_n \left(\frac{1}{n!} |\kappa_n| \right)^{\frac{1}{n}} < \infty$.

Si ces conditions sont vérifiées alors μ est caractérisée par ses moments.

En particulier, une loi à support compact est caractérisée par ses moments.

La formule de Stirling $n! = \sqrt{2\pi n}(n/e)^n(1 + \mathcal{O}(1/n))$ donne $(1/n!)^{1/n} = \mathcal{O}(1/n)$ et donc si $\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} |\kappa_n|^{1/n}$ alors μ est caractérisée par ses moments.

Démonstration. Pour tout n , on a $\int |x|^n d\mu < \infty$ et donc φ est n fois dérivable sur \mathbb{R} . De plus, $\varphi^{(n)}$ est continue sur \mathbb{R} et pour tout $t \in \mathbb{R}$,

$$\varphi^{(n)}(t) = \int_{\mathbb{R}} (ix)^n e^{itx} d\mu(x).$$

En particulier, $\varphi^{(n)}(0) = i^n \kappa_n$, et la série de Taylor de φ en 0 est déterminée par la suite $(\kappa_n)_{n \geq 1}$. Le rayon de convergence r de la série entière $\sum_n a_n z^n$ associée à la suite de nombres complexes $(a_n)_{n \geq 0}$ est donné par la formule de Hadamard $r^{-1} = \overline{\lim}_n |a_n|^{1/n}$. Ainsi, $1 \Leftrightarrow 3$ (prendre $a_n = i^n \kappa_n / n!$). D'autre part, comme pour tout $n \in \mathbb{N}$, $s, t \in \mathbb{R}$,

$$\left| e^{isx} \left(e^{itx} - 1 - \frac{itx}{1!} - \dots - \frac{(itx)^{n-1}}{(n-1)!} \right) \right| \leq \frac{|tx|^n}{n!},$$

on a pour tout $n \in \mathbb{N}$ pair et tous $s, t \in \mathbb{R}$,

$$\left| \varphi(s+t) - \varphi(s) - \frac{t}{1!} \varphi'(s) - \dots - \frac{t^{n-1}}{(n-1)!} \varphi^{(n-1)}(s) \right| \leq \kappa_n \frac{|t|^n}{n!},$$

qui montre que $3 \Rightarrow 2$. Comme $2 \Rightarrow 1$, on a bien équivalence de $1 \Leftrightarrow 2 \Leftrightarrow 3$. Si ces propriétés ont lieu, alors les arguments précédents donnent un $r > 0$ tel que φ est développable en série entière en tout $x \in \mathbb{R}$ avec un rayon de convergence $\geq r$. De proche en proche, on obtient que φ est caractérisée par ses dérivées en zéro. Si μ est à support compact, alors $\sup_n |\kappa_n|^{1/n} < \infty$ et donc 3 a lieu (via Stirling $n! = (n/e)^n \sqrt{2\pi n} (1 + \mathcal{O}(1/n))$). \square

6.3 Moments de la loi du demi-cercle

Comme μ_σ^{DC} est à support compact, elle est caractérisée par ses moments. Les voici :

Théorème 6.3.1 (Moments de la loi du demi-cercle). *Pour tout entier $r \geq 0$,*

$$\int x^{2r+1} d\mu_\sigma^{\text{DC}}(r) = 0 \quad \text{et} \quad \int x^{2r} d\mu_\sigma^{\text{DC}}(r) = \sigma^{2r} \frac{1}{r+1} \binom{2r}{r}.$$

En particulier, la loi du demi-cercle μ_σ^{DC} de support $[-2\sigma, 2\sigma]$ a pour moyenne 0 et variance σ^2 , et les moments pairs de la loi du demi-cercle standard μ_1^{DC} sont les nombres de Catalan.

Démonstration. Les moments impairs sont nuls car μ_σ^{DC} est symétrique. Pour calculer les moments pairs, on observe tout d'abord par dilatation et parité que

$$\int x^{2r} d\mu_\sigma^{\text{DC}}(x) = \frac{\sigma^{2r}}{\pi} \int_0^2 x^{2r} \sqrt{4-x^2} dx.$$

À présent, en utilisant le changement de variable $x = 2 \cos(u)$ et une intégration par parties,

$$\int_0^2 x^{2r} \sqrt{4-x^2} dx = \int_0^{\frac{\pi}{2}} 4^{r+1} \cos^{2r}(u) \sin^2(u) du = \frac{4^{r+1}}{2r+1} \int_0^{\frac{\pi}{2}} \cos^{2(r+1)}(u) du$$

et cette dernière expression n'est rien d'autre qu'une classique intégrale de Wallis. \square

6.4 Convergence des moments et théorème de Wigner faible

Théorème 6.4.1 (Convergence des moments). *Sous les hypothèses du théorème de Wigner faible 6.1.2 on a, pour tout entier $r \geq 1$,*

$$\mathbb{E} \int x^r d\mu_n \xrightarrow{n \rightarrow \infty} \int x^r d\mu_\sigma^{\text{DC}}.$$

Démonstration. Les moments de μ_1^{DC} sont donnés par le théorème 6.3.1. Comme par hypothèse, les coefficients de la matrice aléatoire M sont centrés et réduits, on a

$$\mathbb{E} \int x d\mu_n = \frac{1}{n} \mathbb{E} \sum_{k=1}^n \lambda_{n,k} = \frac{1}{n^{3/2}} \mathbb{E} \text{Tr}(M) = \frac{1}{n^{3/2}} \sum_{i=1}^n \mathbb{E} M_{ii} = 0 = \int x d\mu_1^{\text{DC}}(x)$$

et

$$\mathbb{E} \int x^2 d\mu_n = \frac{1}{n} \mathbb{E} \sum_{k=1}^n \lambda_{n,k}^2 = \frac{1}{n^2} \mathbb{E} \text{Tr}(M^2) = \frac{1}{n^2} \sum_{1 \leq i, j \leq n} \mathbb{E}(M_{ij}^2) = 1 = \int x^2 d\mu_1^{\text{DC}}(x).$$

L'étude du moment d'ordre 3 est un peu plus subtile. On a

$$\mathbb{E} \int x^3 d\mu_n = \frac{1}{n} \mathbb{E} \sum_{k=1}^n \lambda_{n,k}^3 = \frac{1}{n^{1+3/2}} \mathbb{E} \text{Tr}(M^3) = \frac{1}{n^{1+3/2}} \sum_{1 \leq i, j, k \leq n} \mathbb{E}(M_{ij} M_{jk} M_{ki}).$$

Si deux éléments parmi $\{\{i, j\}, \{j, k\}, \{k, i\}\}$ sont distincts alors $\mathbb{E}(M_{ij} M_{jk} M_{ki}) = 0$ par indépendance et centrage. Dans le cas contraire, on a $i = k$ ou $i = j$ ou $k = j$. Si $i = k$ alors $\mathbb{E}(M_{ij} M_{jk} M_{ki}) = \mathbb{E}(M_{ii}^2 M_{ij})$ qui est nul si $j \neq i$ et qui n'a pas de contribution asymptotique si $j = i$. Si $i = j$ alors $\mathbb{E}(M_{ij} M_{jk} M_{ki}) = \mathbb{E}(M_{ii} M_{ik}^2)$ et le raisonnement est identique. Si $k = j$ alors $\mathbb{E}(M_{ij} M_{jk} M_{ki}) = \mathbb{E}(M_{ij}^2 M_{jj})$ et le raisonnement est à nouveau identique. Ainsi, le moment d'ordre 3 de μ_n converge quand $n \rightarrow \infty$ vers 0, qui est le moment d'ordre 3 de μ_1^{DC} . Plus généralement, pour le moment d'ordre $r \geq 1$, on écrit (avec $i_{r+1} := i_1$)

$$\int x^r d\mu_n(x) = \frac{1}{n} \sum_{k=1}^n \lambda_{n,k}^r = \frac{1}{n^{1+r/2}} \text{Tr}(M^r) = \frac{1}{n^{1+r/2}} \sum_{1 \leq i_1, \dots, i_r \leq n} \mathbb{E}(M_{i_1 i_2} \cdots M_{i_r i_{r+1}}).$$

Nous allons associer à chaque r -uplet d'indices i_1, \dots, i_r un graphe orienté, obtenu en

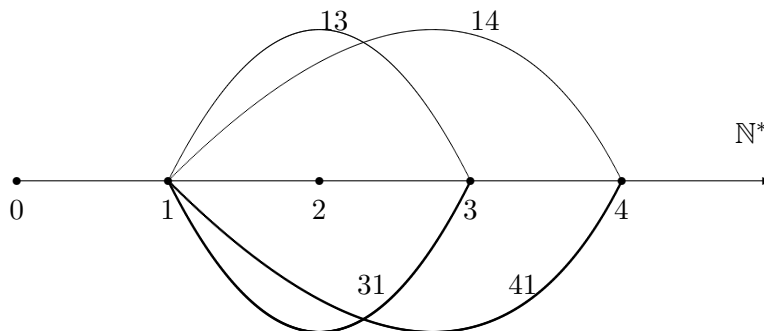


Figure 6.1 – Graphe pour $r = 4$ et $i_1 = i_3 = 1$, $i_2 = 4$, $i_4 = 3$, correspondant à $\mathbb{E}(M_{14} M_{41} M_{13} M_{31})$. Les arrêtes successives sont 14, 41, 13, 31. On a $t = 3$.

positionnant leurs valeurs sur une droite horizontale figurant \mathbb{N}^* , cf. figure. Les sommets du graphe sont les valeurs distinctes prises par ces indices, et on note t leur nombre.

Les arrêtes du graphe sont les liaisons (i_k, i_{k+1}) avec $1 \leq k \leq r+1$ et $i_{r+1} := i_1$. Elles peuvent avoir une multiplicité, et une orientation. On obtient ainsi un graphe cyclique de longueur $r+1$ possédant t sommets distincts, appelé graphe $G(r, t)$. Deux graphes $G(r, t)$ sont équivalents lorsque qu'on peut passer de l'un à l'autre en permutant les indices. Des graphes $G(r, t)$ équivalents donnent la même valeur à $\mathbb{E}(M_{i_1 i_2} \cdots M_{i_r i_{r+1}})$, notée $\mathbb{E}(M_G)$. Il y a $n(n-1) \cdots (n-t+1)$ graphes $G(r, t)$ dans chaque classe d'équivalence (nombre d'arrangements de t objets parmi n objets). Afin de calculer les contributions, on distingue trois types de graphes $G(r, t)$:

- type 1 : ceux pour qui chaque arrête présente ne l'est qu'une fois et dans chaque sens, et le graphe des arrêtes montantes est un arbre (i.e. sans cycles) ;
- type 2 : ceux pour qui une arrête au moins n'apparaît qu'une seule fois ;
- type 3 : ceux qui ne sont pas de type 1 ou de type 2.

Nous abordons à présent la phase finale :

- *Les graphes de type 2 ont une contribution nulle.* En effet, si G est un graphe de type 2 alors $\mathbb{E}(M_G) = 0$ par indépendance et centrage ;
- *Les graphes de type 3 ont une contribution asymptotiquement nulle.* En effet, si G est dans \mathcal{G}_3 alors G contient au moins trois arrêtes de mêmes extrémités ou un cycle d'arrêtes d'extrémités différentes, ce qui implique dans les deux cas l'inégalité $2t \leq r+1$, qui donne la majoration $n(n-1) \cdots (n-t+1) \leq n^t \leq n^{1/2+r/2}$. D'autre part, on a $\text{card}(\mathcal{G}_3) \leq \text{card}(G(r, t)) \leq C^r = \mathcal{O}(1)$, et comme les coefficients de M sont bornés, on a également $\mathbb{E}(M_G) \leq C^r = \mathcal{O}(1)$, d'où enfin

$$\sum_{\text{classes de type 3}} \frac{n(n-1) \cdots (n-t+1)}{n^{1+r/2}} \mathbb{E}(M_G) = \mathcal{O}(n^{-1/2}) = o_{n \rightarrow \infty}(1);$$

- *Seuls les graphes de type 1 contribuent asymptotiquement.* En effet, pour tout graphe G de type 1 on a $\mathbb{E}(M_G) = 1$ par indépendance car les coefficients de M ont une variance de 1. Cela ramène le problème à la détermination de $\text{card}(\mathcal{G}_1)$. Si r est impair alors $\text{card}(\mathcal{G}_1) = 0$, tandis que si r est pair, disons $r = 2s$, alors $t = r/2$ et les classes de type 1 sont en bijections avec les parenthésages ou les excursions de la marche aléatoire simple sur \mathbb{Z} , (théorème 5.1.5), il y en a donc $\frac{1}{s+1} \binom{2s}{s}$, et donc

$$\sum_{\text{cl. t. 1}} \frac{n(n-1) \cdots (n-t+1)}{n^{1+r/2}} \mathbb{E}(M_G) = \frac{n}{n} \cdots \frac{n-s+1}{n} \frac{1}{s+1} \binom{2s}{s} \xrightarrow{n \rightarrow \infty} \frac{1}{1+s} \binom{2s}{s}.$$

□

6.5 Retour aux matrices de covariance empiriques

Considérons le modèle de covariance du début du chapitre. Soit $\lambda_{n,1} \leq \cdots \leq \lambda_{n,d_n}$ le spectre de $\frac{1}{n} Y Y^\top$ ordonné de manière croissante. On définit la mesure de comptage

$$\mu_n := \frac{1}{d_n} \sum_{k=1}^{d_n} \delta_{\lambda_{n,k}}.$$

On dispose alors du théorème de Marchenko-Pastur ci-dessous, qui se démontre avec la même méthode que le théorème de Wigner, bien que la mise en œuvre soit plus lourde.

Théorème 6.5.1 (Marchenko-Pastur). *Si $\Sigma = I_{d_n}$ et si $\lim_{n \rightarrow \infty} \frac{d_n}{n} = \rho \in]0, \infty[$ alors, presque sûrement, pour toute fonction continue et bornée $f : \mathbb{R} \rightarrow \mathbb{R}$,*

$$\int f d\mu_n \xrightarrow{n \rightarrow \infty} \int f d\mu_\rho^{\text{MP}}$$

où μ_ρ^{MP} est la loi de Marchenko-Pastur

$$q\delta_0 + \frac{1}{\rho 2\pi x} \sqrt{(b-x)(x-a)} \mathbb{1}_{[a,b]}(x) dx.$$

avec

$$q = \max(0, (1 - \rho^{-1})) \quad \text{et} \quad a = (1 - \sqrt{\rho})^2 \quad \text{et} \quad b = (1 + \sqrt{\rho})^2.$$

De plus, les moments de μ_ρ^{MP} sont donnés pour tout $r \geq 1$ par

$$\int x^r d\mu_\rho^{\text{MP}}(x) = \sum_{k=0}^{r-1} \frac{\rho^k}{k+1} \binom{r}{k} \binom{r-1}{k}.$$

En particulier, la loi μ_ρ^{MP} a pour moyenne 1 et variance ρ .

La loi μ_ρ^{MP} est un mélange entre une masse de Dirac en 0 et une loi à densité par rapport à la mesure de Lebesgue. L'atome en 0 disparaît lorsque $\rho \leq 1$. Le théorème de Marchenko-Pastur affirme que p.s., pour tout $x \in \mathbb{R}$, avec $x \neq 0$ si $\rho > 1$, en notant $I =] - \infty, x]$,

$$\lim_{n \rightarrow \infty} \mu_n(I) = \mu_\rho^{\text{MP}}(I).$$

Lorsque $\rho = 1$, la matrice $\frac{1}{n}YY^\top$ est en quelque sorte asymptotiquement carrée. Dans ce cas, $a = 0$ et $b = 4$. Par changement de variable, p.s. la mesure de comptage du spectre de $\sqrt{\frac{1}{n}YY^\top}$ converge étroitement quand $n \rightarrow \infty$ vers la loi du quart de cercle de densité $x \mapsto \frac{1}{\pi} \sqrt{4-x^2} \mathbb{1}_{[0,2]}(x)$ (une ellipse à cause de la constante de normalisation).

6.6 Notes et commentaires

L'étude des matrices aléatoires constitue un immense champ de recherche dont l'ampleur provient de la rencontre de deux domaines ubiquitaires : l'algèbre linéaire et les probabilités. Les travaux les plus anciens sont dus principalement à Wishart (années 1920) sur les matrices de covariance empiriques des échantillons gaussiens, à von Neumann et Goldstine (années 1940) sur l'analyse numérique matricielle, et à Wigner, Dyson, et Mehta (années 1950-1960) sur les niveaux d'énergie des noyaux atomiques en physique mathématique. Le théorème de Marchenko-Pastur a été obtenu par Marchenko et Pastur dans les années 1960. Le phénomène d'universalité le plus classique de la théorie des probabilités est sans conteste celui mis en lumière par le théorème central limite. Il se trouve que la loi du demi-cercle constitue à bien des égards un analogue, pour les grandes matrices aléatoires, de la loi gaussienne, comme l'explique la brillante *théorie des probabilités libres* développée à l'origine par Voiculescu. Ces dernières années, plusieurs conjectures importantes concernant des modèles de matrices aléatoires ont été résolues.

Le théorème de Wigner a été obtenu par Wigner dans les années 1950 sous une forme légèrement différente. La version la plus aboutie date de la fin des années 1970. La loi μ_σ^{DC} est centrée et de variance σ^2 . Le moment d'ordre 1 de μ_n converge vers le moment d'ordre 1 de μ_σ^{DC} , mais il n'en va pas de même pour le moment d'ordre 2 si $m \neq 0$. Cela est dû au comportement explosif de la plus grande valeur propre λ_1 lorsque $m \neq 0$. Il n'y a pas de contradiction : la convergence faible de la mesure de comptage μ_n n'empêche pas une fraction asymptotiquement négligeable d'atomes d'exploser quand $n \rightarrow \infty$. Cependant, il est possible d'établir que si $m = 0$ et $\mathbb{E}(M_{11}^4) < \infty$ alors p.s. le support de μ_n converge vers le support limite $[-2\sigma, 2\sigma]$, c'est-à-dire que p.s. $\lim_{n \rightarrow \infty} \lambda_n = -2\sigma$ et $\lim_{n \rightarrow \infty} \lambda_1 = 2\sigma$.

La partie combinatoire de la preuve du théorème de Wigner est détaillée par exemple dans le livre de Bai et Silverstein [BS10], et dans le livre de Anderson, Guionnet, et Zeitouni [AGZ10]. Le théorème de Wigner peut également être obtenu en utilisant la transformée de Cauchy-Stieltjes, faisant apparaître la loi du demi-cercle comme solution d'une équation de point fixe. Notons enfin qu'au delà du théorème 6.2.2, il existe une condition suffisante pour que $\mu \in \mathcal{P}$ soit caractérisée par ses moments $(\kappa_n)_{n \geq 1}$, connue sous le nom de condition de Carleman : $\sum_n \kappa_{2n}^{-1/(2n)} = \infty$. Elle est liée à la quasi-analyticité de la transformée de Fourier (une fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ infiniment dérivable est quasi-analytique lorsqu'elle est caractérisée par la suite de ses dérivées à l'origine, parmi l'ensemble des fonctions infiniment dérivables). L'édition de 1966 du volume 2 du livre de Feller mentionne à tort que la condition de Carleman est nécessaire et suffisante à la caractérisation par les moments, erreur corrigée dans l'édition de 1971.

Séance 7

Processus TCP window-size

Mots-clés. Processus de Poisson ; processus de Markov déterministe par morceaux ; générateur infinitésimal ; couplage en Wasserstein ; processus auto-régressifs ; lois exponentielles ; lois géométriques ; processus de branchement à temps continu de Yule.

7.1 Processus TCP window-size en informatique

Le débit instantané maximal sortant d'un ordinateur connecté à un réseau TCP/IP comme Internet est régulé par l'algorithme suivant : le débit maximal est augmenté de manière déterministe d'une unité à chaque pas de temps, et en cas de signal de congestion, il est multiplié par un facteur entre $[0, 1[$, typiquement $1/2$. Ce mécanisme simple permet à la fois une bonne exploitation du réseau et une réaction efficace en cas de congestion.

Nous allons étudier un modèle markovien idéalisé de ce mécanisme. Si on admet que pendant une durée donnée, chacun des autres ordinateurs du réseau provoque une congestion avec une petite probabilité, indépendamment des autres, alors la loi des petits nombres suggère que les signaux de congestion suivent un processus de Poisson. Soit $N = (N_t)_{t \geq 0}$ un processus de Poisson issu de 0 et d'intensité λ comptant les signaux de congestion. On note $(T_n)_{n \geq 0}$ ses temps de saut successifs, avec la convention $T_0 = 0$. Si $E_n = T_n - T_{n-1}$ pour tout $n \geq 1$ alors $T_n = E_1 + \dots + E_n$ et la suite $E = (E_n)_{n \geq 1}$ est constituée de v.a.r. i.i.d. de loi exponentielle de moyenne $1/\lambda$. On a $N_t = \sum_{n \geq 1} \mathbf{1}_{[0, t]}(T_n)$ pour tout $t \in \mathbb{R}_+$. Soit $Q = (Q_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes à valeurs dans $[0, 1[$ de même loi Q . Soit X_0 une variable aléatoire positive. On suppose que X_0 , E , et Q sont indépendants. On définit le processus $X = (X_t)_{t \geq 0}$ à temps continu et à espace d'états \mathbb{R}_+ de la manière suivante :

$$X_t = \begin{cases} X_{T_n} + t - T_n & \text{si } T_n \leq t < T_{n+1}, \\ Q_{n+1}(X_{T_n} + T_{n+1} - T_n) & \text{si } t = T_{n+1}, \end{cases}$$

où $n = N_t \in \mathbb{N}$ est le nombre de sauts avant l'instant t . Les trajectoires de X sont affines de pente 1 par morceaux, continues à droite avec limite à gauche, et chaque saut correspond à une multiplication par un nombre entre $[0, 1[$. Il appartient à la famille des processus de Markov déterministes par morceaux (Piecewise Deterministic Markov Processes ou PDMP) et plus particulièrement à celles des processus à croissance linéaire et décroissance multiplicative (Additive Increase Multiplicative Decrease ou AIMD). Lorsque $Q = \delta_0$ on dit aussi qu'il s'agit d'un processus Growth Collapse. Le processus X est connu sous le nom de processus TCP window-size car en informatique, le débit maximal sortant est réglé par une taille de « fenêtre TCP » (TCP window-size).

Théorème 7.1.1 (Générateur infinitésimal). *Pour tout $f \in \mathcal{C}^1(\mathbb{R}_+, \mathbb{R})$ et $x \geq 0$ on a*

$$L(f)(x) := \lim_{t \rightarrow 0^+} \frac{\mathbb{E}(f(X_t) | X_0 = x) - f(x)}{t} = f'(x) + \lambda \int_0^1 (f(qx) - f(x)) \mathcal{Q}(dq).$$

Le processus X n'explose pas car N n'explose pas. L'espérance $\mathbb{E}(f(X_t) | X_0 = x)$ a un sens car sur $\{X_0 = x\}$, on a $X_t \leq x + t$, et comme f est \mathcal{C}^1 , la v.a.r. $f(X_t)$ est bornée.

Démonstration. Soit $f \in \mathcal{C}^1(\mathbb{R}_+, \mathbb{R})$ et soit $t \leq 1$. On a, par définition de X et N ,

$$\begin{aligned} \mathbb{E}(f(X_t) \mathbf{1}_{\{N_t=0\}} | X_0 = x) &= f(x+t) \mathbb{P}(N_t = 0) = f(x+t) e^{-\lambda t} \\ &= (f(x) + f'(x)t + o(t))(1 - \lambda t + o(t)) \\ &= f(x) + t(f'(x) - \lambda f(x)) + o(t). \end{aligned}$$

Conditionnellement à $\{N_t = 1\}$, la variable T_1 suit la loi uniforme sur $[0, t]$, donc

$$\begin{aligned} \mathbb{E}(f(X_t) | X_0 = x, N_t = 1) &= \mathbb{E}(f(Q_1(x + T_1) + t - T_1) | N_t = 1) \\ &= \int_0^t \frac{ds}{t} \int_0^1 \mathcal{Q}(dq) f(q(x+s) + t-s) \\ &= \int_0^1 \mathcal{Q}(dq) f(qx) + o(1) \end{aligned}$$

où le $o(1)$ (quand $t \rightarrow 0$) provient de la continuité de f , et donc

$$\begin{aligned} \mathbb{E}(f(X_t) \mathbf{1}_{\{N_t=1\}} | X_0 = x) &= \mathbb{P}(N_t = 1) \left(\int_0^1 \mathcal{Q}(dq) f(qx) + o(1) \right) \\ &= \lambda t \int_0^1 \mathcal{Q}(dq) f(qx) + o(t). \end{aligned}$$

Troisièmement, sur $\{X_0 = x\}$ on a $X_t \leq x + t$ et donc comme $t \leq 1$, on a

$$|\mathbb{E}(f(X_t) \mathbf{1}_{\{N_t \geq 2\}} | X_0 = x)| \leq \left(\sup_{u \in [0, x+1]} |f(u)| \right) \mathbb{P}(N_t \geq 2) = \mathcal{O}(t^2).$$

Finalement, la collecte des trois termes donne bien le résultat voulu. \square

Pour tout $t \geq 0$ on considère l'opérateur P_t qui à toute fonction $f : \mathbb{R}_+ \rightarrow \mathbb{R}$ mesurable et bornée associe la fonction mesurable et bornée

$$x \in \mathbb{R}_+ \mapsto P_t f(x) = \mathbb{E}(f(X_t) | X_0 = x).$$

Cette quantité fait sens pour f mesurable et localement bornée (par exemple continue) car conditionnellement à $X_0 = x$, X_t prend ses valeurs dans l'intervalle borné $[x, x+t]$. On a $P_0 = I$ et la propriété de Markov entraîne que $(P_t)_{t \geq 0}$ est un semi-groupe :

$$P_{t+s} = P_{s+t} = P_t \circ P_s.$$

On dit que ce semi-groupe est markovien car il préserve la positivité et l'unité : $P_t(f) \geq 0$ si $f \geq 0$ et $P_t(1) = 1$. On dit qu'une fonction mesurable et bornée $f : \mathbb{R}_+ \rightarrow \mathbb{R}$ appartient au domaine du générateur L du processus lorsque la limite suivante existe pour tout $x \geq 0$:

$$L(f)(x) = \lim_{t \rightarrow 0^+} \frac{P_t(f)(x) - f(x)}{t} = \partial_{t=0} P_t(f)(x).$$

Tout se passe comme si « $P_t = e^{tL}$ ». On peut établir que la propriété de semi-groupe donne, pour tout $t \geq 0$ et toute fonction f de classe \mathcal{C}^1 ,

$$\partial_t P_t(f) = \lim_{\varepsilon \rightarrow 0} P_t \left(\frac{P_\varepsilon f - f}{\varepsilon} \right) = P_t(Lf) \quad \text{et} \quad \partial_t P_t(f) = \lim_{\varepsilon \rightarrow 0} \frac{P_\varepsilon(P_t f) - P_t f}{\varepsilon} = L(P_t f).$$

On dit qu'il s'agit des équations de Chapman-Kolmogorov *progressive* et *rétrograde* respectivement (*forward* et *backward* en anglais).

Le théorème 7.1.1 affirme que le domaine du générateur L contient les fonctions de classe \mathcal{C}^1 . Il se trouve de plus que le générateur conserve les polynômes sans augmenter leur degré. Cela permet de calculer les moments de manière récursive.

Théorème 7.1.2 (Moments et loi invariante). *Pour tout $n \in \mathbb{N}$, tout $x \geq 0$ et tout $t \geq 0$,*

$$\mathbb{E}(X_t^n | X_0 = x) = \frac{n!}{\prod_{k=1}^n \theta_k} + n! \sum_{m=1}^n \left(\sum_{k=0}^m \frac{x^k}{k!} \prod_{\substack{j=k \\ j \neq m}}^n \frac{1}{\theta_j - \theta_m} \right) e^{-\theta_m t}$$

où on a utilisé la notation $\theta_p = \lambda(1 - \mathbb{E}(Q_1^p))$. De plus, conditionnellement à $\{X_0 = x\}$, la variable X_t converge en loi quand $t \rightarrow \infty$ vers la loi μ sur \mathbb{R}_+ dont les moments sont

$$\int x^n \mu(dx) = \frac{n!}{\prod_{k=1}^n \theta_k}$$

pour tout $n \geq 0$. De plus, tous les moments de X_t convergent vers ceux de μ .

Démonstration. Posons $\alpha_n(t) = \mathbb{E}(X_t^n | X_0 = x) = P_t(f_n)(x)$ avec $f_n(x) = x^n$. Pour tout $n \geq 1$, le théorème 7.1.1 donne

$$L f_n(x) = n x^{n-1} + \lambda x^n \int_0^1 (q^n - 1) \mathcal{Q}(dq) = n f_{n-1}(x) - \theta_n f_n(x)$$

et donc, en utilisant l'équation de Chapman-Kolmogorov progressive (forward),

$$\alpha_n'(t) = \partial_t P_t(f_n)(x) = P_t(L f_n)(x) = P_t(n f_{n-1} - \theta_n f_n)(x) = n \alpha_{n-1}(t) - \theta_n \alpha_n(t).$$

Comme α_0 est constante égale à 1, on obtient l'É.D.O. $\alpha_1'(t) = 1 - \theta_1 \alpha_1(t)$ qui donne α_1 . Une récurrence sur n (système triangulaire !) fournit la formule annoncée pour $(\alpha_n)_{n \geq 0}$. Ainsi, conditionnellement à $\{X_0 = x\}$, chaque moment de X_t converge quand $t \rightarrow \infty$ vers une limite finie. Le fait que cette suite de limites soit la suite des moments d'une loi et que X_t converge en loi vers cette loi découle d'une étude similaire à celle menée dans la preuve du théorème de Wigner par la méthode des moments dans le chapitre 6. On prendra garde au fait qu'on se trouve ici sur \mathbb{R}_+ et non pas sur \mathbb{R} tout entier. \square

Remarque 7.1.3 (Growth-collapse). *Si $\mathcal{Q} = \delta_0$ alors le moment d'ordre n de la mesure invariante μ de X vaut $n!/\lambda^n$ ce qui assure que μ est la loi exponentielle de paramètre λ .*

Remarque 7.1.4 (Loi du processus à un instant donné). *Plaçons-nous dans le cas où $\mathcal{Q} = \delta_q$ avec $0 < q < 1$ et conditionnellement à $\{X_0 = x\}$ avec $x \geq 0$. Alors pour tout $t > 0$, il existe une loi absolument continue μ_t portée par l'intervalle $[0, qx + t]$ telle que*

$$\mathcal{L}(X_t) = e^{-\lambda t} \delta_{x+t} + (1 - e^{-\lambda t}) \mu_t$$

et $\text{dist}(x + t, \text{supp}(\mu_t)) = (1 - q)x$. En effet, conditionnellement à l'événement $\{N_t = n\}$,

$$X_t = \begin{cases} x + t & \text{si } n = 0, \\ q^n x + \sum_{k=1}^{n+1} q^{n-k+1} (U_k - U_{k-1}) & \text{si } n > 0, \end{cases}$$

où $0 = U_0 < U_1 < \dots < U_n < U_{n+1} = t$ est une statistique d'ordre uniforme sur $[0, t]$. Ainsi, $\mathcal{L}(X_t | N_t = 0) = \delta_{x+t}$ tandis que $\mathcal{L}(X_t | N_t = n)$ est absolument continue si $n > 0$, portée par $[q^n x, q^n x + t]$. D'où le résultat avec $\mu_t = \mathcal{L}(X_t | N_t > 0)$. Ceci montre que le processus TCP est asymptotiquement régularisant, sans diffusion. C'est le mécanisme de sauts, seul source d'aléa, qui en est responsable.

Remarque 7.1.5 (Irréductibilité). Pour tout $x \geq 0$ et conditionnellement à l'événement $\{X_0 = x\}$, le processus X entre dans tout intervalle en un temps fini avec probabilité positive. En d'autres termes, sa trajectoire coupe toute bande horizontale en un temps fini avec probabilité positive. Cela est immédiat si la bande est au dessus de x . Si la bande est en dessous de x , il suffit de considérer une trajectoire en dents de scie.

Soit $p \geq 1$ un réel fixé. Si (E, d) est un espace métrique et μ une loi sur E , alors la quantité $\int d(x, y)^p \mu(dy)$ est soit infinie pour tout $x \in E$, soit finie pour tout $x \in E$ et on dit dans ce cas que μ a un moment d'ordre p fini. La distance de Wasserstein¹ W_p sur l'ensemble des lois sur E possédant un moment d'ordre p fini est définie par

$$W_p(\mu_1, \mu_2) = \inf_{\substack{(X, Y) \\ X_1 \sim \mu_1 \\ X_2 \sim \mu_2}} \mathbb{E}(d(X, Y)^p)^{1/p} = \left(\inf_{\Pi \in \mathcal{M}(\mu_1, \mu_2)} \int_{E \times E} d(x, y)^p \Pi(dx, dy) \right)^{1/p} \quad (7.1)$$

où $\mathcal{M}(\mu_1, \mu_2)$ désigne l'ensemble convexe des lois de probabilité sur $E \times E$ qui ont pour lois marginales μ_1 et μ_2 . Il est non vide car il contient la loi produit $\mu_1 \otimes \mu_2$. Un élément Π de cet ensemble constitue un *couplage* de μ_1 et μ_2 . Ici on prend $E = \mathbb{R}$ et $d(x, y) = |x - y|$. On peut montrer qu'une suite $(\mu_n)_n$ converge vers μ pour W_p ssi elle converge faiblement pour \mathcal{C}_b et si tous les moments d'ordre $\leq p$ de μ_n convergent vers ceux de μ .

Théorème 7.1.6 (Comportement en temps long par couplage). Soient $X = (X_t)_{t \geq 0}$ et $Y = (Y_t)_{t \geq 0}$ deux processus TCP de même paramètres λ et Q , construits sur un même espace de probabilité. Si leurs lois initiales $\mathcal{L}(X_0)$ et $\mathcal{L}(Y_0)$ possèdent un moment d'ordre $p \geq 1$ fini alors pour tout $t \geq 0$ et en posant $\theta_p := \lambda(1 - \mathbb{E}(Q_1^p))$ on a

$$W_p(\mathcal{L}(X_t), \mathcal{L}(Y_t)) \leq W_p(\mathcal{L}(X_0), \mathcal{L}(Y_0)) \exp\left(-\frac{\theta_p t}{p}\right).$$

Démonstration. Considérons un couple (X, Y) où X et Y partent de x et y mais utilisent le même N et le même Q (mêmes temps de sauts et coefficients multiplicateurs). Remarquablement, la quantité $|X_t - Y_t|$ reste constante entre deux sauts et au k^e saut elle est multipliée par Q_k . Par conséquent, pour tout $t \geq 0$, on a

$$\begin{aligned} \mathbb{E}(|X_t - Y_t|^p) &= \sum_{k=0}^{\infty} \mathbb{E}(|X_t - Y_t|^p \mathbf{1}_{\{N_t=k\}}) \\ &= \sum_{k=0}^{\infty} \mathbb{E}(|x - y|^p Q_1^p \dots Q_k^p \mathbf{1}_{\{N_t=k\}}) \\ &= |x - y|^p \sum_{k=0}^{\infty} \mathbb{E}(Q_1^p)^k \mathbb{P}(N_t = k) \\ &= |x - y|^p e^{-\lambda t(1 - \mathbb{E}(Q_1^p))}. \end{aligned}$$

1. On parle également de distance de couplage, de transport, de Kantorovich, de Mallows, de Fréchet, etc.

Par conséquent, si à présent X et Y sont deux processus TCP de paramètres λ et Q alors pour tout couplage Π de leurs lois initiales $\mathcal{L}(X_0)$ et $\mathcal{L}(Y_0)$, pour tout $t \geq 0$ et tout $p \geq 1$,

$$W_p(\mathcal{L}(X_t), \mathcal{L}(Y_t))^p \leq e^{-\theta_p t} \int_{\mathbb{R}_+ \times \mathbb{R}_+} |x - y|^p \Pi(dx, dy).$$

Il ne reste plus qu'à prendre l'infimum sur tous les couplages Π pour obtenir le résultat. \square

La chaîne incluse $\widehat{X} := (\widehat{X}_n)_{n \geq 0} = (X_{T_n})_{n \geq 0}$ de X est le processus pris aux instants de sauts. Il s'agit d'une chaîne de Markov homogène à temps discret et à espace d'états continu \mathbb{R}_+ . Elle vérifie $\widehat{X}_0 = X_0$ et constitue un processus auto-regressif linéaire :

$$\widehat{X}_{n+1} = Q_{n+1}(\widehat{X}_n + E_{n+1}).$$

Le noyau de transition \widehat{K} de \widehat{X} est donné pour tout $f \in \mathcal{C}_b(\mathbb{R}_+, \mathbb{R})$ et $x \geq 0$ par

$$\widehat{K}(x, f) = \mathbb{E}(f(\widehat{X}_{n+1}) | \widehat{X}_n = x) = \lambda \int_{[0,1] \times \mathbb{R}_+} e^{-\lambda y} f(q(x+y)) \mathcal{Q}(dq) dy.$$

Théorème 7.1.7 (Convergence en loi de la chaîne incluse). *Quelque soit X_0 la chaîne incluse $(\widehat{X}_n)_{n \geq 0}$ converge en loi vers $\widehat{\mu} = \mathcal{L}(\sum_{n=1}^{\infty} Q_1 \cdots Q_n E_n)$.*

Démonstration. En utilisant la formule auto-regressive et les hypothèses d'indépendance et de stationnarité, on obtient, par récurrence sur n , la formule et l'identité en loi

$$\widehat{X}_n = Q_n \cdots Q_1 X_0 + \sum_{k=1}^n Q_n \cdots Q_{n-k+1} E_{n-k+1} \stackrel{d}{=} Q_1 \cdots Q_n X_0 + \sum_{k=1}^n Q_1 \cdots Q_k E_k.$$

On a $\mathbb{E}(Q_1 \cdots Q_n) = \mathbb{E}(Q)^n$. Or $\mathbb{E}(Q) < 1$ car $\mathbb{P}(0 \leq Q < 1) = 1$ et donc par convergence monotone $\mathbb{E}(\sum_n Q_1 \cdots Q_n) = \sum_n \mathbb{E}(Q)^n < \infty$, donc $\mathbb{P}(\sum_n Q_n < \infty) = 1$ et donc $\mathbb{P}(Q_1 \cdots Q_n \rightarrow 0) = 1$. Par conséquent, $Q_1 \cdots Q_n X_0 \rightarrow 0$ p.s. D'autre part, par convergence monotone, $\sum_{k=1}^n Q_1 \cdots Q_k E_k$ converge p.s. vers la série aléatoire $\sum_{n=1}^{\infty} Q_1 \cdots Q_n E_n$, de loi $\widehat{\mu}$. Par conséquent, la chaîne incluse converge en loi vers $\widehat{\mu}$ quelque soit X_0 . \square

La loi $\widehat{\mu}$ a pour moyenne $\lambda^{-1} \mathbb{E}(Q_1) / \mathbb{E}(1 - Q_1)$. Pour la chaîne incluse, l'oubli de la condition initiale apparaît clairement sur la formule suivante :

$$\mathbb{E}(\widehat{X}_n) = \mathbb{E}(Q)^n \mathbb{E}(X_0) + \mathbb{E}(Q_1) \frac{1 - \mathbb{E}(Q_1)^n}{\lambda(1 - \mathbb{E}(Q_1))}.$$

La loi $\widehat{\mu}$ vérifie l'équation en loi $F \stackrel{d}{=} Q_1(F + E_1)$ où $F \sim \widehat{\mu}$ est indépendante de Q_1, E_1 .

7.2 Intensité des sauts variable selon la position

Il est possible de faire dépendre l'intensité de sauts du chemin parcouru par le processus. On considère une fonction localement intégrable $\lambda : x \in \mathbb{R}_+ \mapsto \lambda(x) \in \mathbb{R}_+$ et une suite $(E_n)_{n \geq 1}$ de v.a.r. i.i.d. de loi exponentielle de paramètre 1. On construit la trajectoire du processus $(\widetilde{X}_t)_{t \geq 0}$ partant de $\widetilde{X}_0 = x$ en posant $\widetilde{X}_s = x_0 + s$ pour tout $0 \leq s < \widetilde{T}_1$ où le premier temps de saut \widetilde{T}_1 est défini par l'équation (instant où l' f atteint E_1)

$$\int_0^{\widetilde{T}_1} \lambda(X_s) ds = \int_0^{\widetilde{T}_1} \lambda(x_0 + s) ds = E_1.$$

Ensuite on pose $\tilde{X}_{\tilde{T}_1} = Q_1 \tilde{X}_{\tilde{T}_1^-} = Q_1(x_0 + \tilde{T}_1)$, et plus généralement, on pose

$$\tilde{X}_t = \begin{cases} \tilde{X}_{\tilde{T}_n} + t - \tilde{T}_n & \text{si } \tilde{T}_n \leq t < \tilde{T}_{n+1}, \\ Q_{n+1}(\tilde{X}_{\tilde{T}_n} + \tilde{T}_{n+1} - \tilde{T}_n) & \text{si } t = \tilde{T}_{n+1}, \end{cases}$$

où la suite croissante des temps de saut $(\tilde{T}_n)_{n \geq 0}$ est définie par $\tilde{T}_0 = 0$ et

$$\int_{\tilde{T}_n}^{\tilde{T}_{n+1}} \lambda(\tilde{X}_s) ds = \int_0^{\tilde{T}_{n+1} - \tilde{T}_n} \lambda(\tilde{X}_{\tilde{T}_n} + s) ds = E_{n+1}.$$

Par exemple, le cas spécial $\lambda(x) = x$ donne

$$\tilde{T}_{n+1} - \tilde{T}_n = \sqrt{2E_{n+1} + (\tilde{X}_{\tilde{T}_n})^2} - \tilde{X}_{\tilde{T}_n}.$$

Cette construction fournit la trajectoire $t \mapsto \tilde{X}_t$ sur l'intervalle $[0, \tilde{T}_\infty[$ où $\tilde{T}_\infty := \lim_{n \rightarrow \infty} \tilde{T}_n$ est le *temps d'explosion* du processus. Si λ est une fonction bornée, alors par comparaison au cas constant, on obtient $\mathbb{P}(\tilde{T}_\infty = \infty) = 1$ et on dit que le processus n'explose pas. Lorsque λ n'est pas bornée, par exemple $\lambda(x) = x$, ce sont les propriétés de \mathcal{Q} qui permettent d'établir la non explosion. On peut montrer que le générateur du processus $(\tilde{X}_t)_{t \geq 0}$ est

$$\tilde{L}(f)(x) = f'(x) + \lambda(x) \int_0^1 (f(qx) - f(x)) \mathcal{Q}(dq).$$

Plus généralement, on peut construire les trajectoires d'un processus de générateur

$$\tilde{L}(f)(x) = G(f)(x) + \lambda(x) \int (f(y) - f(x)) K(x, dy)$$

où K est un noyau markovien, où λ est une fonction positive localement intégrable, et où G est le générateur d'un processus quelconque, qui «agit» entre les temps de sauts.

7.3 Branchement et croissance-fragmentation

Le processus TCP permet de modéliser bien plus de choses que la taille de fenêtre TCP du protocole TCP/IP. On peut penser par exemple à la taille ou masse d'une cellule, qui grossit linéairement puis subit une division cellulaire. On peut également penser à un polymère² dont la longueur augmente linéairement au cours du temps puis subit une dislocation. Du point de vue de la masse, la division ou la dislocation constituent une fragmentation qui conserve la masse. Supposons pour simplifier que les fragments obtenus sont toujours au nombre de deux et évoluent selon le même processus, de manière indépendante. Intéressons nous à la taille N_t de la population à l'instant $t \geq 0$. Cela revient à étudier un arbre binaire où la longueur des segments de branches sont i.i.d. de loi exponentielle de paramètre λ . À l'instant $t \geq 0$, l'arbre possède N_t branches, et $N_0 = 1$. Le processus $(N_t)_{t \geq 0}$ est appelé processus de Yule de paramètre λ . Il constitue une sorte de processus de branchement de Galton-Watson à temps continu de loi de reproduction δ_2 .

Théorème 7.3.1 (Taille de la population). *Pour tout $t \geq 0$ la taille de la population N_t suit la loi géométrique sur \mathbb{N}^* de paramètre $e^{-\lambda t}$. En particulier,*

$$\mathbb{E}(N_t) = e^{\lambda t} \quad \text{et} \quad \text{Var}(N_t) = (1 - e^{-\lambda t})e^{2\lambda t}.$$

2. Longue molécule constituée d'une chaîne de molécules plus petites identiques appelées monomères.

Démonstration. La propriété d'absence de mémoire des lois exponentielles permet d'établir que $(N_t)_{t \geq 0}$ est le processus de comptage issu de $N_0 = 1$ de la suite de v.a.r. $(S_n)_{n \geq 1}$ où $S_n = F_1 + \dots + F_n$ avec $(F_n)_{n \geq 1}$ suite de v.a.r. indépendantes et F_n de loi exponentielle de paramètre $n\lambda$ pour tout $n \geq 1$. On a $N_t - 1 = \sum_{n=1}^{\infty} \mathbf{1}_{[0,t]}(S_n)$. Or la v.a.r. S_n a la loi de $\max(E_1, \dots, E_n)$ où E_1, \dots, E_n sont i.i.d. de loi exponentielle de paramètre λ (propriété également utilisée dans la preuve du théorème 9.2.1). Donc, pour tout entier $n \geq 0$,

$$\mathbb{P}(N_t - 1 \geq n) = \mathbb{P}(S_n \leq t) = \mathbb{P}(E_1 \leq t) \cdots \mathbb{P}(E_n \leq t) = (1 - e^{-\lambda t})^n.$$

□

Considérons le cas où chaque fragmentation donne lieu à $k + 1$ fragments où $k \geq 1$ est fixé. On note $(N_t^{(k)})_{t \geq 0}$ le processus de comptage associé. Pour tout $n \geq 2$, le processus issu de n est la somme de n copies i.i.d. du processus issu de 1. C'est la propriété de superposition ou de branchement. On dit que $(N_t^{(k)})_{t \geq 0}$ est un processus de Yule de loi de reproduction δ_{k+1} . Il s'agit d'un processus de Markov homogène à temps continu d'espace d'état \mathbb{N} et de générateur infinitésimal $L^{(k)}$ donné pour tous $f : \mathbb{N} \rightarrow \mathbb{R}$ et $n \in \mathbb{N}$ par

$$L^{(k)}(f)(n) = \lim_{t \rightarrow 0^+} \frac{\mathbb{E}(f(N_t^{(k)}) | N_0^{(k)} = n) - f(n)}{t} = n\lambda(f(n+k) - f(n)).$$

À comparer, au passage, avec le générateur $\lambda(f(n+1) - f(n))$ du processus de Poisson. Si $f(n) = n$ et $\alpha_n(t) = \mathbb{E}(N_t^{(k)} | N_0^{(k)} = n) = P_t^{(k)}(f)(n)$ alors $(L^{(k)}f)(n) = k\lambda n = k\lambda f(n)$ et l'équation de Chapman-Kolmogorov progressive (forward) donne l'É.D.O.

$$\partial_t \alpha_n(t) = \partial_t P_t^{(k)}(f) = P_t(L^{(k)}f) = k\lambda P_t^{(k)}(f) = k\lambda \alpha_n(t).$$

La condition initiale $\alpha_n(0) = n$ donne alors $\mathbb{E}(N_t^{(k)} | N_0^{(k)} = n) = ne^{k\lambda t}$. L'exploitation systématique de cette méthode conduit au résultat suivant.

Théorème 7.3.2 (Processus de Yule). *Supposons que $N_0^{(k)} \sim \delta_1$. Pour tout réel $t \geq 0$ et tout entier $k \geq 1$ si Z_1, \dots, Z_k sont des v.a.r. i.i.d. de même loi que $N_t^{(k)}$ alors $(Z_1 + \dots + Z_k)/k$ suit la loi géométrique sur \mathbb{N}^* de paramètre $e^{-k\lambda t}$. En particulier,*

$$\mathbb{E}(N_t^{(k)}) = e^{k\lambda t} \quad \text{et} \quad \text{Var}(N_t^{(k)}) = k(1 - e^{-k\lambda t})e^{2k\lambda t}.$$

De plus $e^{-k\lambda t} N_t^{(k)}$ converge en loi quand $t \rightarrow \infty$ vers la loi Gamma($1/k, 1/k$) de densité

$$x \mapsto \frac{(1/k)^{1/k}}{\Gamma(1/k)} x^{1/k-1} e^{-x/k} \mathbf{1}_{\mathbb{R}^+}(x).$$

Il est possible d'établir que le processus $(e^{-k\lambda t} N_t^{(k)})_{t \geq 0}$ est une martingale uniformément intégrable, et converge donc p.s. et dans L^1 vers une v.a.r. de loi Gamma($1/k, 1/k$).

Démonstration. On se ramène à $\lambda = 1$ par changement de temps. Fixons à présent $s \in [0, 1]$ et considérons la fonction $f(n) = s^n$. On a $P_t^{(k)}(f)(n) = \mathbb{E}(s^{N_t^{(k)}} | N_0^{(k)} = n)$. On s'intéresse à la fonction génératrice $g_t(s) := \mathbb{E}(s^{N_t^{(k)}} | N_0^{(k)} = 1) = P_t^{(k)}(f)(1)$. La propriété de branchement donne $P_t^{(k)}(f)(n) = (P_t^{(k)}(f)(1))^n = g_t(s)^n$ et donc

$$L^{(k)}(P_t^{(k)}(f))(n) = n(P_t^{(k)}(f)(n+k) - P_t^{(k)}(f)(n)) = n(g_t(s)^{n+k} - g_t(s)^n).$$

L'équation de Chapman-Kolmogorov rétrograde (backward) donne à présent l'É.D.O.

$$\partial_t g_t(s) = \partial_t P_t^{(k)}(f)(1) = L^{(k)}(P_t^{(k)}(f))(1) = g_t(s)^{k+1} - g_t(s).$$

La condition initiale $g_0(s) = s$ donne enfin $g_t(s) = se^{-t}(1 - (1 - e^{-kt})s^k)^{-1/k}$. En particulier, g_t^k est la fonction génératrice de kG où G est une v.a.r. de loi géométrique sur \mathbb{N}^* de paramètre e^{-kt} . Pour la convergence en loi, pour $\theta < 0$, la transformée de Laplace

$$\begin{aligned}\mathbb{E}\left(\exp\left(\theta e^{-kt} N_t^{(k)}\right)\right) &= \exp\left(\theta e^{-kt}\right) e^{-t} \left(1 - \left(1 - e^{-kt}\right) \exp\left(\theta k e^{-kt}\right)\right)^{-1/k} \\ &= \left(e^{kt} \exp\left(-\theta k e^{-kt}\right) - e^{kt} + 1\right)^{-1/k}\end{aligned}$$

converge, lorsque $t \rightarrow \infty$, vers la transformée de Laplace de la loi $\text{Gamma}(1/k, 1/k)$:

$$(1 - k\theta)^{-1/k} = \left(\frac{1/k}{1/k - \theta}\right)^{1/k}.$$

□

Remarque 7.3.3 (Fragmentation). *La modélisation de la fragmentation peut être faite plus généralement en utilisant une loi sur les partitions de $[0, 1]$ avec un nombre possiblement infini de morceaux. Cela mène à la théorie probabiliste de la fragmentation.*

Supposons que le mécanisme de fragmentation en $k + 1$ fragments consiste en une division équitable de la masse. À l'instant $t \geq 0$, la population de fragments est constituée de $N_t^{(k)}$ individus, positionnés sur un arbre $k + 1$ régulier (chaque individu fait $k + 1$ enfants puis meurt). Pour numéroter les individus, on introduit les étiquettes $\mathcal{U} = \cup_{r=0}^{\infty} (\mathbb{N}^*)^r$ avec la convention $(\mathbb{N}^*)^0 = \emptyset$, et où $(\mathbb{N}^*)^r$ sert à numéroter les individus de la génération r . Les individus vivant à un instant t donné peuvent appartenir à différentes générations. La population à l'instant t est un sous-ensemble aléatoire G_t de \mathcal{U} , et vérifie $N_t^{(k)} = |G_t|$. Si $X_{u,t}$ représente la taille de l'individu $u \in G_t$ vivant à l'instant t alors la répartition des masses à l'instant t dans la population est capturée par la mesure de probabilité empirique $\mu_t = (1/N_t^{(k)}) \sum_{u \in G_t} \delta_{X_{u,t}}$. Cette loi aléatoire donne la masse d'un individu choisi uniformément parmi ceux vivants à l'instant t . Pour une fonction test $f : \mathbb{R}_+ \rightarrow \mathbb{R}$, on a $\mathbb{E}\left((1/N_t^{(k)}) \sum_{u \in G_t} f(X_{u,t})\right) = \mathbb{E}(f(X_{U_t,t}))$ où U_t est de loi uniforme sur G_t . Intuitivement, l'évolution de la masse le long d'une branche quelconque de l'arbre suit le processus TCP. En réalité, il faut préciser ce que l'on entend par « branche quelconque » car un phénomène de biais par la taille peut apparaître³. On peut s'attendre par ailleurs à ce que la limite en grande population de la mesure empirique des masses fasse apparaître une équation d'évolution déterministe de la densité des masses, comme l'équation linéaire de transport-fragmentation de Perthame, à comparer avec l'adjoint L^* du processus TCP.

7.4 Notes et commentaires

Pour le processus TCP lui-même [OK08, LvL08, CMP10], pour la version branchante [BDMT10], pour le processus de Yule [Har02, AN04, BG04], pour la fragmentation [Ber06], et pour les ÉDP de transport-fragmentation [Per07].

Les distances de Wasserstein sont au cœur de la théorie du transport optimal de la mesure, présentée dans le livre de Villani [Vil03]. Si $E = \mathbb{R}$ et $d(x, y) = |x - y|$ alors, en notant F_μ^{-1} l'inverse généralisé de la fonction de répartition de μ , on montre que W_p est une distance L^p inter quantiles :

$$W_p(\mu_1, \mu_2) = \left(\int_0^1 |F_{\mu_1}^{-1}(x) - F_{\mu_2}^{-1}(x)|^p dx\right)^{1/p} = \|F_{\mu_1} - F_{\mu_2}\|_p.$$

3. Il s'agit d'un biais d'échantillonnage classique en statistique : un individu tiré au hasard a plus de chance d'être issu d'une famille nombreuse.

Séance 8

Processus ponctuels de Poisson

Mots clés : transformée de Laplace, processus de Poisson et de branchement

Considérons un processus de Poisson simple $(N_t)_{t \geq 0}$ qui compte des tops espacés par des durées aléatoires i.i.d. de loi exponentielle de paramètre λ . Pour tout $I \subset \mathbb{R}$ on note N_I le nombre de tops dans I . La structure des accroissements du processus fait que pour tous intervalles disjoints $I, J \subset \mathbb{R}$, N_I et N_J sont indépendantes et de loi de Poisson de moyenne $\lambda|I|$ et $\lambda|J|$. Conditionnellement à N_I , les sauts du processus dans I sont indépendants et uniformément distribués. Plus généralement, un *processus ponctuel* modélise un jet de points aléatoire sur un espace mesurable E assez quelconque bien au delà de \mathbb{R}_+ .

On dit que E est un *bon espace* s'il est métrique, localement compact, et possède un recouvrement dénombrable par des compacts. C'est le cas de \mathbb{R}^d et des variétés usuelles. Un *processus ponctuel* X sur E est une application X de $(\Omega, \mathcal{F}, \mathbb{P})$ dans l'ensemble des mesures ponctuelles localement finies (mesures de la forme $\sum_x \delta_x$ qui affectent une masse finie à tout compact) telle que $X(A)$ est une v.a.r. pour tout borélien A de E . Un processus ponctuel X sur E est *simple* lorsque $\mathbb{P}(X(\{a\}) \in \{0, 1\}) = 1$ pour tout $a \in E$. Dans ce cas, l'absence d'atomes répétés fait que $X(A)$ est égal au nombre d'atomes de X appartenant à A .

La loi de X est caractérisée par l'une des manières équivalentes suivantes :

1. la loi de $(X(A_1), \dots, X(A_n))$ pour tout n et A_1, \dots, A_n disjoints et dans $\mathcal{B}(E)$
2. la loi de $\int_E f dX$ pour toute fonction $f =: E \rightarrow \mathbb{R}$ étagée¹
3. la loi de $\int_E f dX$ pour toute fonction $f \in \mathcal{C}_c(E, \mathbb{R})$
4. la transformée de Laplace $f \in \text{Mes}(E, \mathbb{R}_+) \mapsto \mathbb{E}(\exp(-\int_E f dX))$.

La *mesure d'intensité* Λ du *processus ponctuel* X sur E est la mesure sur E suivante :

$$\Lambda(A) := \mathbb{E}(X(A)) \quad \text{pour tout } A \in \mathcal{B}(E).$$

Par linéarité, on obtient la formule de Campbell, valable pour toute $f \in \text{Mes}(E, \mathbb{R}_+)$,

$$\mathbb{E}\left(\int_E f(x) dX(x)\right) = \int_E f(x) d\Lambda(x).$$

Notons que Λ n'est pas forcément localement finie : si par exemple T est une v.a.r. dans \mathbb{N} , non intégrable, alors $\sum_{1 \leq n \leq T} \delta_{1/n}$ constitue un processus ponctuel sur \mathbb{R}_+ dont la mesure d'intensité est infinie sur tout compact de la forme $[0, t]$ avec $t > 0$.

Remarque 8.0.1 (Localisation). *Si X est un processus ponctuel sur E de mesure d'intensité Λ et $A \in \mathcal{B}(E)$ alors la restriction X_A de X à A est un processus ponctuel sur E de mesure d'intensité $\Lambda(A \cap \cdot)$ de densité $\mathbf{1}_A$ par rapport à Λ .*

1. C'est-à-dire de la forme $f = \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i}$ avec les α_i réels et les A_i disjoints et dans $\mathcal{B}(E)$.

8.1 Processus ponctuels de Poisson

Un processus ponctuel X sur E est un processus ponctuel de Poisson lorsque (PPP) :

1. la mesure d'intensité Λ est localement finie
2. si $A_1, \dots, A_d \in \mathcal{B}(E)$ sont disjoints alors $X(A_1), \dots, X(A_d)$ sont indépendantes
3. si $A \in \mathcal{B}(E)$ vérifie $\Lambda(A) < \infty$ alors $X(A)$ suit la loi de Poisson $\text{Poi}(\Lambda(A))$.

En particulier, si X est PPP d'intensité Λ alors pour tout compact K , la v.a.r. $X(K)$ est finie presque sûrement et suit la loi de Poisson $\text{Poi}(\Lambda(K))$.

Exemple 8.1.1 (Famille de PPP sur \mathbb{R}_+). Soient $(E_n)_{n \geq 1}$ des v.a.r. i.i.d. de loi exponentielle $\mathcal{E}(\lambda)$, et $T_n := E_1 + \dots + E_n$. Alors $X = \sum_{n=1}^{\infty} \delta_{T_n}$ est un PPP sur $E = \mathbb{R}_+$ et sa mesure d'intensité Λ est un multiple de la mesure de Lebesgue : $\Lambda(A) := \lambda|A|$, $A \in \mathcal{B}(E)$.

Théorème 8.1.2 (Transformée de Laplace et caractérisation). Soit X un processus ponctuel sur E de mesure d'intensité Λ . Alors X est un PPP ssi pour toute fonction mesurable positive $f : E \rightarrow \mathbb{R}_+$, la transformée de Laplace de X en f est donnée par

$$\mathbb{E} \left(\exp \left(- \int_E f dX \right) \right) = \exp \left(- \int_E (1 - e^{-f}) d\Lambda \right).$$

En particulier, la loi d'un processus ponctuel de Poisson est caractérisée par sa mesure d'intensité, et il est simple ssi la mesure d'intensité est diffuse (i.e. n'a pas d'atomes).

Démonstration. Pour l'expression de la transformée de Laplace, il suffit de considérer f étagée : $f = \sum_{i=1}^d \alpha_i \mathbf{1}_{A_i}$ avec A_1, \dots, A_d disjoints et dans $\mathcal{B}(E)$, et $\alpha_1, \dots, \alpha_d$ dans \mathbb{R} . Dans ce cas, on a en vertu de la propriété d'indépendance sur les ensembles disjoints :

$$\mathbb{E} \left(\exp \left(- \int_E f dX \right) \right) = \mathbb{E} \left(\exp \left(- \sum_{i=1}^d \alpha_i X(A_i) \right) \right) = \prod_{i=1}^d \mathbb{E}(\exp(-\alpha_i X(A_i))).$$

Le résultat découle alors du fait que $X(A_i) \sim \text{Poi}(\Lambda(A_i))$.

Pour la caractérisation de la simplicité, on constate tout d'abord que si la mesure Λ admet une masse $\Lambda(\{x\}) > 0$ en $x \in E$, la variable aléatoire $X(\{x\})$ suit la loi de Poisson $\text{Poi}(\Lambda(\{x\}))$ et en particulier, $\mathbb{P}(X(\{x\}) = 2) > 0$, ce qui entraîne que X n'est pas simple. Réciproquement, supposons que Λ est diffuse, et soit K un sous-ensemble compact de E . Comme Λ est diffuse, pour tout $\varepsilon > 0$, il existe une partition finie de K en boréliens A_1, \dots, A_n tel que $\Lambda(A_i) \leq \varepsilon$ pour tout $1 \leq i \leq n$. D'autre part, si N est une variable aléatoire de Poisson de moyenne λ , alors

$$\mathbb{P}(N \geq 2) = e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^k}{k!} = \lambda^2 e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{(k+2)!} \leq \lambda^2.$$

Par conséquent, en utilisant ces deux propriétés,

$$\mathbb{P}(X \text{ pas simple sur } K) \leq \sum_{i=1}^n \mathbb{P}(X(A_n) \geq 2) \leq \sum_{i=1}^n \Lambda(A_n)^2 \leq \varepsilon \sum_{i=1}^n \Lambda(A_n) = \varepsilon \Lambda(K).$$

En faisant $\varepsilon \rightarrow 0$ on obtient $\mathbb{P}(X \text{ pas simple sur } K) = 0$. Ensuite on considère une suite de compacts qui recouvre le bon espace E , et on utilise la continuité supérieure. \square

Théorème 8.1.3 (Caractérisation par l'indépendance). Soit X un processus ponctuel simple sur E de mesure d'intensité Λ diffuse localement finie, tel que pour tous $A_1, \dots, A_d \in \mathcal{B}(E)$ disjoints, $X(A_1), \dots, X(A_d)$ sont indépendantes. Alors X est un PPP simple.

Démonstration. Comme Λ est diffuse, pour tout sous-ensemble compact K de E , et pour tout entier $n > 0$, il existe une partition finie $A_{n,1}, \dots, A_{n,k_n}$ de K en sous-ensembles relativement compacts tels que pour tout $1 \leq i \leq k_n$,

$$\Lambda(A_{n,i}) < \frac{1}{n} \quad \text{et} \quad A_{n,i} \subset \overline{B}(x_{n,i}, r_{n,i}) \quad \text{avec} \quad r_{n,i} < \frac{1}{n}.$$

Cet argument peut être répété pour chacun des $\overline{A_{n,i}}$. On obtient ainsi par récurrence une suite de partitions emboîtées, avec $(A_{n+1,j})_{j \in I_{n,i}}$ partition de $A_{n,i}$ pour tout $1 \leq i \leq k_n$ et avec $I_{n,1}, \dots, I_{n,k_n}$ disjoints. La v.a. $X(K)$ se décompose de la manière suivante :

$$X(K) = \sum_i X(A_{n,i}) = \sum_i \mathbf{1}_{\{X(A_{n,i})=1\}} + \sum_i X(A_{n,i}) \mathbf{1}_{\{X(A_{n,i}) \geq 2\}} = S_n + S'_n.$$

Montrons à présent que S'_n est négligeable quand $n \rightarrow \infty$ grâce au fait que le processus ponctuel X est simple. On a en utilisant la partition de la partition

$$X(A_{n,i}) \mathbf{1}_{\{X(A_{n,i}) \geq 2\}} = \mathbf{1}_{\{X(A_{n,i}) \geq 2\}} \sum_{j \in I_{n,i}} X(A_{n+1,j}) \geq \sum_{j \in I_{n,i}} X(A_{n+1,j}) \mathbf{1}_{\{X(A_{n+1,j}) \geq 2\}}.$$

Cela implique que $(S'_n)_{n \geq 1}$ est décroissante. Sa limite quand $n \rightarrow \infty$ est nécessairement nulle car sinon, avec probabilité non nulle, il existerait une suite (i_n) telle que $A_{n+1, i_{n+1}} \subset A_{n, i_n}$ et $X(A_{n, i_n}) \geq 2$ pour tout n . Comme les A_{n, i_n} sont inclus dans des boules dont le rayon tend vers 0 quand $n \rightarrow \infty$, la limite des A_{n, i_n} est soit vide soit réduite à un point. La masse de chacun de ces points pour X doit être au moins 2, ce qui contredit le fait que X est simple. Il est donc établi que presque sûrement, (S'_n) tend vers 0. Comme S'_n est bornée par la variable aléatoire intégrable $X(K)$, le théorème de convergence dominée entraîne que (S'_n) converge dans L^1 vers 0. Par conséquent,

$$\lim_{n \rightarrow \infty} \mathbb{E}(S_n) = \Lambda(K).$$

Or $\mathbb{E}(\mathbf{1}_{\{X(A_{n,i})=1\}}) \leq \mathbb{E}(X(A_{n,i})) = \Lambda(A_{n,i}) < 1/n$, et donc les variables aléatoires de Bernoulli $\mathbf{1}_{\{X(A_{n,i})=1\}}$ vérifient les hypothèses de la loi des petits nombres. Par conséquent, (S_n) converge en loi vers $\text{Poi}(\Lambda(K))$. Ainsi, $X(K) \sim \text{Poi}(\Lambda(K))$.

Soit à présent un ensemble $A \in \mathcal{B}(E)$ de mesure finie pour Λ . Comme E est un bon espace, il existe une suite croissante (K_n) de sous-ensembles compacts inclus dans A tels que $\Lambda(A) = \lim_{n \rightarrow \infty} \Lambda(K_n)$. Or $X(K_n) \sim \text{Poi}(\Lambda(K_n))$ et donc $X(K_n)$ converge en loi vers $\text{Poi}(\Lambda(A))$. D'autre part, $\mathbb{E}(|X(A) - X(K_n)|) = \mathbb{E}(X(A \setminus K_n))$ converge vers 0, et donc $X(A) - X(K_n)$ converge vers 0 dans L^1 . Il en découle que $X(A)$ suit la loi $\text{Poi}(\Lambda(A))$. \square

Remarque 8.1.4 (Nécessité des hypothèses). *Le théorème ci-dessus est faux si Λ n'est pas diffuse. En effet, le processus ponctuel trivial sur \mathbb{R} constant et égal à δ_0 est simple et vérifie la propriété d'indépendance mais n'est pas un PPP. Le théorème est également faux lorsque Λ est diffuse mais X n'est pas simple. Si par exemple $\sum_n \delta_{x_n}$ est un PPP simple d'intensité $\Lambda/2$ (existence prouvée plus loin) alors le processus ponctuel $\sum_n n\delta_{x_n}$ a une intensité Λ , hérite de la propriété d'indépendance de X , mais n'est pas un PPP!*

Théorème 8.1.5. *Soit X un PPP sur E de mesure d'intensité Λ . Si $A \in \mathcal{B}(E)$ vérifie $\Lambda(A) = \infty$, alors la variable aléatoire $X(A)$ est presque sûrement infinie.*

Démonstration. Comme E est un bon espace, il existe une suite croissante (K_n) de compacts inclus dans A tels que $\lim_n \Lambda(K_n) = \Lambda(A) = \infty$. Pour tout entier n , on a $X(K_n) \sim \text{Poi}(\Lambda(K_n))$ et donc $\mathbb{P}(X(K_n) \leq r)$ converge vers 0 quand $n \rightarrow \infty$ pour tout réel $r \geq 0$. Or $\mathbb{P}(X(A) \leq r) \leq \mathbb{P}(X(K_n) \leq r)$ car $K_n \subset A$. Le résultat désiré s'en déduit. \square

Théorème 8.1.6 (Stabilité par superposition et restriction). *Si (X_n) est une suite finie ou infinie dénombrable de PPP sur E , indépendants, de mesures d'intensité (Λ_n) , et si $\Lambda := \sum_n \Lambda_n$ est localement finie, alors $\sum_n X_n$ est un PPP sur E de mesure d'intensité Λ .*

Si X est un PPP sur E de mesure d'intensité Λ et si A_1, \dots, A_d sont deux à deux disjoints dans $\mathcal{B}(E)$ alors les processus restreints X_{A_1}, \dots, X_{A_d} constituent des PPP indépendants sur E de mesures d'intensité respectives $\Lambda(A_1 \cap \cdot), \dots, \Lambda(A_d \cap \cdot)$.

Démonstration. La première propriété se traite par transformée de Laplace. Pour la seconde propriété, pour tous $B_1 \subset A_1, \dots, B_d \subset A_d$ dans $\mathcal{B}(E)$, les variables aléatoires $X_{A_1}(B_1), \dots, X_{A_d}(B_d)$ s'écrivent également $X(B_1), \dots, X(B_d)$. Elles sont donc indépendantes et de lois $\text{Poi}(\Lambda(B_1)), \dots, \text{Poi}(\Lambda(B_d))$. Ceci montre que X_{A_1}, \dots, X_{A_d} sont des PPP sur E de mesures d'intensité $\Lambda(A_1 \cap \cdot), \dots, \Lambda(A_d \cap \cdot)$. Les v.a.r. $\int_E f_1 dX_{A_1}, \dots, \int_E f_d dX_{A_d}$ sont indépendantes pour toutes fonctions étagées $f_1, \dots, f_d : E \rightarrow \mathbb{R}_+$, et ceci montre que X_{A_1}, \dots, X_{A_d} sont des processus ponctuels indépendants. \square

Théorème 8.1.7 (Théorème d'existence constructif et simulation). *Toute mesure Λ sur E localement finie est la mesure d'intensité d'un processus ponctuel de Poisson sur E .*

Démonstration. Soit $A \in \mathcal{B}(E)$ vérifiant $0 < \Lambda(A) < \infty$. On note Λ_A la loi de probabilité $\Lambda(A \cap \cdot)/\Lambda(A)$. Soit (Z_n) une suite de v.a.r. i.i.d. de loi Λ_A , et N une v.a.r. de loi de Poisson $\text{Poi}(\Lambda(A))$, indépendante de (Z_n) . Posons à présent $X_A := \delta_{Z_1} + \dots + \delta_{Z_N}$, et calculons sa transformée de Laplace pour une fonction mesurable positive $f : E \rightarrow \mathbb{R}_+$:

$$\mathbb{E}\left(e^{-\int_E f dX_A}\right) = \mathbb{E}\left(\mathbb{E}\left(e^{-\int_E f dX_A} \mid N\right)\right) = \mathbb{E}\left(\mathbb{E}\left(e^{-\sum_{i=1}^N f(Z_i)} \mid N\right)\right).$$

Les hypothèses d'indépendance et de loi sur (Z_n) et N permettent donc d'écrire

$$\begin{aligned} \mathbb{E}\left(e^{-\int_E f dX_A}\right) &= \mathbb{E}\left(\left(\int_E e^{-f} d\Lambda_A\right)^N \mid N\right) \\ &= e^{-\Lambda(A)} \sum_{n=0}^{\infty} \frac{(\Lambda(A) \int_E e^{-f} d\Lambda_A)^n}{n!} = e^{-\int_A (1-e^{-f}) d\Lambda}. \end{aligned}$$

Ainsi, X_A est un PPP de mesure d'intensité $\Lambda(A \cap \cdot)$. Comme E est un bon espace et Λ localement finie, on construit une partition au plus dénombrable (A_n) de E avec des éléments de $\mathcal{B}(E)$ tels que $0 < \Lambda(A_n) < \infty$ pour tout n . La méthode précédente permet de construire une suite (X_n) de PPP sur E , indépendants, de mesures d'intensité $(\Lambda(A_n \cap \cdot))$. Par superposition, leur somme est un PPP de mesure d'intensité $\sum_n \Lambda(A_n \cap \cdot) = \Lambda$. \square

Corollaire 8.1.8 (Structure par localisation). *Si X est un PPP sur E de mesure d'intensité Λ et si $A \in \mathcal{B}(E)$ tel que $0 < \Lambda(A) < \infty$ alors pour tout $n \geq 1$ et conditionnellement à $\{X(A) = n\}$, les atomes de X_A ont la loi de n v.a.r. i.i.d. de loi $\Lambda(A \cap \cdot)/\Lambda(A)$. Si A_1, \dots, A_d dans $\mathcal{B}(E)$ sont deux à deux disjoints avec une mesure finie pour Λ alors*

$$(X(A_1), \dots, X(A_d)) \sim \text{Poi}(\Lambda(A_1)) \otimes \dots \otimes \text{Poi}(\Lambda(A_d)).$$

En particulier, pour tout entier n on a, en posant $p_i = \Lambda(A_i)/\Lambda(A_1 \cup \dots \cup A_d)$,

$$\mathcal{L}((X(A_1), \dots, X(A_d)) \mid X(A_1) + \dots + X(A_d) = n) = \text{Multinom}(n, (p_1, \dots, p_d)).$$

Démonstration. L'usage de la transformée de Laplace montre que X_A est un PPP de mesure d'intensité $\Lambda(A \cap \cdot)$. D'après la preuve du théorème précédent, ce processus a la même loi que le processus $\delta_{Z_1} + \dots + \delta_{Z_N}$ où (Z_n) est une suite de v.a.r. i.i.d. de loi $\Lambda(A \cap \cdot)/\Lambda$ et où

N est une v.a.r. de Poisson de moyenne $\Lambda(A)$ indépendante de (Z_n) . Par conséquent, la loi de X_A sachant $\{X(A) = n\}$ est égale à la loi de $\delta_{Z_1} + \dots + \delta_{Z_N}$ sachant $\{N = n\}$, qui est exactement la loi produit $(\Lambda(A \cap \cdot) / \Lambda(A))^{\otimes n}$. \square

Remarque 8.1.9 (Mesures d'intensité de masse 1). Si X est un PPP sur E dont la mesure d'intensité Λ est une loi de probabilité alors les atomes de X sont des réalisations i.i.d. de loi Λ . Réciproquement, si Λ est une loi de probabilité, et si X_1, \dots, X_n sont des v.a.r. i.i.d. de loi Λ , alors la mesure ponctuelle aléatoire $\sum_{k=1}^n \delta_{X_k}$ suit la même loi qu'un processus ponctuel de Poisson sur E de mesure d'intensité Λ conditionnellement à $\{X(E) = n\}$. La localisation permet de se ramener à ce cas lorsque la mesure d'intensité n'est pas finie.

Théorème 8.1.10 (Stabilité par transformation). Si X est un PPP sur E de mesure d'intensité Λ et $T : E \rightarrow F$ une application mesurable vers un bon espace F telle que la mesure image $\Lambda \circ T^{-1}$ de Λ par T est localement finie, alors la mesure ponctuelle aléatoire $Y := T(X) := \sum_{x \in X} \delta_{T(x)}$ est un PPP sur F de mesure d'intensité $\Lambda \circ T^{-1}$.

Démonstration. La transformée de Laplace de Y s'écrit, pour $f \in \text{Mes}(F, \mathbb{R}_+)$,

$$\mathbb{E} \left(\exp \left(- \int_F f dY \right) \right) = \mathbb{E} \left(\exp \left(- \int_E (f \circ T) dX \right) \right) = \exp \left(- \int_E (1 - e^{-f \circ T}) d\Lambda \right).$$

Remarque : si T n'est pas injective alors Y n'est pas simple en général. \square

Remarque 8.1.11 (Homogénéité et isotropie). Un processus ponctuel X sur \mathbb{R}^d est stationnaire ou homogène lorsque X et $X + t$ ont même loi pour tout $t \in \mathbb{R}^d$ et isotrope lorsque X et RX ont même loi pour tout $R \in \text{SO}(\mathbb{R}^n)$. Un PPP sur \mathbb{R}^d est homogène ssi sa mesure d'intensité est multiple de la mesure de Lebesgue. Il est alors de plus isotrope.

Théorème 8.1.12 (Stabilité par marquage dépendant et amincissement). Soit X un PPP sur E de mesure d'intensité Λ . Conditionnellement à X , on associe à chaque atome x de X une v.a.r. Y_x à valeurs dans un bon espace F , de sorte que ces v.a.r. soient conditionnellement indépendantes sachant X . Dans ce cas, en notant ν_x la loi de Y_x :

1. la mesure ponctuelle aléatoire $Z := \sum_{x \in X} \delta_{(x, Y_x)}$ est un PPP sur $E \times F$ de mesure d'intensité Θ donnée pour toute fonction $f : E \times F \rightarrow \mathbb{R}$ mesurable positive par

$$\int_{E \times F} f(x, y) d\Theta(x, y) := \int_E \left(\int_F f(x, y) d\nu_x(y) \right) d\Lambda(x)$$

(en particulier, si $\nu_x = \nu$ alors $\Theta = \Lambda \otimes \nu$)

2. pour tous A et B disjoints dans $\mathcal{B}(F)$, les mesures ponctuelles aléatoires

$$R_A := \sum_{x \in X; Y_x \in A} \delta_x \quad \text{et} \quad R_B := \sum_{x \in X; Y_x \in B} \delta_x$$

sont des PPP sur E , indépendants, dont les mesures d'intensité sont absolument continues par rapport à Λ , de densité $x \mapsto \nu_x(A) = \mathbb{P}(Y_x \in A)$ et $x \mapsto \nu_x(B) = \mathbb{P}(Y_x \in B)$. En particulier, si $\nu_x = \nu$ alors on trouve les mesures $\nu(A)\Lambda$ et $\nu(B)\Lambda$.

Démonstration. La première propriété s'obtient en calculant la transformée de Laplace de

Z . Z pour une fonction $f : E \times F \rightarrow \mathbb{R}$ mesurable positive quelconque :

$$\begin{aligned} \mathbb{E}\left(\exp\left(-\int_{E \times F} f dZ\right)\right) &= \mathbb{E}\left(\mathbb{E}\left(\exp\left(-\int_{E \times F} f dZ\right) \middle| X\right)\right) \\ &= \mathbb{E}\left(\mathbb{E}\left(\exp\left(-\sum_{x \in X} f(x, Y_x)\right) \middle| X\right)\right) \\ &= \mathbb{E}\left(\mathbb{E}\left(\prod_{x \in X} e^{-f(x, Y_x)} \middle| X\right)\right) \\ &= \mathbb{E}\left(\prod_{x \in X} \mathbb{E}\left(e^{-f(x, Y_x)}\right)\right) \\ &= \mathbb{E}\left(\exp\left(\sum_{x \in X} g(x)\right)\right) \end{aligned}$$

où $g(x) := \log\left(\mathbb{E}\left(e^{-f(x, Y_x)}\right)\right) = \log \int_F e^{-f(x, y)} d\nu_x(y)$. Comme X est un PPP,

$$\mathbb{E}\left(\exp\left(\sum_{x \in X} g(x)\right)\right) = \exp\left(-\int_E (1 - e^{g(x)}) d\Lambda(x)\right).$$

Par conséquent, on obtient

$$\begin{aligned} \mathbb{E}\left(\exp\left(-\int_{E \times F} f dZ\right)\right) &= \exp\left(-\int_E \left(1 - \int_F e^{-f(x, y)} d\nu_x(y)\right) d\Lambda(x)\right) \\ &= \exp\left(-\int_E \left(\int_F (1 - e^{-f(x, y)}) d\nu_x(y)\right) d\Lambda(x)\right) \\ &= \exp\left(-\int_E \left(\int_F (1 - e^{-f(x, y)}) d\Theta(x, y)\right)\right). \end{aligned}$$

Ceci montre que Z est bien un PPP sur $E \times F$ de mesure d'intensité Θ . La mesure Θ est localement finie car Λ l'est et ν_x est une probabilité pour tout x .

Pour la seconde propriété, on constate tout d'abord, en vertu de la propriété de stabilité par restriction, que les restrictions du processus Z aux ensembles *disjoints* $E \times A$ et $E \times B$ constituent des PPP sur $E \times F$, *indépendants*, de mesures d'intensité $\mathbf{1}_{E \times A} \Theta$ et $\mathbf{1}_{E \times B} \Theta$. Maintenant, en vertu de la stabilité par transformation, leurs images R_A et R_B par la projection $\pi : (x, y) \in E \times F \mapsto x \in E$ constituent des PPP sur E , indépendants, dont les mesures d'intensité sont données par les mesures image de $\mathbf{1}_{E \times A} \Theta$ et $\mathbf{1}_{E \times B} \Theta$ par la transformation π , qui ne sont rien d'autre que les mesures absolument continues par rapport à Λ de densités respectives $x \mapsto \nu_x(A)$ et $x \mapsto \nu_x(B)$ (lois marginales !). \square

Exemple 8.1.13 (File d'attente M/M/ ∞). Soit $\lambda, \mu > 0$ et considérons le théorème 8.1.12 avec $E = F = \mathbb{R}_+$ et $d\Lambda(x) = \lambda \mathbf{1}_{\mathbb{R}_+}(x) dx$ où dx désigne la mesure de Lebesgue sur \mathbb{R} , et $\nu = \mathcal{E}(\mu)$. Les atomes de X correspondent à des tops sur \mathbb{R}_+ espacés par des durées aléatoires i.i.d. de loi $\mathcal{E}(\lambda)$. La v.a.r. Y_x correspond à la durée de service de l'atome x . La quantité $x + Y_x$ correspond au temps de fin de service de l'atome x . En vertu du théorème 8.1.12, la mesure ponctuelle aléatoire $Z := \sum_{x \in X} \delta_{(x, Y_x)}$ constitue un PPP d'intensité $\Lambda \otimes \nu$. Soit à présent $t \in \mathbb{R}_+$ et $A_t := \{(u, v) \in \mathbb{R}_+^2; (u, u + v) \in [0, t] \times [t, \infty[\}$. La v.a.r. $Z(A_t)$ représente le nombre d'atomes en cours de service au temps t , c'est-à-dire le nombre de clients en cours de service à l'instant t dans une file d'attente M/M/ ∞ issue de 0 d'intensités

d'arrivée λ et de service μ . On a $Z(A_t) \sim \text{Poi}((\Lambda \otimes \nu)(A_t))$ avec

$$(\Lambda \otimes \nu)(A_t) = \int_{0 \leq u \leq t} \int_{u+v \geq t} \lambda \mathbf{1}_{\mathbb{R}_+}(u) \frac{1}{\mu} e^{-\mu v} \mathbf{1}_{\mathbb{R}_+}(v) dudv = \rho(1 - e^{-\mu t}),$$

avec $\rho := \lambda/\mu$. On retrouve la loi $\text{Poi}(\rho(1 - e^{-\mu}))$ au temps t de la $M/M/\infty$ issue de 0.

8.2 Processus ponctuels de Poisson sur \mathbb{R}_+

Théorème 8.2.1 (Intensité non homogène). Soit $\lambda : \mathbb{R}_+ \rightarrow \mathbb{R}_+^*$ une fonction localement Lebesgue intégrable et strictement positive, et F^{-1} l'inverse de la fonction strictement croissante $F : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ définie par

$$F(t) := \int_0^t \lambda(u) du.$$

Soit $(E_n)_{n \geq 1}$ des v.a.r. i.i.d. de loi exponentielle de moyenne 1 et $T_n := E_1 + \dots + E_n$. Alors la mesure ponctuelle aléatoire $\sum_{n \geq 1} \delta_{F^{-1}(T_n)}$ est un PPP sur \mathbb{R}_+ dont la mesure d'intensité est l'image par F^{-1} de la mesure de Lebesgue sur \mathbb{R}_+ . Elle est absolument continue par rapport à la mesure de Lebesgue sur \mathbb{R}_+ , de densité de Radon-Nikodym λ .

Démonstration. On sait que $\sum_n \delta_{T_n}$ est un PPP sur \mathbb{R}_+ de mesure d'intensité Lebesgue. La stabilité par transformation des PPP mène alors au résultat. D'autre part, on a

$$(F^{-1})' = \frac{1}{\lambda \circ F^{-1}}.$$

□

On a $\{F^{-1}(T_n) > t\} = \{T_n > F(t)\} = \{T_n > \int_0^t \lambda(u) du\}$. Le processus de comptage de $(F^{-1}(T_n))_{n \geq 1}$, associé à la mesure ponctuelle aléatoire $\sum_{n \geq 1} \delta_{F^{-1}(T_n)}$, constitue un processus de Markov sur \mathbb{N} , inhomogène lorsque λ n'est pas constante. Si $(N_t)_{t \geq 0}$ désigne le processus de comptage de $(T_n)_{n \geq 1}$, alors $(N_{F(t)})_{t \geq 0}$ est le processus de comptage de $(F^{-1}(T_n))_{n \geq 1}$. En quelque sorte, le processus inhomogène $(N_{F(t)})_{t \geq 0}$ s'obtient à partir du processus homogène en distordant le temps par la fonction d'intensité λ .

Lorsque λ est constante, alors $F(t) = \lambda t$ et $F^{-1}(t) = t/\lambda$. Par conséquent, dans ce cas, $F^{-1}(T_n) = T_n/\lambda = E_1/\lambda + \dots + E_n/\lambda$, et on retrouve le processus de Poisson d'intensité λ associé au comptage de tops espacés par des durées i.i.d. de loi $\mathcal{E}(\lambda)$.

8.3 Processus auto-excités

Les processus ponctuels auto-excités mêlent mécanisme ponctuel spatial et de branche-ment. Ils permettent de modéliser les répliques de tremblement de terres ou de gènes sur l'ADN, les amas de galaxies, etc (épidémiologie, fiabilité, ...). Soit X_0 un PPP de mesure d'intensité Λ_0 sur un groupe E . Soit Λ une mesure positive sur E vérifiant² $\Lambda(E) < 1$. Conditionnellement à X_0 , chaque atome x de X_0 fait, indépendamment des autres, des enfants avec un PPP de mesure d'intensité $\delta_x * \Lambda$, et donc la génération 1 est un PPP de mesure d'intensité $X_1 = \sum_{x \in X_0} \delta_x * \Lambda$. Plus généralement, conditionnellement à X_0, \dots, X_n , la génération $n+1$ est un PPP de mesure d'intensité $\sum_{x \in X_n} \delta_x * \Lambda$. Le processus *auto-excité* est la superposition des générations : $X := \sum_{n \geq 0} X_n$ (il serait possible d'incorporer un

2. Par exemple $\Lambda = mL$ car x fait Z enfants (avec $m = \mathbb{E}(Z) < 1$) en des positions i.i.d. de loi L sur E .

mécanisme d'immigration à chaque génération). Conditionnellement à X_0, \dots, X_n , la v.a.r. $X_{n+1}(E)$ suit la loi de Poisson de moyenne

$$\sum_{x \in X_n} (\delta_x * \Lambda)(E) = \Lambda(E)X_n(E).$$

Il en découle que $\mathbb{E}(X_n(E)) = \Lambda(E)^n \Lambda_0(E)$ pour tout $n \geq 0$. Comme $\Lambda(E) < 1$ on obtient

$$\mathbb{E}(X) = \Lambda_0(E) \sum_{n \geq 0} \Lambda(E)^n = \frac{\Lambda_0(E)}{1 - \Lambda(E)}.$$

Notons que $(X_n(E))_{n \geq 0}$ est un Galton-Watson sous-critique de loi de reproduction $\text{Poi}(\Lambda(E))$ issu de $X_0(E)$ (bouquet d'arbres de G.-W.). Calculons la transformée de Laplace de X :

$$\mathbb{E} \left(\exp \left(- \int_E f dX_{n+1} \right) \mid X_n \right) = \exp \left(- \sum_{x \in X_n} \int_E (1 - e^{f(x+\cdot)}) d\Lambda \right) = \exp \left(- \int_E g dX_n \right)$$

où $g(x) = \int_E (1 - e^{f(x+y)}) d\Lambda(y)$. Si X_n était un PPP de mesure d'intensité Λ_n alors

$$\mathbb{E} \left(\exp \left(- \int_E f dX_{n+1} \right) \right) = \exp \left(- \int_E (1 - e^g) d\Lambda_n \right)$$

qui n'est pas $\exp \left(- \int_E (1 - e^f) d\Lambda_{n+1} \right)$ pour une mesure Λ_{n+1} , et X n'est donc pas PPP.

Notes et commentaires

Ce chapitre est inspiré principalement des premières pages du livre de Robert [Rob00] sur les réseaux stochastiques et les files d'attente, qui contient également une mise en œuvre du concept de mesure de Palm, ignoré ici par souci de simplicité. De nombreux ouvrages abordent le concept de processus ponctuel, comme par exemple les livres [Kin93] et [Jac06]. Les processus auto-excités, été étudiés par Hawkes et Oakes à la fin des années 1960 [HO74] – voire également le livre [DVJ03] – permettent de modéliser un grand nombre de phénomènes comme la position des galaxies, les répliques des tremblements de terre, les copies d'un gène du code génétique, etc. La modélisation des répartitions de points conduit également à d'autres généralisations, comme par exemple les processus de Cox, mélanges de processus ponctuels obtenus en rendant aléatoire la mesure d'intensité.

Les atomes des PPP sont indépendants à numérotation près. Au delà de l'indépendance, il est possible de définir des processus ponctuels dont les atomes sont dépendants. Plus précisément, on dit qu'un processus ponctuel μ sur \mathbb{R} , à n atomes (n est fixé ici), est *déterminantal* de noyau $K : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ lorsque ses atomes ont pour densité (à numérotation près) $x \in \mathbb{R}^n \mapsto \det((K(x_i, x_j))_{1 \leq i, j \leq n})$. Ces processus ont été étudiés dans les années 1970 par Macchi [Mac75], voire également le livre [SKM87] et l'article de synthèse [HKPV06]. Un exemple classique de processus déterminantal est donné par $K(x, y) = \sum_{k=0}^{n-1} \varphi_k(x) \varphi_k(y)$ où $\varphi_k(u) = \exp(-\frac{1}{2}u^2) H_k(u)$ et où $(H_k)_{k \geq 1}$ sont les polynômes de Hermite, orthogonaux pour la loi gaussienne $\mathcal{N}(0, 1)$. Dans ce cas, un calcul permet d'établir que la densité des atomes du processus déterminantal est proportionnelle à $\exp(-x_1^2 - \dots - x_n^2) \prod_{1 \leq i < j \leq n} (x_i - x_j)^2$. Il s'avère que cette densité est exactement celle des valeurs propres de matrices hermitiennes aléatoires de densité gaussienne proportionnelle à $H \mapsto \exp(-\frac{1}{2} \text{Tr}(H^2))$ (ont dit GUE pour Gaussian Unitary Ensemble). Le déterminant de Vandermonde exprime une répulsion entre les variables. Le remplacement du déterminant par le permanent fournit un modèle à attraction plutôt qu'à répulsion.

Séance 9

Généalogies et coalescence

Mots-clés. Arbres ; lois géométriques ; lois exponentielles.

On s'intéresse à l'évolution d'une population de taille finie N constante au fil des générations. On suppose que les générations ne se chevauchent pas, et que pour tout $n \geq 0$, la génération n meurt en donnant naissance à la génération $n + 1$. Le mécanisme de transition est markovien : la génération $n + 1$ dépend de la génération n et d'une source d'aléa indépendante. À chaque génération, on numérote les individus de 1 à N . Pour tout $n \geq 0$, soit a_i^{n+1} le numéro du parent de l'individu i de la génération $n + 1$. Ce parent appartient à la génération n . On suppose que les variables aléatoires $a_1^{n+1}, \dots, a_N^{n+1}$ sont i.i.d. de loi uniforme sur $\{1, \dots, N\}$. On note ν_i^n le nombre de descendants à la génération $n + 1$ de l'individu i vivant à la génération n . Les variables aléatoires $(\nu_i^n)_{1 \leq i \leq N}$ ne sont pas indépendantes puisqu'elles doivent vérifier la relation $\nu_1^n + \dots + \nu_N^n = N$. On a

$$\nu_i^n = \sum_{j=1}^N \mathbf{1}_{\{a_j^{n+1}=i\}}.$$

Par conséquent, le vecteur aléatoire $\nu^n = (\nu_1^n, \dots, \nu_N^n)$ suit la loi multinomiale de taille N et de paramètre $(1/N, \dots, 1/N)$, c'est-à-dire

$$\mathbb{P}(\nu_1^n = n_1, \dots, \nu_N^n = n_N) = \frac{N!}{n_1! \dots n_N!} \left(\frac{1}{N}\right)^N \mathbf{1}_{\{n_1 + \dots + n_N = N\}}.$$

Si chaque individu est identifié à un allèle A ou B d'un gène à deux allèles et si le nombre d'allèles A à la génération n est égal à x , alors la loi du nombre d'allèles A à la génération $n + 1$ suivra une loi binomiale $\text{Bin}(N, x/N)$. On retrouve donc le modèle de Fisher-Wright (le lien avec la loi multinomiale a déjà été observé au moment de l'étude du modèle de Cannings). Nous avons construit ici le modèle de Fisher-Wright **à rebours en remontant les générations**. Chaque individu de la génération $n + 1$ possède un ancêtre à la génération n choisi indépendamment des autres et selon la loi uniforme sur $\{1, \dots, N\}$. Certains membres de la génération n n'ont donc aucun descendant à la génération $n + 1$ et c'est ce qui va entraîner le phénomène dit de dérive génétique (fixation d'un allèle).

Remarque 9.0.1 (Des poissons à la multinomiale). Si Y_1, \dots, Y_N sont des variables aléatoires indépendantes de loi de Poisson de paramètres respectifs $\theta_1, \dots, \theta_N$ alors

$$\mathcal{L}((Y_1, \dots, Y_N) | Y_1 + \dots + Y_N = N) = \text{Mul}(N, (p_1, \dots, p_N))$$

où $p_k = \theta_k / (\theta_1 + \dots + \theta_N)$ pour tout $1 \leq k \leq N$. En particulier, si $\theta_1 = \dots = \theta_N$ alors $p_1 = \dots = p_N = 1/N$. Ainsi le modèle de Fisher-Wright est un modèle de Galton-Watson de loi de reproduction Poisson conditionné à avoir une taille de population fixée (chapitre 4).

Le modèle de Wright-Fisher sans mutation est connu pour faire apparaître un phénomène dérive génétique, c'est-à-dire qu'un allèle l'emporte sur l'autre et que la population devient homozygote après un nombre aléatoire mais fini (presque sûrement) de générations. De même, après un certain nombre de générations, tous les individus proviendront donc d'un même ancêtre. C'est ce phénomène qu'il s'agit à présent de décrire et quantifier.

Comme les individus choisissent leur parent de manière i.i.d. et uniforme, la probabilité que deux individus fixé aient deux parents distincts (à la génération précédente) est égale à $N(N-1)/N^2 = 1 - 1/N$. Ces parents choisissant eux-mêmes leurs parents indépendamment, la probabilité pour que le premier ancêtre commun à deux individus donnés remonte à plus de r générations vaut $(1 - 1/N)^r$. Le temps T_2 d'apparition de l'**Ancêtre Commun le Plus Récent** (ACPR) suit donc une loi géométrique sur \mathbb{N}^* de paramètre $p_2 = 1/N$. Si l'on considère à présent trois individus, notons T'_3 l'instant d'apparition de l'ACPR pour deux individus au moins (premier temps de coalescence parmi trois individus). On a

$$\mathbb{P}(T'_3 > 1) = \frac{N(N-1)(N-2)}{N^3} = 1 - \frac{3}{N} + \frac{2}{N^2}.$$

Comme les choix des parents sont indépendants à chaque génération, il vient, pour $k \geq 1$,

$$\mathbb{P}(T'_3 > k) = \left(1 - \frac{3}{N} + \frac{2}{N^2}\right)^k.$$

La loi de T'_3 est donc la loi géométrique de paramètre $p_3 = 3/N - 2/N^2$. On note $T_3 = T'_3 + T'_2$ le temps d'apparition d'un ACPR pour les trois individus (ici $T'_2 := T_3 - T'_3$). Remarquons que si l'ACPR de deux individus est l'ACPR des trois individus alors $T'_2 = 0$. Déterminons la loi de T'_2 . La probabilité que trois individus distincts aient le même père est $1/N^2$. La probabilité que T'_2 soit nul sachant que $T'_3 = k$ est donc égale à

$$\mathbb{P}(T'_2 = 0 | T'_3 = k) = \frac{\mathbb{P}(T'_2 = 0, T'_3 = k)}{\mathbb{P}(T'_3 = k)} = \frac{(1 - p_3)^{k-1} (1/N^2)}{p_3 (1 - p_3)^{k-1}} = \frac{1}{3N - 2}.$$

Comme les choix des parents sont indépendants d'une génération à l'autre, on déduit de la première partie que la loi de T'_2 sachant que $T'_3 = k$ et $T'_2 > 0$ est la loi géométrique de paramètre p_2 . Cela implique que sachant $T'_3 = k$, la variable T'_2 est égale en loi à UV où U et V sont indépendantes de lois respectives $\text{Bern}(1 - 1/(3N - 2))$ et $\text{Geom}_{\mathbb{N}^*}(p_2)$.

Plus généralement, le nombre de parents distincts d'un groupe de k individus peut être vu comme le nombre d'urnes occupées après que l'on a lancé k balles, indépendamment et à chaque fois uniformément, dans N urnes. Pour tout $j \in \{1, \dots, k\}$, on notera la probabilité pour que ce nombre soit j de la manière suivante :

$$g_{k,j}^{(N)} = \mathbb{P}(k \text{ individus ont } j \text{ parents distincts}) = \frac{N(N-1) \cdots (N-j+1) S_{k,j}}{N^k},$$

où $S_{k,j}$ est le nombre de Stirling de seconde espèce¹. En effet, il y a $N(N-1) \cdots (N-j+1)$ façons de choisir j parents distincts parmi N , et $S_{k,j}$ façons d'associer à ces j parents k enfants, et enfin, N^k est le nombre de façons d'assigner k enfants à leurs parents. On définit à présent le processus ancestral $(A_n^N(r))_{r \in \mathbb{N}}$ en notant $A_n^N(r)$ le nombre d'ancêtres distincts à la génération $-r$ pour un groupe de taille n au temps 0 (le temps remonte ici). La suite $(A_n^N(r))_{r \in \mathbb{N}}$ est une chaîne de Markov sur $\{1, \dots, n\}$ de matrice de transition

$$G_N = \left(g_{j,k}^{(N)} \mathbf{1}_{\{j \leq k\}} \right)_{1 \leq j, k \leq n}.$$

1. C'est le nombre de façons de découper un ensemble à k éléments en j ensembles non vides, déjà utilisé pour étudier le collectionneur de coupons par exemple

Pour cette chaîne, l'état 1 est absorbant tandis que les états $2, \dots, n$ sont transitoires puisqu'ils mènent tous à 1. Il est difficile d'étudier les propriétés fines de cette chaîne, comme par exemple des propriétés sur le temps d'atteinte de l'état 1. Nous allons donc remplacer ce modèle à temps discret par un modèle plus simple à temps continu.

9.1 Modèle généalogique à temps continu

Considérons la chaîne de Markov à temps continu $(A(t))_{t \geq 0}$ sur \mathbb{N}^* de générateur

$$L(j, k) = \begin{cases} \binom{j}{2} & \text{si } j \geq 2 \text{ et } k = j - 1, \\ -\binom{j}{2} & \text{si } j \geq 2 \text{ et } k = j, \\ 0 & \text{sinon.} \end{cases}$$

Il s'agit d'un processus de mort pur sur \mathbb{N} (sauts de n à $n - 1$ seulement) pour lequel 1 est absorbant. Le temps de séjour en n suit la loi exponentielle de paramètre $\binom{n}{2}$. Ceci s'interprète de la manière suivante : chaque couple d'individus se cherche un père indépendamment des autres et y arrive au bout d'un temps exponentiel de paramètre 1. Pour n individus, on a $\binom{n}{2}$ couples distincts. Or le minimum de $\binom{n}{2}$ v.a.r. i.i.d. de loi $\mathcal{E}(1)$ est la loi $\mathcal{E}(\binom{n}{2})$. On note $(A_n(t))_{t \geq 0}$ le processus issu de n , qui est à valeurs dans $\{1, \dots, n\}$.

Montrons à présent que le processus à temps discret renormalisé converge vers le processus à temps continu. On s'intéresse au cas limite où N est grand et où l'unité de temps se compte en N générations. La date d'apparition de l'ACPR de deux individus donnés est donc T'_2/N où T'_2 suit la loi géométrique sur \mathbb{N}^* de paramètre $p_2 = 1/N$.

Lemme 9.1.1 (Loi exponentielle comme limite de lois géométriques renormalisées). *Soit $(V_n)_n$ une suite de variables aléatoires de loi géométrique de paramètres respectifs $(\mu_n)_n$ telle que $n\mu_n$ converge vers $\mu > 0$ quand $n \rightarrow \infty$. Alors la suite $(n^{-1}V_n)_n$ converge en loi vers une variable aléatoire V de loi exponentielle de paramètre μ .*

Démonstration. Il suffit d'utiliser les fonctions caractéristiques. Si $V \sim \mathcal{E}(\mu)$, alors

$$\varphi_{V_n/n}(t) = \frac{\mu_n e^{it/n}}{1 - (1 - \mu_n)e^{it/n}} = \frac{n\mu_n}{n\mu_n - n(1 - e^{it/n})} \xrightarrow{n \rightarrow \infty} \frac{\mu}{\mu - it} = \varphi_V(t),$$

Alternativement, pour tout $x \geq 0$, $\mathbb{P}(n^{-1}V_n \geq x) = \mathbb{P}(V_n \geq nx) = e^{-\mu_n \lfloor nx \rfloor} \rightarrow e^{-\mu x}$. □

Dans le cas d'une population comportant N individu et dans une échelle de temps où l'unité de temps est égale à N générations, le temps T_2 d'apparition de l'ACPR pour deux individus donnés a approximativement pour loi la loi exponentielle de paramètre 1. En particulier $\mathbb{E}(T_2) = 1$. On a donc montré que

$$(A_2^N(\lfloor Nt \rfloor))_{t \geq 0} \xrightarrow[N \rightarrow \infty]{\mathcal{L}} (A_2(t))_{t \geq 0}.$$

Considérons l'arbre généalogique de trois individus. Quand N tend vers l'infini, T'_2/N converge en loi vers $\mathcal{E}(1)$. De plus, on montre de même que $(T'_2/N, T'_3/N)$ converge vers (T_2, T_3) où T_2 et T_3 sont indépendantes de lois exponentielles respectives $\mathcal{E}(1)$ et $\mathcal{E}(3)$:

$$\begin{aligned} \varphi_{(T'_2/N, T'_3/N)}(s, t) &= \mathbb{E}(\exp(isT'_2/N + itT'_3/N)) \\ &= \mathbb{E}[\mathbb{E}(\exp(isT'_2/N) | T'_3) \exp(itT'_3/N)] \\ &= \frac{1}{3N - 2} \mathbb{E}[e^{itT'_3/N}] + \frac{3N - 3}{3N - 2} \mathbb{E}[e^{itV/N}] \mathbb{E}[e^{itT'_3/N}] \end{aligned}$$

et le lemme 9.1.1 permet de conclure que $(A_3^N([Nt]))_{t \geq 0}$ converge en loi vers le processus $(A_3(t))_{t \geq 0}$ quand N tend vers l'infini. En d'autres termes, lorsque N tend vers l'infini et que le temps est mesuré en unités de N générations, le mécanisme d'apparition de l'ACPR d'un groupe composé de trois individus distincts pour le processus limite est le suivant :

- après un temps exponentiel $T_3 \sim \mathcal{E}(3)$, un ancêtre commun à 2 individus apparaît,
- l'ancêtre commun apparaît alors après un temps $T_2 \sim \mathcal{E}(1)$,
- les variables aléatoires T_2 et T_3 sont indépendantes.

En particulier, à la limite, la probabilité que le premier ancêtre commun soit commun aux trois individus est nulle. On peut généraliser le résultat comme suit.

Théorème 9.1.2 (Du discret au continu). *Pour tout $n \geq 2$, au sens des lois marginales*

$$(A_n^N([Nt]))_{t \geq 0} \xrightarrow[N \rightarrow \infty]{\mathcal{L}} (A_n(t))_{t \geq 0}.$$

Démonstration. On esquisse seulement les grandes lignes. Considérons dans un premier temps le comportement de la matrice de transition de G_N lorsque N est grand. Puisque $S_{k,k-1} = \binom{k}{2}$, pour tout $2 \leq k \leq n$,

$$g_{k,k-1}^{(N)} = S_{k,k-1} \frac{N(N-1) \cdots (N-k+1)}{N^k} = \binom{k}{2} \frac{1}{N} + O(N^{-2})$$

tandis que pour $j < k-1$, on a

$$g_{k,j}^{(N)} = S_{k,j} \frac{N(N-1) \cdots (N-j+1)}{N^k} = O(N^{-2}).$$

Enfin, on a

$$g_{k,k}^{(N)} = N^{-k} N(N-1) \cdots (N-k+1) = 1 - \binom{k}{2} \frac{1}{N} + O(N^{-2}).$$

Ainsi $G_N = I + N^{-1}Q + O(N^{-2})$. La suite consiste à identifier la limite de $(G_N)^{[Nt]}$. . . \square

9.2 Longueur de l'arbre généalogique

Le processus à temps continu $(A_n(t))_{t \geq 0}$ d'espace d'état $\{1, \dots, n\}$ est un processus de mort pur, c'est-à-dire que ses trajectoires continues par morceaux sont décroissantes. Il est issu de $A_n(0) = n$ et décroît uniquement avec des sauts d'amplitude -1 . Pour tout $k \geq 2$, sachant que le processus est dans l'état k , le temps d'attente T_k avant de passer dans l'état $k-1$ (seule transition possible) suit une loi $\mathcal{E}(\binom{k}{2})$. De plus, les variables aléatoires $(T_k)_{2 \leq k \leq n}$ sont indépendantes. Si W_n est le temps d'apparition de l'ACPR alors

$$W_n = T_n + \cdots + T_2,$$

où les $(T_k)_{2 \leq k \leq n}$ sont les temps de séjours dans les états $2, \dots, n$. En particulier, on a donc

$$\mathbb{E}(W_n) = 2 \sum_{k=2}^n \frac{1}{k(k-1)} = 2 \sum_{k=2}^n \left(\frac{1}{k-1} - \frac{1}{k} \right) = 2 \left(1 - \frac{1}{n} \right) \rightarrow 2$$

et les T_2, \dots, T_n étant indépendantes, on a de plus

$$\begin{aligned} \text{Var}(W_n) &= \sum_{k=2}^n \text{Var}(T_k) \\ &= \sum_{k=2}^n \frac{4}{k^2(k-1)^2} \\ &= 4 \sum_{k=2}^n \left(\frac{1}{k^2} + \frac{1}{(k-1)^2} - \frac{2}{k(k-1)} \right) \\ &= 8 \sum_{k=1}^{n-1} \frac{1}{k^2} + \frac{4}{n^2} - 4 - 8 \left(1 - \frac{1}{n} \right) \end{aligned}$$

On a $\text{Var}(W_n) \nearrow \frac{4\pi^2}{3} - 12 \simeq 1.16$ (notons que $\text{Var}(T_2) = 1$). Il est possible d'obtenir une expression explicite de la densité de W_n en remarquant que $\mathbb{P}(W_n \leq t) = \mathbb{P}(A_n(t) = 1)$ et en calculant la matrice de transition du processus ancestral (assez peu passionnant).

On appelle arbre généalogique d'un groupe de n individus l'ensemble de tous leurs ancêtres toutes générations comprises, eux compris, jusqu'au premier ancêtre commun à tous les individus. On note L_n , et l'on appelle *longueur de l'arbre généalogique* la variable aléatoire égale à la somme des temps de vie de tous les individus de l'arbre. La longueur de l'arbre L_n s'exprime en fonction des temps d'apparition des ancêtres commun :

$$L_n = 2T_2 + \dots + nT_n.$$

En particulier, comme T_2, \dots, T_n sont indépendantes avec $T_k \sim \mathcal{E}\left(\frac{k}{2}\right)$,

$$\mathbb{E}(L_n) \sim_{n \rightarrow \infty} 2 \log(n) \quad \text{et} \quad \text{Var}(L_n) \sim_{n \rightarrow \infty} \frac{2\pi^2}{3}.$$

Théorème 9.2.1 (Longueur de l'arbre ancestral). *La variable aléatoire L_n suit la loi du maximum de $n - 1$ v.a.r. i.i.d. de loi exponentielle de paramètre $1/2$. De plus, on a*

$$\frac{L_n}{\log(n)} \xrightarrow[n \rightarrow \infty]{p.s.} 2 \quad \text{et} \quad L_n - 2 \log(n) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} G_{\frac{1}{2}}$$

où $G_{\frac{1}{2}}$ est la loi de Gumbel de fonction de répartition $F(t) = \exp(-\exp(-\frac{1}{2}t))$.

Démonstration. La v.a.r. L_n est la somme des v.a.r. indépendantes $2T_2, \dots, nT_n$ qui suivent les lois $\mathcal{E}(1/2), \dots, \mathcal{E}((n-1)/2)$. Le premier résultat découle alors de la propriété suivante : si E_1, \dots, E_n sont indépendantes de lois $\mathcal{E}(\lambda), \mathcal{E}(2\lambda), \dots, \mathcal{E}(n\lambda)$ alors $S_n := E_1 + \dots + E_n$ à la même loi que $M_n := \max(F_1, \dots, F_n)$ où F_1, \dots, F_n sont i.i.d. de loi $\mathcal{E}(\lambda)$. La densité de M_n est

$$f_n(x) := (\mathbb{P}(M_n \leq x))' = ((1 - e^{-\lambda x})^n \mathbf{1}_{\mathbb{R}_+}(x))' = n\lambda(1 - e^{-\lambda x})^{n-1} e^{-\lambda x} \mathbf{1}_{\mathbb{R}_+}(x).$$

Montrons par récurrence sur n que S_n a pour densité f_n . C'est vrai pour $n = 1$. Si cela est vrai pour n , alors la densité de S_{n+1} est, en notant g_λ la densité de $\mathcal{E}(\lambda)$,

$$\begin{aligned} f_n * g_{(n+1)\lambda}(y) &= (n+1)\lambda \int_{-\infty}^y f_n(x) e^{-\lambda(n+1)(y-x)} dx \\ &= \lambda(n+1)n\lambda e^{-\lambda(n+1)y} \int_0^y (e^{\lambda x} - 1)^{n-1} e^{\lambda x} dx \\ &= \lambda(n+1)e^{-\lambda(n+1)y} (e^{\lambda y} - 1)^n \\ &= \lambda(n+1)e^{-\lambda y} (1 - e^{-n\lambda y})^n = f_{n+1}(y). \end{aligned}$$

Pour établir le second résultat (convergence p.s.) on utilise le fait suivant : si $(X_n)_n$ sont des v.a.r. indépendantes, centrées, de carré intégrable, avec $\sum_n \mathbb{E}(X_n^2) < \infty$ alors $\sum_n X_n$ converge p.s. et dans L^2 (cas particulier du théorème de convergence des martingales bornée dans L^2). En effet, avec $X_n = nT_n - \mathbb{E}(nT_n) = nT_n - 2/(n-1)$ on a $\mathbb{E}(X_n) = 0$ et

$$\sum_{n \geq 2} \mathbb{E}(X_n^2) = \sum_{n \geq 2} \text{Var}(X_n) = \sum_{n \geq 2} \frac{4}{(n-1)^2} < \infty$$

donc $\sum_n X_n$ converge p.s. et donc

$$\frac{L_n}{\log(n)} = \frac{1}{\log(n)} \sum_{k=2}^n X_k + \frac{1}{\log(n)} \sum_{k=2}^n \mathbb{E}(kT_k) = \frac{1}{\log(n)} \sum_{k=2}^n X_k + \frac{2}{\log(n)} \sum_{k=1}^{n-1} \frac{1}{k} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} 2.$$

Le troisième résultat (fluctuation Gumbel) vient de $L_n \stackrel{d}{=} M_n$ car pour tout $t \in \mathbb{R}$,

$$\mathbb{P}(L_n - 2 \log(n) \leq t) = \left(1 - e^{-\frac{1}{2}(t+2 \log(n))}\right)^{n-1} = \left(1 - \frac{1}{n} e^{-\frac{1}{2}t}\right)^n \xrightarrow[n \rightarrow \infty]{} e^{-e^{-\frac{1}{2}t}}.$$

□

9.3 Mutations

Pour rendre compte de la diversité des individus dans une population, il faut tenir compte des possibilités de mutation de l'ADN, en particulier lors de la réplication. Les mutations entraînent une diversification des enfants d'un même individu. Le taux de mutation d'une base est très faible. On suppose pour simplifier que ce taux ne dépend pas de la base concernée, ni de sa position dans l'ADN et qu'il est constant au cours du temps. Ainsi, à chaque génération, un individu a une probabilité μ d'avoir une mutation qui le différencie de son parent. Si l'on considère une séquence de m bases, la probabilité que deux mutations aient lieu au même endroit est $1/m$. Comme les probabilités de mutation sont faibles, cela arrive très rarement. On fera donc l'hypothèse que l'on a une infinité d'allèles possibles et que chaque nouvelle mutation affecte un site différent. Ainsi, **une mutation donne toujours un nouvel allèle**. Le temps d'apparition d'une mutation dans la lignée ancestrale d'un individu est donc une loi géométrique de paramètre μ . On pose $\theta = 2\mu N$ et on suppose que θ est d'ordre 1. Ainsi, lors du passage à la limite en N tend vers l'infini, le processus de mutation devient un processus de Poisson. Plus exactement, **on munit chaque branche de l'arbre d'un processus de Poisson de paramètre $\theta/2$** qui comptera les mutations. Les deux processus de coalescence et mutation sont d'origines aléatoires différentes. Nous les considérerons indépendants.

Théorème 9.3.1 (Loi du nombre d'allèles dans un échantillon). *Si K_n désigne le nombre d'allèles distincts dans un groupe de n personnes alors K_n a la loi de $\eta_1 + \dots + \eta_n$ où $(\eta_k)_{k \geq 1}$ sont indépendantes de loi de Bernoulli de paramètres $(\theta/(k-1+\theta))_{k \geq 1}$.*

Notons que $\eta_1 = 1$ p.s. La v.a.r. K_n correspond au nombre de tables dans le processus des restaurants chinois (chapitre 3). En particulier, on sait calculer les deux premiers moments, et on connaît le comportement asymptotique (convergence et fluctuation).

Lemme 9.3.2 (Horloges exponentielles en compétition). *Si E_1, \dots, E_n sont des v.a.r. exponentielles indépendantes de paramètres respectifs $\lambda_1, \dots, \lambda_n$ alors $M := \min(E_1, \dots, E_n)$ suit la loi exponentielle de paramètre $\lambda_1 + \dots + \lambda_n$. De plus, presque sûrement le minimum dans la définition de M est atteint en un unique entier I aléatoire de $\{1, \dots, n\}$, indépendant de M , de loi $\mathbb{P}(I = i) = \mathbb{P}(M = E_i) = \lambda_i / (\lambda_1 + \dots + \lambda_n)$ pour tout $i \in \{1, \dots, n\}$.*

Preuve du lemme 9.3.2. Pour tout $i \in \{1, \dots, n\}$ et tout $x \in \mathbb{R}_+$,

$$\mathbb{P}(\min_{k \neq i}(E_k) > E_i \geq x) = \lambda_i \int_x^\infty \prod_{k \neq i} \mathbb{P}(E_k > y) e^{-\lambda_i y} dy = \frac{\lambda_i}{\lambda_1 + \dots + \lambda_n} e^{-(\lambda_1 + \dots + \lambda_n)x}$$

C'est la formule pour $\mathbb{P}(I = i, M \geq x)$, qui donne la loi du couple (I, M) . \square

Preuve du théorème 9.3.1. Pour un groupe de n individus, on s'intéresse au premier temps (dans le passé) d'apparition d'une coalescence ou d'une mutation. Tout se passe comme si U_n était le minimum entre T_n , premier temps de coalescence (de loi $\mathcal{E}(\binom{n}{2})$) et R_1, \dots, R_n premiers temps de mutation des individus $1, \dots, n$ de même loi $\mathcal{E}(\theta/2)$. Puisqu'elles sont indépendantes, leur minimum est encore une loi exponentielle et son paramètre est égal à

$$\binom{n}{2} + n \frac{\theta}{2} = \frac{n(n-1+\theta)}{2}.$$

De plus, la probabilité que ce premier phénomène soit une coalescence est (lemme 9.3.2)

$$\frac{\binom{n}{2}}{\frac{n(n-1+\theta)}{2}} = \frac{n-1}{n-1+\theta}.$$

De même, la probabilité pour que ce phénomène soit une mutation est donc

$$1 - \frac{n-1}{n-1+\theta} = \frac{\theta}{n-1+\theta}.$$

On pose $\eta_n = 1$ si ce premier phénomène est une mutation et $\eta_n = 0$ si c'est une coalescence. La variable aléatoire η_n suit donc la loi de Bernoulli de paramètre $\theta/(n-1+\theta)$. Étudions ce qui se passe après ce temps U_n :

- Si $\eta_n = 0$ alors le nombre d'ancêtres à l'instant U_n est $n-1$ et le nombre d'allèles distincts dans ce groupe de $n-1$ individus est $K_{n-1} = K_n$.
- Si $\eta_n = 1$ alors l'individu qui a muté à l'instant U_n est à l'origine d'un allèle différent. Il correspond, dans la population initiale de n individus, à la présence d'un allèle distinct de tous les autres. De plus, cet allèle est présent une seule fois dans la population des n ancêtres. On obtient donc une population de $n-1$ individus possédant $K_{n-1} = K_n - 1$ allèles distincts.

Dans tous les cas, on a obtenu $K_n = K_{n-1} + \eta_n$ et on considère à partir de l'instant U_n une population de $n-1$ individus possédant K_{n-1} allèles différents. Par récurrence, on obtient bien la décomposition de K_n en somme de variables aléatoires de Bernoulli. La propriété de Markov du processus de coalescence et l'absence de mémoire du processus de mutation assure l'indépendance des variables aléatoires $(\eta_k)_k$. \square

On peut en fait décrire plus précisément la structure des allèles d'une population de n individus. Dans le modèle à nombre d'allèles infini, un échantillon de taille n peut être représenté par une configuration $c = (c_1, \dots, c_n)$ où c_i est le nombre d'allèles représentés i fois et $|c| = c_1 + 2c_2 + \dots + nc_n = n$. On notera $e_i = (0, 0, \dots, 0, 1, 0, \dots, 0)$ le $i^{\text{ème}}$ vecteur unitaire. Il s'agit à présent d'établir une équation satisfaite par les probabilités $(q(c))$ où $q(c)$ est la probabilité qu'un échantillon de taille $|c|$ tiré sous la probabilité stationnaire ait la configuration c . On pose $q(e_1) = 1$. Supposons que la configuration soit c . En remontant la généalogie de l'échantillon jusqu'au premier changement, nous trouvons une mutation ou une coalescence (ancêtre commun à deux individus). Analysons ces deux événements :

— Le premier événement (mutation) a une probabilité (lemme 9.3.2)

$$\frac{n\theta/2}{n\theta/2 + n(n-1)/2} = \frac{\theta}{\theta + n - 1}.$$

La configuration \mathbf{b} qui a conduit à la \mathbf{c} est

- $\mathbf{b} = \mathbf{c}$ si la mutation a eu lieu sur l'un des c_1 singletons (probabilité c_1/n),
 - $\mathbf{b} = \mathbf{c} - 2e_1 + e_2$ si la mutation a eu lieu sur un des allèles représenté 2 fois (probabilité $2b_2/n = 2(c_2 + 1)/n$),
 - $\mathbf{b} = \mathbf{c} - e_1 - e_{j-1} + e_j$ si la mutation a eu lieu sur un des allèles représenté j fois (probabilité $jb_j/n = j(c_j + 1)/n$)
- Le second événement (coalescence) a lieu avec une probabilité $(n-1)/(\theta + n - 1)$ et dans ce cas, la configuration précédente était de la forme $\mathbf{b} = \mathbf{c} + e_j - e_{j+1}$: un allèle présent j fois (parmi les $c_j + 1$) s'est dédoublé, ramenant de $c_j + 1$ à c_j le nombre d'allèles présent j fois et de $c_{j+1} - 1$ à c_{j+1} le nombre d'allèles présent $j + 1$ fois. Cet événement a une probabilité de $jb_j/(n-1) = j(c_j + 1)/(n-1)$.

On retrouve le processus des restaurants chinois du chapitre 3, qui nous enseigne que \mathbf{c} suit la loi d'Ewens (on parle de formule d'échantillonnage d'Ewens en génétique² : pour tous entiers a_1, \dots, a_n vérifiant $a_1 + 2a_2 + \dots + na_n = n$,

$$q(\mathbf{c}) = \mathbb{P}(\mathbf{c}_1(n) = a_1, \dots, \mathbf{c}_n(n) = a_n) = \frac{n!}{\theta(\theta + 1) \cdots (\theta + n - 1)} \prod_{j=1}^n \left(\frac{\theta}{j}\right)^{c_j} \frac{1}{c_j}.$$

9.4 Coalescent de Kingman

Le «coalescent de Kingman» est un processus à temps continu à valeurs dans l'ensemble \mathcal{E}_n des relations d'équivalence sur $[n] = \{1, \dots, n\}$ (c'est-à-dire l'ensemble des partitions de $[n]$). Il décrit la généalogie d'une population de taille n . Jusqu'à présent, nous avons proposé un modèle pour l'évolution temporelle du nombre d'ancêtres d'une population initiale donnée. Il s'agit à présent d'être plus précis en gardant trace des liens de parentés entre tous les individus. Numérotons les individus de 1 à n . À un instant t , on définit la relation d'équivalence (aléatoire) \sim sur $[n]$ par $i \sim j$ si et seulement si les individus i et j ont le même ancêtre commun au temps t . Chaque classe d'équivalence correspond à un ancêtre de la population initiale et les éléments de cette classe sont tous les descendants de cet individu. Notons $C(t)$ cette partition (aléatoire). Quelle est la dynamique de $(C(t))_{t \geq 0}$? Supposons que $C(0) = \alpha \in \mathcal{E}_n$ et notons k le nombre de ses classes (on écrira $|\alpha| = k$). Lorsque t augmente, nous progressons vers le passé et le processus reste constant jusqu'à l'apparition d'un ancêtre commun à deux des ancêtres de chaque classe. Lorsque ceci se produit, les deux ancêtres, et par conséquent tous leurs descendants, partagent cet ancêtre commun. Les deux classes d'équivalence sont donc regroupées (coalescent). Le taux d'apparition de cet événement est 1. Il est intuitivement clair que C est markovien.

Le processus $(C(t))_{t \geq 0}$ est le processus de Markov à temps continu sur \mathcal{E}_n d'état initial $C(0) = \Delta = \{\{1\}, \{2\}, \dots, \{n\}\}$ (personne n'est relié à personne), et de générateur Q

$$Q(\alpha, \beta) = \begin{cases} -\binom{k}{2} & \text{si } \alpha = \beta \text{ et } |\alpha| = k, \\ 1 & \text{si } \alpha \sim \beta, \\ 0 & \text{sinon,} \end{cases}$$

2. ESF : Ewens Sampling Formula.

où la notation $\alpha \sim \beta$ signifie que la partition β peut être obtenue à partir de la partition α en fusionnant deux de ses classes d'équivalence. Le processus C est appelé coalescent ou n -coalescent. Pour déterminer sa loi, il faut tout d'abord étudier la chaîne incluse $(C_k^{\text{inc}})_{k=n, n-1, \dots, 1}$. Cette chaîne est issue de Δ et admet pour transitions :

$$\mathbb{P}(C_{k-1}^{\text{inc}} = \beta \mid C_k^{\text{inc}} = \alpha) = \begin{cases} \frac{1}{\binom{k}{2}} & \text{si } \alpha \sim \beta \text{ et } |\alpha| = k, \\ 0 & \text{sinon.} \end{cases}$$

Le processus C évolue donc selon une suite $\Delta = C_n^{\text{inc}} \sim C_{n-1}^{\text{inc}} \sim \dots \sim C_1^{\text{inc}} = \Theta$, où Θ est la partition triviale et passe en la partition C_k^{inc} un temps exponentiel de paramètre $\binom{k}{2}$. Les taux de transition ne dépendent de la partition qu'au travers de son cardinal et

$$|C(t)| = A_n(t),$$

puisque les classes de $C(t)$ sont en bijection avec les ancêtres au temps t . Ainsi, les processus $(A_n(t))_{t \geq 0}$ et $(C_k^{\text{inc}})_k$ sont indépendants et $C(t) = C_{A_n(t)}^{\text{inc}}$ pour tout $t \geq 0$. Plus précisément,

$$\mathbb{P}(C(t) = \alpha) = \mathbb{P}(A_n(t) = |\alpha|) \mathbb{P}(C_{|\alpha|}^{\text{inc}} = \alpha).$$

Théorème 9.4.1 (Loi du coalescent de Kingman). *Soit $1 \leq j \leq n$ et α une partition de $[n]$ dont les classes d'équivalence admettent les cardinaux $\lambda_1, \dots, \lambda_j$. Alors,*

$$\mathbb{P}(C_j^{\text{inc}} = \alpha) = \frac{(n-j)!j!(j-1)!}{n!(n-1)!} \lambda_1! \cdots \lambda_j!.$$

Démonstration. On procède par récurrence descendante. Le résultat est évident pour $j = n$ (si α n'est pas la partition Δ , sa probabilité d'apparition est nulle, et Δ est la seule partition à n classes, toutes singleton). Supposons que le résultat soit vrai pour $j \geq 2$. Alors

$$p_{j-1}(\beta) := \mathbb{P}(C_{j-1}^{\text{inc}} = \beta) = \sum_{\alpha \in \mathcal{E}_n} p_j(\alpha) \mathbb{P}(C_{j-1}^{\text{inc}} = \beta \mid C_j^{\text{inc}} = \alpha) = \sum_{\alpha \sim \beta} p_j(\alpha) \frac{2}{j(j-1)}.$$

Notons $\lambda_1, \dots, \lambda_{j-1}$ les tailles des classes d'équivalence de β . Celles de α sont

$$\lambda_1, \dots, \lambda_{l-1}, m, \lambda_l - m, \lambda_{l+1}, \dots, \lambda_{j-1}$$

pour un certain $1 \leq l \leq j-1$ et $1 \leq m \leq \lambda_l - 1$. Grâce à l'hypothèse de récurrence³,

$$\begin{aligned} p_{j-1}(\beta) &= \frac{1}{2} \sum_{l=1}^{j-1} \sum_{m=1}^{\lambda_l-1} \frac{2(n-j)!j!(j-1)!}{j(j-1)n!(n-1)!} \lambda_1! \cdots \lambda_{l-1}! m! (\lambda_l - m)! \lambda_{l+1}! \cdots \lambda_{j-1}! \binom{\lambda_l}{m} \\ &= \frac{(n-j)!(j-1)!(j-2)!}{n!(n-1)!} \lambda_1! \cdots \lambda_{j-1}! \sum_{l=1}^{j-1} \sum_{m=1}^{\lambda_l-1} 1 \\ &= \frac{(n-j+1)!(j-1)!(j-2)!}{n!(n-1)!} \lambda_1! \cdots \lambda_{j-1}! \end{aligned}$$

car $\sum_{l=1}^{j-1} \sum_{m=1}^{\lambda_l-1} 1 = \lambda_1 + \dots + \lambda_{j-1} - j + 1 = n - j + 1$. □

9.5 Notes et commentaires

Livres [DJ06], [Ewe04], [Kin93], [Dur08], [Ber09].

3. Le facteur 1/2 est dû à la permutation de m et $\lambda_l - m$. Considérer l'exemple $\lambda_l = 2$, puis $\lambda_l = 3$.

Séance 10

Simulation de lois discrètes

Mots-clés. Lois discrètes ; simulation ; chaîne de Markov.

Ce chapitre est consacré à quelques algorithmes remarquables de simulations de loi discrètes. Les concepts et les techniques abordés vont parfois bien au delà du cadre considéré, qui a le mérite toutefois de nécessiter assez peu de technologie.

10.1 Algorithme basique pour les lois discrètes

Considérons le problème de la simulation d'une loi discrète $\mu = \sum_{a \in E} \mu(a) \delta_a$ où E est au plus dénombrable. Nous pouvons numéroter les atomes en choisissant une bijection $\varphi : E \rightarrow \{1, 2, \dots\}$. À présent, si on partitionne l'intervalle réel $[0, 1]$ en blocs de mesures de Lebesgue respectives $\mu(\varphi^{-1}(1)), \mu(\varphi^{-1}(2)), \dots$, par exemple en utilisant les intervalles

$$I_1 = [0, \mu(\varphi^{-1}(1))[, I_2 = [\mu(\varphi^{-1}(1)), \mu(\varphi^{-1}(1)) + \mu(\varphi^{-1}(2))[, \dots,$$

et si U est une variable aléatoire uniforme sur $[0, 1]$, alors $\mathbb{P}(U \in I_{\varphi(a)}) = \mu(a)$. L'algorithme basique de simulation de μ est alors le suivant : on génère tout d'abord une réalisation u de U , ensuite si $u \leq \mu(\varphi^{-1}(1))$ alors on décide $\varphi^{-1}(1)$; sinon, si $u \leq \mu(\varphi^{-1}(1)) + \mu(\varphi^{-1}(2))$, alors on décide $\varphi^{-1}(2)$, etc. Si F est la fonction de répartition de $\mu \circ \varphi^{-1}$, d'inverse généralisé F^{-1} , alors $\varphi^{-1}(F^{-1}(U)) \sim \mu$. Il s'agit d'un cas spécial de la méthode de simulation par inversion.

Le coût de cet algorithme est le nombre N de tests utilisés. Ce nombre est aléatoire, de loi $\mu \circ \varphi^{-1}$. En particulier, $\mathbb{P}(N < \infty) = 1$, et le coût moyen est $\mathbb{E}(N) = \sum_{a \in E} \varphi(a) \mu(a)$, qui peut très bien être infini si $\mu \circ \varphi^{-1}$ n'a pas d'espérance (ne peut se produire que si E est infini) ! Si φ minimise le coût moyen $\mathbb{E}(N)$ alors $\mu(\varphi^{-1}(1)) \geq \mu(\varphi^{-1}(2)) \geq \dots$. Pour la loi géométrique (de moyenne quelconque) et pour la loi de Poisson (de moyenne 1), la numérotation naturelle est déjà à poids décroissants. D'autre part, si $\text{card}(E)$ est petit, il est facile de déterminer l'ordre à poids décroissants en utilisant un algorithme de tri.

Les lois discrètes usuelles (binomiale, géométrique, Poisson, etc) sont simulables par divers algorithmes dédiés tirant partie de leurs propriétés spéciales. À ce sujet, signalons qu'il est possible de simuler la loi de Poisson de moyenne quelconque λ à partir d'un générateur de la loi de Poisson de moyenne 1. Il suffit en effet d'utiliser un amincissement¹. Plus précisément, on simule $\lceil \lambda \rceil$ variables aléatoires $X_1, \dots, X_{\lceil \lambda \rceil}$ i.i.d. de loi $\text{Poi}(1)$, puis, conditionnellement à leur somme $S = X_1 + \dots + X_{\lceil \lambda \rceil}$, on simule S variables aléatoires i.i.d. de loi de Bernoulli de moyenne $\lambda / \lceil \lambda \rceil$, et on tire parti du fait que $B_1 + \dots + B_S \sim \text{Poi}(\lambda)$.

1. Si (X, Y) est un couple tel que $\mathcal{L}(X) = \text{Poi}(\lambda)$ et $\mathcal{L}(Y|X = n) = \text{Bin}(n, p)$ pour tout n alors $Y \sim \text{Poi}(p\lambda)$.

```

function [sums,indexes] = generator_finite_init(p)
    [p_sorted,indexes] = sort(p)
    sums = cumsum(p_sorted)
endfunction

function generator_finite(sums,indexes)
    u = rand
    for i = 1 to length(sums) do
        if u < sums(i) then return indexes(i) endif
    endfor
endfunction

```

Figure 10.1 – Algorithme basique de simulation de loi discrète finie en pseudo-code.

```

function [sums, indexes] = generator_poisson_init()
    sums = []
    indexes = []
    p = 1/e
    s = p
    n = 0
    while p > 0
        sums = [sums , s]
        indexes = [indexes, n]
        n = n + 1
        p = 1/(e*gamma(n+1))
        s = s + p
    endwhile
endfunction

function generator_poisson(sums,indexes,lambda)
    c = ceil(lambda)
    s = 0
    for i = 1 to c do s = s + generator_finite(sums,indexes) endfor
    if c == lambda then return s else return sum(rand(1,s) < lambda/c) endif
endfunction

```

Figure 10.2 – Algorithme de simulation de loi de Poisson en pseudo-code (optimisable).

Le cas de la loi uniforme sur un ensemble E fini est justiciable d'une approche directe : $\varphi^{-1}(\lceil \text{card}(E)U \rceil)$ suit cette loi ! Cependant, l'algorithme basique est impraticable lorsque E est difficile à énumérer et donc φ est difficile d'accès, ou lorsque $\text{card}(E)$ est très grand et μ est peu concentrée. Nous étudions par la suite deux exemples de ce type : les ensembles des permutations et des partitions de $\{1, \dots, n\}$, pour lesquels nous présentons des algorithmes spécifiques efficaces. D'autre part, dans de nombreuses situations concrètes, E est complexe et les poids des atomes de μ ne sont connus qu'à une constante multiplicative près, c'est-à-dire que seuls leurs rapports sont connus. Dans ces situations délicates, il est en revanche bien souvent possible d'utiliser un algorithme markovien comme l'algorithme de Metropolis-Hastings (approché) ou celui de Propp-Wilson (exact) abordés plus loin.

10.2 Permutations aléatoires

Certaines situations nécessitent de permuter aléatoirement une liste finie d'objets : construction de plans d'expériences dans les sciences expérimentales, anonymisation, etc. Cela conduit au problème de la simulation de la loi uniforme $\sum_{\sigma \in \mathcal{S}_n} \text{card}(\mathcal{S}_n)^{-1} \delta_\sigma$ sur l'ensemble fini \mathcal{S}_n des permutations de $\{1, \dots, n\}$ (groupe symétrique). Or $\text{card}(\mathcal{S}_n) = n! \sim \sqrt{2\pi n} (n/e)^n$ est un nombre à environ $n \log(n)$ chiffres, et cette réalité combinatoire disqualifie très vite l'algorithme basique de simulation des lois discrètes. Il est possible de simuler la loi uniforme sur \mathcal{S}_n en réordonnant un désordre symétrique : si U_1, \dots, U_n sont des variables aléatoires i.i.d. de loi uniforme sur $[0, 1]$ et si σ est une permutation aléatoire telle que $U_{\sigma(1)} \leq \dots \leq U_{\sigma(n)}$, alors σ suit la loi uniforme sur \mathcal{S}_n . La complexité est celle de l'algorithme de tri utilisé². Un algorithme plus naïf pour simuler une permutation σ uniforme consiste à tirer uniformément et sans remise les valeurs de $\sigma(1), \dots, \sigma(n)$ dans $\{1, \dots, n\}$. Cependant, cet algorithme d'apparence séduisante, a une complexité plus élevée que celui du tri, car il faut tenir compte dans l'implémentation des éléments déjà tirés.

L'un des meilleurs algorithmes pour simuler la loi uniforme sur \mathcal{S}_n est connu sous le nom de *Fisher-Yates shuffle* ou de *Knuth shuffle*. Sa complexité est d'ordre n .

Théorème 10.2.1 (Algorithme de Fisher-Yates-Knuth). *Si U_1, \dots, U_n sont des variables aléatoires indépendantes avec U_i de loi uniforme sur $\{1, \dots, i\}$ pour tout $1 \leq i \leq n$ alors le produit de transpositions aléatoires $(1, U_1) \cdots (n, U_n)$ suit la loi uniforme sur \mathcal{S}_n .*

```

function shuffle(v)
  for k from length(v) downto 2 do
    i = ceil(k*rand)
    swap(v[i], v[k])
  endfor
  return v
endfunction

```

Figure 10.3 – Algorithme de Fisher-Yates-Knuth en pseudo-code.

L'inversion étant bijective, et les transpositions étant leur propre inverse, il en découle que le produit renversé $(n, U_n) \cdots (1, U_1)$ suit également la loi uniforme sur \mathcal{S}_n . Notons que la transposition $(1, U_1)$ est triviale (élément neutre de \mathcal{S}_n), et il n'est pas nécessaire d'en tenir compte si $n \geq 2$ (elle rend cependant la formule valable pour $n = 1$).

2. De l'ordre de $n \log(n)$ avec grande probabilité et n^2 au pire pour l'algorithme *Quick Sort*.

Démonstration. On procède par récurrence sur n . La propriété est triviale pour $n = 1$. Pour tout $\sigma \in \mathcal{S}_n$, on note encore σ l'élément de \mathcal{S}_{n+1} obtenu à partir de σ en ajoutant le cycle $(n+1)$ (point fixe). Supposons que $\sigma_n = (1, U_1) \cdots (n, U_n)$ suit la loi uniforme sur \mathcal{S}_n . Soit U_{n+1} une variable aléatoire indépendante de σ_n , de loi uniforme sur $\{1, \dots, n+1\}$. Montrons que $\sigma_{n+1} = \sigma_n(n+1, U_{n+1})$ suit la loi uniforme sur \mathcal{S}_{n+1} . Pour tout $\sigma \in \mathcal{S}_{n+1}$, on a tout d'abord

$$\mathbb{P}(\sigma_{n+1} = \sigma) = \sum_{i=1}^{n+1} \mathbb{P}(\sigma_n = \sigma(n+1, i)) \mathbb{P}(U_{n+1} = i) = \frac{1}{n+1} \sum_{i=1}^{n+1} \mathbb{P}(\sigma_n = \sigma(n+1, i)).$$

Comme $n+1$ est point fixe de σ_n , et n'est point fixe de $\sigma(n+1, i)$ que pour une et une seule valeur de i , notée i_σ , image réciproque de $n+1$ par σ , il en découle finalement que

$$\mathbb{P}(\sigma_{n+1} = \sigma) = \frac{1}{n+1} \mathbb{P}(\sigma_n = \sigma(n+1, i_\sigma)) = \frac{1}{n+1} \frac{1}{n!} = \frac{1}{(n+1)!}.$$

□

10.3 Partitions aléatoires

Intéressons nous à la simulation de la loi uniforme sur l'ensemble des partitions Π_n de $\{1, \dots, n\}$. Cette loi affecte le même poids $1/B_n$ à chaque élément de Π_n , où $B_n = \text{card}(\Pi_n)$. En combinatoire, la suite $(B_n)_{n \geq 1}$ constitue les *nombre de Bell*. On a $B_1 = 1$, $B_2 = 2$, et plus généralement, en utilisant la convention $B_0 = 1$, on a la formule de récurrence triangulaire

$$B_{n+1} = \sum_{k=0}^n \binom{n}{k} B_k,$$

où k s'interprète comme le nombre d'éléments qui ne sont pas dans le bloc de $n+1$. Il en découle que la série formelle $G(X) = \sum_{n=0}^{\infty} \frac{B_n}{n!} X^n$ vérifie $G'(X) = \exp(X)G(X)$, d'où $G(X) = \exp(\exp(X) - 1)$. On reconnaît la transformée de Laplace de la loi Poi(1). Les nombres de Bell sont donc les moments de cette loi, d'où la formule dite de Dobinski :

$$B_n = \frac{1}{e} \sum_{k=1}^{\infty} \frac{k^n}{k!}.$$

Revenons au problème de la simulation de la loi uniforme sur Π_n . On pourrait croire que la décomposition en cycles d'une permutation aléatoire uniforme de $\{1, \dots, n\}$ fournirait une partition de loi uniforme sur Π_n . Il n'en est rien. L'un des algorithmes les plus élégants de simulation de la loi uniforme sur Π_n est dû à Stam, et fait appel à la formule de Dobinski.

Théorème 10.3.1 (Algorithme de Stam). *Soit K un entier aléatoire valant k avec probabilité $k^n/(k!eB_n)$ pour tout $k \geq 0$. Sachant K , soient C_1, \dots, C_n des variables aléatoires i.i.d. de loi uniforme sur $\{1, \dots, K\}$. Soit P la partition aléatoire de $\{1, \dots, n\}$ obtenue en décidant que i, j sont dans le même bloc ssi $C_i = C_j$. Alors P suit la loi uniforme sur Π_n .*

La loi de K est bien définie grâce à la formule de Dobinski. Il est commode d'interpréter C_1, \dots, C_n comme des couleurs, les blocs de P regroupant donc les éléments par couleur.

Démonstration. Il suffit d'observer que pour tout $p \in \Pi_n$, en notant b son nombre de blocs,

$$\mathbb{P}(P = p) = \sum_{k=b}^{\infty} \mathbb{P}(P = p | K = k) \mathbb{P}(K = k) = \sum_{k=b}^{\infty} \frac{k(k-1) \cdots (k-b+1)}{k^n} \frac{k^n}{k!eB_n} = \frac{1}{B_n}.$$

□

```

function [sums, indexes] = generator_stam_init(n)
    sums = []
    indexes = []
    p = 1/(e*bell(n))
    s = p
    k = 1
    while p > 0
        sums = [sums , s]
        indexes = [indexes, k]
        k = k + 1
        p = p*(1+1/k)^n/k
        s = s + p
    endwhile
endfunction

function generator_stam(sums,indexes,n)
    m = generator_finite(sums,indexes)
    c = ceil(m*rand(1,n))
    partition = []
    for i = 1 to n do partition(c(i)).append(i) endfor
    return partition
endfunction

```

Figure 10.4 – Algorithme de Stam en pseudo-code (optimisable).

10.4 Algorithme de Metropolis-Hastings

Soit μ une loi de probabilité sur E au plus dénombrable, chargeant tous les éléments. L'algorithme de Metropolis-Hastings consiste à construire et à simuler une chaîne de Markov récurrente apériodique $(X_n)_{n \geq 0}$ sur E de loi invariante μ , qui fournit alors un générateur approché de μ , car comme $(X_n)_{n \geq 0}$ converge en loi vers μ , on a $\mathcal{L}(X_n) \approx \mu$ pour $n \gg 1$. Malheureusement, au delà de catégories d'exemples concrets, aucune borne quantitative générale n'est connue, rendant souvent hasardeux le contrôle pratique de l'erreur.

Pour construire le noyau \mathbf{P} de $(X_n)_{n \geq 0}$, on se donne tout d'abord un noyau de transition irréductible \mathbf{Q} sur E vérifiant $\mathbf{Q}(x, y) = 0$ ssi $\mathbf{Q}(y, x) = 0$. Ensuite, pour tous $x \neq y$, on pose

$$\alpha(x, y) = \begin{cases} \min \left(1, \frac{\mu(y)\mathbf{Q}(y, x)}{\mu(x)\mathbf{Q}(x, y)} \right) & \text{si } \mathbf{Q}(x, y) > 0; \\ 0 & \text{sinon.} \end{cases}$$

On pose alors, pour tous $x, y \in E$,

$$\mathbf{P}(x, y) = \begin{cases} \mathbf{Q}(x, y)\alpha(x, y) & \text{si } x \neq y; \\ 1 - \sum_{z \neq x} \mathbf{P}(x, z) & \text{si } x = y. \end{cases}$$

Théorème 10.4.1 (Noyau de Metropolis-Hastings). \mathbf{P} est un noyau de transition sur E , irréductible récurrent positif apériodique admettant μ pour unique loi invariante réversible.

Démonstration. \mathbf{P} est un noyau car $0 \leq \mathbf{P}(x, y) \leq \mathbf{Q}(x, y)$ pour tous $x \neq y$. Il est irréductible car \mathbf{P} hérite du squelette de \mathbf{Q} , et apériodique car la diagonale de \mathbf{P} est strictement positive. D'autre part, la loi μ est symétrique pour \mathbf{P} , c'est-à-dire que $\mu(x)\mathbf{P}(x, y) = \mu(y)\mathbf{P}(y, x)$ pour tous $x, y \in E$, et elle est donc invariante et réversible. Comme \mathbf{P} est irréductible, il est alors irréductible récurrent positif car il possède une loi invariante, qui est unique. \square

Tout l'intérêt du noyau \mathbf{P} réside dans le fait que sa construction ne dépend de μ qu'à travers les rapports $\mu(x)/\mu(y)$. Ainsi, il suffit de connaître μ à une constante multiplicative près. Typiquement, la loi μ s'écrit $\mu(x) = Z^{-1}e^{-H(x)}$ où $H : E \rightarrow \mathbb{R}$ est une fonction qui attribue à chaque état une « énergie », et où $Z = \sum_{x \in E} e^{-H(x)}$ est une constante de normalisation (difficile à calculer). On dit souvent que μ est une mesure de Boltzmann-Gibbs. La loi μ favorise les états de faible énergie. Le rapport $\mu(x)/\mu(y) = e^{H(y)-H(x)}$ ne dépend que de la différence d'énergie $H(y) - H(x)$.

La simulation des trajectoires de la chaîne $(X_n)_{n \geq 0}$ de noyau \mathbf{P} par la méthode récursive revient à simuler la loi discrète $\mathbf{P}(x, \cdot)$ pour un x quelconque. Pour ce faire, soit Y une variable aléatoire qui suit la loi discrète $\mathbf{Q}(x, \cdot)$ sur E , et U une variable aléatoire de loi uniforme sur $[0, 1]$, indépendante de Y . Soit Z la variable aléatoire définie par $Z = Y$ si $U < \alpha(x, Y)$ et $Z = x$ sinon. Alors pour tous $y \neq x$,

$$\mathbb{P}(Z = y) = \mathbb{P}(U < \alpha(x, Y), Y = y) = \alpha(x, y)\mathbf{Q}(x, y) = \mathbf{P}(x, y).$$

Par conséquent, la variable aléatoire Z suit la loi $\mathbf{P}(x, \cdot)$. Il est remarquable que l'évaluation de $\mathbf{P}(x, x)$ soit inutile dans la simulation. La fonction α est appelée *fonction d'acceptation-rejet* tandis que le noyau \mathbf{Q} est parfois appelé *noyau d'exploration ou de proposition*.

```

iter_max = big_value
iter = 1
x = x_init
while iter < iter_max
  y = generator(Q(x, :))
  alpha = min(1, exp(h(x)-h(y))*Q(y,x)/Q(x,y))
  if rand < alpha(x,y) then y = x endif
  iter = iter + 1
endwhile
return x

```

Figure 10.5 – Exemple élémentaire d'algorithme de Metropolis-Hastings en pseudo-code.

Remarque 10.4.2 (Choix du noyau d'exploration). *Lorsque les éléments de E représentent des états complexes (emplois du temps d'un lycée par exemple), les transitions du noyau \mathbf{Q} correspondent typiquement à des transformations élémentaires (garantissant toutefois l'irréductibilité). Le choix de \mathbf{Q} dépend de la nature de l'espace. Lorsque E est muni d'une distance d_E vérifiant $n_x = \text{card}\{y \in E : d_E(x, y) = 1\} > 0$ pour tout $x \in E$, il est commode de prendre par exemple $\mathbf{Q}(x, y) = 1/n_x$ si $d_E(x, y) = 1$ et $\mathbf{Q}(x, y) = 0$ sinon. C'est typiquement le cas des marches aléatoires aux plus proches voisins sur les graphes.*

Remarque 10.4.3 (Algorithme du recuit simulé). *Intéressons-nous au problème du voyageur de commerce. On se donne un espace métrique fini (M, d_M) et on considère l'ensemble E des chemins qui passent par tous les points de M . On note $H(x)$ la longueur d'un chemin $x \in E$. On souhaite trouver un élément de E minimisant H . Il est possible de définir une distance d_E sur E correspondant à des modifications locales des chemins. Un algorithme déterministe de type gradient explorant E par petites modifications serait piégé par les minima locaux de H (imaginer H comme un paysage montagneux sur E). Pour dépasser cette difficulté, on peut adopter une stratégie probabiliste. Pour tout $\beta > 0$, on note μ_β la loi de probabilité sur E définie par $\mu_\beta(x) = Z_\beta^{-1}e^{-\beta H(x)}$ où $Z_\beta = \sum_{x \in E} e^{-\beta H(x)}$. Le calcul de Z_β n'est pas envisageable, mais en revanche, le calcul de $H(y) - H(x)$ est facile si x et y diffèrent peu. Lorsque $\beta \rightarrow \infty$, la loi μ_β converge vers une loi portée par les minima de H . Il est commode d'interpréter β comme l'inverse d'une température, ou comme l'inverse*

d'une variance par analogie gaussienne. L'idée de l'algorithme du recuit simulé consiste à utiliser la chaîne de Metropolis-Hastings associée à μ_β , en augmentant à chaque transition le paramètre β . La chaîne devient donc inhomogène, et la suite $(\beta_n)_{n \geq 1}$ utilisée est appelée schéma de température. Si au temps n , la chaîne est en x , alors le noyau d'exploration Q fait une proposition y , acceptée avec une probabilité $\alpha(x, y) = \min(1, e^{\beta_n(H(x)-H(y))})$. En particulier, la transition est acceptée à coup sûr si $H(y) \leq H(x)$, et avec probabilité non nulle mais décroissante en n si $H(y) > H(x)$. Ce mécanisme subtil permet à la chaîne d'accepter parfois de faire pire (i.e. augmenter H) pour échapper aux minima locaux de H , avec une probabilité décroissante au fil du temps. En pratique, le réglage de la décroissance de la température $1/\beta_n$ est délicat car une décroissance trop rapide peut empêcher l'atteinte d'un minimum global. Comme pour beaucoup d'algorithmes d'optimisation globale, l'algorithme du recuit simulé souffre de l'absence de critère d'arrêt satisfaisant. D'autre part, si H est trop complexe, ou si la notion de modification locale n'a pas de sens sur E , alors l'algorithme du recuit perd de son intérêt, et ne fait pas mieux que des tirages i.i.d. dans E .

```

iter_max = big_value
iter = 1
beta = 1
h_min = +infinity
x = x_init
h_x = h(x)
while iter < iter_max
  y = generator(Q(x, :))
  h_y = h(y)
  if h_y < h(x)
    then
      x = y
      h_min = h_y
    else
      alpha = exp(beta(h(x)-h(y)))
      if rand < alpha then x = y endif
    endif
  iter = iter + 1
  beta = cooling(beta, iter)
endwhile
return h_min

```

Figure 10.6 – Exemple élémentaire d'algorithme du recuit simulé en pseudo-code.

10.5 Algorithme de Propp-Wilson

L'algorithme de Metropolis-Hasting a l'inconvénient de ne pas être exact. Cela conduit au problème suivant : comment simuler de manière exacte la loi invariante μ d'un noyau de transition \mathbf{P} sur un ensemble au plus dénombrable E ?

Adoptons l'interprétation des chaînes de Markov sous forme de suites récurrentes aléatoires. Soit donc $g : E \times [0, 1] \rightarrow E$ une fonction telle que $g(x, U) \sim \mathbf{P}(x, \cdot)$ pour tout $x \in E$, où U est une variable aléatoire uniforme sur $[0, 1]$. On note $G : E \rightarrow E$ la fonction aléatoire définie par $G(x) = g(x, U)$. Soit $(G_n)_{n \geq 0}$ une suite i.i.d. de fonctions aléatoires de E dans E , de même loi que G , construites à la manière de G en utilisant une suite $(U_n)_{n \geq 0}$ de variables aléatoires i.i.d. de loi uniforme sur $[0, 1]$. Pour tout n , on construit les applications aléatoires

$A_n : E \rightarrow E$ et $B_n : E \rightarrow E$ par composition de la manière suivante :

$$A_n = G_n \circ \dots \circ G_1 \quad \text{et} \quad B_n = G_1 \circ \dots \circ G_n.$$

On convient que A_0 et B_0 sont égales à l'application identité de E . Pour tout $x \in E$ et tout n , les variables aléatoires $A_n(x)$ et $B_n(x)$ suivent la loi $\mathbf{P}^n(x, \cdot)$. Les suites $(A_n(x))_{n \geq 0}$ et $(B_n(x))_{n \geq 0}$ ont les mêmes lois marginales de dimension 1, mais n'ont pas la même loi en général. La suite $(A_n(x))_{n \geq 0}$ est une chaîne de Markov sur E de noyau \mathbf{P} et de loi initiale δ_x . En revanche, $(B_n(x))_{n \geq 0}$ n'est pas une chaîne de Markov car le temps est « inversé ». Nous définissons à présent les temps de contraction T_A et T_B à valeurs dans $\mathbb{N} \cup \{\infty\}$ par

$$T_A = \inf\{n \geq 0 : \text{card}(A_n(E)) = 1\} \quad \text{et} \quad T_B = \inf\{n \geq 0; \text{card}(B_n(E)) = 1\}.$$

On dit que G est *contractante* lorsque $p = \mathbb{P}(\text{card}(G(E)) = 1) > 0$.

Théorème 10.5.1 (Convergence vers μ à la coalescence pour la suite renversée). *Supposons que G est contractante. Alors $\mathbb{P}(T_B < \infty) = 1$ et il existe une variable aléatoire x_B sur E telle que $B_n(x) = x_B$ pour tout $x \in E$ et tout $n \geq T_B$. De plus, si μ_B est la loi de x_B , alors $\lim_{n \rightarrow \infty} \mathbf{P}^n(x, y) = \mu_B(y)$ pour tous x et y , et en particulier, $\mu_B = \mu$.*

Démonstration. Les événements $C_n = \{\text{card}(G_n(E)) = 1\}$ sont indépendants et de même probabilité $p = \mathbb{P}(\text{card}(G(E)) = 1) > 0$. Par le lemme de Borel-Cantelli, presque sûrement, $\text{card}(G_n(E)) = 1$ pour une infinité de valeurs de n , et en particulier $\mathbb{P}(T_B < \infty) = 1$. Le temps aléatoire T_B définit presque sûrement un singleton aléatoire $\{x_B\}$ tel que $B_{T_B}(x) = x_B$ pour tout $x \in E$. Il en découle que $B_n(x) = x_B$ pour tout $x \in E$ et tout $n \geq T_B$ car

$$B_n(x) = B_{T_B}((G_{T_B+1} \circ \dots \circ G_n)(x)) = x_B.$$

Cela correspond à une *coalescence* des trajectoires de $(B_n(x))_{n \geq 0}$ quel que soit l'état initial x . Si μ_B désigne la loi de x_B , alors le théorème de convergence dominée entraîne que pour tout $x \in E$ et toute fonction bornée $f : E \rightarrow \mathbb{R}$,

$$\lim_{n \rightarrow \infty} \mathbb{E}[f(B_n(x))] = \lim_{n \rightarrow \infty} \mathbb{E}[f(B_n(x))\mathbf{I}_{\{T_B \leq n\}}] = \mathbb{E}[f(x_B)] = \mu_B f.$$

Comme $\mathbb{E}[f(B_n(x))] = \mathbf{P}^n(x, \cdot)f$, on obtient $\lim_{n \rightarrow \infty} \mathbf{P}^n(x, y) = \mu_B(y)$ pour tous x et y . □

Lorsque G est contractante, on a également $\mathbb{P}(T_A < \infty) = 1$ et T_A définit presque sûrement un singleton aléatoire $\{x_A\}$, vérifiant $A_{T_A}(x) = x_A$ pour tout $x \in E$. Cependant, la définition de $(A_n)_{n \geq 0}$ n'assure pas que $A_n(x) = x_A$ pour $n > T_A$ et tout $x \in E$, et le raisonnement utilisé pour μ_B ne peut pas être utilisé pour la loi μ_A de x_A . Observons que pour tous x et y dans E , la suite $((A_n(x), A_n(y)))_{n \geq 0}$ est un couplage coalescent (les deux composantes sont des chaînes de noyau \mathbf{P} , et restent collées après leur première rencontre), de loi initiale $\delta_x \otimes \delta_y$. Si T_C désigne le temps de couplage associé, alors $T_C \leq T_A$. La suite $((B_n(x), B_n(y)))_{n \geq 0}$ quant à elle est parfois qualifiée de *couplage par le passé*.

En pratique, il n'est pas commode de déterminer le temps de coalescence T_B car il fait intervenir tous les états initiaux possibles. Le résultat suivant fournit une alternative.

Théorème 10.5.2 (Contrôle géométrique du temps de coalescence par monotonie). *Supposons que E soit ordonné et que $g(x, \cdot) : [0, 1] \rightarrow E$ soit constante par morceaux et croissante pour tout $x \in E$. Supposons également que E possède un plus petit élément x_* vérifiant $\varepsilon = \inf_{x \in E} \mathbf{P}(x, x_*) > 0$. Soit σ le temps aléatoire de loi $\mathcal{G}(\varepsilon)$ défini par*

$$\sigma = \inf\{n \geq 1; U_n < \varepsilon\}.$$

Alors $B_n(x) = B_{\sigma-1}(x_)$ pour tout $x \in E$ et tout $n \geq \sigma$. Si μ désigne la loi de $B_{\sigma-1}(x_*)$, alors $\lim_{n \rightarrow \infty} \mathbf{P}^n(x, y) = \mu(y)$ pour tous x et y dans E . Ainsi, μ est l'unique loi invariante de \mathbf{P} .*

La variable aléatoire σ suit la loi géométrique sur \mathbb{N}^* de moyenne $1/\varepsilon$.

Démonstration. On a $\mathbb{P}(G(E) = \{x_*\}) \geq \varepsilon > 0$ par définition de ε et donc G est contractante. La mesure de Lebesgue du premier morceau de la fonction $g(x, \cdot)$ vaut $\mathbf{P}(x, x_*)$. Par conséquent, $g(x, u) = x_*$ pour $u < \varepsilon$ et tout $x \in E$. Notons que $\mathbb{P}(\sigma < \infty) = 1$. La définition de σ assure alors que $G_\sigma(x) = g(x, U_\sigma) = x_*$ pour tout $x \in E$. Ainsi, $G_\sigma(E) = \{x_*\}$ et donc $B_\sigma(E) = \{B_{\sigma-1}(x_*)\}$ est un singleton. Ainsi, $B_n(x) = B_{\sigma-1}(x_*)$ pour tout $x \in E$ et tout $n \geq \sigma$. Il suffit alors de procéder comme dans la preuve du théorème 10.5.1. \square

Comme E est identifiable à \mathbb{N} , on peut le munir d'un ordre total et d'un plus petit élément x_* . On prend pour $g(x, \cdot) : [0, 1] \rightarrow E$ l'inverse de la fonction de répartition de $\mathbf{P}(x, \cdot)$. Lorsque E est fini, l'irréductibilité et l'apériodicité de \mathbf{P} donnent $\varepsilon = \min_{x,y} \mathbf{P}^r(x, y) > 0$ pour un entier r . Le théorème 10.5.2, appliqué au noyau \mathbf{P}^r au lieu de \mathbf{P} , assure que la loi μ de $B_{\sigma-1}(x_*)$ constitue l'unique loi invariante de \mathbf{P}^r . Comme la loi invariante de \mathbf{P} est également invariante pour \mathbf{P}^r , il en découle que μ est l'unique loi invariante de \mathbf{P} . On dispose donc d'un algorithme de simulation exacte de μ , dont la complexité est géométrique. Cependant, si $\text{card}(E)$ est petit ou si \mathbf{P} possède une structure favorable, il est parfois plus efficace de calculer μ en déterminant le vecteur propre à gauche de \mathbf{P} associé à la valeur propre 1 par un algorithme d'algèbre linéaire, puis de simuler μ avec l'algorithme basique.

Réciproquement, partant d'une loi μ sur E fini, il est toujours possible de construire un noyau \mathbf{P} irréductible et apériodique pour lequel μ est invariante (par exemple en utilisant la construction de Metropolis-Hastings). Si la fonction g associée est bien choisie, alors la loi μ peut être simulée en simulant la variable aléatoire $B_{\sigma-1}(x)$ construite à partir d'une puissance \mathbf{P}^r adéquate et d'un état initial x quelconque.

FIXME: ajouter code ici:

Figure 10.7 – Exemple élémentaire d'algorithme de Propp-Wilson en pseudo-code.

10.6 Notes et commentaires

Une analyse probabiliste de la complexité de l'algorithme de tri randomisé *Quick Sort* se trouve par exemple dans le livre de Motwani et Raghavan [MR95]. Le livre monumental de Knuth [Knu05] constitue une référence incontournable pour l'analyse des algorithmes et la simulation de la loi uniforme sur les ensembles classiques comme les permutations ou les partitions. C'est dans la première édition de ce livre datant des années 1960 que Knuth a popularisé l'algorithme de simulation de la loi uniforme sur les permutations, reprenant un article antérieur de Durstenfeld. L'algorithme remonte en fait à Fisher et Yates [FY48]. Il fait désormais partie du folklore, et a été implémenté en standard dans les logiciels de calcul. Cet algorithme correspond exactement au processus des restaurants chinois du chapitre 3.

Pour tout $n \geq 1$ fixé, la marche aléatoire sur le groupe symétrique \mathcal{S}_n dont les pas sont i.i.d. de loi uniforme sur les transpositions, étudiée par Diaconis et Shahshahani [DS81], converge vers la loi uniforme sur \mathcal{S}_n de manière abrupte après environ $n \log(n)$ étapes, comme pour celle du mélange de cartes du chapitre 5. Si $\sigma \in \mathcal{S}_n$ et $\tau = (i, j)$ est une transposition, alors la décomposition en cycles de $\sigma\tau$ s'obtient à partir de celle de σ en fusionnant les cycles de σ contenant i et j s'ils sont différents, ou bien en fissionnant le cycle de σ contenant i et j dans le cas contraire. La chaîne de Diaconis et Shahshahani, traduite sur Π_n en considérant la partition donnée par les supports des cycles, est une chaîne de fragmentation-coalescence. Il est possible de concevoir son noyau de transition de la

manière suivante : sachant que la chaîne est en $P \in \Pi_n$, on tire au hasard uniformément et avec remise i et j dans $\{1, \dots, n\}$, puis on fusionne les blocs de P contenant i et j s'ils sont différents, ou bien on fissionne uniformément le bloc de P contenant i et j dans le cas contraire. Cette chaîne sur Π_n est considérée dans [DMWZZ04]. Plus généralement, au delà des transpositions aléatoire, Berestycki, Schramm et Zeitouni ont établi dans [BSZ11] que pour tout pour tous $n \geq k \geq 2$, la marche aléatoire sur \mathcal{S}_n dont les pas sont des k -cycles i.i.d. uniformes converge vers la loi uniforme sur \mathcal{S}_n après environ $(1/k)n \log(n)$ étapes.

L'algorithme de Stam de simulation de la loi uniforme sur les partitions a été proposé dans [Sta83]. On prendra garde à ne pas confondre les partitions d'un ensemble fini avec la notion de partition d'entier, qui est reliée aux diagrammes de Young ou de Ferrers.

La partie sur les algorithmes de Metropolis-Hastings et de Propp-Wilson est une version abrégée de [BC07, Chapitre 4]. L'algorithme de Metropolis-Hastings a été introduit par Metropolis [MRR⁺53] puis généralisé par Hastings [Has70], et fait aujourd'hui partie des méthodes MCMC (Monte Carlo Markov Chains), qui consistent à utiliser des chaînes de Markov pour approcher des espérances par la méthode de Monte Carlo. L'échantillonneur de Gibbs est un cas particulier de l'algorithme de Metropolis-Hastings. Les méthodes MCMC sont au cœur des implémentations quantitatives de la statistique bayésienne.

En métallurgie, le procédé du recuit consiste à recuire le métal pour échapper aux minima locaux d'énergie et obtenir une structure métallique de basse énergie, garantissant une meilleure solidité. L'algorithme du recuit simulé (simulated annealing) s'en inspire. Une bonne analyse de l'algorithme se trouve par exemple dans le panorama de Catoni [Cat99].

L'algorithme de Propp-Wilson a été proposé par Propp et Wilson [PW96], voir également [PW98] et [LPW09]. D'autres algorithmes de simulation exacte de la loi invariante ont été développés, comme par exemple l'algorithme de Fill [Fil98], qui ne fait pas appel au couplage par le passé, mais plutôt à une méthode du rejet.

Séance 11

Problème du voyageur de commerce

Mots-clés. Concentration ; poissonisation.

Voici trois problèmes phares en optimisation combinatoire (randomisée ou non) :

- Problème du voyageur de commerce (Traveling Salesman Problem ou TSP)
- Arbre couvrant minimal (Minimum Spanning Tree ou MST)
- Appariement euclidien minimal (Minimum Euclidean Matching ou MEM)

Parmi les méthodes utilisées, on compte la sous-additivité, la concentration, la poissonisation, et la méthode objective. Nous allons aborder le premier problème, et mettre en oeuvre certaines techniques. Le problème du voyageur de commerce consiste à trouver une tournée (c'est-à-dire un chemin circulaire¹) de longueur minimale passant par des points prescrits $X_1, \dots, X_n \in \mathbb{R}^d$, $n, d \geq 2$. Cela revient à résoudre sur le groupe symétrique

$$\min_{\sigma \in \Sigma_n} \sum_{k=1}^n |X_{\sigma(k)} - X_{\sigma(k+1)}|$$

où $|\cdot|$ désigne la norme euclidienne de \mathbb{R}^d , et avec la convention $\sigma(n+1) = \sigma(1)$. Il est clair que la notion de distance choisie a une influence sur la solution du problème.

On peut chercher à déterminer la valeur du minimum et/ou un point où il est (presque) atteint. Nous allons nous intéresser à la valeur du minimum, lorsque $(X_k)_{k \geq 1}$ sont des variables aléatoires indépendantes et de même loi μ sur \mathbb{R}^d . On note $L_n = L_n(X_1, \dots, X_n)$ la longueur minimale de la tournée, qui est une fonction de X_1, \dots, X_n .

Théorème 11.0.1 (Bearwood-Halton-Hammersley). *Il existe une constante $0 < \gamma_d < \infty$ qui dépend de $d \geq 2$ telle que si μ est à support compact alors*

$$\frac{L_n(X_1, \dots, X_n)}{n^{(d-1)/d}} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \gamma_d \int_{\mathbb{R}^d} f(x)^{(d-1)/d} dx$$

où f est la densité de la partie absolument continue de μ par rapport à la mes. de Lebesgue.

En particulier $L_n(X_1, \dots, X_n)$ est d'ordre \sqrt{n} en dimension $d = 2$. Nous allons établir ce théorème lorsque μ est la loi uniforme sur le cube $[0, 1]^d$. Nous allons établir que la variable aléatoire $L_n(X_1, \dots, X_n)$ est d'autant plus concentrée autour de son espérance $\mathbb{E}(L_n(X_1, \dots, X_n))$ que n est grand, puis établir que cette espérance est d'ordre $n^{(d-1)/d}$.

1. On parle parfois de circuit hamiltonien.

11.1 Concentration pour le cas uniforme

Théorème 11.1.1 (Concentration autour de la moyenne). *Si μ est la loi uniforme sur le cube $[0, 1]^d$ alors pour tout $n \geq 1$ et tout réel $t \geq 0$ on a*

$$\mathbb{P}(|L_n - \mathbb{E}(L_n)| \geq t) \leq 2 \exp \begin{cases} -\frac{ct^2}{\log(n)} & \text{si } d = 2, \\ -\frac{ct^2}{n^{(d-2)/d}} & \text{si } d > 2, \end{cases}$$

où $c > 0$ est une constante qui ne dépend que de la dimension d . Ainsi,

$$\mathbb{P}\left(\left|\frac{L_n}{n^{(d-1)/d}} - \mathbb{E}\left(\frac{L_n}{n^{(d-1)/d}}\right)\right| \geq t\right) \leq 2 \exp \begin{cases} -\frac{cnt^2}{\log(n)} & \text{si } d = 2, \\ -cnt^2 & \text{si } d > 2, \end{cases}$$

La preuve du théorème 11.1.1 nécessite les deux lemmes suivants.

On dit qu'une variable est concentrée lorsqu'elle reste proche de sa moyenne avec grande probabilité. Une telle propriété peut être obtenue en contrôlant la queue de distribution de la variable, par exemple au moyen de moments, comme dans l'inégalité quadratique de Tchebychev basée sur la variance, l'inégalité exponentielle de Chernoff basée sur la transformée de Laplace (moments exponentiels), ou l'inégalité exponentielle de Bernstein basée sur la variance. L'inégalité de Azuma-Hoeffding fait partie des outils les plus simples. Elle exploite une information sur l'oscillation de la variable (diamètre du support). On note $\text{osc}(f) = \sup f - \inf f = \text{diamètre}(\text{support}(f))$, et on a alors $\text{osc}(f) \leq 2\|f\|_\infty$.

Lemme 11.1.2 (Inégalité de concentration de Azuma-Hoeffding). *Si $Y : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R}$ est une variable aléatoire intégrable, alors pour tout $r \geq 0$,*

$$\mathbb{P}(|Y - \mathbb{E}(Y)| \geq r) \leq 2 \exp\left(-\frac{2r^2}{\text{osc}(d_1)^2 + \dots + \text{osc}(d_n)^2}\right)$$

pour la décomposition télescopique de Doob en somme de différences de martingales

$$Y - \mathbb{E}(Y) = \sum_{k=1}^n \mathbb{E}(Y | \mathcal{F}_k) - \mathbb{E}(Y | \mathcal{F}_{k-1}) = \sum_{k=1}^n d_k$$

associée à une filtration arbitraire d'interpolation $\{\emptyset, \Omega\} = \mathcal{F}_0 \subset \mathcal{F}_1 \subset \dots \subset \mathcal{F}_n = \mathcal{F}$.

Démonstration. Soit U une v.a.r. telle que $\mathbb{E}(U) = 0$ et $a \leq U \leq b$. La convexité de $u \mapsto e^u$ donne, pour tout $t \geq 0$ et tout $a \leq x \leq b$,

$$e^{tx} \leq \frac{x-a}{b-a} e^{tb} + \frac{b-x}{b-a} e^{ta}.$$

En posant $p = -a/(b-a)$ et $f(u) = -pu + \log(1-p+pe^u)$ il vient donc

$$\mathbb{E}(e^{tU}) \leq \frac{b}{b-a} e^{ta} - \frac{a}{b-a} e^{tb} = e^{ta} \left((1-p) + pe^{t(b-a)} \right) = e^{f(t(b-a))}.$$

À présent on a

$$f'(u) = -p + \frac{pe^u}{1-p+pe^u} \quad \text{et} \quad f''(u) = p(1-p) \frac{e^u}{(1-p+pe^u)^2} \leq \frac{1}{4}.$$

Comme $f(0) = f'(0) = 0$, on en déduit que $f(u) \leq u^2/8$ et donc

$$\mathbb{E}(e^{tU}) \leq e^{\frac{t^2}{8}(b-a)^2}.$$

Appliquée à $U = d_k = \mathbb{E}(Y | \mathcal{F}_k) - \mathbb{E}(Y | \mathcal{F}_{k-1})$ sachant \mathcal{F}_{k-1} , cela donne

$$\mathbb{E}(e^{td_k} | \mathcal{F}_{k-1}) \leq e^{\frac{t^2}{8} \text{osc}(d_k)^2}.$$

Ensuite, en écrivant la somme télescopique $Y - \mathbb{E}(Y) = d_n + \dots + d_1$ on obtient

$$\mathbb{E}(e^{t(Y - \mathbb{E}(Y))}) = \mathbb{E}(e^{t(d_{n-1} + \dots + d_1)}) \mathbb{E}(e^{td_n} | \mathcal{F}_{n-1}) \leq \dots \leq e^{\frac{t^2}{8} (\text{osc}(d_1)^2 + \dots + \text{osc}(d_n)^2)}.$$

À présent, pour tout $t, r > 0$, avec $c := \text{osc}(d_1)^2 + \dots + \text{osc}(d_n)^2$, par l'inégalité de Markov,

$$\begin{aligned} \mathbb{P}(Y - \mathbb{E}(Y) \geq r) &= \mathbb{P}(e^{t(Y - \mathbb{E}(Y))} \geq e^{tr}) \\ &\leq e^{-tr} \mathbb{E}(e^{t(Y - \mathbb{E}(Y))}) \\ &\leq e^{-tr + \frac{ct^2}{8}} \\ &\leq e^{\inf_{t>0} (-tr + \frac{ct^2}{8})} \\ &= e^{-\frac{2r^2}{c}}. \end{aligned}$$

En utilisant cela pour Y et $-Y$, on obtient le résultat souhaité pour $\mathbb{P}(|Y - \mathbb{E}(Y)| \geq r)$. \square

Lemme 11.1.3 (Lemme géométrique). *Il existe une constante $c_d > 0$ telle que si X_1, \dots, X_k sont i.i.d. de loi uniforme sur $[0, 1]^d$ alors pour tout $x \in [0, 1]^d$,*

$$g_k(x) := \mathbb{E} \left(\min_{1 \leq i \leq k} |X_i - x| \right) \leq c_d k^{-1/d}.$$

Démonstration. Si $B(x, r)$ désigne la boule de centre x et de rayon $r > 0$ dans \mathbb{R}^d , le volume minimal de $B(x, r) \cap [0, 1]^d$ quand x parcourt $[0, 1]^d$ est atteint lorsque x est un coin du cube $[0, 1]^d$. Lorsque $r \leq 1$, la valeur du minimum est $2^{-d} |B(0, r)| = 2^{-d} |B(0, 1)| r^d$ (dessin). Si $1 < r \leq \sqrt{d}$ alors le volume minimal est atteint quand x est un coin du cube, mais la valeur du minimum est difficile à calculer. Elle reste supérieure ou égale à celle du cas $r = 1$. Ainsi, pour tout $0 < r \leq \sqrt{d}$, ce volume minimal est $\geq 2^{-d} |B(0, 1)| (r/\sqrt{d})^d = a_d r^d$. Donc pour tout $x \in [0, 1]^d$ et tout $0 < r \leq \sqrt{d}$, en utilisant à la fin $(1 - u)^k \leq e^{-ku}$,

$$\begin{aligned} \mathbb{P} \left(\min_{1 \leq i \leq k} |X_i - x| \geq r \right) &= \prod_{i=1}^k \mathbb{P}(X_i \in B(x, r)^c) \\ &= \left(1 - |B(x, r) \cap [0, 1]^d| \right)^k \\ &\leq \left(1 - a_d r^d \right)^k \\ &\leq \exp \left(-a_d k r^d \right). \end{aligned}$$

Ceci reste valable si $r > \sqrt{d}$ car dans ce cas $\mathbb{P}(\min_{1 \leq i \leq k} |X_i - x| \geq r) = 0$. À présent,

$$\mathbb{E} \left(\min_{1 \leq i \leq k} |X_i - x| \right) = \int_0^\infty \mathbb{P} \left(\min_{1 \leq i \leq k} |X_i - x| \geq r \right) dr \quad \text{et} \quad \int_0^\infty e^{-br^d} dr = \frac{\Gamma(1/d)}{db^{1/d}}.$$

\square

Preuve du théorème 11.1.1. Utilisons l'inégalité d'Azuma-Hoeffding du lemme 11.1.2 appliqué à la variable aléatoire $Y = L_n(X_1, \dots, X_n)$ et à la filtration $\mathcal{F}_k = \sigma(X_1, \dots, X_k)$ avec $\mathcal{F}_0 = \{\emptyset, \Omega\}$. Cela ramène le problème à la majoration uniforme de

$$d_k = \mathbb{E}(L_n(X_1, \dots, X_n) | \mathcal{F}_k) - \mathbb{E}(L_n(X_1, \dots, X_n) | \mathcal{F}_{k-1}).$$

Si X'_1, \dots, X'_n est une copie indépendante de X_1, \dots, X_n on remarque que

$$\mathbb{E}(L_n(X_1, \dots, X_k, \dots, X_n) | \mathcal{F}_{k-1}) = \mathbb{E}(L_n(X_1, \dots, X'_k, \dots, X_n) | \mathcal{F}_k)$$

et donc

$$d_k := \mathbb{E}(L_n(X_1, \dots, X_k, \dots, X_n) - L_n(X_1, \dots, X'_k, \dots, X_n) | \mathcal{F}_k).$$

Nous allons maintenant majorer $\|d_k\|_\infty$. Soit L la fonction qui associe à un ensemble fini de \mathbb{R}^d la longueur minimale de la tournée. On a, pour tout $S \subset \mathbb{R}^d$ fini et tout $x \in \mathbb{R}^d$,

$$L(S) \leq L(S \cup \{x\}) \leq L(S) + 2 \min_{y \in S} |x - y|.$$

En appliquant cette inégalité à $S = \{x_1, \dots, x_n\} \setminus \{x_k\}$ et à $x = x_k$ et $x = x'_k$ on obtient

$$\begin{aligned} |L_n(x_1, \dots, x_k, \dots, x_n) - L_n(x_1, \dots, x'_k, \dots, x_n)| &\leq 2 \min_{i \neq k} |x'_k - x_i| + 2 \min_{i \neq k} |x_k - x_i| \\ &\leq 2 \min_{i > k} |x'_k - x_i| + 2 \min_{i > k} |x_k - x_i|. \end{aligned}$$

En posant $g_k(x) := \mathbb{E} \left(\min_{1 \leq i \leq k} |X_i - x| \right)$ on obtient pour $1 \leq k \leq n-1$,

$$|d_k| \leq 2 \mathbb{E} \left(\min_{i > k} |X_k - X_i| + \min_{i > k} |X'_k - X_i| \middle| \mathcal{F}_k \right) = 2g_{n-k}(X_k) + 2\mathbb{E}(g_{n-k}(X'_k)).$$

Le lemme géométrique 11.1.3 donne, pour $d \geq 2$ et $1 \leq k \leq n-1$,

$$\|d_k\|_\infty \leq c_d(n-k)^{-1/d}.$$

Comme $\|d_n\|_\infty \leq c_d$ pour une autre constante car les X_i sont bornées, on obtient

$$\sum_{k=1}^n \|d_k\|_\infty^2 \leq \begin{cases} c_d \log(n) & \text{pour } d = 2, \\ c_d n^{(d-2)/d} & \text{pour } d > 2. \end{cases}$$

□

11.2 Évaluation de la moyenne du cas uniforme

Lemme 11.2.1 (Un bon début). *Si μ est la loi uniforme sur $[0, 1]^d$ alors pour tout $n \geq 1$,*

$$c_d^- n^{(d-1)/d} \leq \mathbb{E}(L_n(X_1, \dots, X_n)) \leq c_d^+ n^{(d-1)/d}$$

où $0 < c_d^\pm < \infty$ sont des constantes qui ne dépendent que de la dimension d .

Démonstration. Le cube $[0, 1]^d$ est l'union de $(1/\varepsilon)^d$ petits cubes isométriques à $[0, \varepsilon]^d$. Avec $\varepsilon = n^{-1/d}$ on obtient que $[0, 1]^d$ peut être recouvert par $\mathcal{O}(n)$ petits cubes de diamètre $\mathcal{O}(n^{-1/d})$. Le principe des tiroirs² entraîne alors que pour tous $x_1, \dots, x_n \in [0, 1]^d$ on a

$$\min_{1 \leq i \neq j \leq n} |x_i - x_j| \leq c_d n^{-1/d}$$

2. *Pigeonhole principle* en anglais : si on dispose n objets dans m boîtes avec $n > m$ alors au moins l'une des boîtes contient deux objets ou plus.

où c_d est une constante qui peut dépendre de la dimension d . Par conséquent, si ℓ_n est le maximum sur x_1, \dots, x_n de la longueur de tournée minimale pour x_1, \dots, x_n , alors

$$\ell_n \leq \ell_{n-1} + 2c_d n^{-1/d}$$

qui donne, pour une nouvelle constante c_d qui peut dépendre de la dimension d ,

$$\|L_n(X_1, \dots, X_n)\|_\infty \leq c_d n^{-1/d+1} = c_d n^{(d-1)/d}. \quad (11.1)$$

Par ailleurs, en reprenant la preuve du lemme 11.1.3, on voit que $|B(x, r) \cap [0, 1]^d|$ est maximal quand x est au centre du cube, d'où, pour tout $x \in [0, 1]^d$ et tout $0 < r \leq 1/2$,

$$\mathbb{P}\left(\min_{1 \leq i \leq n-1} |X_i - x| \geq r\right) \geq (1 - \omega_d r^d)^{n-1}$$

avec $\omega_d := |B(0, 1)|$, d'où, en utilisant l'inégalité $(1 - u)^\alpha \geq 1 - \alpha u$ pour $0 \leq u \leq 1/\alpha$,

$$\mathbb{E}\left(\min_{1 \leq i \leq n-1} |X_i - x|\right) \geq \int_0^{1/2} (1 - \omega_d r^d)^{n-1} dr \geq c_d n^{-1/d}$$

d'où

$$\min_{1 \leq j \leq n} \mathbb{E}\left(\min_{1 \leq i \neq j \leq n} |X_i - X_j|\right) = \min_{1 \leq j \leq n} \mathbb{E}\mathbb{E}\left(\min_{1 \leq i \neq j \leq n} |X_i - X_j| \mid X_j\right) \geq c_d n^{-1/d}$$

où c_d est une constante qui dépend de la dimension d . Or le circuit optimal à travers X_1, \dots, X_n possède n arêtes et passe par chacun des n sommets. On dispose donc d'une écriture $L_n(X_1, \dots, X_n) = \sum_{i=1}^n |X_i - X_{j_i}|$ pour des entiers aléatoires j_1, \dots, j_n , ce qui donne

$$\mathbb{E}(L_n(X_1, \dots, X_n)) \geq \sum_{j=1}^n \mathbb{E}\left(\min_{1 \leq i \neq j \leq n} |X_i - X_j|\right) \geq n c_d n^{-1/d} = c_d n^{(d-1)/d}.$$

□

Théorème 11.2.2 (Le bon résultat). *Si μ est la loi uniforme sur $[0, 1]^d$ alors*

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E}(L_n(X_1, \dots, X_n))}{n^{(d-1)/d}} = \gamma_d$$

où $0 < \gamma_d < \infty$ est un réel qui peut dépendre de d .

Nous savons déjà que $a_n := \mathbb{E}(L_n) \approx n^{(d-1)/d}$ ce qui rend naturel de chercher à établir que $n^{d/(d-1)} a_n$ converge quand n tend vers l'infini. Ce comportement non linéaire empêche l'usage direct d'une technique de sous-additivité. Il est cependant possible de linéariser le problème par poissonisation puis dépoissonisation. L'heuristique est la suivante : si N est une v.a.r. à valeurs entières alors $\mathbb{E}(a_N) \approx \mathbb{E}(N^{(d-1)/d}) \approx \mathbb{E}(N)^{(d-1)/d}$ qui est linéaire en t lorsque $N \sim \text{Poi}(t^{d/(d-1)})$. Par ailleurs si $N \sim \text{Poi}(n)$ alors $a_n \approx \mathbb{E}(a_N)$.

Démonstration. Poissonisation. On note $L(S)$ la longueur minimale de la tournée pour un ensemble fini de points $S = \{x_1, \dots, x_n\} \subset \mathbb{R}^d$, avec la convention $L(S) = 0$ si $\text{Card}(S) \leq 2$. Soit P un processus ponctuel de Poisson sur \mathbb{R}^d de mesure d'intensité Lebesgue. Soit $(Z_t)_{t \geq 0}$ le processus défini par $Z_t = L(P \cap [0, t]^d)$ c'est-à-dire la longueur minimale de la tournée pour les atomes du processus de Poisson P se trouvant dans le cube $[0, t]^d$. Pour tout $n \geq 0$,

$$\text{Loi}(P \mid \text{Card}(P \cap [0, t]^d) = n) = \text{Unif}([0, t]^d)^{\otimes n}.$$

D'autre part, $L(tS) = tL(S)$ et $\text{Card}(P \cap [0, t]^d) \sim \text{Poi}(t^d)$, ce qui donne

$$\mathbb{E}(Z_t) = \mathbb{E}(\mathbb{E}(Z_t | \text{Card}(P \cap [0, t]^d))) = e^{-t^d} \sum_{n=0}^{\infty} t a_n \frac{t^{dn}}{n!} \quad \text{où } a_n := \mathbb{E}(L_n(X_1, \dots, X_n)).$$

Le facteur t devant a_n vient du fait que a_n concerne $[0, 1]^d$ et non pas $[0, t]^d$. Le lemme 11.2.1 entraîne que $a_n = \mathcal{O}(n^{(d-1)/d})$ et donc $t \mapsto \mathbb{E}(Z_t)$ est continue et même analytique. La suite $(a_n)_{n \geq 1}$ peut donc s'obtenir à partir de la donnée de $t \mapsto \mathbb{E}(Z_t)$ au voisinage de $t = 0$. Cependant, nous allons plutôt montrer que $\lim_{t \rightarrow \infty} t^{-1} \mathbb{E}(Z_t)$ existe, ce qui suffira.

Sous additivité. Soit C_1, \dots, C_{k^d} une partition de $[0, t]^d$ en k^d cubes similaires à $[0, t/k]^d$. Pour tous $x_1, \dots, x_n \in \mathbb{R}^d$, les k^d tournées des ensembles $S_i = \{x_1, \dots, x_n\} \cap C_i$ peuvent être concaténées pour former une tournée pour $\{x_1, \dots, x_n\} \cap [0, t]^d$, avec un coût supplémentaire $\mathcal{O}(k^{d-1})$: choisir un point dans chacune des k^d tournées et les connecter avec une tournée de coût $\mathcal{O}(k^{d-1})$ grâce à la borne (11.1) ce qui donne une grande tournée en chapelet. Ainsi, il existe une constante $C > 0$ telle que pour tout $k \geq 1$ et tout $t \geq 0$,

$$L(\{x_1, \dots, x_n\} \cap [0, t]^d) \leq \sum_{i=1}^{k^d} L(\{x_1, \dots, x_n\} \cap C_i) + C t k^{d-1}.$$

Cela donne $\mathbb{E}(Z_t) \leq k^d \mathbb{E}(Z_{t/k}) + C t k^{d-1}$, d'où l'inégalité

$$\frac{\mathbb{E}(Z_{tk})}{(tk)^d} \leq \frac{\mathbb{E}(Z_t)}{t^d} + C t^{1-d} \quad (11.2)$$

En prenant $t = 1$ on obtient

$$0 \leq \gamma := \liminf_{k \rightarrow \infty} \frac{\mathbb{E}(Z_k)}{k^d} \leq \mathbb{E}(Z_1) + C < \infty.$$

Par définition de γ , pour tout $\varepsilon > 0$ on peut choisir k_0 assez grand pour que

$$\frac{\mathbb{E}(Z_{k_0})}{k_0^d} + C k_0^{1-d} \leq \gamma + \varepsilon.$$

Comme $t \mapsto \mathbb{E}(Z_t)$ est continue, on peut choisir $\delta > 0$ tel que pour tout $k_0 < t < k_0 + \delta$,

$$\frac{\mathbb{E}(Z_t)}{t^d} + C t^{1-d} \leq \gamma + 2\varepsilon. \quad (11.3)$$

Grâce à (11.2) on voit que (11.3) a lieu pour tout $k k_0 < t < k(k_0 + \delta)$. Or pour $k > k_0/\delta$ les intervalles $I_k :=]k k_0, k(k_0 + \delta)[$ et I_{k+1} se recouvrent. On en déduit que (11.3) a lieu pour tout $t \geq \lceil k_0^2/\delta \rceil$, ce qui implique en particulier

$$\overline{\lim}_{t \rightarrow \infty} \frac{\mathbb{E}(Z_t)}{t^d} \leq \liminf_{t \rightarrow \infty} \frac{\mathbb{E}(Z_t)}{t^d} + 2\varepsilon = \gamma + 2\varepsilon.$$

Comme $\varepsilon > 0$ est arbitraire, on obtient donc

$$\lim_{t \rightarrow \infty} \frac{\mathbb{E}(Z_t)}{t^d} = \gamma.$$

Ainsi, le développement en série de $\mathbb{E}(Z_t)$ donne, après le changement de variable $u = t^d$,

$$\sum_{k=0}^{\infty} e^{-u} a_k \frac{u^k}{k!} \sim_{u \rightarrow \infty} \gamma u^{(d-1)/d}. \quad (11.4)$$

Dépoissonisation. On commence par montrer que $n \mapsto a_n$ est assez régulière. Comme il est possible de concaténer un circuit pour X_1, \dots, X_m et un circuit pour X_{m+1}, \dots, X_{m+n} avec un coût inférieur à $2\sqrt{d}$ (diamètre du cube) on en déduit que

$$a_{n+m} \leq a_n + a_m + 2\sqrt{d}$$

ce qui donne $0 \leq a_n - a_k \leq a_{n-k} + 2\sqrt{d}$ pour tout $0 \leq k \leq n$. Or le lemme 11.2.1 donne $a_n = \mathcal{O}(n^{(d-1)/d})$ et on en déduit

$$|a_n - a_k| \leq |a_{n-k} + 2\sqrt{d}| \leq c|n - k|^{(d-1)/d}.$$

Si N est une v.a.r. de Poisson de moyenne n alors $\mathbb{E}(a_N) = \sum_{k=0}^{\infty} a_k e^{-n} \frac{n^k}{k!}$ et donc

$$\begin{aligned} |a_n - \mathbb{E}(a_N)| &\leq \sum_{k=0}^{\infty} |a_n - a_k| e^{-n} \frac{n^k}{k!} \\ &\leq c \sum_{k=0}^{\infty} |n - k|^{(d-1)/d} e^{-n} \frac{n^k}{k!} \\ &= c \mathbb{E}(|N - \mathbb{E}(N)|^{(d-1)/d}) \\ &\leq c \mathbb{E}(|N - \mathbb{E}(N)|^2)^{(d-1)/(2d)} \\ &= c \text{Var}(N)^{(d-1)/(2d)} \\ &= \mathcal{O}(n^{(d-1)/(2d)}) = o(n^{(d-1)/d}). \end{aligned}$$

Or (11.4) avec $t = n$ donne $\mathbb{E}(a_N) = \sum_{k=0}^{\infty} a_k e^{-n} n^k / k! \sim_{n \rightarrow \infty} \gamma n^{(d-1)/d}$ et donc

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E}(L_n(X_1, \dots, X_n))}{n^{(d-1)/d}} = \lim_{n \rightarrow \infty} \frac{a_n}{n^{(d-1)/d}} = \gamma.$$

Le fait que $\gamma > 0$ découle de la borne inférieure du lemme 11.2.1. □

11.3 Preuve du cas uniforme

On se place dans le cas où μ suit la loi uniforme sur le cube $[0, 1]^d$. Par le théorème 11.1.1 et le premier lemme de Borel-Cantelli, presque sûrement,

$$L_n - \mathbb{E}L_n = \begin{cases} O_{n \rightarrow \infty}(\log(n)) & \text{si } d = 2, \\ O_{n \rightarrow \infty}(n^{(d-2)/(2d)} \sqrt{\log(n)}) & \text{si } d > 2 \end{cases}$$

tandis que par le théorème 11.2.2 on a $\mathbb{E}(L_n(X_1, \dots, X_n)) \sim_{n \rightarrow \infty} \gamma n^{(d-1)/d}$ pour un $0 < \gamma < \infty$. Ceci achève la preuve du théorème 11.0.1 dans le cas uniforme.

11.4 Courbes qui remplissent le plan

Une manière d'aborder le problème en dimension $d = 2$ consiste à utiliser une courbe qui remplit l'espace. Soit donc $\psi : [0, 1] \rightarrow [0, 1]^2$ une fonction surjective pour laquelle pour tout $x \in [0, 1]^2$ on peut rapidement calculer un $t \in [0, 1]$ tel que $\psi(t) = x$. Pour $x_1, \dots, x_n \in [0, 1]^2$ on considère le circuit $x_{\sigma(1)}, \dots, x_{\sigma(n)}$ où $\sigma \in \Sigma_n$ est la permutation associée la statistique d'ordre $t_{\sigma(1)} \leq \dots \leq t_{\sigma(n)}$ où $t_1, \dots, t_n \in [0, 1]$ sont tels que $\psi(t_i) = x_i$ pour

chaque $1 \leq i \leq n$. Le circuit obtenu est d'autant plus intéressant que ψ est régulière. On dit que ψ est Hölder lorsqu'il existe des réels $c_\psi, h_\psi > 0$ tels que pour tous $x, y \in [0, 1]$,

$$|\psi(s) - \psi(t)| \leq c_\psi |x - y|^{h_\psi}.$$

Comme ψ est surjective, l'image de l'union d'intervalles $[0, 1] = \cup_{i=0}^{k-1} [i/k, (i+1)/k]$ par ψ est $[0, 1]^2$ tout entier, et donc au moins un de ces intervalles a une image de diamètre supérieur à $2/\sqrt{k\pi}$, ce qui montre que nécessairement $h \leq 1/2$. Il se trouve qu'on sait construire des fonctions $\psi : [0, 1] \rightarrow [0, 1]^2$ surjectives et Hölder d'ordre $h_\psi = 1/2$. On peut alors montrer qu'il existe une constante $c = c_\psi^2$, telle que pour tous $x_1, \dots, x_n \in [0, 1]^2$, il existe $\sigma \in \Sigma_n$ telle que (avec la convention $\sigma(n+1) = \sigma(1)$)

$$\sum_{k=1}^n |x_{\sigma(k)} - x_{\sigma(k+1)}|^2 < c.$$

Notons que parmi n points dans $[0, 1]^2$ il y a forcément deux points à distance $\mathcal{O}(n^{-1/2})$ ce qui donne seulement $\sum_{k=1}^n |x_{\sigma(k)} - x_{\sigma(k+1)}|^2 = \mathcal{O}(\log(n))$. Cependant, les deux méthodes donnent $L_n(x_1, \dots, x_n) = \mathcal{O}(\sqrt{n})$ (inégalité de Cauchy-Schwarz pour la première).

11.5 Notes et commentaires

Extrait de [Ste97], voir aussi [Yuk98]. Pour Azuma, voir [McD89, McD98]. Jillian Beardwood et John H. Halton étaient des étudiants en thèse de John M. Hammersley (1920 – 2004) à Oxford. Le théorème 11.0.1 date de 1959. On sait que $\gamma_2 \approx 0,7$ et que $\gamma_d \sim_{d \rightarrow \infty} \sqrt{d}2\pi e$. Avram et Yao ont établi en 1991 que le résultat reste valable si μ possède un moment fini d'ordre³ $> (d-1)/d$. Rhee a montré en 1993 que la finitude du moment d'ordre $(d-1)/d$ ne suffit pas. Le théorème 11.0.1 peut être démontré par réductions successives en se ramenant au cas où μ est uniforme sur le cube $[0, 1]^d$.

En théorie de la complexité algorithmique, la complexité d'un algorithme est le coût d'exécution en fonction de la taille (n pour TSP) du problème. Un algorithme polynomial est préférable à un algorithme exponentiel au delà d'une certaine taille. On dit qu'un problème est NP lorsqu'il est possible de vérifier la validité d'une solution avec un algorithme polynomial. L'explosion combinatoire fait qu'un problème NP n'est pas automatiquement résoluble par un algorithme polynomial en testant toutes les solutions. On dit qu'un problème est NP-complet lorsque le problème est au moins aussi difficile à résoudre que tout autre problème NP, c'est-à-dire que tout problème NP se réduit à celui-ci avec un algorithme polynomial. Les problèmes NP-complets sont donc des problèmes clés. À l'heure actuelle, tous les algorithmes connus pour résoudre les problèmes NP-complets sont exponentiels, ce qui les rend assez rapidement inexploitable. La question ouverte la plus fameuse de l'informatique consiste à trouver un algorithme polynomial pour résoudre un problème NP complet. On sait démontrer que TSP est NP-complet, et on ne connaît pas d'algorithme de résolution polynomial. Il est cependant possible de rechercher une solution approchée, par exemple avec un algorithme stochastique générique comme l'algorithme du recuit simulé. Contrairement à TSP, les deux autres problèmes phares de l'optimisation combinatoire (MST et EMM) sont résolubles en temps polynomial.

3. C'est-à-dire $\mathbb{E}(\|X_1\|^r) < \infty$ pour un $r > (d-1)/d$.

Bibliographie

- [ABR08] L. Addario-Berry et B. A. Reed – « Ballot theorems, old and new », *Horizons of combinatorics*, Bolyai Soc. Math. Stud., vol. 17, Springer, Berlin, 2008, p. 9–35.
- [ABT03] R. Arratia, A. D. Barbour et S. Tavaré – *Logarithmic combinatorial structures : a probabilistic approach*, EMS Monographs in Mathematics, European Mathematical Society (EMS), Zürich, 2003.
- [AD86] D. Aldous et P. Diaconis – « Shuffling cards and stopping times », *Amer. Math. Monthly* **93** (1986), no. 5, p. 333–348.
- [AF01] D. J. Aldous et J. A. Fill – « Reversible Markov Chains and Random Walks on Graphs », cours non publié mais disponible sous forme électronique <http://www.stat.berkeley.edu/~aldous/RWG/book.html>, 1994-2001.
- [AGZ10] G. W. Anderson, A. Guionnet et O. Zeitouni – *An introduction to random matrices*, Cambridge Studies in Advanced Mathematics, vol. 118, Cambridge University Press, Cambridge, 2010.
- [Ald85] D. J. Aldous – « Exchangeability and related topics », École d’été de probabilités de Saint-Flour, XIII—1983, Lecture Notes in Math., vol. 1117, Springer, Berlin, 1985, p. 1–198.
- [AN04] K. B. Athreya et P. E. Ney – *Branching processes*, Dover Publications Inc., Mineola, NY, 2004, Reprint of the 1972 original [Springer, New York; MR0373040].
- [And] D. André – « Solution directe du problème résolu par M. Bertrand ».
- [BC07] B. Bercu et D. Chafaï – *Modélisation stochastique et simulation, cours et applications*, Collection Sciences Sup, Dunod, 2007, Mathématiques appliquées/SMAI.
- [BD92] D. Bayer et P. Diaconis – « Trailing the dovetail shuffle to its lair », *Ann. Appl. Probab.* **2** (1992), no. 2, p. 294–313.
- [BDMT10] V. Bansaye, J.-F. Delmas, L. Marsalle et V. C. Tran – « Limit theorems for Markov processes indexed by continuous time Galton-Watson trees », prépublication arXiv:0911.1973 [math.PR], 2010.
- [Ber06] J. Bertoin – *Random fragmentation and coagulation processes*, Cambridge Studies in Advanced Mathematics, vol. 102, Cambridge University Press, Cambridge, 2006.
- [Ber09] N. Berestycki – *Recent progress in coalescent theory*, Ensaios Matemáticos [Mathematical Surveys], vol. 16, Sociedade Brasileira de Matemática, Rio de Janeiro, 2009.
- [BG04] J. Bertoin et C. Goldschmidt – « Dual random fragmentation and coagulation and an application to the genealogy of Yule processes », Mathematics and computer science. III, Trends Math., Birkhäuser, Basel, 2004, p. 295–308.
- [BHJ92] A. D. Barbour, L. Holst et S. Janson – *Poisson approximation*, Oxford Studies in Probability, vol. 2, The Clarendon Press Oxford University Press, New York, 1992, Oxford Science Publications.
- [BS10] Z. Bai et J. W. Silverstein – *Spectral analysis of large dimensional random matrices*, second éd., Springer Series in Statistics, Springer, New York, 2010.
- [BSZ11] N. Berestycki, O. Schramm et O. Zeitouni – « Mixing times for random k-cycles and coalescence-fragmentation chains », *Ann. Probab.* **39** (2011), no. 5, p. 1815–1843.
- [Cat99] O. Catoni – « Simulated annealing algorithms and Markov chains with rare transitions », Séminaire de Probabilités, XXXIII, Lecture Notes in Math., vol. 1709, Springer, Berlin, 1999, p. 69–119.
- [CDGC80] M. Cottrell, C. Duhamel et V. Genon-Catalot – *Exercices de probabilités*, Collections DIA., 1980, Avec rappels de cours. [With course reviews], With a preface by Gabriel Ruget.
- [CDM12] D. Chafaï, Y. Doumerc et F. Malrieu – « Processus des restaurants chinois et loi d’ewens », article didactique sous à RMS – La revue de la filière mathématiques, 2012.
- [CMP10] D. Chafaï, F. Malrieu et K. Paroux – « On the long time behavior of the TCP window size process », *Stochastic Processes and their Applications* **8** (2010), no. 120, p. 1518–1534.

- [Dem05] A. Dembo – « Favorite points, cover times and fractals », Lectures on probability theory and statistics, Lecture Notes in Math., vol. 1869, Springer, Berlin, 2005, p. 1–101.
- [DFP92] P. Diaconis, J. A. Fill et J. Pitman – « Analysis of top to random shuffles », *Combin. Probab. Comput.* **1** (1992), no. 2, p. 135–155.
- [DJ06] J.-F. Delmas et B. Jourdain – *Modèles aléatoires*, Mathématiques & Applications (Berlin) [Mathematics & Applications], vol. 57, Springer-Verlag, Berlin, 2006, Applications aux sciences de l'ingénieur et du vivant. [Applications to engineering and the life sciences].
- [DM47] A. Dvoretzky et T. Motzkin – « A problem of arrangements », *Duke Math. J.* **14** (1947), p. 305–313.
- [DMP95] P. Diaconis, M. McGrath et J. Pitman – « Riffle shuffles, cycles, and descents », *Combinatorica* **15** (1995), no. 1, p. 11–29.
- [DMWZZ04] P. Diaconis, E. Mayer-Wolf, O. Zeitouni et M. P. W. Zerner – « The Poisson-Dirichlet law is the unique invariant distribution for uniform split-merge transformations », *Ann. Probab.* **32** (2004), no. 1B, p. 915–938.
- [DP13] A. V. Dumas et V. G. Papanicolaou – « Asymptotics of the rising moments for the coupon collector's problem », *Electron. J. Probab.* **18** (2013), p. no. 41, 15.
- [DS81] P. Diaconis et M. Shahshahani – « Generating a random permutation with random transpositions », *Z. Wahrsch. Verw. Gebiete* **57** (1981), no. 2, p. 159–179.
- [DS84] P. G. Doyle et J. L. Snell – *Random walks and electric networks*, Carus Mathematical Monographs, vol. 22, Mathematical Association of America, Washington, DC, 1984.
- [Dur08] R. Durrett – *Probability models for DNA sequence evolution*, second éd., Probability and its Applications (New York), Springer, New York, 2008.
- [DVJ03] D. J. Daley et D. Vere-Jones – *An introduction to the theory of point processes. Vol. I*, second éd., Probability and its Applications (New York), Springer-Verlag, New York, 2003, Elementary theory and methods.
- [Dwa69] M. Dwass – « The total progeny in a branching process and a related random walk. », *J. Appl. Probability* **6** (1969), p. 682–686.
- [Ewe04] W. J. Ewens – *Mathematical population genetics. I*, second éd., Interdisciplinary Applied Mathematics, vol. 27, Springer-Verlag, New York, 2004, Theoretical introduction.
- [Fel68] W. Feller – *An introduction to probability theory and its applications. Vol. I*, Third edition, John Wiley & Sons Inc., New York, 1968.
- [Fel71] —, *An introduction to probability theory and its applications. Vol. II.*, Second edition, John Wiley & Sons Inc., New York, 1971.
- [Fil98] J. A. Fill – « An interruptible algorithm for perfect sampling via Markov chains », *Ann. Appl. Probab.* **8** (1998), no. 1, p. 131–162.
- [FY48] R. A. Fisher et F. Yates – *Statistical Tables for Biological, Agricultural and Medical Research*, Oliver and Boyd, London, 1948, 3d ed.
- [Har02] T. E. Harris – *The theory of branching processes*, Dover Phoenix Editions, Dover Publications Inc., Mineola, NY, 2002, Corrected reprint of the 1963 original [Springer, Berlin ; MR0163361 (29 #664)].
- [Has70] W. K. Hastings – « Monte Carlo sampling methods using Markov chains and their applications », *Biometrika* (1970), no. 57, p. 97–109.
- [HJV07] P. Haccou, P. Jagers et V. A. Vatutin – *Branching processes : variation, growth, and extinction of populations*, Cambridge Studies in Adaptive Dynamics, Cambridge University Press, Cambridge, 2007.
- [HKPV06] J. B. Hough, M. Krishnapur, Y. Peres et B. Virág – « Determinantal processes and independence », *Probab. Surv.* **3** (2006), p. 206–229.
- [HO74] A. G. Hawkes et D. Oakes – « A cluster process representation of a self-exciting process », *J. Appl. Probability* **11** (1974), p. 493–503.
- [Hol01] L. Holst – « Extreme value distributions for random coupon collector and birthday problems », *Extremes* **4** (2001), no. 2, p. 129–145 (2002).
- [Jac06] M. Jacobsen – *Point process theory and applications*, Probability and its Applications, Birkhäuser Boston Inc., Boston, MA, 2006, Marked point and piecewise deterministic processes.
- [JKB97] N. L. Johnson, S. Kotz et N. Balakrishnan – *Discrete multivariate distributions*, Wiley Series in Probability and Statistics : Applied Probability and Statistics, John Wiley & Sons Inc., New York, 1997, A Wiley-Interscience Publication.

- [Kin80] J. F. C. Kingman – *Mathematics of genetic diversity*, CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 34, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pa., 1980.
- [Kin93] —, *Poisson processes*, Oxford Studies in Probability, vol. 3, The Clarendon Press Oxford University Press, New York, 1993, Oxford Science Publications.
- [Knu05] D. E. Knuth – *The art of computer programming. Vol. 1–4*, Addison-Wesley, 2005.
- [LPW09] D. A. Levin, Y. Peres et E. L. Wilmer – *Markov chains and mixing times*, American Mathematical Society, Providence, RI, 2009, With a chapter by James G. Propp and David B. Wilson.
- [LvL08] A. H. Löpker et J. S. H. van Leeuwen – « Transient moments of the TCP window size process », *J. Appl. Probab.* **45** (2008), no. 1, p. 163–175.
- [Mac75] O. Macchi – « The coincidence approach to stochastic point processes », *Advances in Appl. Probability* **7** (1975), p. 83–122.
- [Mar04] J.-F. Marckert – « Structures arborescentes, algorithmes », Habilitation à diriger des recherches, disponible sur la page de l’auteur, 2004.
- [McD89] C. McDiarmid – « On the method of bounded differences », *Surveys in combinatorics*, 1989 (Norwich, 1989), London Math. Soc. Lecture Note Ser., vol. 141, Cambridge Univ. Press, Cambridge, 1989, p. 148–188.
- [McD98] —, « Concentration », *Probabilistic methods for algorithmic discrete mathematics*, Algorithms Combin., vol. 16, Springer, Berlin, 1998, p. 195–248.
- [MR95] R. Motwani et P. Raghavan – *Randomized algorithms*, Cambridge University Press, Cambridge, 1995.
- [MRR⁺53] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller et E. Teller – « Equations of State Calculations by Fast Computing Machines », *Journal of Chemical Physics* (1953), no. 21, p. 1087–1092.
- [MV12] S. Méléard et D. Villemonais – « Quasi-stationary distributions and population processes », *Probab. Surv.* **9** (2012), p. 340–410.
- [Nor98] J. R. Norris – *Markov chains*, Cambridge Series in Statistical and Probabilistic Mathematics, vol. 2, Cambridge University Press, Cambridge, 1998, Reprint of 1997 original.
- [NS60] D. J. Newman et L. Shepp – « The double dixie cup problem », *Amer. Math. Monthly* **67** (1960), p. 58–61.
- [OK08] T. J. Ott et J. H. B. Kemperman – « Transient behavior of processes in the TCP paradigm », *Probab. Engng. Inform. Sci.* **22** (2008), no. 3, p. 431–471.
- [Per07] B. Perthame – *Transport equations in biology*, Frontiers in Mathematics, Birkhäuser Verlag, Basel, 2007.
- [Pit06] J. Pitman – *Combinatorial stochastic processes*, Lecture Notes in Mathematics, vol. 1875, Springer-Verlag, Berlin, 2006, Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002, With a foreword by Jean Picard.
- [PW96] J. D. Propp et D. B. Wilson – « Exact sampling with coupled Markov chains and applications to statistical mechanics », *Proceedings of the Seventh International Conference on Random Structures and Algorithms (Atlanta, GA, 1995)*, vol. 9, 1996, p. 223–252.
- [PW98] —, « Coupling from the past : a user’s guide », *Microsurveys in discrete probability* (Princeton, NJ, 1997), DIMACS Ser. Discrete Math. Theoret. Comput. Sci., vol. 41, Amer. Math. Soc., Providence, RI, 1998, p. 181–192.
- [Rob00] P. Robert – *Réseaux et files d’attente : méthodes probabilistes*, Mathématiques & Applications (Berlin) [Mathematics & Applications], vol. 35, Springer-Verlag, Berlin, 2000.
- [Rug01] C. Ruget (éd.) – *Mathématiques en situation*, SCOPOS, vol. 11, Springer-Verlag, Berlin, 2001, Issues de l’épreuve de modélisation de l’agrégation. [From the Examination in Modelling for the Agrégation].
- [SKM87] D. Stoyan, W. S. Kendall et J. Mecke – *Stochastic geometry and its applications*, Wiley Series in Probability and Mathematical Statistics : Applied Probability and Statistics, John Wiley & Sons Ltd., Chichester, 1987, With a foreword by D. G. Kendall.
- [Sta83] A. J. Stam – « Generation of a random partition of a finite set by an urn model », *J. Combin. Theory Ser. A* **35** (1983), no. 2, p. 231–240.
- [Ste94] J. M. Steele – « Le Cam’s inequality and Poisson approximations », *Amer. Math. Monthly* **101** (1994), no. 1, p. 48–54.

- [Ste97] — , *Probability theory and combinatorial optimization*, CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 69, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997.
- [Tav04] S. Tavaré – « Ancestral inference in population genetics », Lectures on probability theory and statistics, Lecture Notes in Math., vol. 1837, Springer, Berlin, 2004, p. 1–188.
- [Vil03] C. Villani – *Topics in optimal transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003.
- [Yuk98] J. E. Yukich – *Probability theory of classical Euclidean optimization problems*, Lecture Notes in Mathematics, vol. 1675, Springer-Verlag, Berlin, 1998.