Recueil de modèles stochastiques

Version préliminaire datée du 29 octobre 2014
À haute teneur en coquilles

Djalil Chafaï, professeur à l'Université Paris-Dauphine, France
Florent Malrieu, professeur à l'Université de Tours, France

Recueil de modèles stochastiques – Page i/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Recueil de modèles stochastiques – Page ii/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Recueil de modèles stochastiques – Page i/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Avant-propos

« *…I prefer concrete things and I don't like to learn more about abstract stuff than I absolutely have to. …* ». Marc Kac (1914 – 1984), in *Enigmas of Chance : An Autobiography* (1985).

Ce recueil de modèles stochastiques puise sa source dans les cours de Master de mathématiques et de préparation à l'agrégation de mathématiques, que nous avons dispensés, pendant plusieurs années, aux étudiants des universités de Toulouse, Rennes, Marne-la-Vallée, Paris-Dauphine, et Tours. Le parti pris est de polariser la rédaction par les modèles plutôt que par les outils, et de consacrer chaque séance ou chapitre à un modèle. Bien que parfois reliés, les chapitres sont essentiellement autonomes. Ils ne contiennent pas de rappels de cours sur les outils fondamentaux comme les théorèmes limites, les martingales, et les chaînes de Markov, pour lesquels d'excellentes références sont disponibles. Chaque chapitre commence par une liste de mots-clés et se termine par quelques notes et commentaires. La liste des thèmes abordés n'a rien de canonique ni d'exhaustif, mais constitue un panorama varié, que nous espérons enrichissant.

La qualité de ce recueil, aussi bien sur le fond que sur la forme, doit beaucoup à nos tous premiers lecteurs, petits et grands. Grand merci, donc, aux étudiants des universités de Toulouse, Rennes, Marne-la-Vallée, Paris-Dauphine, et Tours, ainsi qu'à Amine Asselah, Jean-Baptiste Bardet, Marie-Line Chabanol, Bertrand Cloez, Yan Doumerc, Pierre Tarrès, Marie-Noémie Thai, Frédérique Watbled, et Pierre-André Zitt, FIXME : **`mettre à jour cette liste!`**.

Bonne lecture !

Djalil Chafaï et Florent Malrieu
Vincennes et Tours, automne 2014

Recueil de modèles stochastiques – Page ii/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Recueil de modèles stochastiques – Page iii/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Table des matières

Table des matières **iii**

1 Pile, face, coupons **1**
1.1 Jeu de pile ou face . . . 1
1.2 Approximation de la loi binomiale par la loi normale . . . 3
1.3 Distance en variation totale . . . 4
1.4 Approximation de la loi binomiale par la loi de Poisson . . . 6
1.5 Problème du collectionneur de coupons . . . 7
1.6 Notes et commentaires . . . 11

2 Marches aléatoires **13**
2.1 Marche aléatoire simple sur la droite . . . 13
2.2 Marche aléatoire simple symétrique dans l'espace . . . 18
2.3 Problème de Dirichlet et champ libre gaussien . . . 20
2.4 Marche aléatoire sur le groupe symétrique . . . 23
2.5 Notes et commentaires . . . 25

3 Modèle d'Ehrenfest **27**
3.1 Modèle microscopique . . . 27
3.2 Urne d'Ehrenfest . . . 31
3.3 Espérance et variance de la pression au cours du temps . . . 32
3.4 Fluctuations autour de la moyenne à l'équilibre . . . 33
3.5 Quand reverra-t-on le vide ? . . . 34
3.6 Notes et commentaires . . . 35

4 Modèle de Fisher-Wright **37**
4.1 Modèle de Fisher-Wright et fixation . . . 38
4.2 Modèle de Fisher-Wright avec mutations . . . 41
4.3 Modèle de Fisher-Wright avec sélection . . . 43
4.4 Modèle de Cannings . . . 44
4.5 Modèle de Moran . . . 46
4.6 Notes et commentaires . . . 46

5 Branchement **49**
5.1 Extinction et phénomène de seuil . . . 50
5.2 Étude des trois cas possibles . . . 52
5.3 Taille de l'arbre en régimes critique et sous-critique . . . 55
5.4 Immigration . . . 58

5.5 Notes et commentaires . . . 59

6 Percolation **61**
6.1 Percolation dans un graphe . . . 61
6.2 Graphe de Bethe (arbre régulier) . . . 63
6.3 Graphe euclidien (grille) . . . 64
6.4 Graphe complet et modèle de Erdős-Rényi . . . 67
6.5 Notes et commentaires . . . 69

7 Renforcement **71**
7.1 Urne de Pólya . . . 71
7.2 Graphe à attachement préférentiel de Barabási-Albert . . . 74
7.3 Marche aléatoire renforcée . . . 76
7.4 Théorème de Rubin sur les urnes de Pólya généralisées . . . 79
7.5 Notes et commentaires . . . 80

8 Simulation de lois discrètes **83**
8.1 Algorithme basique pour les lois discrètes . . . 83
8.2 Permutations et partitions aléatoires . . . 84
8.3 Graphes aléatoires . . . 86
8.4 Algorithme de Metropolis-Hastings . . . 88
8.5 Algorithme du recuit simulé . . . 89
8.6 Algorithme de Propp-Wilson . . . 90
8.7 Notes et commentaires . . . 92

9 Restaurants chinois **95**
9.1 Lois d'Ewens . . . 96
9.2 Nombre de tables . . . 98
9.3 Tables extrêmes . . . 100
9.4 Compléments de combinatoire . . . 102
9.5 Notes et commentaires . . . 103

10 Généalogies et coalescence **105**
10.1 Modèle généalogique à temps continu . . . 107
10.2 Longueur de l'arbre généalogique . . . 109
10.3 Mutations . . . 111
10.4 Coalescent de Kingman . . . 113
10.5 Notes et commentaires . . . 114

11 Agrégation limitée par diffusion interne **115**
11.1 Modèle d'agrégation par diffusion . . . 115
11.2 Modèle unidimensionnel . . . 116
11.3 Convergence presque sûre en dimension 1 . . . 116
11.4 Convergence en loi en dimension 1 . . . 117
11.5 Dimensions supérieures . . . 119
11.6 Notes et commentaires . . . 120

12 Chaînes de Markov cachées **123**
12.1 Algorithme «forward-backward» . . . 123
12.2 Filtre de Kalman . . . 127
12.3 Notes et commentaires . . . 129

13 Algorithme EM et mélanges **131**

Recueil de modèles stochastiques – Page iv/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

13.1 Nids de mouettes . . . 131
13.2 Situation favorable mais irréaliste . . . 133
13.3 Résolution du vrai problème . . . 133
13.4 Notes et commentaires . . . 135

14 Records, extrêmes, et recrutements ! 137
14.1 Élitisme . . . 137
14.2 Au-dessus de la moyenne . . . 138
14.3 Cas de la loi exponentielle . . . 140
14.4 Cas de la loi de Pareto . . . 142
14.5 Notes et commentaires . . . 143

15 Polymères dirigés en environnement aléatoire 145
15.1 Modèle et résultats principaux . . . 145
15.2 Fonction de partition normalisée . . . 148
15.3 Borne inférieure en grandes dimensions . . . 150
15.4 Énergie libre moyenne . . . 151
15.5 Borne supérieure sur le paramètre critique . . . 153
15.6 Notes et commentaires . . . 154

16 Matrices aléatoires 155
16.1 Théorème de Wigner . . . 156
16.2 Convergence des moments et convergence étroite . . . 158
16.3 Moments de la loi du demi-cercle . . . 159
16.4 Convergence des moments et théorème de Wigner faible . . . 160
16.5 Retour aux matrices de covariance empiriques . . . 162
16.6 Quelques outils . . . 163
16.7 Notes et commentaires . . . 165

17 Problème du voyageur de commerce 167
17.1 Concentration pour le cas uniforme . . . 169
17.2 Évaluation de la moyenne du cas uniforme . . . 171
17.3 Preuve du cas uniforme . . . 174
17.4 Notes et commentaires . . . 174

18 File d'attente M/M/Infini 177
18.1 Lois exponentielles . . . 177
18.2 Temps de demi-vie d'un fragment de métal radioactif . . . 178
18.3 Loi du nombre d'atomes . . . 179
18.4 Comportement en temps long . . . 182
18.5 Notes et commentaires . . . 184

19 Ruine d'une compagnie d'assurance 187
19.1 Processus de Poisson composé . . . 188
19.2 La ruine est-elle presque sûre ? . . . 189
19.3 Expression de la probabilité de ruine . . . 190
19.4 Sinistres exponentiels . . . 191
19.5 Notes et commentaires . . . 192

20 Naissances et assassinats 193
20.1 Condition suffisante de stabilité . . . 194
20.2 Condition suffisante d'instabilité . . . 195
20.3 Notes et commentaires . . . 197

Recueil de modèles stochastiques – Page v/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

21 Croissance et fragmentation **199**
21.1 Processus TCP window-size en informatique . . . 199
21.2 Intensité des sauts variable selon la position . . . 204
21.3 Branchement et croissance-fragmentation . . . 205
21.4 Notes et commentaires . . . 207

22 Modèle du télégraphe **209**
22.1 Aspects markoviens du processus complet . . . 210
22.2 Lien entre équation et processus . . . 212
22.3 Modification ergodique . . . 212
22.4 Notes et commentaires . . . 215

23 Problème de Dirichlet **217**
23.1 Interprétation probabiliste . . . 219
23.2 Cas de la boule unité . . . 220
23.3 Approches numériques . . . 221
23.4 Demi-plan supérieur : du discret au continu . . . 221
23.5 Récurrence, transience, et fonctions harmoniques . . . 223
23.6 Notes et commentaires . . . 224

24 Processus d'Ornstein-Uhlenbeck **225**
24.1 Décomposition spectrale . . . 228
24.2 Entropie, couplage, et temps long . . . 230
24.3 Inégalité de Sobolev logarithmique . . . 231
24.4 Notes et commentaires . . . 233

25 Modèles de diffusion **235**
25.1 Modèle de Langevin . . . 235
25.2 Processus de Langevin confiné . . . 236
25.3 Notes et commentaires . . . 242

26 Des chaînes de Markov aux processus de diffusion **245**
26.1 Paresse et échantillonnage . . . 245
26.2 Diffusion sur un intervalle . . . 246
26.3 Convergence d'une chaîne de Markov vers une diffusion . . . 247
26.4 Application aux processus de Fisher-Wright . . . 249
26.5 Schéma d'Euler . . . 253
26.6 Notes et commentaires . . . 254

Table des figures **255**

Bibliographie **259**

Index **271**

Recueil de modèles stochastiques – Page vi/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Chapitre 1

Pile, face, coupons

Recueil de modèles stochastiques – Page 1/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Mots-clés. Combinatoire ; loi discrète ; loi des grands nombres ; théorème limite central ; distance en variation totale ; loi des petits nombres ; loi de Gumbel.

Le *jeu de pile ou face* et le problème du *collectionneur de coupons* constituent deux éléments probabilistes importants 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 $(X_n)_{n\geqslant 1}$ 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 $S_n = X_1 + \cdots + X_n$ suit la loi binomiale $\mathrm{Bin}(n,p)$ de taille n et de paramètre p, donnée pour tout $k = 0, 1, \ldots, n$ 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

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

Si $p > 0$, alors le nombre de lancers pour obtenir un succès $T = \inf\{n \geqslant 1 : X_n = 1\}$ suit la loi géométrique $\mathrm{Geo}(p)$ 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

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

Le nombre d'échecs avant le premier succès $T' = \inf\{n \geqslant 0 : X_{n+1} = 1\} = T - 1$ suit la loi géométrique $\mathrm{Geo}_{\mathbb{N}}(p)$ 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

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

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 $\mathrm{Geo}(p)$. La variable aléatoire T_r suit la loi de Pascal ou loi binomiale-négative $\mathrm{Geo}(p)^{*r}$. On a pour tout $k \geqslant r$,

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

et

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

Le *processus de Bernoulli* $(S_n)_{n \geqslant 0}$ a des trajectoires constantes par morceaux, avec des sauts d'amplitude $+1$, et les temps de saut sont donnés par $(T_r)_{r \geqslant 1}$ (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, $(S_n)_{n \geqslant 0}$ 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 $(S_n - np)_{n \geqslant 0}$ est une martingale.

La *loi forte des grands nombres* (LGN) et le *théorème limite central* (TLC) s'écrivent ici

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

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 $U_{(1)} \leqslant \dots \leqslant U_{(n)}$ est leur réordonnement croissant alors $U_{(k)}$ suit la loi $\mathrm{Beta}(k, n-k+1)$ sur $[0, 1]$ de densité $t \in [0, 1] \mapsto (\int_0^1 s^{k-1}(1-s)^{n-k}\, ds)^{-1} t^{k-1}(1-t)^{n-k}$, et

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

Remarque 1.1 (Motifs répétés et lois du zéro-un)**.** *Lorsque* $0 < p < 1$*, le* 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 1.2 (Lien avec la loi uniforme sur $[0, 1]$)**.** *Une variable aléatoire* $U = \sum_{n=1}^{\infty} b_n 2^{-n}$ *suit la loi uniforme sur* $[0, 1]$ *si et seulement si les bits* $(b_n)_{n \geqslant 1}$ *de son écriture en base* 2 *sont des variables aléatoires de Bernoulli indépendantes et de paramètre* $1/2$*.*

Remarque 1.3 (Algorithme de von Neumann)**.** *Soit* $(X_n)_{n \geqslant 1}$ *une suite de variables aléatoires de Bernoulli indépendantes et de paramètre* $0 < p < 1$ *inconnu. On fabrique la suite* $(Y_n)_{n \geqslant 1}$ *de variables aléatoires indépendantes et de même loi sur* $\{0, 1, 2\}$ *comme suit*

$$\underbrace{X_1 X_2}_{Y_1} \underbrace{X_3 X_4}_{Y_2} \cdots$$

en posant $Y_n = 0$ *si* $(X_{2n-1}, X_{2n}) = (0, 1)$*,* $Y_n = 1$ *si* $(X_{2n-1}, X_{2n}) = (1, 0)$*, et* $Y_n = 2$ *sinon. La suite* $(Z_n)_{n \geqslant 1}$ *obtenue à partir de* $(Y_n)_{n \geqslant 1}$ *en effaçant les* 2 *est constituée de variables aléatoires de Bernoulli indépendantes de paramètre* $1/2$*. La production de chaque terme de la suite* $(Z_n)_{n \geqslant 1}$ *nécessite un nombre aléatoire géométrique de termes de la suite* $(X_n)_{n \geqslant 1}$*.*

Recueil de modèles stochastiques – Page 2/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

1.2 Approximation de la loi binomiale par la loi normale

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

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

Le théorème de Berry-Esseen fournit une borne quantitative : pour tous $t \in \mathbb{R}$, $n \geqslant 1$,

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

où $\tau^3 = \mathbb{E}(|X_1 - \mathbb{E}(X_1)|^3)$. Lorsque $X_1, X_2, \ldots$ 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

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

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. Pour ces régimes extrêmes, il est plus approprié d'utiliser une approximation par la loi de Poisson, abordée dans la section qui suit.

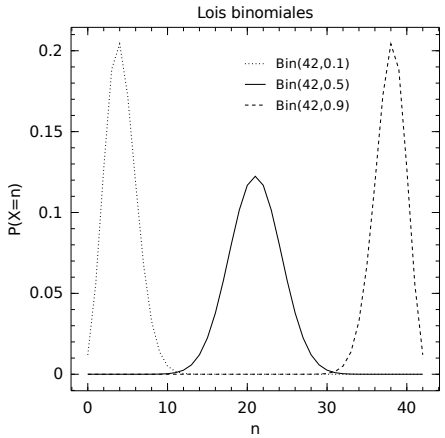


FIGURE 1.1 – Lois binomiales de même taille, pour trois valeurs du paramètre, illustrant la pertinence de l'approximation par une loi normale et par une loi de Poisson selon les cas.

Lorsque $X_1, X_2, \ldots$ sont des variables de Bernoulli de paramètre $p \in]0,1[$, le TLC suggère d'approcher la loi binomiale $\mathrm{Bin}(n,p)$ de S_n par la loi normale $\mathcal{N}(nm, n\sigma^2)$ lorsque n est grand. Le théorème de De Moivre et Laplace précise que pour tous $-\infty < a < b < +\infty$,

$$\lim_{n\to\infty} \sqrt{n} \sup_{k\in I_n(a,b)} \left| \mathbb{P}(S_n = k) - \frac{1}{\sqrt{2\pi np(1-p)}} \exp\left(-\frac{(k-np)^2}{2np(1-p)}\right) \right| = 0$$

où $I_n(a,b) = \left\{0 \leqslant k \leqslant n : \frac{k-np}{\sqrt{np(1-p)}} \in [a,b]\right\}$, qui redonne le TLC par intégration :

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{S_n - np}{\sqrt{np(1-p)}} \in [a,b]\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}}\, dx.$$

Recueil de modèles stochastiques – Page 3/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

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*

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

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

Théorème 1.4 (Expressions alternatives)**.** *Si μ et ν sont des lois sur E alors*[1]

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

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

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

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

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

Démonstration. La seconde égalité provient de l'inégalité

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

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

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

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 1.5 (Convergence en loi)**.** *Si (X_n) est une suite de v.a. 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. $\lim_{n\to\infty}\int f\,d\mu_n = \int f\,d\mu$ *pour toute fonction bornée* $f : E\to\mathbb{R}$ *;*
2. $\lim_{n\to\infty}\mu_n(x) = \mu(x)$ *pour tout* $x\in E$ *;*
3. $\lim_{n\to\infty} d_{VT}(\mu_n,\mu) = 0$.

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

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

Recueil de modèles stochastiques – Page 4/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

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

Démonstration. Pour déduire *1.* de *3.* il suffit d'utiliser l'expression variationnelle fonctionnelle de $d_{\mathrm{VT}}(\cdot,\cdot)$. 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 $n \geqslant N$. Par ailleurs, on peut contrôler le second terme du membre de droite de la manière suivante :

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

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

$$\sum_{x\in A^c} |\mu_n(x) - \mu(x)| \leqslant \sum_{x\in A} |\mu_n(x) - \mu(x)| + 2\sum_{x\in A^c} \mu(x).$$

Puisque $\mu \in \mathcal{P}$, pour tout $\varepsilon'' > 0$, on peut choisir A fini tel que $\mu(A^c) \leqslant \varepsilon''$. Ainsi, on obtient $\lim_{n\to\infty} \sum_{x\in E} |\mu_n(x) - \mu(x)| = 0$, qui n'est rien d'autre que *3.* d'après le théorème 1.4. □

Remarque 1.6 (Dispersion à l'∞)**.** *Si (μ_n) sont des lois et $\mu(x) := \lim_{n\to\infty} \mu_n(x)$ 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 $\{1\}, \ldots, \{n\}$, ce qui donne μ identiquement nulle.*

Théorème 1.7 (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

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

□

Théorème 1.8 (Couplage)**.** *Si μ et ν sont des lois sur E alors*

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

*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 $\mathbb{P}(X = x, Y = x) \leqslant \mu(x) \wedge \nu(x)$ pour tout $x \in E$ on a

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

Recueil de modèles stochastiques – Page 5/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

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

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

Cas où $p = 0$. On a alors $d_{\mathrm{VT}}(\mu, \nu) = 1$ 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 $d_{\mathrm{VT}}(\mu, \nu) = 0$ 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 $\mathrm{Ber}(p)$ 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 $\mathrm{Bin}(n, p)$ 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 $\lim_{n\to\infty} np = \lambda$ alors la loi de S_n tend vers $\mathrm{Poi}(\lambda)$. La distance en variation permet de quantifier cette convergence en loi : l'inégalité de poissonisation de Le Cam du théorème 1.9 ci-dessous donne (utile si np^2 est petit)

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

Théorème 1.9 (Inégalité de poissonisation de Le Cam)**.** *Soient $X_1, \ldots, X_n$ des v.a. indépendantes de lois de Bernoulli $\mathrm{Ber}(p_1), \ldots, \mathrm{Ber}(p_n)$. Soit μ_n la loi de $S_n = X_1 + \cdots + X_n$ et soit $\nu_n = \mathrm{Poi}(p_1 + \cdots + p_n)$ la loi de Poisson de même moyenne que S_n. Alors on a*

$$d_{VT}(\mu_n, \nu_n) \leqslant p_1^2 + \cdots + p_n^2.$$

Démonstration. On commence par établir par récurrence sur n que si $\alpha_1, \ldots, \alpha_n$ et $\beta_1, \ldots, \beta_n$ sont des lois de probabilité sur $\mathbb{N}$ alors on a l'inégalité sous-additive

$$d_{\mathrm{VT}}(\alpha_1 * \cdots * \alpha_n, \beta_1 * \cdots * \beta_n) \leqslant d_{\mathrm{VT}}(\alpha_1, \beta_1) + \cdots + d_{\mathrm{VT}}(\alpha_n, \beta_n).$$

Ensuite on établit que $d_{\mathrm{VT}}(\mathrm{Ber}(p), \mathrm{Poi}(p)) \leqslant p^2$. Rappel : $\mathrm{Poi}(a) * \mathrm{Poi}(b) = \mathrm{Poi}(a + b)$. □

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

$$\lim_{n\to\infty} p_{n,1} + \cdots + p_{n,n} = \lambda \quad \text{et} \quad \lim_{n\to\infty} \max_{1\leqslant k\leqslant n} p_{n,k} = 0,$$

alors $X_{n,1} + \cdots + X_{n,n}$ converge en loi vers $\mathrm{Poi}(\lambda)$ quand $n \to \infty$.

Recueil de modèles stochastiques – Page 6/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

1.5 Problème du collectionneur de coupons

Le problème du collectionneur de coupons est à ranger dans la même boîte à outils que le jeu de pile ou face ou le jeu de dé, auquel il est intimement relié. Un grand nombre de situations concrètes cachent un 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 $r \geqslant 2$ faces, combien de fois faut-il lancer le dé pour voir apparaître les r faces différentes ? On modélise cela, pour un entier fixé $r \geqslant 2$, en considérant la variable aléatoire

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

où $(X_n)_{n\geqslant 1}$ 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 1.10 (Expression combinatoire de la loi)**.** *On a $T \geqslant r$ et pour tout $n \geqslant r$,*

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

où la notation en accolades désigne le nombre de Stirling de seconde espèce [2]*.*

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 $n \geqslant r$, 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. □

Le théorème 1.10 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 1.11 (Décomposition)**.** *On a $T = G_1 + \cdots + G_r$ où $G_1, \dots, G_r$ sont des variables aléatoires indépendantes avec $G_i \sim \mathrm{Geo}(\pi_i)$ où $\pi_i := (r-i+1)/r$. En particulier, on a*

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

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

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

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

On a $\mathrm{card}(\{X_1, \dots, X_{G_i}\}) = i$ pour tout $1 \leqslant i \leqslant n$. Les variables aléatoires $G_1, G_1 + G_2, \dots, G_1 + \cdots + G_r$ 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

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

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

Recueil de modèles stochastiques – Page 7/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

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

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

□

Théorème 1.12 (Queue de distribution)**.** *Pour tout $n \geqslant 1$,*

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

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 1.12 en raison de l'alternance des signes.

Théorème 1.13 (Déviation)**.** *Pour tout entier $m \geqslant 0$,*

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

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

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

Comme $\mathbb{P}(E_{n,i}) = (1 - 1/r)^n \leqslant e^{-n/r}$, 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 $e^{-m} \leqslant \alpha$, par exemple $m = -\log(\alpha)$, ce qui donne l'intervalle de prédiction $[r, r\log(r) + rm]$ de niveau α pour T.

Théorème 1.14 (Comportement asymptotique)**.**

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

Démonstration. Pour tout $t > 0$ fixé, l'inégalité de Markov et le lemme 1.11 donnent

$$\mathbb{P}\left(\left|\frac{T}{r\log(r)} - 1\right| > t\right) \leqslant \frac{\mathbb{E}((T - r\log(r))^2)}{t^2 r^2 \log(r)^2} = \frac{\mathbb{V}(T) + (\mathbb{E}(T) - r\log(r))^2}{t^2 r^2 \log(r)^2} = \mathcal{O}\left(\frac{1}{\log(r)^2}\right).$$

On dit qu'il s'agit d'une preuve par la « méthode de second moment ». Dans le même esprit, une méthode de quatrième moment est utilisée pour le théorème 18.3 page 178. □

Recueil de modèles stochastiques – Page 8/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

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-Cantelli*[3]. 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, $(T - \mathbb{E}(T))/\sqrt{\mathbb{V}(T)}$ converge quand $r \to \infty$ 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 1.15 (Fluctuations asymptotiques)**.** *On a*

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

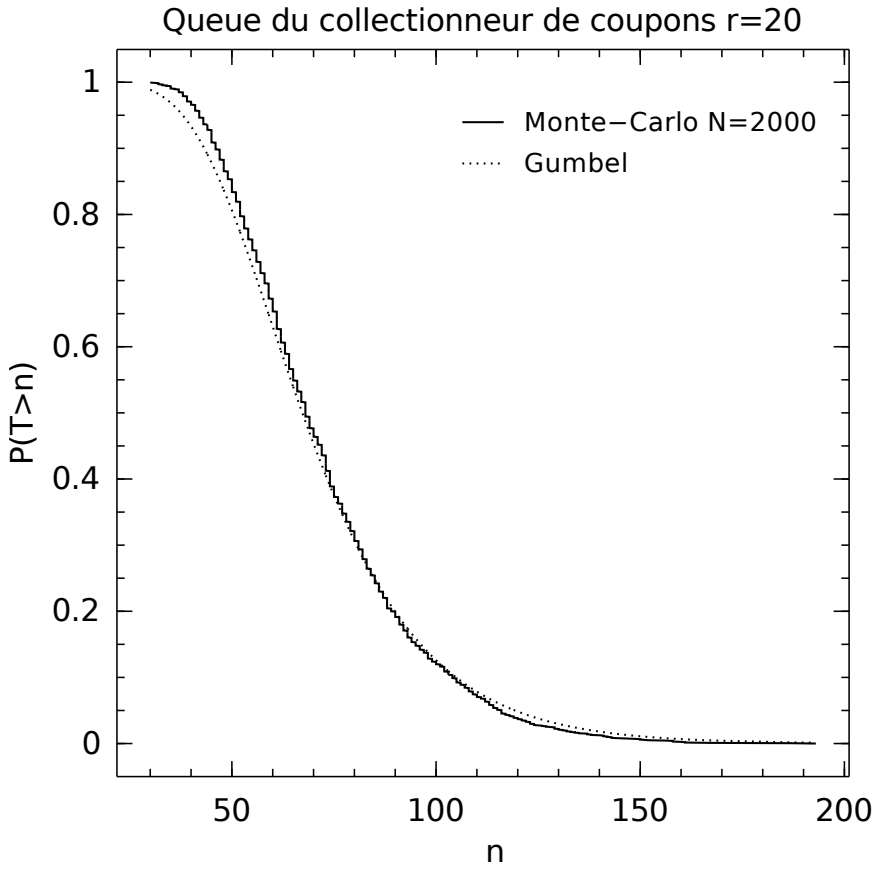


FIGURE 1.2 – Approximations de la queue de T par celle de $r\log(r) + rG$ où G suit la loi de Gumbel en vertu du théorème 1.15, ainsi que par une méthode de Monte-Carlo avec N tirages en utilisant la représentation de T en somme de v.a. géométriques indépendantes.

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

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

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

Recueil de modèles stochastiques – Page 9/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

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 1.12 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 $\binom{r}{k} \leqslant r^k/k!$ et $1 - u \leqslant e^{-u}$ pour tout $u \geqslant 0$, on a

$$\binom{r}{k} \left(1 - \frac{k}{r}\right)^{n_{t,r}} \underset{r\to\infty}{\overset{\leqslant}{\longrightarrow}} \frac{e^{-tk}}{k!}.$$

Enfin, par convergence dominée, on obtient

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

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 $k \geqslant 1$ fixé, la suite $(H_k)_{r\geqslant k}$ 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 $k \geqslant 1$, il existe une constante $C > 0$ telle que pour r assez grand,

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

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

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

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

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

Comme dans le théorème 10.4, on observe que $E_1 + \cdots + E_r$ a même loi que $\max(F_1, \ldots, F_r)$ où $F_1, \ldots, F_r$ sont des v.a.r. i.i.d. de loi exponentielle de paramètre 1, puis on utilise le fait que $(\max(F_1, \ldots, F_r) - \log(r))_{r\geqslant 1}$ converge en loi vers la loi de Gumbel. □

Le théorème 1.15 fournit un intervalle de prédiction pour T : pour tous réels $b \geqslant a$,

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

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

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

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

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

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

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

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

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

□

Recueil de modèles stochastiques – Page 10/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

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 et les principes d'invariance, abordées dans le livre de William 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). Un joli problème de la plus longue sous-suite commune à deux jeux de pile ou face est étudié dans le livre d'optimisation combinatoire randomisée de Michael Steele [Ste97]. 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

$$d_{\mathrm{VT}}\left(\mathrm{Ber}(p_1) * \cdots * \mathrm{Ber}(p_n), \mathrm{Poi}(p_1 + \cdots + p_n)\right) \leqslant (1 - e^{-(p_1+\cdots+p_n)}) \frac{p_1^2 + \cdots + p_n^2}{p_1 + \cdots + p_n}.$$

À titre de comparaison, une version inhomogène du théorème de Berry-Esseen (tirée du livre de William Feller) affirme que si $(X_n)_{n\geqslant 1}$ sont des variables aléatoires 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

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

D'autre part, on peut voir $d_{\mathrm{VT}}(\cdot,\cdot)$ comme une distance de couplage de 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 lois marginales μ et ν,

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

Le problème du collectionneur de coupons est notamment abordé dans les livres [Fel68, Fel71, MR95], et dans les articles [Hol01, DP13]. Donald Newman et Lawrence Shepp on 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 $r\log(r) + (m-1)r\log(\log(r)) + \mathcal{O}(r)$. On trouvera d'autres variantes dans le livre [BPPC99] par exemple. Le théorème 1.15 révélant une fluctuation asymptotique de loi de Gumbel a été obtenu par Paul Erdős et Alfréd Rényi. Une analyse du cas non uniforme associé à une probabilité discrète $(p_1, \ldots, p_r)$ est menée dans un article de Lars Holst [Hol01]. Lorsque r n'est pas connu, on dispose de l'estimateur $\widehat{r}_n := \mathrm{card}\{X_1, \ldots, X_n\}$. Si $e_1, \ldots, e_r$ est la base canonique de $\mathbb{R}^r$ alors le vecteur aléatoire $C_n := (C_{n,1}, \ldots, C_{n,r}) := e_{X_1} + \cdots + e_{X_n}$ de $\mathbb{N}^r$ suit la loi multinomiale de taille n et de paramètre $(p_1, \ldots, p_r)$, et on a $\widehat{r}_n = \sum_{i=1}^r \mathbb{1}_{\{C_{n,i}>0\}} = r - \sum_{i=1}^r \mathbb{1}_{\{C_{n,i}=0\}}$. En particulier,

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

Comme les r types sont ordonnés, on dispose aussi de l'estimateur $\max\{X_1, \ldots, X_n\}$, qui estime le bord droit du support. Il s'agit de l'analogue discret de la loi uniforme sur $[0, \theta]$.

Le collectionneur de coupons est un cas particulier du problème du recouvrement abordé dans les livres [AF01, LPW09, Dem05] : si $(X_n)_{n\geqslant 1}$ est une suite de variables aléatoires, pas nécessairement indépendantes ou de même loi, et prenant leurs valeurs dans un même ensemble fini E, alors le *temps de recouvrement*[4] de E est $T := \inf\{n \geqslant 1 : \{X_1, \ldots, X_n\} = E\}$.

4. En anglais : *covering time*.

Recueil de modèles stochastiques – Page 11/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Chapitre 2

Marches aléatoires

Mots-clés. Marche aléatoire ; chaîne de Markov ; martingale ; temps de sortie ; ruine du joueur ; loi multinomiale ; problème de Dirichlet ; fonctions harmoniques ; champ libre gaussien ; groupe symétrique ; collectionneur de coupons.

2.1 Marche aléatoire simple sur la droite

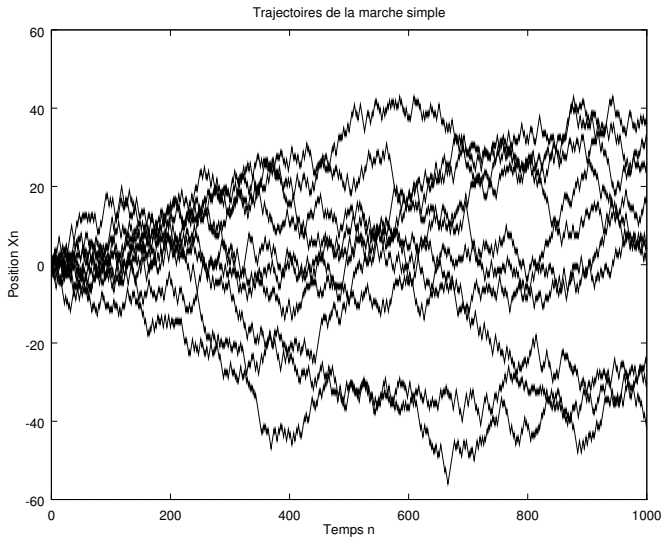


Figure 2.1 – Trajectoires issues de 0 de la marche aléatoire simple sur $\mathbb{Z}$.

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

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

pour tout $n \geqslant 0$, où $(\varepsilon_n)_{n\geqslant 1}$ 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 $(X_n)_{n\geqslant 0}$ est une chaîne de Markov d'espace d'états $\mathbb{N}$ 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 $n \geqslant 0$,

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

Recueil de modèles stochastiques – Page 13/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Théorème 2.1 (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 $n \geqslant 0$. 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. $X_n = X_0 + n(2p - 1 + o_{n\to\infty}(1))$, donc $X_n \to +\infty$ p.s. si $p > 1/2$ et $X_n \to -\infty$ p.s. si $p < 1/2$, ce qui interdit toute récurrence. □

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). On adopte la notation conditionnelle traditionnelle $\mathbb{P}_x(\cdot) := \mathbb{P}(\cdot \,|\, X_0 = x)$ et $\mathbb{E}_x(\cdot) = \mathbb{E}(\cdot \,|\, X_0 = x)$.

Théorème 2.2 (Sortie de boîte[1] ou ruine du joueur)**.** *Soient $a < b$ dans $\mathbb{Z}$, et les temps d'arrêt*

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

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

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

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 $a \leqslant x \leqslant b$. 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 2.2 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 $Y = (Y_n)_{n\geqslant 0}$ d'espace d'états fini $\{a,\dots,b\}$ de mêmes transitions que $(X_n)_{n\geqslant 0}$ 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 $a \leqslant x \leqslant b$.

Démonstration. Montrons que $\mathbb{E}_x(\tau) < \infty$ pour tout $a \leqslant x \leqslant b$. 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 4.2. Pour tout $a \leqslant x \leqslant b$, il existe un chemin ℓ_x de longueur $|\ell_x| \leqslant (b-a)$ qui mène de x à a ou b. On a

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

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

Recueil de modèles stochastiques – Page 14/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

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

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

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 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

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

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 2.3 (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. $(X_n)_{n\geqslant 1}$ 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 $n \geqslant 1$, en posant $I_n = \frac{1}{\sqrt{n}}\,]a, b[$, on a*

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

Recueil de modèles stochastiques – Page 15/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

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

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

Remarque 2.4 (Martingales)**.** *Si* $(Z_n)_{n\geqslant 0}$ *est une chaîne de Markov sur* E *fini de noyau* $P = L + I$ *alors pour toute fonction* $f : E \to \mathbb{R}$, *la suite* $(M_n)_{n\geqslant 0}$ *définie par*

$$M_n = f(Z_n) - f(Z_0) - \sum_{k=0}^{n-1} (Lf)(Z_k)$$

est une martingale pour la filtration naturelle de la suite $(Z_n)_{n\geqslant 0}$. *La formule ci-dessus peut être vue comme une formule d'Itô discrète. Lorsque* f *est harmonique pour* L, *c'est-à-dire que* $Lf = 0$, *alors* $(f(Z_n) - f(Z_0))_{n\geqslant 0}$ *est une martingale. Il se trouve que le vecteur des probabilités d'atteinte d'un ensemble clos est harmonique. Il est possible de retrouver les formules du théorème 2.2 en utilisant des martingales bien choisies et le théorème d'arrêt de Doob-Wald. Par exemple, la martingale* $(X_n - nm)_{n\geqslant 0}$ *donne* $\mathbb{E}_0(X_\tau) = m\mathbb{E}_0(\tau)$ *tandis que la martingale exponentielle* $(\rho^{X_n})_{n\geqslant 0}$ *donne* $\mathbb{E}_0(\rho^{X_\tau}) = 1$. *Cette méthode à base de martingales s'adapte au cadre des processus à temps et espace continus, et permet notamment d'obtenir des formules pour le temps de sortie pour un processus de diffusion sur* $\mathbb{R}$ *(chapitre 26). D'autre part, les équations satisfaites par les fonctions* r *et* R *dans la preuve du théorème 2.2 sont des cas particulier du problème de Dirichlet du théorème 2.9.*

Théorème 2.5 (Nombres de Catalan)**.** *Si* $\tau = \inf\{n \geqslant 1 : X_n = 0\}$ *alors pour tout* $n \geqslant 0$,

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

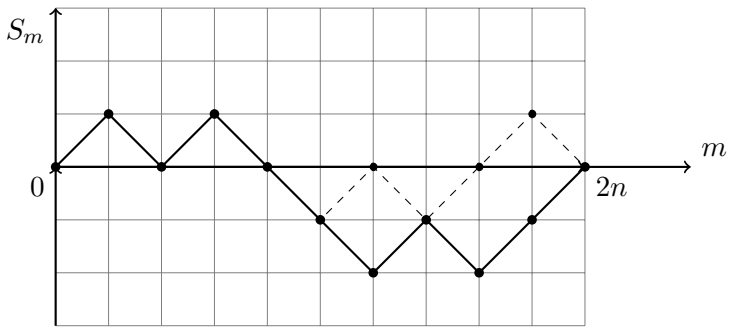


Figure 2.2 – Réflexion de Désiré André (preuve du théorème 2.5).

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 2.5. 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

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

Recueil de modèles stochastiques – Page 16/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

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. Un exemple d'illustration est donné dans la figure 2.2. 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 $\binom{2n}{n-1}$. Cette bijection donne donc $C_n = \binom{2n}{n} - \binom{2n}{n-1} = \frac{1}{n+1}\binom{2n}{n}$ (formule de Désiré André). □

Théorème 2.6 (Théorème du scrutin [2]). *Si $n = a+b$ et $k = a-b$ avec $0 \leqslant b \leqslant a$ alors*

$$\mathbb{P}(S_1 > 0, \ldots, S_n > 0 \,|\, S_0 = 0, S_n = k) = \frac{k}{n}.$$

Ainsi par exemple, lors d'une élection avec deux candidats A et B et n votants, dans laquelle A obtient a votes et B obtient $b \leqslant a$ votes, la probabilité que A soit devant B tout le long du dépouillement des n bulletins de votes est $(a-b)/(a+b)$.

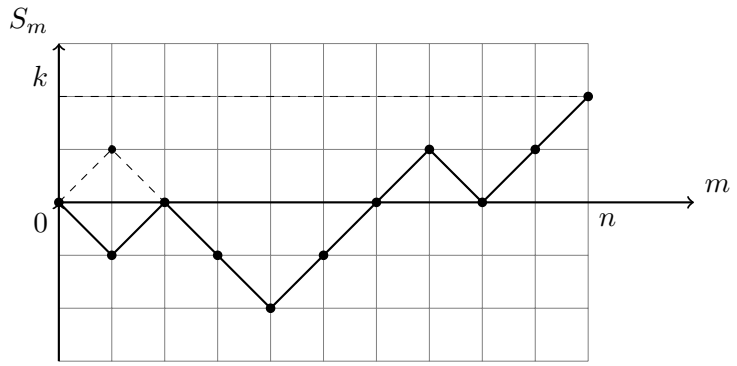


FIGURE 2.3 – Réflexion de Désiré André (preuve du théorème 2.6, $n = 10, k = 2$).

Preuve du théorème 2.6. Notons que $0 \leqslant k \leqslant n$ et $n-k$ est impair. Soit $P_{n,k}$ l'ensemble des chemins de la marche aléatoire simple, de longueur n, partant de $(0,0)$ et finissant en (n,k). Soit $P^+_{n,k}$ l'ensemble de ces chemins strictement positifs après $(0,0)$. On a

$$\mathbb{P}(S_1 > 0, \ldots, S_n > 0 \,|\, S_0 = 0, S_n = k) = |P^+_{n,k}| \frac{p^a(1-p)^b}{\mathbb{P}(S_n = k \,|\, S_0 = 0)}.$$

D'un autre côté $\mathbb{P}(S_n = k \,|\, S_0 = 0) = |P_{n,k}|\, p^a(1-p)^b$ et donc

$$\mathbb{P}(S_1 > 0, \ldots, S_n > 0 \,|\, S_0 = 0, S_n = k) = \frac{|P^+_{n,k}|}{|P_{n,k}|}$$

(formule sans p !). L'ensemble $P_{n-1,k-1}$ est en bijection avec les éléments de $P_{n,k}$ qui commencent par un incrément -1. D'un autre côté, cet ensemble est en bijection avec les éléments de $P_{n,k} \setminus P^+_{n,k}$ qui commencent avec un incrément $+1$. Pour le voir, il suffit d'utiliser une réflexion sur la portion du chemin située avant l'atteinte de 0. Une illustration est fournie par la

2. «Ballot theorem» en anglais.

Recueil de modèles stochastiques – Page 17/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

figure 2.3. On reconnaît là l'astuce de Désiré André, utilisée dans la preuve du théorème 2.5. Il découle de ces bijections l'identité $|P_{n,k}| - |P^+_{n,k}| = 2|P_{n-1,k-1}|$, qui donne $|P^+_{n,k}| = (k/n)|P_{n,k}|$, ce qui conduit immédiatement au résultat annoncé. Notons que la formule à base de nombres de Catalan du théorème 2.5 s'écrit ici

$$\mathbb{P}(S_1 > 0, \dots, S_{2n-1} > 0 \,|\, S_0 = 0, S_{2n} = 0) = \frac{1}{n+1}\binom{2n}{n} = C_n.$$

□

2.2 Marche aléatoire simple symétrique dans l'espace

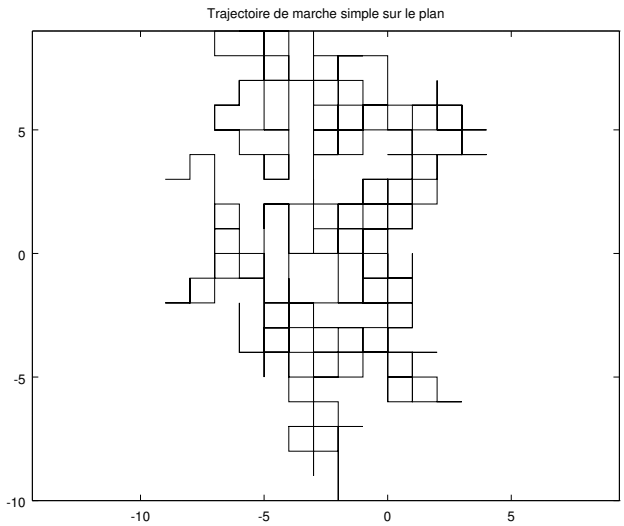


Figure 2.4 – Début de trajectoire de la marche aléatoire simple sur $\mathbb{Z}^2$.

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

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

où $(\varepsilon_n)_{n\geqslant 1}$ est une suite de v.a.r. i.i.d. , indépendantes de la position initiale X_0, et de loi uniforme sur $\{\pm e_1, \dots, \pm e_d\}$, où $e_1, \dots, e_d$ désigne la base canonique de $\mathbb{R}^d$. La suite $(X_n)_{n\geqslant 0}$ est une chaîne de Markov d'espace d'état $\mathbb{Z}^d$ et de noyau de transition $\mathbf{P}(x,y) = \frac{1}{2d}\mathbb{1}_{|x-y|_1=1}$ où $|x|_1 := |x_1| + \cdots + |x|_d$. On parle également de marche aléatoire symétrique aux plus proches voisins au sens de la distance associée à la norme $|\cdot|_1$. Les incréments $(\varepsilon_n)_{n\geqslant 1}$ sont de loi uniforme sur la sphère unité pour la norme $|\cdot|_1$. 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 la *loi multinomiale*

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

par l'application

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

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 coupons* étudiée dans le chapitre 1.

Recueil de modèles stochastiques – Page 18/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Théorème 2.7 (Récurrence de la marche aléatoire simple symétrique)**.** *La marche aléatoire simple symétrique sur $\mathbb{Z}^d$ est récurrente nulle si $d \leqslant 2$ et transitoire si $d \geqslant 3$.*

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

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

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

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

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

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 première et seconde bissectrices (pentes ± 1). On vérifie que $(U_n)_{n\geqslant 0}$ et $(V_n)_{n\geqslant 0}$ 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 $(X_n)_{n\geqslant 0}$ (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 $\binom{n}{r_1\ r_2\ r_3} \leqslant \binom{n}{m\ m\ m}$ et donc, grâce à la formule du trinôme de taille n et de paramètre $(1/3,1/3,1/3)$, on obtient, pour $n = 3m$,

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

Ainsi, $\sum_m \mathbf{P}^{6m}(0,0) < \infty$. Comme $(1/6)^{2k}\mathbf{P}^{6m}(0,0) \leqslant \mathbf{P}^{6m+2k}(0,0)$ 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 $d \geqslant 3$ (méthode directe). La formule de la multinomiale donne

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

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

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

Le lemme 2.8 donne $m_{n+1}/m_n = (n+1)/(q+1) \leqslant d$ et donc $(m_n/d^n)_{n\geqslant 1}$ décroît, d'où

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

Recueil de modèles stochastiques – Page 19/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

À présent, le lemme 2.8 et la formule de Stirling donnent, lorsque $q \to \infty$,

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

Or $d \geqslant 3$, 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 $X' = (X'_n)_{n\geqslant 0}$ la chaîne projetée sur $\mathbb{Z}^3$ c'est-à-dire constituée par les trois premières coordonnées de $X = (X_n)_{n\geqslant 0}$. 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 transcience de X'' implique celle de X', et donc X est transitoire. □

Lemme 2.8 (Coefficient multinomial central)**.** *Pour tout $d \geqslant 1$ et tout $n \geqslant 1$,*

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

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

Démonstration. Si $n_1 + \cdots + n_d = n$ et si $n_1 < q$ alors on a nécessairement $n_k \geqslant q + 1$ pour un $1 \leqslant k \leqslant d$, 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 $q \leqslant n_k \leqslant q + 1$ pour tout $1 \leqslant k \leqslant n$, par exemple $n_1 = \cdots = n_r = q + 1$ et $n_{r+1} = \cdots = n_d = q$, ce qui donne le résultat annoncé. □

2.3 Problème de Dirichlet et champ libre gaussien

Soit $(X_n)_{n\geqslant 0}$ la marche aléatoire simple sur $\mathbb{Z}^d$, et P son noyau de transition en tant que chaîne de Markov. Pour toute fonction bornée $f : \mathbb{Z}^d \to \mathbb{R}$ et tout $x \in \mathbb{Z}^d$ on a, en notant $\Delta := P - I$,

$$\mathbb{E}(f(X_{n+1}) \mid X_n = x) - f(x) = (Pf)(x) - f(x) = (\Delta f)(x).$$

L'opérateur $\Delta = P - I$ est le générateur de la marche aléatoire simple symétrique sur $\mathbb{Z}^d$. Il s'agit d'un opérateur laplacien discret, qui calcule l'écart à la moyenne sur les voisins :

$$(\Delta f)(x) = \left(\frac{1}{2d}\sum_{y:|y-x|_1=1} f(y)\right) - f(x) = \frac{1}{2d}\sum_{y:|y-x|_1=1}(f(y) - f(x)).$$

On dit que f est harmonique sur $A \subset \mathbb{Z}^d$ lorsque $\Delta f = 0$ sur A, ce qui signifie qu'en tout point de A la valeur de f est égale à la moyenne de ses valeurs sur les $2d$ voisins. L'opérateur Δ est local en ce sens que $(\Delta f)(x)$ ne dépend que des valeurs de f en x et ses plus proches voisins. Ainsi la valeur de Δf sur A ne dépend que des valeurs de f sur $\bar{A} := A \cup \partial A$ où $\partial A := \{y \notin A : \exists x \in A, |x - y|_1 = 1\}$ est le bord extérieur de A. Le *problème de Dirichlet* consiste à trouver une fonction harmonique sur A dont la valeur sur ∂A est prescrite. Il s'agit en fait d'un problème d'algèbre linéaire, pour lequel le théorème 2.9 ci-dessous fournit une expression probabiliste de la solution, basée sur $\tau_{\partial A} := \inf\{n \geqslant 0 : X_n \in \partial A\}$. En physique, la quantité $f(x)$ modélise la température au point x, l'harmonicité de f sur A et la condition au bord ∂A traduisent respectivement un équilibre thermique local et un thermostat au bord.

Recueil de modèles stochastiques – Page 20/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Théorème 2.9 (Problème de Dirichlet)**.** *Soit $A \subset \mathbb{Z}^d$ un ensemble non vide fini. Alors pour tout $x \in A$ on a $\mathbb{P}_x(\tau_{\partial A} < \infty) = 1$. De plus, pour toute fonction $g : \partial A \to \mathbb{R}$, la fonction $f : \bar{A} \to \mathbb{R}$ définie par $f(x) = \mathbb{E}_x(g(X_{\tau_{\partial A}}))$ pour tout $x \in \bar{A}$ est l'unique solution du système*

$$\begin{cases} f = g & \text{sur } \partial A, \\ \Delta f = 0 & \text{sur } A. \end{cases}$$

Lorsque $d = 1$ on retrouve la fonction r étudiée dans la preuve du théorème 2.2 sur la ruine du joueur. D'autre part, d'après la remarque 2.4, l'image d'une chaîne de Markov par une fonction harmonique pour son générateur est une martingale, ce qui explique a posteriori la formule probabiliste pour la solution du problème de Dirichlet grâce au théorème d'arrêt de Doob.

La quantité $f(x) = \mathbb{E}_x(g(\tau_{\partial A})) = \sum_{y \in \partial A} g(y)\mathbb{P}_x(\tau_{\partial A} = y)$ est la moyenne de g pour la loi μ_x sur ∂A, appelée *mesure harmonique*, définie par $\mu_x(y) = \mathbb{P}_x(X_{\tau_{\partial A}} = y)$ pour tout $y \in \partial A$.

Démonstration. La propriété $\mathbb{P}_x(\tau_{\partial A} < \infty) = 1$ pour tout $x \in A$ peut être établie en procédant comme dans la remarque 2.3, ou encore comme dans la preuve du théorème 2.2, qui fournit de plus une borne sous-géométrique pour la queue de la loi de τ.

Vérifions que la fonction f proposée est bien solution. Pour tout $x \in \partial A$ on a $\tau_{\partial A} = 0$ sur $\{X_0 = x\}$ et donc $f = g$ sur ∂A. Montrons à présent que $\Delta f = 0$ sur A. On se ramène tout d'abord par linéarité au cas où $g = \mathbb{1}_{\{z\}}$ avec $z \in \partial A$. Ensuite on écrit, pour tout $y \in \bar{A}$,

$$\begin{aligned} f(y) = \mathbb{P}_y(X_{\tau_{\partial A}} = z) &= \sum_{n=0}^{\infty} \mathbb{P}_y(X_n = z, \tau_{\partial A} = n) \\ &= \mathbb{1}_{y=z} + \mathbb{1}_{y \in A} \sum_{n=1}^{\infty} \sum_{x_1,\dots,x_{n-1} \in A} P(y, x_1)P(x_1, x_2) \cdots P(x_{n-1}, z). \end{aligned}$$

D'autre part, comme $f = 0$ sur $\partial A \setminus \{z\}$ et Δ est local, on a, pour tout $x \in A$,

$$(Pf)(x) = \sum_{y \in \mathbb{Z}^d} P(x, y)f(y) = P(x, z)f(z) + \sum_{y \in A} P(x, y)f(y).$$

Par conséquent, pour tout $x \in A$,

$$\begin{aligned} (Pf)(x) &= P(x, z)f(z) + \sum_{n=1}^{\infty} \sum_{y, x_1,\dots,x_{n-1} \in A} P(x, y)P(y, x_1) \cdots P(x_{n-1}, z) \\ &= P(x, z)f(z) + (f(x) - (\mathbb{1}_{x=z} + P(x, z))), \end{aligned}$$

égal à $f(x)$ car $\mathbb{1}_{x=z} = 0$ et $f(z) = 1$. Ainsi $Pf = f$ sur A, c'est-à-dire que $\Delta f = 0$ sur A.

Pour établir l'unicité de la solution, on se ramène par linéarité à établir que $f = 0$ est l'unique solution lorsque $g = 0$. Or si $f : \bar{A} \to \mathbb{R}$ est harmonique sur A, l'interprétation de $\Delta f(x)$ comme écart à la moyenne sur les plus proches voisins permet d'établir qu'à la fois le minimum et le maximum de f sur $\bar{A}$ sont (au moins) nécessairement atteints sur le bord ∂A. Or comme par ailleurs f est nulle sur ∂A, il en découle que f est nulle sur $\bar{A}$. □

Le théorème 2.9 se généralise de la manière suivante :

Théorème 2.10 (Problème de Dirichlet et fonction de Green)**.** *Si $A \subset \mathbb{Z}^d$ est un ensemble non vide fini alors pour toutes fonctions $g : \partial A \to \mathbb{R}$ et $h : A \to \mathbb{R}$, la fonction $f : \bar{A} \to \mathbb{R}$ définie par $f(x) = \mathbb{E}_x(g(X_{\tau_{\partial A}}) + \sum_{n=0}^{\tau_{\partial A}-1} h(X_n))$ pour tout $x \in \bar{A}$ est l'unique solution de*

$$\begin{cases} f = g & \text{sur } \partial A, \\ \Delta f = -h & \text{sur } A. \end{cases}$$

Recueil de modèles stochastiques – Page 21/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Lorsque $d = 1$ et $h = 1$ on retrouve la fonction R de la preuve du théorème 2.2. Le théorème 2.9 correspond au cas spécial où $h = 0$. Pour tout $x \in A$ on a

$$f(x) = \sum_{y \in \partial A} g(y) \mathbb{P}_x(\tau_{\partial A} = y) + \sum_{y \in A} h(y) G_A(x, y)$$

où $G_A(x, y)$ est le nombre moyen de passages en y en partant de x et avant de sortir de A :

$$G_A(x, y) := \mathbb{E}_x \left(\sum_{n=0}^{\tau_{\partial A} - 1} \mathbb{1}_{X_n = y} \right) = \sum_{n=0}^{\infty} \mathbb{P}_x(X_n = y, n < \tau_{\partial A}).$$

On dit que G_A est la *fonction de Green* de la marche aléatoire simple symétrique sur A tuée au bord ∂A. C'est l'inverse de la restriction $-\Delta_A$ de $-\Delta$ aux fonctions sur $\bar{A}$ nulles sur ∂A :

$$G_A = -\Delta_A^{-1}.$$

En effet, pour $g = 0$ et $h = \mathbb{1}_{\{y\}}$ on trouve $f(x) = G_A(x, y)$, d'où $\Delta_A G_A = -I_A$.

Preuve du théorème 2.10. Grâce au théorème 2.9 il suffit par linéarité de vérifier que $f(x) = \mathbb{1}_{x \in A} G_A(x, z)$ est solution lorsque $g = 0$ et $h = \mathbb{1}_{\{z\}}$ avec $z \in A$. Or pour tout $x \in A$,

$$\begin{aligned} f(x) &= \mathbb{1}_{\{x=z\}} + \sum_{n=1}^{\infty} \mathbb{P}_x(X_n = z, n < \tau_{\partial A}) \\ &= \mathbb{1}_{\{x=z\}} + \sum_{y : |x-y|_1 = 1} \sum_{n=1}^{\infty} \mathbb{P}(X_n = z, n < \tau_{\partial A} \mid X_1 = y) P(x, y) \\ &= \mathbb{1}_{\{x=z\}} + \sum_{u : |x-y|_1 = 1} f(y) P(x, y) \end{aligned}$$

grâce à la propriété de Markov. On a bien $f = h + Pf$ sur A, c'est-à-dire $\Delta f = -h$ sur A. □

Les preuves des théorèmes 2.9 et 2.10 restent valables pour des marches aléatoires asymétriques sur $\mathbb{Z}^d$, à condition de remplacer le générateur Δ de la marche symétrique par le générateur $L := P - I$, qui est un opérateur local : $|x - y| > 1 \Rightarrow L(x, y) = P(x, y) = 0$. Le dépassement de ce cadre nécessite l'adaptation de la notion de bord : $\{y \notin A : \exists x \in A, L(x, y) > 0\}$.

Le *champ libre gaussien*[3] est un modèle d'interface aléatoire lié à la marche aléatoire simple symétrique et au problème de Dirichlet. Soit $A \subset \mathbb{Z}^d$ un sous-ensemble non-vide fini. Une interface est une fonction de hauteur $f : \bar{A} \to \mathbb{R}$ qui associe à chaque site $x \in \bar{A}$ une hauteur $f(x)$, également appelée *spin*. Pour simplifier, on impose la condition au bord $f = 0$ sur le bord extérieur ∂A de A. On note $\mathcal{F}_A$ l'ensemble des interfaces f sur $\bar{A}$ nulles sur le bord ∂A, qu'on peut identifier à $\mathbb{R}^A$. L'énergie $H_A(f)$ de l'interface $f \in \mathcal{F}_A$ est définie par

$$H_A(f) = \frac{1}{4d} \sum_{\substack{\{x,y\} \subset \bar{A} \\ |x-y|_1 = 1}} (f(x) - f(y))^2,$$

où on a posé $f = 0$ sur ∂A. L'énergie $H_A(f)$ est d'autant plus petite que l'interface f est «plate». En notant $\langle u, v \rangle_A := \sum_{x \in A} u(x) v(x)$ il vient, pour tout $f \in \mathcal{F}_A$,

$$H_A(f) = \frac{1}{4d} \sum_{x \in A} \sum_{\substack{y \in \bar{A} \\ |x-y|_1 = 1}} (f(x) - f(y)) f(x) = \frac{1}{2} \langle -\Delta f, f \rangle_A.$$

3. «Gaussian free field» (GFF) en anglais.

Recueil de modèles stochastiques – Page 22/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Comme $H_A(0) = 0$ et $H_A(f) = 0$ entraîne $f = 0$, la forme quadratique H_A n'est pas dégénérée, et on peut définir une loi gaussienne Q_A sur $\mathcal{F}_A$ favorisant les faibles énergies :

$$Q_A(df) = \frac{1}{Z_A} e^{-H_A(f)}\, df \quad \text{où} \quad Z_A := \int_{\mathcal{F}_A} e^{-H_A(f)}\, df.$$

Cette loi gaussienne, appelée champ libre gaussien, est caractérisée par sa moyenne $m_A : A \to \mathbb{R}$ et sa matrice de covariance $C_A : A \times A \to \mathbb{R}$, données pour tous $x, y \in A$ par

$$m_A(x) := \int f_x\, Q_A(df) = 0 \quad \text{et} \quad C_A(x,y) := \int f_x f_y\, Q_A(df) - m_A(x) m_A(y) = -(\Delta_A^{-1})(x,y) = G_A(x,y),$$

où f_x désigne l'application coordonnée $f_x : f \in \mathcal{F}_A \mapsto f(x) \in \mathbb{R}$.

2.4 Marche aléatoire sur le groupe symétrique

On considère $r \geqslant 2$ 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 est codée par un élément σ du *groupe symétrique* Σ_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, appelée «*top to random shuffle*» en anglais, qui constitue sans doute le *battage de cartes* le plus simple. 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 $k+1^{\text{e}}$ 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 $(k, k-1, \ldots, 1)\sigma$ où

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

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

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

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

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

1	$\sigma^{-1}(1)$
$\vdots$	$\vdots$
r	$\sigma^{-1}(r)$
Configuration initiale	Configuration après mélange

FIGURE 2.5 – La carte en position k se déplace en position $\sigma(k)$.

Théorème 2.11 (Convergence en loi)**.** *La chaîne $(X_n)_{n \geqslant 0}$ 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 $(X_n)_{n \geqslant 0}$ converge en loi vers μ quelque soit la loi initiale $\mathcal{L}(X_0)$.

Recueil de modèles stochastiques – Page 23/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

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 transition[4] 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. □

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 $1 \leqslant k \leqslant r-1$, on note T_k le temps que cette carte passe en position $r-k$. Pour tout $k \geqslant 1$, on a, avec la convention $T_1 + \cdots + T_{k-1} = 0$ si $k = 1$,

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

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

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

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

Théorème 2.12 (Bon mélange après remontée)**.** *Pour tout $n \geqslant 0$, 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).

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 $T_1 + \cdots + T_{r-1} = T - 1$, 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 $n \geqslant 0$. □

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 $d_{\mathrm{VT}}(\mathcal{L}(X_n), \mu)$ est proche de

4. 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 !

Recueil de modèles stochastiques – Page 24/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

zéro, où $d_{\mathrm{VT}}(\cdot,\cdot)$ est la distance en variation totale (section 1.3). Il s'avère que pour r assez grand, la quantité $d_{\mathrm{VT}}(\mathcal{L}(X_n),\mu)$ passe de 1 à 0 de manière abrupte[5] autour de $n = r\log(r)$. Ce phénomène de convergence abrupte est quantifié par le théorème suivant.

Théorème 2.13 (Convergence abrupte autour de $n = r\log(r)$)**.** *Pour tout réel $c > 0$,*

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

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

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

Démonstration. Grâce au théorème 2.12, on a, pour tout $A \subset \Sigma_r$,

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

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

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

Le résultat voulu découle à présent du théorème 1.16 sur le collectionneur de coupons. □

Le graphique de la figure 1.2 page 1 illustre le phénomène de convergence abrupte.

2.5 Notes et commentaires

La marche aléatoire simple sur $\mathbb{Z}^d$, étudiée notamment par George Pólya puis par Frank Spitzer, fait partie des modèles de base des probabilités. Elle fait l'objet d'une attention particulière dans les livres [Fel68, Spi70, DS84, Nor98]. L'analyse de Fourier constitue une alternative puissante à la combinatoire pour l'étude de la récurrence et de la transcience, comme expliqué par exemple dans les livres [Law13, LL10]. L'astuce combinatoire de Désiré André se trouve dans [And87]. La physique statistique a inspiré nombre de modèles de marches aléatoires, comme par exemple les marches aléatoires en milieu aléatoire, en paysage aléatoire, en auto interaction (évitement, renforcement, excitation, …), etc. Ce sujet historique est d'une grande richesse et fait toujours l'objet de recherches à l'heure actuelle. Les chapitres 5, 11, et 7 font intervenir des marches aléatoires. Par ailleurs le *théorème limite central* permet de concevoir le mouvement brownien comme un analogue en temps et en espace continus de la marche aléatoire simple, obtenu comme limite d'échelle de modèles discrets (εZ^d approche $\mathbb{R}^d$). Le problème de Dirichlet discret est étudié en détail dans le livre [Law13]. Le problème de Dirichlet possède une version à temps et espace continus, étudiée dans le chapitre 23, qui

5. On parle de «cutoff phenomenon» en anglais.

Recueil de modèles stochastiques – Page 25/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

est une limite d'échelle du problème de Dirichlet discret. Au niveau du processus, la marche aléatoire simple symétrique devient le mouvement brownien grâce au *théorème limite central*, tandis qu'au niveau du générateur, le laplacien discret devient l'opérateur différentiel laplacien grâce à une formule de Taylor. Le champ libre gaussien constitue un objet fondamental en physique mathématique, abordé par exemple dans le livre [GJ87], tandis que sa limite d'échelle est présentée dans [She07].

Le modèle de battage de cartes étudié dans ce chapitre, bien que peu réaliste, a le mérite de mener très simplement au phénomène important de convergence abrupte à l'équilibre, commun à la plupart des manières de mélanger les cartes qui ont été étudiées, comme par exemple le «*riffle shuffle*» : on coupe le paquet en deux et on fusionne les deux sous-paquets en intercalant leurs cartes. La convergence abrupte dans les battages de cartes fait partie plus généralement du thème de la convergence à l'équilibre des chaînes de Markov, très étudié par Persi Diaconis et ses collaborateurs, et brillamment présenté dans les livres [AF01, LPW09].

Le collectionneur de coupons intervient également dans l'analyse de la marche aléatoire sur l'hypercube (section 3.1) 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'états 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. On peut enfin définir une marche aléatoire sur un groupe en utilisant des «incréments» i.i.d.

Recueil de modèles stochastiques – Page 26/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Chapitre 3

Modèle d'Ehrenfest

Recueil de modèles stochastiques – Page 27/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Mots-clés. Urne ; chaîne de Markov ; mesure invariante ; mesure réversible ; temps de retour ; convergence à l'équilibre.

Le *second principe de la thermodynamique* postule que pour tout système isolé, une grandeur macroscopique extensive appelée *entropie* augmente au cours du temps. Dans la *théorie cinétique des gaz de Boltzmann*, le second principe est une conséquence de la mécanique classique appliquée aux constituants microscopiques. Bien que séduisante, cette approche fut critiquée par les adversaires de l'atomisme. En effet, comment lever la contradiction entre la *réversibilité microscopique* héritée de la mécanique classique et l'*irréversiblité macroscopique* de la thermodynamique ? Ce paradoxe est résolu par le *modèle d'Ehrenfest*.

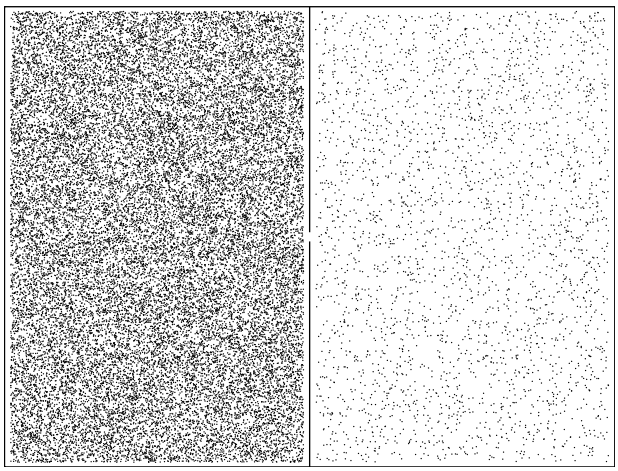


Figure 3.1 – Répartition des particules dans les deux urnes.

3.1 Modèle microscopique

Considérons deux urnes A et B dans lesquelles sont réparties a boules numérotées de 1 à a. On associe à une configuration, c'est-à-dire à une répartition des a boules un élément $x = (x(1), \ldots, x(a))$ de l'hypercube discret $F = \{0,1\}^a$ où $x(i) = 1$ si la particule i est dans

l'urne A et $x(i) = 0$ sinon. On dit que deux configurations x et y de F sont voisines, et on note $x \sim y$, si elles ne diffèrent que d'une coordonnée. Cette relation de voisinage définit sur F une structure de graphe non orienté dont les sommets sont les points de F et les arêtes sont les couples de sommets voisins. La *marche aléatoire sur l'hypercube discret* $(Y_n)_{n\geqslant 0}$ est la chaîne de Markov sur F dont la matrice de transition est donnée, pour tous $x, y \in F$, par

$$\mathbf{Q}(x,y) = \begin{cases} \dfrac{1}{a} & \text{si } x \sim y, \\ 0 & \text{sinon.} \end{cases}$$

Cette matrice est symétrique et en particulier *doublement stochastique*.

Lemme 3.1 (Propriétés)**.** *La chaîne Y de matrice de transition $\mathbf{Q}$ est irréductible, récurrente, de période 2, et la loi uniforme π sur F est invariante et réversible*[1]*. De plus, le temps moyen de retour en $x \in F$ ne dépend pas de x et vaut 2^a.*

La loi du temps d'atteinte de y en partant de x ne dépend que de la distance

$$d(x,y) = \sum_{i=1}^{a} |x_i - y_i| = \|x - y\|_1,$$

qui compte le nombre de coordonnées différentes entre x et y. Notons m_d (en oubliant un moment la dépendance en a) le temps d'atteinte moyen de y partant de x lorsque $d(x,y) = d$.

Théorème 3.2 (Temps moyens)**.** *Pour $1 \leqslant d \leqslant a$, on a*

$$m_d = \sum_{i=1}^{d} \mathbf{Q}^a_{a-i} \quad \text{où} \quad \mathbf{Q}^a_i = \sum_{k=0}^{i} \frac{\binom{a}{k}}{\binom{a-1}{i}} \quad \text{si } 1 \leqslant i \leqslant a-1.$$

Démonstration. D'après la propriété de Markov, la suite $(m_d)_{0\leqslant d\leqslant a+1}$ vérifie

$$\begin{cases} m_d = 1 + \dfrac{d}{a} m_{d-1} + \dfrac{a-d}{a} m_{d+1} & \text{si } 1 \leqslant d \leqslant a, \\ m_0 = m_{a+1} = 0, \end{cases}$$

et ce système d'équations de récurrence possède une unique solution. □

Comme Y est de période 2, la loi de Y_n ne converge pas vers la loi invariante π quand $n \to \infty$. Pour rendre la chaîne apériodique, on peut la rendre paresseuse[2]. Cela consiste à modifier la matrice de transition en posant, pour tous $x, y \in F$

$$\widetilde{\mathbf{Q}}(x,y) = \frac{1}{2}\mathbf{1}_{\{y=x\}} + \frac{1}{2}\mathbf{Q}(x,y).$$

La chaîne $(Z_n)_{n\geqslant 0}$ associée à la matrice de transition $\widetilde{\mathbf{Q}}$ reste sur place avec probabilité $1/2$ ou se déplace avec la même probabilité vers l'un des sommets voisins choisi selon la loi uniforme. Cette chaîne est irréductible récurrente apériodique et réversible pour la loi π. La loi de Z_n converge donc vers π quelle que soit la loi de Z_0. Nous présentons deux méthodes pour quantifier cette convergence. La première fait appel à une méthode de couplage et au collectionneur de coupons de la section 1.5. La seconde, plus fine, utilise la décomposition spectrale de $\widetilde{\mathbf{Q}}$. Note note $d_{\mathrm{VT}}(\cdot,\cdot)$ la distance en variation totale étudiée dans la section 1.3.

1. Sur les trajectoires de Y, cela signifie que si $Y_0 \sim \pi$ alors $(Y_0, \ldots, Y_n) \overset{\text{loi}}{=} (Y_n, \ldots, Y_0)$ pour tout $n \in \mathbb{N}$.
2. «*Lazy chain*» en anglais.

Recueil de modèles stochastiques – Page 28/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Théorème 3.3 (Temps long et collectionneur de coupons)**.** *Pour tout $n \geqslant 0$,*

$$\max_{x \in F} d_{VT}\left(\widetilde{\mathbf{Q}}^n(x,\cdot),\pi\right) \leqslant \mathbb{P}(T_a \geqslant n),$$

où T_a est le temps de complétion d'une collection de a coupons.

Démonstration. On construit simultanément deux chaînes de Markov Z et Z' de matrice de transition $\widetilde{\mathbf{Q}}$ de lois initiales respectives $\delta_{(0,\ldots,0)}$ et $\delta_{(1,\ldots,1)}$ dont les trajectoires sont égales après un temps aléatoire que l'on sait contrôler. Pour cela, on se donne deux suites indépendantes de variables aléatoires indépendantes $(U_n)_{n\geqslant 0}$ et $(V_n)_{n\geqslant 0}$ de lois respectives de Bernoulli $\mathrm{Ber}(1/2)$ et uniforme sur $\{1,\ldots,a\}$. On pose alors pour tout $n \geqslant 0$ et $1 \leqslant i \leqslant a$,

$$\begin{cases} Z_{n+1}(V_n) = U_n, \\ Z_{n+1}(i) = Z_n(i) \text{ si } i \neq V_n, \end{cases} \qquad \begin{cases} Z'_{n+1}(V_n) = U_n, \\ Z'_{n+1}(i) = Z'_n(i) \text{ si } i \neq V_n. \end{cases}$$

En d'autres termes, à chaque instant une coordonnée est choisie uniformément et elle est mise à jour à la même valeur pour les deux chaînes. Si les coordonnées étaient égales, elles le restent. Sinon, elles le deviennent. Le temps de couplage est alors le premier instant où toutes les coordonnées ont été mises à jour au moins une fois. Si les variables aléatoires $(T_i)_{1\leqslant i\leqslant a}$ désignent les temps successifs de mise à jour d'une nouvelle coordonnée alors les variables aléatoires $(G_i)_{1\leqslant i\leqslant a}$ définies par $G_1 = 1$ et $G_i = T_i - T_{i-1}$ pour $2 \leqslant i \leqslant a$ sont indépendantes et de lois géométriques de paramètres respectifs $(a+1-i)/a$. On retrouve ici le collectionneur de coupons de la section 1.5, d'où le résultat. □

Le théorème 1.15 assure que

$$\frac{T_a - a\log(a)}{a} \xrightarrow[a\to\infty]{\text{loi}} G,$$

où G est une variable aléatoire de loi de Gumbel, de fonction de répartition $r \in \mathbb{R} \mapsto e^{-e^{-r}}$. Le temps de mélange de la chaîne associée à la matrice $\widetilde{\mathbf{Q}}$ est donc de l'ordre de $a\log(a)$. On peut améliorer ce résultat d'un facteur 2.

Théorème 3.4 (Temps long)**.** *Si $n = (1/2)a\log(a) + ca$ avec $c > -(1/2)\log(\log(2))$, alors*

$$\max_{x \in F} d_{VT}\left(\widetilde{\mathbf{Q}}^n(x,\cdot),\pi\right) \leqslant \frac{e^{-c}}{\sqrt{2}}.$$

Démonstration. On munit $\mathbb{R}^F \equiv \mathbb{R}^{2^a}$ du produit scalaire usuel noté $\langle\cdot,\cdot\rangle$. La matrice stochastique $\widetilde{\mathbf{Q}}$ est symétrique[3], son spectre est réel et inclus dans $[-1,1] = \mathbb{R} \cap \{z \in \mathbb{C} : |z| \leqslant 1\}$. De plus, il existe une base orthonormée de $\mathbb{R}^F$ formée de vecteurs propres $(f_j)_{1\leqslant j\leqslant 2^a}$ de $\widetilde{\mathbf{Q}}$ associés aux valeurs propres $(\lambda_j)_{1\leqslant j\leqslant 2^a}$ classées par ordre décroissant. En particulier, λ_1 vaut 1 et f_1 est la fonction constante égale à $2^{-a/2}$. Pour toute fonction f de F dans $\mathbb{R}$ et x de F,

$$\widetilde{\mathbf{Q}}^n(f)(x) = \sum_{j=1}^{2^a} \langle f, f_j\rangle f_j(x)\lambda_j^n = \pi(f) + \sum_{j=2}^{2^a} \langle f, f_j\rangle f_j(x)\lambda_j^n.$$

En particulier, si $f(z) = \mathbf{1}_y(z)$ pour $z \in F$, on obtient la décomposition spectrale

$$\widetilde{\mathbf{Q}}^n(x,y) = \frac{1}{2^a} + \sum_{j=2}^{2^a} f_j(x)f_j(y)\lambda_j^n.$$

3. Pour toute matrice de transition sur un espace d'états fini, la loi uniforme est invariante (respectivement réversible) si et seulement si la matrice est doublement stochastique (respectivement symétrique).

Recueil de modèles stochastiques – Page 29/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

D'après le théorème 1.4 et l'inégalité de Cauchy-Schwarz,

$$4d_{\mathrm{VT}}\left(\widetilde{\mathbf{Q}}^n(x,\cdot),\pi(\cdot)\right)^2=\left(\sum_{y\in F}\left|\widetilde{\mathbf{Q}}^n(x,y)-\pi(y)\right|\right)^2\leqslant 2^a\sum_{y\in F}\left|\widetilde{\mathbf{Q}}^n(x,y)-\pi(y)\right|^2$$

Puisque $(f_i)_i$ est une base orthonormée, la décomposition spectrale précédente assure que

$$\sum_{y\in F}\left|\widetilde{\mathbf{Q}}^n(x,y)-\pi(y)\right|^2=\sum_{i,j=2}^{2^a}f_i(x)f_j(x)\langle f_i,f_j\rangle\lambda_i^n\lambda_j^n=\sum_{i=2}^{2^a}f_i(x)^2\lambda_i^{2n}.$$

Par symétrie de F et $\widetilde{\mathbf{Q}}$, la quantité $\sum_{y\in F}\left|\widetilde{\mathbf{Q}}^n(x,y)-\pi(y)\right|^2$ ne dépend pas de $x\in F$. En sommant sur $x\in F$, on obtient

$$2^a\sum_{y\in F}\left|\widetilde{\mathbf{Q}}^n(x,y)-\pi(y)\right|^2=\sum_{i=2}^{2^a}\lambda_i^{2n},$$

qui fournit la majoration

$$4d_{\mathrm{VT}}\left(\widetilde{\mathbf{Q}}^n(x,\cdot),\pi(\cdot)\right)\leqslant\sum_{i=2}^{2^a}\lambda_i^{2n}.$$

Or on peut établir que les valeurs propres de $\widetilde{\mathbf{Q}}$ sont les réels $(1-k/a)_{0\leqslant k\leqslant a}$ avec les multiplicités respectives $\binom{a}{k}$ pour $0\leqslant k\leqslant a$. On a donc

$$4d_{\mathrm{VT}}\left(\widetilde{\mathbf{Q}}^n(x,\cdot),\pi\right)\leqslant\sum_{k=1}^{a}\left(1-\frac{k}{a}\right)^{2n}\binom{a}{k}\leqslant\sum_{k=1}^{a}e^{-2nk/a}\binom{a}{k}=\left(1+e^{-2n/a}\right)^a-1.$$

Le choix $n=(1/2)a\log(a)+ca$ fournit

$$4d_{\mathrm{VT}}\left(\widetilde{\mathbf{Q}}^n(x,\cdot),\pi\right)^2\leqslant\left(1+\frac{e^{-2c}}{a}\right)^a-1\leqslant e^{e^{-2c}}-1.$$

On a $e^{e^{-2c}}-1\leqslant 2e^{-2c}$ dès que $c>-(1/2)\log(\log(2))$, ce qui conclut la preuve. □

Une autre façon de contourner le problème de la périodicité de la matrice $\mathbf{Q}$ est d'introduire une version à temps continu $(Z_t)_{t\geqslant 0}$ de la chaîne $(Y_n)_{n\geqslant 0}$. On suppose que chaque coordonnée est mise à jour (de 0 à 1 ou de 1 à 0) à des instants aléatoires. Plus précisément, soient $(N^{(i)})_{1\leqslant i\leqslant a}$ des *processus de Poisson* indépendants de même intensité $1/a$. Posons

$$Z_t(i)=Z_0(i)+N_t^{(i)}\ \mathrm{mod}2.$$

Chaque coordonnée est mise à jour indépendamment des autres à des temps exponentiels de paramètre $1/a$. Cette dynamique peut être résumée par le générateur infinitésimal du processus $\mathbf{L}=\mathbf{Q}-(1-1/a)\mathbf{I}$: pour toute fonction f de F dans $\mathbb{R}$ et tous $x,y\in F$,

$$\mathbf{L}f(x)=\frac{1}{a}\sum_{y\sim x}(f(y)-f(x)).$$

La loi uniforme π est solution de $\pi\mathbf{L}=0$ et constitue ici encore la loi invariante de Z. On peut construire un couplage (Z,Z') comme suit : les couples de coordonnées $(Z_\cdot(i),Z'_\cdot(i))_{1\leqslant i\leqslant a}$ évoluent indépendamment et pour chaque i, $Z_t(i)$ et $Z'_t(i)$ sont indépendants tant qu'ils sont différents et restent collés lorsqu'ils se touchent. Le temps de couplage S_a s'écrit alors

$$S_a=\max_{1\leqslant i\leqslant a}(E_i),$$

où les variables aléatoires $(E_i)_{1\leqslant i\leqslant a}$ sont indépendantes de loi exponentielle $\mathrm{Exp}(2/a)$. Le temps aléatoire S_a est plus facile à étudier que le temps T_a de complétion d'une collection même s'ils ont le même comportement asymptotique comme le montre le résultat suivant.

Recueil de modèles stochastiques – Page 30/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Théorème 3.5 (Temps long pour la chaîne à temps continu)**.**

$$\frac{S_a - (a/2)\log(a)}{a/2} \xrightarrow[a\to\infty]{\text{loi}} G,$$

où G est la loi de Gumbel de fonction de répartition $r \in \mathbb{R} \mapsto e^{-e^{-r}}$.

On retrouve le phénomène de convergence abrupte pour S_a autour de la valeur $(a/2)\log(a)$ avec une fenêtre de largeur a.

Démonstration. Pour tout $a \geqslant 1$ et $r \geqslant -\log(a)/2$,

$$\begin{aligned}\mathbb{P}\left(\frac{S_a - (a/2)\log(a)}{a/2} \leqslant r\right) &= \mathbb{P}\left(\max_{1\leqslant i\leqslant a}(E_i) \leqslant \frac{ar}{2} + \frac{a\log(a)}{2}\right)\\ &= \left(1 - \mathbb{P}\left(E_1 \leqslant \frac{a}{2}(r+\log(a))\right)\right)^a\\ &= \left(1 - \frac{e^{-r}}{a}\right)^a.\end{aligned}$$

Ainsi, pour tout $r \in \mathbb{R}$,

$$\mathbb{P}\left(\frac{S_a - (a/2)\log a}{a/2} \leqslant r\right) \xrightarrow[a\to\infty]{} e^{-e^{-r}},$$

ce qui est la convergence annoncée. □

3.2 Urne d'Ehrenfest

Observer l'évolution de la chaîne Y demanderait de pouvoir déterminer la position de chaque particule ce qui est physiquement impossible car a est de l'ordre du nombre d'Avogadro [4]. La grandeur macroscopique qui nous intéresse, et que l'on peut mesurer, est la proportion ou le nombre de particules situées dans l'urne A. Pour simplifier certaines expressions dans la suite, on sera amené à se placer parfois dans le cas où a est pair. La lettre b désignera alors toujours dans la suite l'entier $a/2$. À la chaîne de Markov Y qui décrit le système microscopiquement on associe le processus $(X_n)_{n\geqslant 0}$ à valeurs dans $E = \{0, 1, \ldots, a\}$ où $X_n := \sum_{i=1}^a Y_n(i)$ est la somme des coordonnées de Y_n, qui décrit le système macroscopiquement.

Théorème 3.6. *Le processus $(X_n)_{n\geqslant 0}$ est une chaîne de Markov sur E de matrice de transition*

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & \ldots & 0\\ 1/a & 0 & (a-1)/a & 0 & \ldots & \vdots\\ 0 & 2/a & 0 & (a-2)/a & \ddots & \vdots\\ \vdots & \ddots & \ddots & \ddots & \ddots & 0\\ \vdots & \ldots & 0 & (a-1)/a & 0 & 1/a\\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

irréductible, récurrente, de période 2. La loi binomiale $\mathrm{Bin}(a, 1/2)$ *est invariante et réversible.*

4. Ce nombre approximativement égal à 6.022×10^{23} définit la mole et correspond au nombre d'atomes dans 12 grammes de carbone 12. C'est aussi environ le nombre de molécules contenues dans 22.4 litres d'un gaz parfait dans les conditions normales de température (20 °C) et de pression (1 Pa).

Recueil de modèles stochastiques – Page 31/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Démonstration. Soit $\mathcal{F}_n$, respectivement $\mathcal{G}_n$, la tribu engendrée par $(X_k)_{0\leqslant k\leqslant n}$, respectivement $(Y_k)_{0\leqslant k\leqslant n}$, et soit $S(y) = y(1) + \cdots + y(a)$ pour tout $y \in F$. Puisque $X_n = S(Y_n)$, la tribu $\mathcal{F}_n$ est incluse dans la tribu $\mathcal{G}_n$. On a donc, pour toute fonction f de E dans $\mathbb{R}$,

$$\mathbb{E}(f(X_{n+1}) \,|\, \mathcal{F}_n) = \mathbb{E}(\mathbb{E}(f(X_{n+1}) \,|\, \mathcal{G}_n) \,|\, \mathcal{F}_n).$$

Par définition de X et puisque Y est une chaîne de Markov,

$$\mathbb{E}(f(X_{n+1}) \,|\, \mathcal{G}_n) = \mathbb{E}(f(S(Y_{n+1})) \,|\, \mathcal{G}_n) = \mathbb{E}(f(S(Y_{n+1})) \,|\, Y_n).$$

Enfin,

$$\mathbb{E}(f(S(Y_{n+1})) \,|\, Y_n) = \left(1 - \frac{S(Y_n)}{a}\right) f(S(Y_n) + 1) + \left(\frac{S(Y_n)}{a}\right) f(S(Y_n) - 1).$$

On a donc

$$\begin{aligned}\mathbb{E}(f(X_{n+1}) \,|\, \mathcal{F}_n) &= \mathbb{E}\left(\left(1 - \frac{X_n}{a}\right) f(X_n + 1) + \left(\frac{X_n}{a}\right) f(X_n - 1) \,|\, \mathcal{F}_n\right) \\ &= \left(1 - \frac{X_n}{a}\right) f(X_n + 1) + \left(\frac{X_n}{a}\right) f(X_n - 1),\end{aligned}$$

ce qui prouve que X est une chaîne de Markov de matrice de transition $\mathbf{P}$. La chaîne est irréductible puisque i et $i + 1$ communiquent pour tout $0 \leqslant i \leqslant a$. Comme l'espace d'états est fini, la matrice est donc récurrent (positive). Pour tout n, X_n et X_{n+1} n'ont pas la même parité et $\mathbf{P}^2_{0,0} > 0$ et la chaîne est de période 2.

Note : en général, l'image d'une chaîne de Markov n'est pas une chaîne de Markov. □

3.3 Espérance et variance de la pression au cours du temps

La pression au temps n dans l'urne A est de l'ordre de $P_n = X_n/a$. On s'intéresse ici aux deux premiers moments de cette variable aléatoire.

Théorème 3.7 (Pression)**.** *Pour tout $n \in \mathbb{N}$, en notant $\alpha = 1 - 2/a$ et $\beta = 1 - 4/a$,*

$$\begin{aligned}\mathbb{E}(P_n) &= \frac{1}{2} + \left(\mathbb{E}(P_0) - \frac{1}{2}\right)\alpha^n, \\ \mathbb{V}(P_n) &= \frac{1 - \beta^n}{4a} + \mathbb{V}(P_0)\beta^n - \left(\mathbb{E}(P_0) - \frac{1}{2}\right)^2 (\alpha^{2n} - \beta^n).\end{aligned}$$

Démonstration. On raisonne par conditionnement :

$$\mathbb{E}(X_{n+1} \,|\, X_n) = (X_n + 1)(1 - \frac{X_n}{a}) + (X_n - 1)\frac{X_n}{a} = \alpha X_n + 1.$$

On en déduit, en prenant l'espérance, une relation de récurrence pour la suite $(\mathbb{E}(P_n))_n$ qui donne immédiatement la formule pour la moyenne. De même,

$$\mathbb{E}(X_{n+1}^2 \,|\, X_n) = (X_n + 1)^2\left(1 - \frac{X_n}{a}\right) + (X_n - 1)^2 X_n/a = \left(1 - \frac{4}{a}\right) X_n^2 + 2X_n + 1.$$

Ceci implique

$$\mathbb{V}(X_{n+1}) = \beta\mathbb{V}(X_n) + \frac{4}{a^2}\mathbb{E}(X_n)(a - \mathbb{E}(X_n)) = \beta\mathbb{V}(X_n) + 1 - \frac{4}{a^2}\left(\mathbb{E}(X_0) - \frac{a}{2}\right)^2 \alpha^{2n}.$$

Un simple calcul fournit alors la formule pour la variance. □

Recueil de modèles stochastiques – Page 32/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Remarque 3.8 (Valeurs et vecteurs propres)**.** *Soit f la fonction sur E définie par $f(x) = x-b$ où $b = a/2$, qui s'identifie avec le vecteur colonne $(-b, -b+1, \ldots, b-1, b)^T$. Un calcul immédiat montre que $\mathbf{P}f = (1-2/a)f$. Ainsi, pour tout $n \geqslant 0$ et tout $x \in E$,*

$$\mathbb{E}(X_n \,|\, X_0 = x) - b = \mathbb{E}(f(X_n) \,|\, X_0 = x) = \mathbf{P}^n f(x) = \alpha^n f(x) = \alpha^n (x-b).$$

On retrouve la relation pour la moyenne du théorème 3.7 en prenant l'espérance par rapport à la loi de X_0. Plus généralement, on peut montrer que si $a = 2b$ est pair alors les valeurs propres de $\mathbf{P}$ sont les rationnels $(i/b)_{-b \leqslant i \leqslant b}$.

D'après le théorème 3.7, si le vide est fait dans l'urne A à l'instant initial, alors $X_0 = 0$, $P_0 = 0$, et la pression moyenne $\mathbb{E}(P_n)$ au temps n converge géométriquement vers $1/2$ quand $n \to \infty$. D'autre part $\mathbb{V}(P_n) \leqslant 1/(4a)$ puisque $\mathbb{V}(P_0) = 0$ et $\alpha^2 \geqslant \beta$, et l'inégalité de Bienaymé-Tchebychev donne

$$\mathbb{P}(|P_n - \mathbb{E}(P_n)| \geqslant \varepsilon) \leqslant \frac{1}{4a\varepsilon^2}.$$

Comme a est très grand, la variable P_n est quasiment déterministe, ce qui correspond à l'expérience. La figure 3.2 donne la trajectoire de $(P_n)_{n \geqslant 0}$ pour différentes conditions initiales.

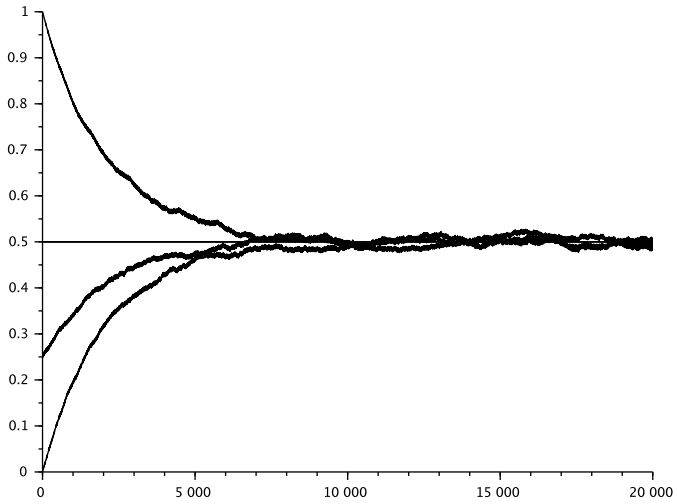


FIGURE 3.2 – Évolution de la proportion de particules dans l'urne A pour différentes valeurs initiales 0, 1000, et 4000, pour $a = 4000$ particules.

3.4 Fluctuations autour de la moyenne à l'équilibre

Supposons que le système soit à l'équilibre, c'est-à-dire que X soit de loi binomiale $\mathrm{Bin}(a, 1/2)$. Alors, $P = X/a$ vérifie, en vertu du *théorème limite central*,

$$\mathbb{P}\left(|P - 1/2| \geqslant \frac{x}{2\sqrt{a}}\right) \underset{a \text{ grand}}{\simeq} 2(1 - \Phi(x)) \leqslant \frac{2}{x\sqrt{2\pi}} e^{-x^2/2},$$

où Φ est la fonction de répartition de la loi gaussienne standard $\mathcal{N}(0,1)$. Une autre façon d'estimer cette probabilité consiste à utiliser l'inégalité de Chernov. Si $(Z_i)_{i \geqslant 1}$ sont des variables

Recueil de modèles stochastiques – Page 33/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

aléatoires indépendantes et identiquement distribuées de loi de Rademacher $(1/2)(\delta_{-1}+\delta_1)$ alors, pour tout $r \geqslant 0$ et tout $\lambda > 0$,

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} Z_i \geqslant r\right) = \mathbb{P}\left(\exp\left(\lambda \sum_{i=1}^{n} Z_i\right) \geqslant e^{\lambda n r}\right) \leqslant e^{-\lambda n r}\mathrm{ch}(\lambda)^n.$$

En remarquant que $\mathrm{ch}(\lambda) \leqslant \exp\left(\lambda^2/2\right)$, grâce par exemple à un développement en série, et en optimisant en λ, ce qui conduit à choisir $\lambda = r$, on obtient

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} Z_i \geqslant r\right) \leqslant e^{-nr^2/2}.$$

Ceci assure que

$$\mathbb{P}(|P - 1/2| \geqslant r) \leqslant 2e^{-2ar^2}.$$

Si $a = 5 \times 10^{23}$ alors la probabilité de trouver une fluctuation supérieure au milliardième est inférieure à $2e^{-10^5}$. Cette probabilité extrêmement petite rend l'événement inaccessible.

3.5 Quand reverra-t-on le vide ?

La chaîne d'Ehrenfest $(X_n)_{n\geqslant 0}$ est irréductible à espace d'états fini, et elle visite donc tous les états, et les temps d'atteinte de chaque état est intégrable. Ceci peut sembler paradoxal si l'on pense à l'irréversibilité macroscopique. Essayons donc d'estimer ces temps d'atteinte moyens. Le plus simple est de s'intéresser tout d'abord aux *temps de retour*. On définit le temps T_{ii} de premier retour en i de la manière suivante :

$$T_{ii} = \inf\{n \geqslant 1,\ X_n = i \,|\, X_0 = i\}.$$

Théorème 3.9 (Temps de retours moyens)**.** *Si a est pair et $b = a/2$ alors*

$$\mathbb{E}(T_{00}) = 2^a \quad \textit{et} \quad \mathbb{E}(T_{bb}) \sim \sqrt{\pi b}.$$

Démonstration. L'espérance du temps de retour en i est égale à $1/\pi(i)$ où π est la loi invariante, qui n'est rien d'autre que $\mathrm{Bin}(a, 1/2)$. On a $\pi(i) = 2^{-a}\binom{a}{i}$, et le résultat est alors immédiat pour $i = 0$ et découle de la formule de Stirling pour $i = b$. □

Il faut donc attendre un temps immense avant que l'urne A initialement vide, ne le redevienne ! On souhaite à présent plus de précision et obtenir des estimations des temps d'atteinte de 0 et b partant de 0 ou b. Soit m_{ij} le temps moyen d'atteinte de j par la chaîne partant de i. Partant de $i+1$, la chaîne doit passer par i pour aller en 0, donc pour $0 \leqslant i < a$,

$$m_{i+1,0} = m_{i+1,i} + m_{i,0}.$$

Par symétrie, pour $0 \leqslant i < a$,

$$m_{i,i+1} = m_{a-i,a-i-1} = m_{a-i,0} - m_{a-i-1,0}.$$

Il suffit donc de connaître les réels $(m_{i,0})_i$ pour déterminer les réels $(m_{ij})_{ij}$. Remarquons à présent que $X_n = 0$ si et seulement si $Y_n = (0, \ldots, 0)$. Donc le temps moyen que met X à se rendre de i à 0 est le temps moyen que met Y d'aller de n'importe quel élément de F situé à une distance i de $(0, \ldots, 0)$ à $(0, \ldots, 0)$, autrement dit m_i^a. Le théorème 3.2 assure alors que

$$m_{i,i+1} = m_{a-i,0} - m_{a-i-1,0} = \sum_{k=1}^{a-i} \mathbf{Q}^a_{a-k} - \sum_{k=1}^{a-i-1} \mathbf{Q}^a_{a-k} = \mathbf{Q}^a_i \quad \text{pour } 0 \leqslant i < a-1.$$

Recueil de modèles stochastiques – Page 34/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Corollaire 3.10 (Temps d'atteinte moyens)**.** *On a*

$$m_{ij} = \begin{cases} 2^a / \binom{a}{i} & \text{si } i = j, \\ \sum_{k=i}^{j-1} \mathbf{Q}_k^a & \text{si } i < j, \\ \sum_{k=a-i}^{a-j-1} \mathbf{Q}_k^a & \text{si } i > j. \end{cases}$$

Théorème 3.11 (Temps d'atteinte moyens)**.** *Si a est pair, alors, en notant $b = a/2$,*

$$a(2\log(2) - 1/2) - 2 \leqslant m_{0,b} \leqslant a(1/4 + (1/2)\log(a)),$$

et

$$2^a\left(1 + \frac{1}{a-1}\right) \leqslant m_{0,a} \leqslant 2^a\left(1 + \frac{2}{a-1}\right).$$

Démonstration. On se contente d'établir la première estimation. Pour tout $0 \leqslant i \leqslant a-1$,

$$m_{i,i+1} + m_{i+1,i} = \frac{1}{\binom{a-1}{i}}\left(\sum_{k=0}^{i}\binom{a}{k} + \sum_{k=i+1}^{a}\binom{a}{k}\right) = \frac{2^a}{\binom{a-1}{i}}.$$

Donc

$$m_{0,b} + m_{b,0} = 2^a \sum_{i=0}^{b-1} \frac{1}{\binom{a-1}{i}}.$$

De plus, puisque $m_{b,0} = m_{b,a}$, on a

$$m_{0,a} = m_{0,b} + m_{b,a} = 2^a \sum_{i=0}^{b-1} \frac{1}{\binom{a-1}{i}}.$$

Remarquons à présent, en séparant les deux premiers termes de la somme des autres, que

$$1 + \frac{1}{a-1} \leqslant \sum_{i=0}^{b-1} \frac{1}{\binom{a-1}{i}} \leqslant 1 + \frac{1}{a-1} + \frac{b-1}{\binom{a-1}{2}}$$

ce qui fournit l'encadrement souhaité. □

Le théorème 3.11 assure que le temps mis par le système pour passer d'un état de total déséquilibre à celui d'équilibre parfait est extrêmement négligeable devant le temps mis pour l'évolution inverse. Pour un nombre de particules égal à 100, $m_{0,50}$ est majoré par 256 tandis que $m_{50,0}$ est de l'ordre de 10^{30} soit quelques mille milliards de milliards de milliards.

3.6 Notes et commentaires

Le temps de mélange pour la marche aléatoire sur l'hypercube vérifie une propriété de convergence abrupte autour de $(a/2)\log(a)$ de fenêtre a : dans un intervalle de longueur de l'ordre de a autour de $(a/2)\log a$ la distance en variation totale entre $\widetilde{\mathbf{Q}}^n(x,\cdot)$ et π passe d'un voisinage de 1 à un voisinage de 0. La borne supérieure est fournie par le théorème 3.4, qu'on peut compléter par l'estimation suivante qui se trouve dans le livre David Levin, Yuval Peres, et Elizabeth Wilmer [LPW09] : si $n = (1/2)a\log(a) - ca$ avec $c > 0$, alors

$$\max_{x\in F} d_{\mathrm{VT}}\left(\widetilde{\mathbf{Q}}^n(x,\cdot), \pi\right) \geqslant 1 - 8e^{-(2c+1)}.$$

Recueil de modèles stochastiques – Page 35/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Ce résultat semble provenir d'un travail de Persi Diaconis et Mehrad Shahshahani [DS87]. Les valeurs propres et vecteurs propres de la matrice de transition $\mathbf{P}$ de la chaîne d'Ehrenfest X sont obtenus par Mark Kac dans [Kac47]. On peut également les obtenir à partir de ceux de la matrice de transition $\widetilde{\mathbf{Q}}$ du modèle microscopique (marche sur l'hypercube) qui sont décrits dans le livre [LPW09, exemple 12.15].

La relation classique entre espérance du temps de retour et mesure invariante utilisée dans la preuve du théorème 3.9 est établie par exemple dans le livre de James Norris [Nor98]. Les théorèmes 3.2 et 3.11 sont tirés du livre de John Kemeny et Laurie Snell [KS76].

Les particules passent très vite d'une urne à l'autre. Ainsi, le temps microscopique est différent du temps macroscopique. Considérons qu'une unité de temps macroscopique corresponde à a changements d'urnes. Quelle est alors l'évolution de la pression dans cette nouvelle échelle de temps ? Le théorème 3.7 entraîne immédiatement que, pour tout $t > 0$,

$$\mathbb{E}\big(P_{\lfloor at \rfloor}\big) \xrightarrow[a\to\infty]{} \frac{1}{2} + \left(\mathbb{E}(P_0) - \frac{1}{2}\right)e^{-2t} \quad \text{et} \quad \mathbb{E}\big(V_{\lfloor at \rfloor}\big) \xrightarrow[a\to\infty]{} \mathbb{V}(X_0)e^{-4t},$$

puisque $\lfloor at \rfloor / a$ tend vers t. Cette convergence est la partie émergée d'un iceberg : convenablement renormalisée en temps et en espace, la chaîne d'Ehrenfest converge en loi vers le *processus de diffusion d'Ornstein-Uhlenbeck*, comme présenté dans le chapitre 26.

Recueil de modèles stochastiques – Page 36/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Chapitre 4

Modèle de Fisher-Wright

Recueil de modèles stochastiques – Page 37/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Mots-clés. Chaîne de Markov ; martingale ; absorption ; invariance.

Ce chapitre est consacré à des modèles d'évolution de génotypes au fil des générations dans une population de taille constante. Le point de départ est la loi de Hardy-Weinberg liée à la théorie de Mendel, qui nous conduit au modèle de Fisher-Wright, le coeur du chapitre, puis aux modèles de Cannings et de Moran, inspirés du modèle de Fisher-Wright.

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

Théorème 4.1 (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 hermaphrodite*[1] *diploïde*[2] *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

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

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

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

Elle est donc bien constante au cours du temps. □

La *loi de Hardy-Weinberg* est généralisable à un locus [3] 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.

Recueil de modèles stochastiques – Page 38/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

4.1 Modèle de Fisher-Wright et fixation

Le *modèle de Fisher-Wright* est 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 $(X_n)_{n\geqslant 0}$ comme une suite récurrente aléatoire $X_{n+1}=\sum_{k=1}^N \mathbb{1}_{\{U_{n,k}\leqslant X_n/N\}}$ où $(U_{n,k})_{n\geqslant 1,1\leqslant k\leqslant N}$ 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

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

Le processus $(X_n)_{n\geqslant 0}$ 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

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

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 $(X_n)_{n\geqslant 0}$ 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 4.2 (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 $\mathbb{P}(X_{n+1}\in\{0,N\}\,|\,X_n=x)=\mathbf{P}(x,\{0,N\})\geqslant p_*$ où*

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

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

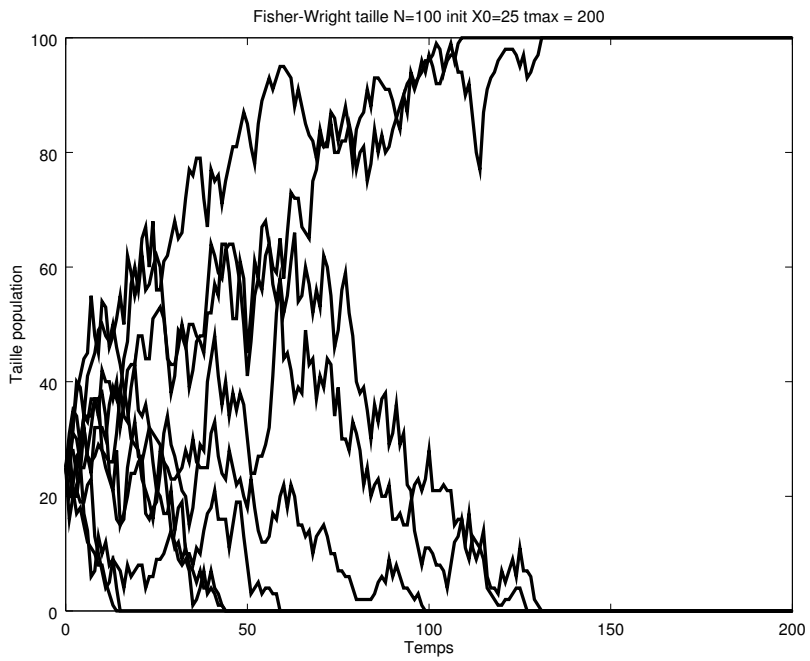


FIGURE 4.1 – Quelques trajectoires, jusqu'au temps $t = 200$, de la chaîne de Fisher-Wright pour une population de taille $N = 100$, toutes issues de $X_0 = N/4 = 25$.

(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}. *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 $(I_{x,y})_{0 \leqslant y \leqslant N}$ une partition de l'intervalle $[0, 1]$, avec $I_{x,y}$ de longueur $|I_{x,y}| = \mathbf{P}(x, y)$, et soit $(U_n)_{n \geqslant 1}$ une suite de v.a.r. i.i.d. de loi uniforme sur $[0, 1]$, indépendante de X_0. On réalise $(X_n)_{n \geqslant 0}$ 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 $p_* = \min_{0 \leqslant x \leqslant N} |I_{x,0} \cup I_{x,N}|$ et*

$$T := \inf\{n \geqslant 0 : X_n \in \{0, N\}\} \leqslant T_* := \inf\{n \geqslant 1 : \mathbb{1}_{\{U_n \leqslant p_*\}} = 1\}$$

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

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

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

Recueil de modèles stochastiques – Page 39/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

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 $(X_n)_{n\geqslant 0}$ 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 $\lim_{n\to\infty} \mathbb{E}_x(X_n) = \mathbb{E}(X_T) = 0\mathbb{P}_x(X_T = 0) + N\mathbb{P}_x(X_T = N) = N\mathbb{P}_x(X_T = N)$. Or grâce à la loi conditionnelle binomiale $\mathcal{L}(X_n|\mathcal{F}_{n-1}) = \mathcal{L}(X_n|X_{n-1}) = \mathrm{Bin}(N, \frac{X_{n-1}}{N})$, on a $\mathbb{E}(X_n) = \mathbb{E}(\mathbb{E}(X_n|\mathcal{F}_{n-1})) = \mathbb{E}(N\frac{X_{n-1}}{N}) = \mathbb{E}(X_{n-1}) = \cdots = \mathbb{E}(X_0) = x$, d'où le résultat. □

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 $(X_n)_{n\geqslant 0}$.

Remarque 4.5 (Martingale)**.** *En reprenant la preuve du théorème 4.4, 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 $(X_n)_{n\geqslant 0}$ est une martingale pour sa filtration naturelle $(\mathcal{F}_n)_{n\geqslant 0}$ où $\mathcal{F}_n = \sigma(X_0, \ldots, X_n)$. En particulier, on a bien sûr la conservation $\mathbb{E}(X_0) = \mathbb{E}(X_n)$ pour tout $n \geqslant 1$. Le théorème de Doob indique que $(X_{n\wedge T})_{n\geqslant 0}$ 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 quand $n \to \infty$ car T est fini p.s. et X est uniformément bornée. Comme T est un temps de fixation, on a en fait $X_n = X_{n\wedge T}$. Il est également possible d'appliquer le théorème de convergence p.s. et dans L^p des martingales bornées dans L^p avec $p > 1$.

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

$$H_n = \frac{2X_n(N - X_n)}{N(N-1)} = \frac{2}{N-1}\mathbb{V}(X_{n+1} \,|\, X_n).$$

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

Théorème 4.6 (Décroissance exponentielle de l'hétérozygotie)**.** *Pour tout $n \geqslant 1$,*

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

et

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

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

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

On a

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

On écrit alors

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

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

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

□

Recueil de modèles stochastiques – Page 40/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

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 :

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

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 $Y^{(N)} = (\frac{1}{N}X_{\lfloor Nt \rfloor})_{t\in\mathbb{R}_+}$ 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 $\log(N)$. Les approximations ci-dessus sont rendues rigoureuse en établissant que la suite de processus $Y^{(N)}$ converge en loi quand $N \to \infty$ vers un processus de diffusion Y sur $[0,1]$, solution de l'équation différentielle stochastique $dY_t = \sqrt{Y_t(1-Y_t)}dB_t$, comme abordé dans le chapitre 26.

4.2 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 $(X_n)_{n\geqslant 0}$ à valeurs dans $\{0,1,\ldots,N\}$ qui vérifie $\mathcal{L}(X_{n+1} \,|\, X_0,\ldots,X_n) = \mathrm{Bin}(N,\psi_{X_n})$ où

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

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

— 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 $\{1,\ldots,N-1\}$ 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 $p_x = x/N$;

Recueil de modèles stochastiques – Page 41/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

— si $u = 0$ et $0 < v \leqslant 1$ 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 $0 < u \leqslant 1$ 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 $0 < u \leqslant 1$ et $0 < v < 1$ ou si $0 < u < 1$ et $0 < v \leqslant 1$ 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

$$\lim_{n\to\infty} \frac{\mathrm{card}\{0 \leqslant k \leqslant n : X_k = x\}}{n+1} = \mu(x).$$

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 $x \in \{0, 1, \dots, N\}$,

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

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 $\mathbf{P}(x,\cdot) = \mathrm{Bin}(N, \frac{(1-u)x+v(N-x)}{N})$, 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

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

Pour la variance σ^2 de μ, un calcul permet d'établir que

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

Notons que si $u = v$ alors $m = N/2$ ne dépend plus de u et v. Au delà des deux premiers moments, on peut établir, comme présenté dans le chapitre 26, que $Y^{(N)} = (\frac{1}{N}X_{\lfloor Nt \rfloor})_{t\in\mathbb{R}_+}$ converge en loi quand $(N, Nu, Nv) \to (\infty, \alpha, \beta)$ 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 $\mathrm{Beta}(2\beta, 2\alpha)$, dont la moyenne est $\frac{\beta}{\alpha+\beta}$ et la variance $\frac{\alpha\beta}{(\alpha+\beta)^2(2(\alpha+\beta)+1)}$.

Recueil de modèles stochastiques – Page 42/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

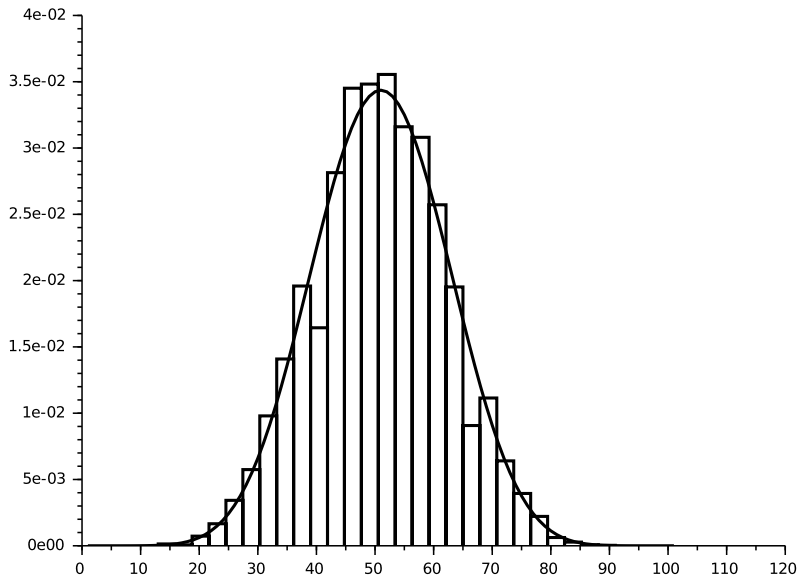


FIGURE 4.2 – Loi invariante, et histogramme d'un échantillon de taille 10000 de X_{100} avec $X_0 = 50$, pour une population de taille $N = 100$, avec taux de mutation $u = v = 0.05$.

4.3 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 $\mathcal{L}(X_{n+1} \,|\, X_0, \ldots, X_n) = \mathcal{L}(X_{n+1} \,|\, X_n) = \mathrm{Bin}(N, \psi_{X_n})$. Le processus $(X_n)_{n\geqslant 0}$ 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 $\{1, \ldots, N-1\}$ 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 $Y^{(N)} = (\frac{1}{N} X_{\lfloor Nt \rfloor})_{t\in\mathbb{R}_+}$ 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

Recueil de modèles stochastiques – Page 43/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

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. Comme présenté dans le chapitre 26, il est possible de rendre rigoureuse cette analyse heuristique en établissant que $Y^{(N)}$ converge en loi quand $N \to \infty$ et $Ns \to \alpha$ 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.

4.4 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 $y_1, \ldots, y_N$ désignent le nombre d'enfants de chaque individu de la génération n (numérotés de 1 à N) alors on a $y_1 + \cdots + y_N = N$ et $(y_1, \ldots, y_N)$ suit la loi multinomiale symétrique de taille N et de paramètre $(1/N, \ldots, 1/N)$. 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 $(y_1, \ldots, y_N)$ vérifie $y_1 + \cdots + y_N = N$ et possède une loi échangeable, c'est-à-dire que pour toute permutation $\tau \in \mathcal{S}_N$,

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

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

$$0 = \mathbb{V}(y_1 + \cdots + y_N) = N\mathbb{V}(y_1) + N(N-1)\mathrm{Cov}(y_1, y_2).$$

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 $(X_n)_{n\geqslant 0}$ d'espace d'états $\{0, 1, \ldots, N\}$ donnée pour tout $x \in \{0, 1, \ldots, N\}$ par

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

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 $(X_n)_{n\geqslant 0}$ est une martingale. De plus,

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

Remarque 4.7 (Loi multinomiale et ses marges)**.** *Soit $n, d \in \mathbb{N}^*$ et $(p_1, \ldots, p_d) \in [0,1]^d$ avec $p_1 + \cdots + p_d = 1$. La loi multinomiale de taille n et de paramètre $(p_1, \ldots, p_d)$ est la loi sur*

Recueil de modèles stochastiques – Page 44/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

l'ensemble fini $\{(n_1,\dots,n_d)\in\mathbb{N}^d : n_1+\cdots+n_d=n\}$ *donnée par*

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

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* $p_1=\cdots=p_d=1/d$. *Si* $(X_1,\dots,X_d)\sim\mathrm{Mul}(n,(p_1,\dots,p_d))$ *et si* $I_1,\dots,I_r$ *est une partition de* $\{1,\dots,n\}$ *alors*

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

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

Théorème 4.8 (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* $j\in\{1,\dots,N\}$,

$$\lambda_j=\mathbb{E}(y_1y_2\cdots y_j).$$

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

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

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

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

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

$$\mathbb{E}(X_1^j\,|\,X_0=i)=\mathbb{E}((y_1+\ldots+y_i)^j)=i^{[j]}\mathbb{E}(y_1y_2\cdots y_j)+\sum_{k=0}^{j-1}b_{i,k}i^k,$$

où $i^{[j]}=i(i-1)(i-2)\cdots(i-j+1)$. 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 $T_{0,0}=1$, $T_{j,j}=\mathbb{E}(y_1\cdots y_j)$ 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. □

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 $\lambda_2=\mathbb{E}(y_1y_2)=\mathrm{Cov}(y_1,y_2)+\mathbb{E}(y_1)^2=1-\sigma^2/(N-1)$.

Recueil de modèles stochastiques – Page 45/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

4.5 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 $(y_1, \ldots, y_N)$ 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 $x, y \in \{0, 1, \ldots, N\}$ 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 $x \in \{1, \ldots, N-1\}$. 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 $N \geqslant 1$, posons $x_N = \lfloor pN \rfloor$. On a alors

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

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.

4.6 Notes et commentaires

Les mécanismes de la génétique ont été considérablement discutés depuis leur découverte par Mendel au dix-neuvième siècle. La loi de Godfrey Harold Hardy et de Wilhelm Weinberg a été découverte de manière indépendante par ces deux scientifiques au tout début du vingtième siècle. Le modèle de Ronald Aylmer Fisher et de Sewall Wright date des années 1930. Le modèle de John Moran date des années 1950 tandis que le modèle de Chris Cannings date

Recueil de modèles stochastiques – Page 46/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

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 26.

Recueil de modèles stochastiques – Page 47/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Chapitre 5

Branchement

Recueil de modèles stochastiques – Page 49/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Mots-clés. Arbre ; branchement ; généalogie ; fonction génératrice ; martingale ; chaîne de Markov ; marche aléatoire.

Le processus de Galton-Watson décrit l'évolution d'une population asexuée au fil de générations qui ne se recoupent pas. Soit $P = p_0\delta_0 + p_1\delta_1 + \cdots$ 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 $(X_{n,k})_{n\geqslant 1, k\geqslant 1}$ sont i.i.d. de loi P, indépendantes de Z_0. On adopte la convention $\sum_{\varnothing} = 0$ de sorte que si $Z_n = 0$ alors $Z_{n+1} = 0$. Pour tous $n \in \mathbb{N}$ et $z_0, \ldots, z_n \in \mathbb{N}$, on a

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

La suite $(Z_n)_{n\geqslant 0}$ 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

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

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 5.1 (Arbre de Galton-Watson)**.** *La suite $(Z_n)_{n\geqslant 0}$ 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 (ancêtre général) possède r voisins tandis que tous les autres sommets (descendants de l'ancêtre général) possèdent $r+1$ voisins. L'arbre a pour* profondeur *ou* hauteur $T = \inf\{n \in \mathbb{N} : Z_n = 0\}$*, pour* largeur $\max_{n\in\mathbb{N}} Z_n$*, et pour* taille totale $N = Z_0 + \cdots + Z_{T-1}$*. L'arbre de la figure 5.2 est de profondeur* 3*, largeur* 5*, et taille totale* 12*. Lorsque $Z_0 > 1$, on parle parfois de* forêt aléatoire*.*

On effectue à présent les observations suivantes :

— Le comportement de $(Z_n)_{n\geqslant 0}$ se ramène au cas $Z_0 = 1$ car[1] conditionnellement à $\{Z_0 = z\}$, la suite $(Z_n)_{n\geqslant 0}$ 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 $Z_{n+1} = zZ_n = \cdots = z^{n+1}Z_0$;
— Si $p_0 = 0$ et $p_1 < 1$ alors $\mathbb{P}(Z_n \nearrow \infty \,|\, Z_0 = 1) = 1$ ($Z_n \nearrow$, comparer à pile ou face) ;
— Si $p_0 + p_1 = 1$ et $p_0 \neq 0$, alors $\mathbb{P}(Z_n \searrow 0 \,|\, Z_0 = 1) = 1$ ($Z_n \searrow$, T est géométrique).

Ces observations nous conduisent à supposer que dans toute la suite :

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

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

5.1 Extinction et phénomène de seuil

Nous commençons par les deux premiers moments de Z_n.

Théorème 5.2 (Moyenne et variance)**.** *Si m existe alors $\mathbb{E}(Z_n) = m^n$ pour tout $n \geqslant 0$ et si $\sigma^2 < \infty$ alors $\mathbb{V}(Z_n) = \sigma^2(m^n(m^n - 1))/(m^2 - m)$ prolongé en $n\sigma^2$ si $m = 1$.*

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

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 $\mathbb{V}(Z_n)$ 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 $\lim_{s\to 1} g^{(r)}(s) = \mathbb{E}(X_{1,1}(X_{1,1}-1)\cdots(X_{1,1}-r+1))$ (*moment factoriel d'ordre r de P*). En particulier, $g'(1^-) = m$ et $g''(1^-) = \sigma^2 + m^2 - m$.

Théorème 5.3 (Génératrice)**.** *Pour tout $n \geqslant 1$ 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 $n \geqslant 1$ et tout $s \in [0,1]$,

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

□

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

Recueil de modèles stochastiques – Page 50/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Remarque 5.4 (Liens entre Galton-Watson et Fisher-Wright)**.** *Considérons le cas où la loi de reproduction P est une loi de Poisson* $\mathrm{Poi}(\lambda)$*. Pour tout entier $N \geqslant 1$ fixé,*

$$\mathcal{L}((X_{n+1,1}, \dots, X_{n+1,N}) \,|\, X_{n+1,1} + \cdots + X_{n+1,N} = N) = \mathrm{Mul}(N, (1/N, \dots, 1/N)).$$

Ainsi, le processus de Fisher-Wright du chapitre 4 est lié au processus de Galton-Watson de loi de reproduction Poisson conditionné à être de taille constante (exercice 10.1). Le processus de Galton-Watson décrit l'évolution de la taille de la population, et ne dit rien à l'échelle individuelle. Réciproquement, si $(Y_1, \dots, Y_N) \sim \mathrm{Mul}(N, (1/N, \dots, 1/N))$ *alors pour tout $k \geqslant 1$ fixé,* $(Y_1, \dots, Y_k)$ *converge en loi quand $N \to \infty$ vers la loi produit* $\mathrm{Poi}(1)^{\otimes k}$*, 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 5.5 (Dichotomie)**.** *Presque sûrement, soit le processus s'éteint c'est-à-dire* $\lim_{n\to\infty} Z_n = 0$ *soit il tend vers l'infini c'est-à-dire* $\lim_{n\to\infty} Z_n = \infty$*. Autrement dit, p.s.*

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

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 $\mathbb{N}^*$. □

Le théorème 5.5 suggère de considérer la probabilité d'extinction $\mathbb{P}(T < \infty)$.

Théorème 5.6 (Probabilité d'extinction)**.** *La probabilité d'extinction* $\mathbb{P}(T < \infty)$ *est point fixe de g. Si $m \leqslant 1$ 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 $\{T < \infty\} = \{\lim_{n\to\infty} Z_n = 0\} = \cup_{n\geqslant 0}\{Z_n = 0\}$. La suite $(\{Z_n = 0\})_{n\geqslant 0}$ est croissante et la *probabilité d'extinction* vérifie donc

$$\mathbb{P}(T < \infty) = \mathbb{P}(\lim_{n\to\infty} Z_n = 0) = \mathbb{P}(\cup_{n\geqslant 0}\{Z_n = 0\}) = \lim_{n\to\infty} \mathbb{P}(Z_n = 0).$$

Or du théorème 5.3 on tire $\mathbb{P}(Z_n = 0) = g_n(0) = g^{\circ n}(0)$ et donc

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

La probabilité d'extinction $\mathbb{P}(T < \infty) = \lim_{n\to\infty} g^{\circ n}(0)$ est un point fixe de g (i.e. $g(s) = s$) car $g : [0, 1] \to [0, 1]$ 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 $g'(1^-) = m \leqslant 1$ 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 5.1 illustre des cas particuliers tirés des exemples 5.7 et 5.8. □

Recueil de modèles stochastiques – Page 51/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

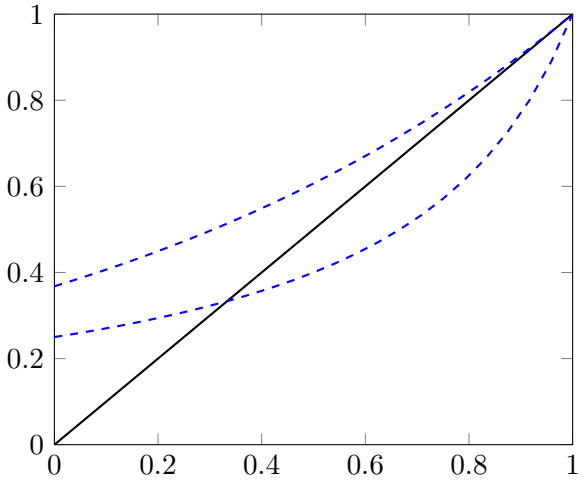


Figure 5.1 – En pointillés, la fonction génératrice de la loi géométrique de paramètre 1/4 (deux points fixes : 1/3, attractif, et 1, répulsif) ainsi que celle de la loi de Poisson de paramètre 1 (unique point fixe en 1). La ligne continue est la première bissectrice.

Exemple 5.7 (Reproduction poissonnienne)**.** *Si* $P = \mathrm{Poi}(\lambda)$ *alors* $m = \lambda$ *et* $g(s) = e^{\lambda(s-1)}$ *et donc si* $\lambda \leqslant 1$ *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 5.8 (Reproduction géométrique)**.** *Si* $P = \mathrm{Geo}_{\mathbb{N}}(p) = \sum_{n\geqslant 0} q^n p \delta_n$ *alors* $m = q/p$ *et* $g(s) = p/(1-sq)$ *et donc si* $p \geqslant 1/2$ *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 5.6 :
— 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 $(Z_n)_{n\geqslant 0}$, on introduit la suite $(Y_n)_{n\geqslant 0}$ 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 :

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

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

Comme $Y_n = 0$ si $n \geqslant T$, on obtient $\{T < \infty\} \subset \{Y_\infty = 0\}$ p.s.. D'après le théorème 5.6, 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 $m \leqslant 1$ alors $\sup_{n\geqslant 0} \frac{Z_n}{m^n}$ n'est pas intégrable, car sinon, par convergence dominée, on aurait $0 = \mathbb{E}(Y_T) = \mathbb{E}(\lim_{n\to\infty} Y_n) = \lim_{n\to\infty} \mathbb{E}(Y_n) = 1$.

5.2 Étude des trois cas possibles

Cas sous-critique

Si $m < 1$ alors le théorème 5.6 donne $\mathbb{P}(T < \infty) = 1$ et $Z_n = Z_{n\wedge T} \to Z_T = 0$ p.s. tandis que le théorème 5.2 donne $\mathbb{E}(Z_n) = m^n \to 0$ et $\mathbb{V}(Z_n) \to 0$ (si $\sigma^2 < \infty$).

Recueil de modèles stochastiques – Page 52/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Théorème 5.9 (Cas sous-critique (théorème de Yaglom)). *Si $m < 1$ alors $\mathcal{L}(Z_n \,|\, Z_n > 0)$ converge lorsque $n \to \infty$ 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 $(h_n(1))_{n\geqslant 1}$ est constante et égale à 1 et converge donc vers $h(1) := 1$. Pour tout $s \in [0,1[$, on a

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

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 $g_n(s) \to 1$ et $g_n(0) \to 1$ quand $n \to \infty$ car 1 est l'unique point fixe de g car $m < 1$ (théorème 5.6). Comme les fonctions $s \mapsto r(s) := (1 - g(s))/(1 - s)$ et g_n sont croissantes sur $[0,1[$, on a

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

Par conséquent, la suite positive $(1 - h_n(s))_{n\geqslant 1}$ 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

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

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 $(\mu_n)_{n\geqslant 1}$ converge (étroitement) vers μ. Comme $(h_n)_{n\geqslant 1}$ 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. □

Cas critique

Si $m = 1$ alors le théorème 5.6 donne $\mathbb{P}(T < \infty) = 1$ et $Z_n = Z_{n\wedge T} \to Z_T = 0$ p.s. tandis que le théorème 5.2 donne $\mathbb{E}(Z_n) = 1$ et $\mathbb{V}(Z_n) \to \infty$ (si $\sigma^2 < \infty$). La «largeur» de l'arbre $\sup_{n\geqslant 0} Z_n$ n'est pas intégrable car cela entraînerait par convergence dominée que $1 = \lim_{n\to\infty} \mathbb{E}(Z_n) = \mathbb{E}(\lim_{n\to\infty} Z_n) = 0$ 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 $(Z_n)_{n\geqslant 0} = (Y_n)_{n\geqslant 0}$ et on retrouve au passage que $Y_\infty = 0$ p.s.

Théorème 5.10 (Cas critique). *Si $m = 1$ et $\sigma^2 < \infty$ alors*

1. $\lim_{n\to\infty} n\mathbb{P}(Z_n > 0) = 2\sigma^{-2}$
2. $\lim_{n\to\infty} \frac{1}{n}\mathbb{E}(Z_n \,|\, Z_n > 0) = \frac{1}{2}\sigma^2$
3. $\mathcal{L}(n^{-1}Z_n \,|\, Z_n > 0)$ *tend quand $n \to \infty$ 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 $(Z_n)_{n\geqslant 1}$ 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 16 pour la preuve du théorème de Wigner.

Recueil de modèles stochastiques – Page 53/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

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 $\lim_{s\to 1} \alpha(s) = 0$. 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 $(g_k)_{k\geqslant 1}$ converge uniformément vers 1, il vient

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

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

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

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

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

□

Cas sur-critique

Si $m > 1$ alors le théorème 5.6 donne $0 < \mathbb{P}(T < \infty) < 1$, sur $\{T < \infty\}$ on a $Z_n \to 0$ tandis que sur $\{T = \infty\}$ on a $Z_n \to \infty$ (théorème 5.5), et d'autre part, le théorème 5.2 donne $\mathbb{E}(Z_n) = m^n \to \infty$ et $\mathbb{V}(Z_n) \to \infty$. Enfin, si $m > 1$, on sait que sur $\{T < \infty\}$ on a $Z_n \to 0$ tandis que sur $\{Y_\infty > 0\}$ on a $Z_n \sim_{n\to\infty} m^nY_\infty$. 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 5.11 (Cas sur-critique – Dichotomie)**.** *On a $\{Y_\infty = 0\} = \{T < \infty\}$ p.s. et donc p.s. soit $Y_\infty = 0$ et $\lim_{n\to\infty} Z_n = 0$ soit $Y_\infty > 0$ et $Z_n \sim_{n\to\infty} m^nY_\infty \nearrow +\infty$.*

Démonstration. Comme $\mathbb{E}(Y_\infty) = 1$ et $Y_\infty \geqslant 0$ on en déduit que $\mathbb{P}(Y_\infty > 0) > 0$. Sur $\{Y_\infty > 0\}$ on a $Z_n \sim_{n\to\infty} m^nY_\infty \to \infty$ 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

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

Recueil de modèles stochastiques – Page 54/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

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

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

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 5.6. □

Le lemme de Fatou donne $\mathbb{E}(Y_\infty) \leqslant \liminf_{n\to\infty} \mathbb{E}(Y_n) = 1$, ce qui assure que Y_∞ est intégrable, mais la convergence de $(Y_n)_{n\geqslant 0}$ 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 5.12 (Cas sur-critique – Loi de Y_∞)**.** *Si $m > 1$ et $\sigma^2 < \infty$ alors la martingale $(Y_n)_{n\geqslant 0}$ converge p.s. et dans L^2 vers une v.a.r. $Y_\infty \geqslant 0$ de moyenne $\mathbb{E}(Y_\infty) = 1$ et de variance $\mathbb{V}(Y_\infty) = \sigma^2/(m^2 - m)$. 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 :*

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

Démonstration. La martingale $(Y_n)_{n\geqslant 0}$ est bornée dans L^2 car par le théorème 5.2 :

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

qui converge car $m > 1$. Par conséquent, $(Y_n)_{n\geqslant 0}$ 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 5.2 on obtient, pour tout $n, k \geqslant 0$,

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

Comme $m > 1$, ceci montre que $(Y_n)_{n\geqslant 0}$ 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 $k \to \infty$. Par convergence monotone, on obtient $\mathbb{E}(\sum_{n=0}^\infty (Y_n - Y_\infty)^2) < \infty$ et donc $(Y_n)_{n\geqslant 0}$ 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 $n \geqslant 0$

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

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

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

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

Du théorème 5.2 on déduit que la population totale à l'instant n a pour moyenne

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

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

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

Recueil de modèles stochastiques – Page 55/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

d'où, par convergence monotone,

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

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

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

où les v.a. $(N^{(i)})_{i\geqslant 1}$ 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 5.13 (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*

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

Exemple 5.14 (Reproduction géométrique)**.** *Si $P = \mathrm{Geo}_{\mathbb{N}}(p) = \sum_{n\geqslant 0} q^n p \delta_n$ avec $p \geqslant 1/2$, alors $f(s)$ est solution de l'équation $qf(s)^2 - f(s) + sp = 0$. On a donc*

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

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 5.15 (Loi de la taille de l'arbre et marche aléatoire)**.** *On a l'égalité en loi*

$$N \overset{\text{loi}}{=} T_{-1},$$

où $T_{-1} := \inf\{n \geqslant 1 : S_n = -1\}$ est le temps d'atteinte de -1 d'une marche aléatoire $(S_n)_{n\geqslant 0}$ sur $\mathbb{Z}$ issue de $S_0 = 0$ et d'incréments $(U_i)_{i\geqslant 1}$ tels que $1 + U_i \sim P$. De plus, pour tout $n \geqslant 1$,

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

La loi de N peut bien sûr s'obtenir par un calcul direct, sans faire appel à l'identité en loi. Cependant, cette si belle identité en loi permet notamment un contrôle de la queue de distribution, ce qui s'avère utile pour l'étude du graphe aléatoire de Erdős-Rényi par exemple (théorème 6.14).

Démonstration. L'idée consiste à associer bijectivement 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 $1 \leqslant i \leqslant n$, et on pose

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

La figure 5.2 donne un exemple d'association. Le mode de numérotation étant fixé, l'arbre se reconstruit aisément à partir de la trajectoire de S. La suite $(S_i)_{0\leqslant i\leqslant n}$ a la loi d'une marche aléatoire d'accroissement $(U_i)_{i\geqslant 1}$ 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 l'égalité en loi $N \overset{\text{loi}}{=} T_{-1}$. Enfin, la loi de T_{-1} est fournie par le lemme 5.16.

Recueil de modèles stochastiques – Page 56/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

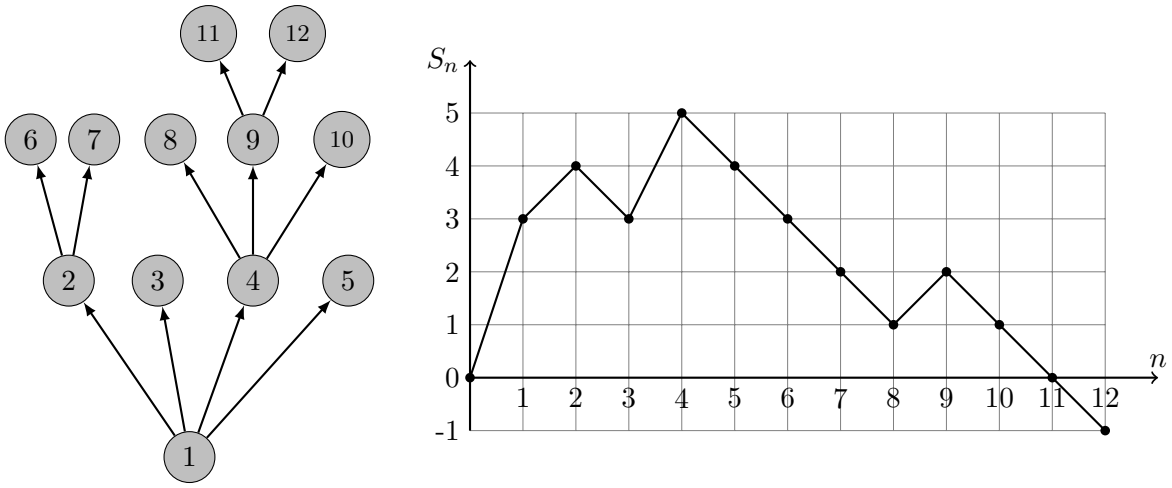


FIGURE 5.2 – Arbre et marche aléatoire associée.

Lemme 5.16 (Principe de rotation)**.** *Soit* $(S_n)_{n\geqslant 0}$ *une marche aléatoire issue de 0 d'incréments* $(U_n)_{n\geqslant 1}$ *i.i.d. à valeurs entières, et* $T_{-1} := \inf\{n \geqslant 1 : S_n = -1\}$*. Pour tout* $n \geqslant 1$*,*

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

Preuve du lemme 5.16. Soit $n \geqslant 1$ et $U_1, \ldots, U_n$ 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 $l = 1, 2, \ldots, n$, 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 5.3 fournit un exemple. Remarquons que pour tout l, $S_n^{(l)} = S_n = -1$. Cependant, il

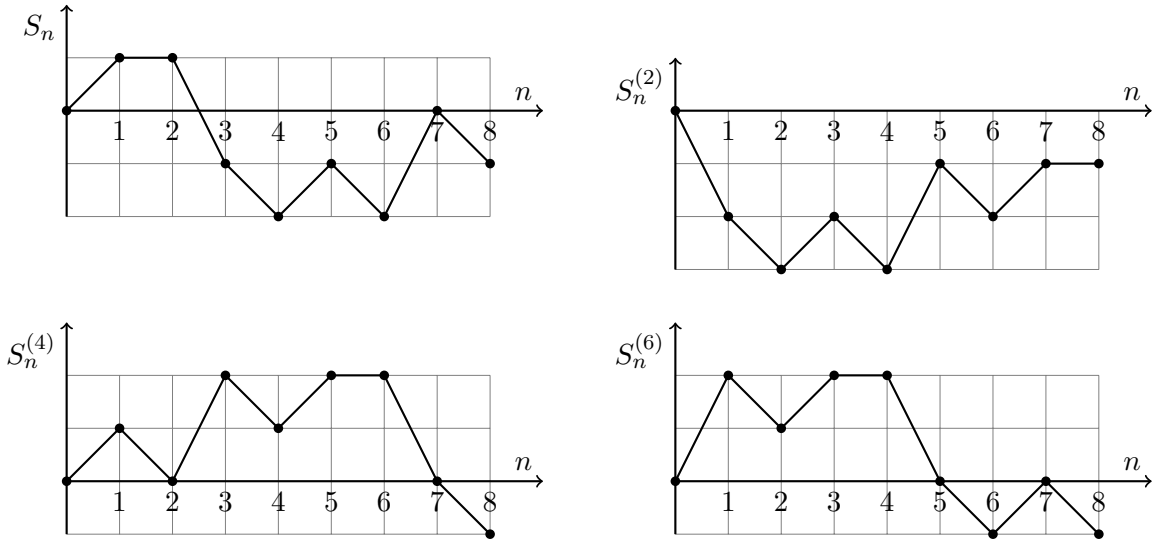


FIGURE 5.3 – Une marche parmi n reste positive avant d'atteindre -1.

existe une et une seule trajectoire $(S_i^{(l)})_{0\leqslant i\leqslant n}$ 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. □

□

Recueil de modèles stochastiques – Page 57/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Exemple 5.17 (Reproduction poissonnienne)**.** *Si la loi de reproduction P est la loi de Poisson* $\mathrm{Poi}(\lambda)$ *avec $\lambda < 1$, alors la loi de U_1 est la loi* $\mathrm{Poi}(\lambda) * \delta_{-1}$ *et, pour tout $n \geqslant 1$,*

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

puisque S_n suit la loi $\mathrm{Poi}(n\lambda) * \delta_{-n}$*. D'autre part, la série génératrice f de N est solution de l'équation fonctionnelle $f(s) = se^{\lambda(f(s)-1)}$ qui n'est pas facile à résoudre…*

5.4 Immigration

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

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

pour tout $n \geqslant 0$, où $(X_{n,k})_{n\geqslant 1, k\geqslant 1}$ sont i.i.d. de loi P sur $\mathbb{N}$, $(I_n)_{n\geqslant 1}$ i.i.d. de loi P_+ sur $\mathbb{N}$, toutes ces variables formant avec Z_0 une famille indépendante. On suppose que P et P_+ ont pour moyenne m et m_+. On note σ^2 et σ_+^2 leur variance lorsqu'elle existe. On suppose que $Z_0 = 1$. Soit $\mathcal{F}_n$ la tribu engendrée par $Z_0, (X_{i,j})_{1\leqslant i\leqslant n, j\geqslant 1}, I_1, \ldots, I_n$. Pour tout $n \geqslant 0$,

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

d'où on tire (récurrence linéaire) que pour tout $n \geqslant 1$,

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

Pour la variance, on a

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

ce qui donne

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

Si $m < 1$ alors $\mathbb{E}(Z_n) \to m_+/(1 - m)$ et $\mathbb{V}(Z_n) \to (\sigma^2 m_+/(1 - m) + \sigma_+^2)/(1 - m^2)$. Notons que $Z_n = Z_{n,0} + \sum_{k=1}^n Z_{n,k}$ où $Z_{n,0}$ est le nombre de descendants à l'instant n de l'individu présent au temps 0 (processus de Galton-Watson sans immigration !), tandis que $Z_{n,k}$ avec $1 \leqslant k \leqslant n$ est le nombre de descendants à l'instant n des individus immigrés à l'instant k. Toutes ces variables sont indépendantes. On observe que $Z_{n,k}$ a la même loi que la somme de I_k copies indépendantes de $Z_{n-k,0}$. Bien qu'il soit possible d'utiliser ces observations pour étudier le cas $m \leqslant 1$, nous nous contentons par souci de simplicité du cas $m > 1$.

Théorème 5.18 (Cas sur-critique avec immigration)**.** *Si $m > 1$ alors $(Y_n)_{n\geqslant 0} = (Z_n/m^n)_{n\geqslant 0}$ converge p.s. vers une v.a.r. $Y_\infty \geqslant 0$. Si de plus $\sigma < \infty$ et $\sigma_+ < \infty$ alors la convergence a également lieu dans L^2 et*[2] $\mathbb{E}(Y_\infty) = 1 + m_+/(m - 1)$.

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

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

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

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

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

Recueil de modèles stochastiques – Page 58/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Comme $m > 1$ et $Y_n \geqslant 0$ 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 $(Y_n)_{n\geqslant 0}$ converge p.s. vers une v.a.r. $Y_\infty \geqslant 0$ intégrable (la convergence a lieu dans L^1 si la suite est uniformément intégrable). Supposons que $\sigma < \infty$ et $\sigma_+ < \infty$ et montrons que la convergence a lieu dans L^2. On a

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

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

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

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

5.5 Notes et commentaires

C'est vers 1875 que Francis Galton et Henri 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 immigration se trouve dans les livres [Rug01, AN04]. Le théorème 5.9 de Akiva Yaglom constitue un bon point de départ pour l'étude plus générale des distributions quasi-stationnaires des processus de populations, présentée dans le survol de Sylvie Méléard et Denis Villemonais [MV12].

Le lemme 5.16 est obtenu dans un article de Aryeh Dvoretzky et Theodore Motzkin [DM47]. L'expression pour la taille d'un arbre de Galton-Watson est établie dans un article de Meyer Dwass [Dwa69] avant que le lien entre les deux situations ne 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 scrutin de Luigi Addario-Berry et Bruce Reed [ABR08]. Au delà du théorème 5.15, la combinatoire foisonne de bijections entre collections d'objets de natures différentes.

Recueil de modèles stochastiques – Page 59/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Chapitre 6

Percolation

Recueil de modèles stochastiques – Page 61/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Mots-clés. Percolation ; loi du zéro-un ; phénomène de seuil ; couplage ; graphe aléatoire ; arbre aléatoire ; arbre de Bethe ; réseau euclidien ; graphe complet ; graphe aléatoire de Erdős-Rényi.

Dans un percolateur, l'eau ne traverse le marc de café que si celui-ci est suffisamment mouillé. La modélisation du marc de café par une grille dont les arêtes sont mouillées aléatoirement ouvre la voie à une étude mathématique du phénomène physique de la percolation. Nous commençons par introduire la notion de percolation sur un graphe général, puis nous abordons le cas plutôt simple du graphe de Bethe (arbre), puis le cas plus physique du graphe carré. Nous terminons par quelques mots sur un modèle de graphe complet.

6.1 Percolation dans un graphe

Un graphe (non-orienté) est un couple $G = (V, E)$ où V est un ensemble non-vide fini ou infini dénombrable, et où $E \subset \mathcal{P}_2(V)$, où $\mathcal{P}_2(V)$ est l'ensemble des parties de V à deux éléments. Les éléments de V et E sont respectivement appelés[1] *sommets* et *arêtes* du graphe G. Si l'arête $\{x, y\}$ appartient à E, on dit que x et y sont voisins et on note $x \sim y$. L'exemple le plus immédiat est celui du graphe euclidien $\mathcal{E}_d = (V, E)$ pour lequel

$$V = \mathbb{Z}^d \quad \text{et} \quad E = \{\{x, y\} : x, y \in \mathbb{Z}^d, |x - y|_1 := |x_1 - y_1| + \cdots + |x_d - y_d| = 1\}.$$

Un *chemin* de longueur n dans le graphe (V, E) reliant un sommet x à un sommet y est une suite finie $x = x_0 \sim \cdots \sim x_n = y$. S'il existe un chemin reliant x à y, on dit que x et y *communiquent* (ou sont connectés) et on note $x \leftrightarrow y$. On a jamais $x \sim x$, mais on convient que $x \leftrightarrow x$ pour tout $x \in V$. La relation binaire $\leftrightarrow$ sur V est une relation d'équivalence et la classe d'équivalence (ou composante connexe) du point x est notée $C(x)$. Un chemin est dit *auto-évitant* si tous les sommets qui le composent sont distincts. Si $x \leftrightarrow y$ et $x \neq y$ alors il existe un chemin auto-évitant de x à y, construit à partir d'un chemin quelconque en effaçant les boucles. Un chemin reliant un sommet à lui-même est appelée *cycle*. On dit que (V, E) est un *arbre* lorsqu'il est connexe et n'a pas de cycles, ou de manière équivalente lorsque deux sommets distincts sont toujours reliés par un unique chemin auto-évitant.

Dans la suite V est, sauf mention du contraire, infini dénombrable. Soit $F \subset E$. On dit qu'il y a *percolation* en $x \in V$ (pour F) si la classe $C(x)$ de x dans le graphe (V, F) est de taille infinie i.e. $|C(x)| = \infty$, et on note $x \leftrightarrow \infty$. On identifie l'ensemble des parties de E à l'ensemble

$$\Omega := \{0, 1\}^E.$$

1. On dit «vertices» et «edges» en anglais, d'où la notation.

Dans la suite, on choisit F de manière aléatoire dans Ω, et ainsi (V, F) est un *graphe aléatoire*. On décide de fabriquer F à partir de E en tirant à pile ou face pour savoir si on conserve ou si on efface chaque arête de E. Cela revient à introduire la loi de probabilité

$$\mathbb{P}_p = \bigotimes_{e \in E} \mathrm{Ber}(p)$$

sur Ω muni de sa tribu cylindrique $\mathcal{F}$, où $p \in [0,1]$ est un paramètre fixé. On note $\mathbb{E}_p$ l'espérance sous $\mathbb{P}_p$. Pour étudier l'influence du paramètre p sur le phénomène de percolation sous $\mathbb{P}_p$, on distingue un sommet particulier $\varnothing \in V$, que nous appelons *racine* – on dit qu'on enracine (V, E) en $\varnothing$ – et on définit la *probabilité de percolation* sous $\mathbb{P}_p$ pour $\varnothing$ par

$$\theta(p) := \mathbb{P}_p(\varnothing \leftrightarrow \infty) = \mathbb{P}_p(|C(\varnothing)| = \infty).$$

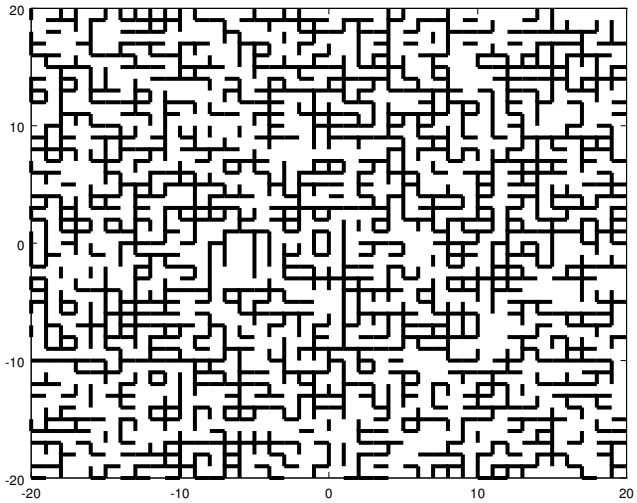


Figure 6.1 – Réalisation du graphe aléatoire (V, F) du graphe euclidien $\mathcal{E}_2$ sous $\mathbb{P}_{1/2}$.

Lemme 6.1 (Monotonie)**.** *On a $\theta(0) = 0$, $\theta(1) = 1$, et $p \in [0,1] \mapsto \theta(p)$ est croissante.*

La monotonie de θ conduit à introduire la notion de *probabilité critique* $p_c \in [0,1]$ par

$$p_c = \sup\{p \in [0,1] \ : \ \theta(p) = 0\}.$$

Démonstration. Si $p = 0$ alors F est vide et donc $\theta(0) = 0$ tandis que si $p = 1$ alors $F = E$ et $\theta(1) = 1$. Pour établir la monotonie en p, on procède en couplant toute les lois $(\mathbb{P}_p)_{p \in [0,1]}$. Soient $(U_e)_{e \in E}$ des v.a.r. i.i.d. de loi uniforme sur $[0,1]$ (nécesairement définies sur un même espace de probabilité). Pour tout $p \in [0,1]$ on réalise $\mathbb{P}_p$ comme la loi de $F_p := (\mathbb{1}_{\{U_e \leqslant p\}})_{e \in E}$. À présent Si $p \leqslant p'$, alors $F_p \subset F_{p'}$. Toute arête de (V, F_p) est aussi une arête de $(V, F_{p'})$. Si $x \leftrightarrow y$ pour (V, F_p) alors c'est aussi le cas pour $(V, F_{p'})$. Par conséquent on obtient

$$\theta(p) = \mathbb{P}_p(\varnothing \leftrightarrow \infty) \leqslant \mathbb{P}_{p'}(\varnothing \leftrightarrow \infty) = \theta(p').$$

□

Remarque 6.2 (Couplage et monotonie)**.** *Munissons Ω de l'ordre partiel issu de l'inclusion dans $\mathcal{P}_2(V)$, ce qui signifie que $F \leqslant F'$ ssi toute arête de (V, F) est aussi une arête de (V, F'). Pour toute $f : \Omega \to \mathbb{R}$ bornée, l'argument de couplage utilisé dans la preuve du lemme 6.1 assure que si f est croissante alors $p \in [0,1] \mapsto \mathbb{E}_p(f)$ l'est aussi.*

Recueil de modèles stochastiques – Page 62/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

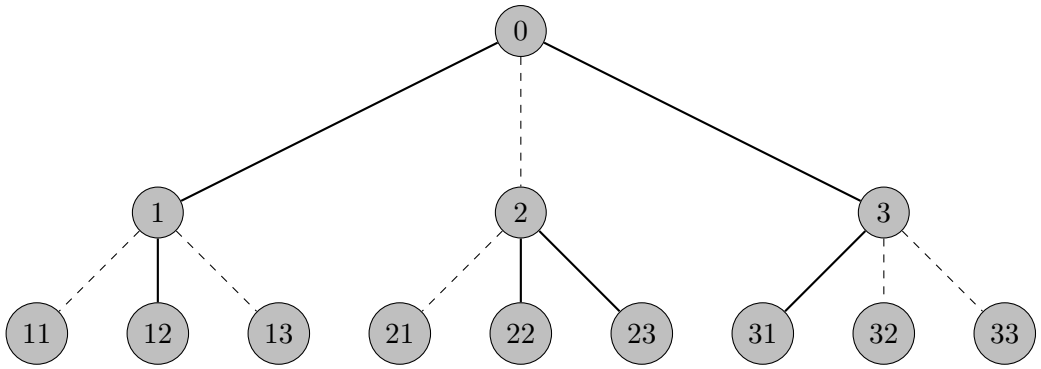


FIGURE 6.2 – Les premières branches de $\mathcal{B}_4$ avec des arêtes supprimées par $\mathbb{P}_p$ (en pointillé).

6.2 Graphe de Bethe (arbre régulier)

On considère dans cette section le cas où (V, E) est le *graphe de Bethe* $\mathcal{B}_r$ de degré $r \geqslant 2$, c'est-à-dire l'arbre infini dont chaque sommet a r voisins sauf un qui n'en a que $r-1$, noté 0 et appelé *racine* de l'arbre (voir la figure 6.2). Pour tout $p \in [0,1]$, sous $\mathbb{P}_p$, le graphe aléatoire (V, F) est en quelque sorte une forêt d'arbres. La probabilité de percolation et la probabilité critique pour le sommet racine $\varnothing = 0$ dans $\mathcal{B}_r$ sous $\mathbb{P}_p$ sont notées

$$\theta_r(p) := \mathbb{P}_p(0 \leftrightarrow \infty) = \mathbb{P}_p(|C(0)| = \infty) \quad \text{et} \quad p_c(r) := \sup\{p \in [0,1] \,:\, \theta_r(p) = 0\}.$$

Théorème 6.3 (Phénomène de seuil pour la percolation sur le graphe de Bethe)**.** *Pour tout $r \geqslant 2$ et tout $p \in [0,1]$ la quantité $\theta_r(p)$ est la plus grande racine en $\theta \in [0,1]$ de l'équation $1-\theta = (1-p\theta)^{r-1}$, tandis que $p_c(r) = 1/(r-1)$. De plus, la fonction θ_r est continue, nulle sur $[0, p_c(r)]$, strictement croissante sur $[p_c(r), 1]$, et pour $r \geqslant 3$ on a*

$$\lim_{p \to p_c(r)^+} \frac{\theta_r(p)}{p - p_c(r)} = \frac{2}{p_c(r)(1-p_c(r))} = \frac{2(r-1)^2}{r-2}.$$

Si $r \geqslant 3$ alors $0 < p_c(r) < 1$ et on a un *phénomène de seuil* pour la percolation sur $\mathcal{B}_r$.

Démonstration. Le graphe de Bethe $\mathcal{B}_r$ coïncide avec l'*arbre de Galton-Watson* de loi de reproduction δ_{r-1}. Sous $\mathbb{P}_p$ on obtient un arbre aléatoire dans lequel chaque sommet de l'arbre est relié à chacun de ses $r-1$ enfants avec une probabilité p. Le nombre de ces connexions est distribué selon la loi de reproduction $\mathrm{Bin}(r-1, p)$ de moyenne $m = (r-1)p$. D'après le chapitre 5, cet arbre est sous-critique si $m < 1$, critique si $m = 1$, et sur-critique si $m > 1$. D'après le théorème 5.6 la probabilité d'extinction de la population est la plus petite racine s_p de l'équation $g_p(s) = s$ où g_p est la fonction génératrice de la loi de reproduction : $g_p(s) = (ps + 1 - p)^{r-1}$ pour $s \in [0,1]$. L'extinction de la population est synonyme de la finitude de la composante connexe de la racine. On a donc $\theta_r(p) = 1 - s_p$ et $\theta_r(p)$ est ainsi la plus grande racine dans $[0,1]$ de l'équation $1-\theta = (1-p\theta)^{r-1}$.

Soit h_p la fonction définie sur $[0,1]$ par $h_p(\theta) = (1-p\theta)^{r-1} - 1 + \theta$. Elle est nulle en 0, strictement positive en 1, strictement convexe si $r \geqslant 3$ et affine si $r = 2$. De plus $h_p'(0)$ a le signe de $1 - p(r-1)$. Ainsi, pour $p(r-1) \leqslant 1$, h_p ne s'annule qu'en 0 et $\theta_r(p) = 0$. En revanche, h_p s'annule exactement une fois sur $]0,1[$ dès que $p(r-1) > 0$. En conclusion, la probabilité $\theta_r(p)$ est nulle pour $p \leqslant (r-1)^{-1}$ et strictement positive pour $p > (r-1)^{-1}$. La fonction $p \mapsto h_p(\theta)$ est strictement décroissante et régulière. Ceci assure que la fonction θ_r est continue sur $[0,1]$ et strictement croissante sur $[p_c(r), 1]$.

On a enfin $1 - \theta = 1 - (r-1)p\theta + \frac{1}{2}(r-1)(r-2)p^2\theta^2 + o(\theta^2)$. Puisque $\theta_r(p_c(r)) = 0$ et $\theta_r(p)$ est strictement positive pour $p > p_c = (r-1)^{-1}$, on obtient après simplification

$$\theta_r(p) = 2\frac{(r-1)p - 1}{(r-1)(r-2)p^2} + o(\theta_r(p)) = \frac{2}{(r-2)p^2}(p - p_c) + o(\theta_r(p)).$$

Recueil de modèles stochastiques – Page 63/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

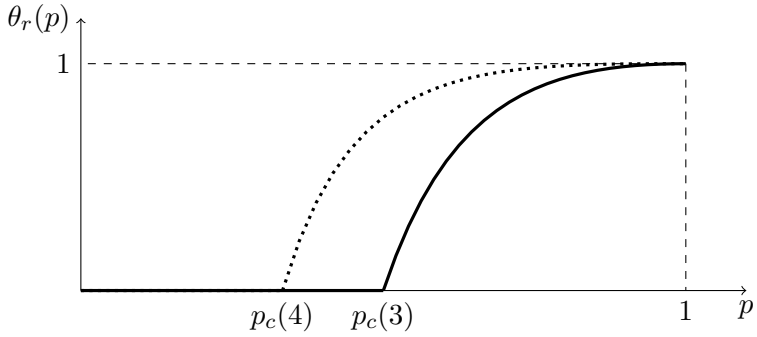


Figure 6.3 – Percolation pour $\mathcal{B}_r$ avec $r = 3$ (trait plein) et $r = 4$ (trait en pointillé).

On conclut en remarquant que $(r-2)p_c = 1 - p_c$. □

Remarque 6.4 (Caractère explicite des cas $r \leqslant 4$)**.** *Pour $r = 2$, le graphe de Berthe $\mathcal{B}_2$ est identifiable au graphe de $\mathbb{N}$ et sous $\mathbb{P}_p$ la v.a. $|C(0)|$ suit la loi géométrique de paramètre $1-p$, et $\theta_2 = \mathbb{1}_1$ et $p_c(2) = 1$. Pour $r = 3$ et $r = 4$ l'équation vérifiée par $\theta_r(p)$ fournie par le théorème 6.3 est résoluble explicitement (voir la figure 6.2) :*

$$\theta_3(p) = \begin{cases} 0 & \text{si } p \leqslant 1/2, \\ \dfrac{2p-1}{p^2} & \text{si } p > 1/2, \end{cases} \quad \text{et} \quad \theta_4(p) = \begin{cases} 0 & \text{si } p \leqslant 1/3, \\ \dfrac{3p^2 - \sqrt{4p^3 - 3p^4}}{2p^3} & \text{si } p > 1/3. \end{cases}$$

Remarque 6.5 (Taille moyenne de la composante de la racine)**.** *Reprenons la preuve du théorème 6.3, et supposons que $p \leqslant p_c(r)$. Alors $|C(0)| < \infty$ p.s. car la lignée de la racine s'éteint p.s. . De plus d'après la section 5.3 la fonction génératrice h de $|C(0)|$ est solution de l'équation $h(s) = s(ph(s) + 1 - p)^{r-1}$. En particulier on obtient*

$$\mathbb{E}_p(|C(0)|) = \begin{cases} \dfrac{1}{1 - p(r-1)} & \text{si } p < p_c(r), \\ \infty & \text{si } p = p_c(r). \end{cases}$$

D'autre part, si $p \in]p_c, 1[$ alors il existe une infinité de classes de taille infinie. Cette situation est très différente de celle du graphe euclidien étudié dans la suite.

6.3 Graphe euclidien (grille)

On considère dans cette section le cas où (V, E) est le graphe euclidien $\mathcal{E}_d$, avec $V = \mathbb{Z}^d$, $d \geqslant 1$, et $E = \left\{\{x, y\} \in \mathcal{P}_2(\mathbb{Z}^d) \ : \ |x - y|_1 = \sum_{i=1}^d |x_i - y_i| = 1\right\}$. La probabilité de percolation et la probabilité critique pour le sommet racine $\varnothing = 0$ dans $\mathcal{E}_d$ sous $\mathbb{P}_p$ sont notées

$$\theta_d(p) := \mathbb{P}_p(0 \leftrightarrow \infty) = \mathbb{P}_p(|C(0)| = \infty) \quad \text{et} \quad p_c(d) := \sup\{p \in [0,1] \ : \ \theta_d(p) = 0\}.$$

Pour $d = 1$, le graphe $\mathcal{E}_1$ est identifiable au graphe $\mathbb{Z}$, pour tout $x \in \mathbb{Z}$, sous $\mathbb{P}_p$, la v.a. $|C(x)|$ a la loi de $G_- + G_+ - 1$ où G_- et G_+ sont des v.a. indépendantes de loi géométrique de paramètre $1-p$. Ainsi pour $d = 1$ on obtient $\theta_1 = \mathbb{1}_1$ et $p_c(1) = 1$.

Dès que $d > 1$, la probabilité de l'évènement $\{x \leftrightarrow y\}$ pour $\mathcal{E}_d$ sous $\mathbb{P}_p$ n'est pas explicite même pour $x \sim y$ car toutes les arêtes du graphe sont susceptibles de jouer un rôle, contrairement au cas des graphes de Bethe. Du point de vue de la connectivité, la grille est beaucoup plus complexe qu'un arbre car la présence de cycles permet de faire des détours.

Recueil de modèles stochastiques – Page 64/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Lemme 6.6 (Probabilité d'une petite classe)**.** *Pour tous $d \geqslant 1$, $x \in \mathbb{Z}^d$, $p \in [0,1]$,*

$$\mathbb{P}_p(C(x) = \{x\}) = (1-p)^{2d} \quad \textit{et} \quad \mathbb{P}_p(|C(x)| = 2) = p(1-p)^{2(2d-1)}.$$

Démonstration. On a $|C(x)| = 1$ si et ssi x n'est relié à aucun de ses $2d$ voisins, tandis que $|C(x)| = 2$ ssi x n'est relié qu'à un seul de ses $2d$ voisins, lui-même n'étant relié qu'à x. □

Théorème 6.7 (Phénomène de seuil pour percolation sur graphe euclidien)**.** *Pour tous $2 \leqslant d \leqslant d'$ on a $\theta_d \leqslant \theta_{d'}$ et $p_c(d) \leqslant p_c(d')$. De plus on a $0 < p_c(d) < 1$ et plus précisément*

$$\frac{1}{2d-1} \leqslant \frac{1}{\kappa(d)} \leqslant p_c(d) \leqslant 1 - \frac{1}{\kappa(2)} \leqslant \frac{2}{3},$$

où la constante $\kappa(d)$ est définie dans le lemme 6.8.

La constante $\kappa(d)$ est appelée *constante de connectivité* de $\mathcal{E}_d$. Pour plus de clarté, la démonstration du théorème 6.7 est découpée en quatre lemmes.

Lemme 6.8 (Constante de connectivité)**.** *Pour tous $d \geqslant 2$ et $n \geqslant 1$ soit $\kappa_n(d)$ le cardinal de l'ensemble $\mathcal{C}_n^d$ des chemins auto-évitants dans $\mathcal{E}_d$ de longueur n issus de l'origine 0. On a*

$$d^n \leqslant \kappa_n(d) \leqslant 2d(2d-1)^{n-1}.$$

De plus, la suite $(\kappa_n(d)^{1/n})_{n\geqslant 1}$ converge et sa limite $\kappa(d)$ vérifie

$$d \leqslant \kappa(d) \leqslant 2d-1.$$

Preuve du lemme 6.8. La première arête d'un chemin auto-évitant issu de 0 peut être choisie parmi les $2d$ arêtes issues de 0. Pour les pas suivants, les retours en arrière sont proscrits, d'où $k_n(d) \leqslant 2d(2d-1)^{n-1}$. La borne inférieure $k_n(d) \geqslant d^n$ est obtenue en remarquant que tout chemin issu de 0 dont les coordonnées des sommets successifs sont croissantes est auto-évitant. La suite $(\kappa_n(d))_{n\geqslant 1}$ est sous-multiplicative : pour tous $n, m \geqslant 1$,

$$\kappa_{n+m}(d) \leqslant \kappa_n(d)\kappa_m(d),$$

car si $(0, x_1, \ldots, x_{n+m}) \in \mathcal{C}_{n+m}^d$ alors $(0, x_1, \ldots, x_n) \in \mathcal{C}_n^d$ et $(0, x_{n+1}-x_n, \ldots, x_{n+m}-x_n) \in \mathcal{C}_m^d$. Ceci implique que la suite $(\log \kappa_n(d))_{n\geqslant 1}$ est sous-additive, et le lemme de Fekete donne

$$\lim_{n\to\infty} \frac{\log \kappa_n(d)}{n} = \lambda \quad \text{où} \quad \lambda := \inf_{n\geqslant 1} \frac{\log \kappa_n(d)}{n} \in [-\infty, \infty[.$$

Ceci donne $\kappa(d) = e^\lambda$, et l'encadrement de $\kappa(d)$ découle de celui de $\kappa_n(d)$. □

Lemme 6.9 (Minoration)**.** *Pour tout $d \geqslant 2$ on a $p_c(d) \geqslant 1/\kappa(d)$.*

Preuve du lemme 6.9. Soit $n \geqslant 1$ et N_n la v.a. égale au nombre d'éléments de $\mathcal{C}_n^d$ inclus dans l'ensemble d'arêtes aléatoire $C(0)$ sous $\mathbb{P}_p$. Si $|C(0)| = \infty$, alors $N_n \geqslant 1$. D'autre part chaque élément de $\mathcal{C}_n^d$ a une probabilité p^n d'avoir toutes ses arêtes dans $C(0)$. Ainsi

$$\theta_d(p) \leqslant \mathbb{P}_p(N_n \geqslant 1) = \mathbb{E}_p(\mathbb{1}_{N_n\geqslant 1}) \leqslant \mathbb{E}_p(N_n) \leqslant p^n \kappa_n(d).$$

À présent le lemme 6.8 donne $\kappa_n(d)^{1/n} = \kappa(d) + o_n(1)$, d'où $\theta_d(p) \leqslant (p\kappa(d) + o_n(1))^n$. Par conséquent, si $p\kappa(d) < 1$ alors on obtient $\theta_d(p) = 0$ quand $n \to \infty$. □

Lemme 6.10 (Décroissance)**.** *Si $2 \leqslant d \leqslant d'$ alors $\theta_d \leqslant \theta_{d'}$ et $p_c(d') \leqslant p_c(d)$.*

Recueil de modèles stochastiques – Page 65/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Preuve du lemme 6.10. L'idée est de procéder par couplage un peu comme dans la preuve du lemme 6.1. Comme le graphe $\mathcal{E}_d$ est identifiable au graphe de $\mathcal{E}_{d'}$ correspondant par exemple aux d premières coordonnées dans $\mathbb{Z}^{d'}$, on peut coupler, pour tout $p \in [0,1]$ fixé, les graphes aléatoires $(\mathbb{Z}^d, F)$ et $(\mathbb{Z}^{d'}, F')$ de $\mathcal{E}_d$ et $\mathcal{E}_{d'}$ sous $\mathbb{P}_p$ de sorte que

$$\theta_d(p) = \mathbb{P}_p(0 \leftrightarrow +\infty \text{ dans } \mathcal{E}_d) \leqslant \mathbb{P}_p(0 \leftrightarrow +\infty \text{ dans } \mathcal{E}_{d'}) = \theta_{d'}(p).$$

Il en découle en particulier que $p_c(d) \leqslant p_c(d')$. □

Lemme 6.11 (Majoration)**.** *Si $d = 2$ alors $p_c(2) \leqslant 1 - 1/\kappa(2)$.*

Preuve du lemme 6.11. Le graphe dual $\mathcal{E}_2^* = (V^*, E^*)$ de $\mathcal{E}_2 = (V, E)$ est le translaté de $\mathcal{E}_2$ par le vecteur $(1/2, 1/2)$ (voir la figure 6.3). Chaque arête e du graphe initial rencontre une et une seule arête e_* du graphe dual. On construit le graphe aléatoire (V^*, F^*) de $\mathcal{E}_2^*$ à partir du graphe aléatoire (V, F) de $\mathcal{E}_2$ sous $\mathbb{P}_p$ en décidant que $e_* \in F^*$ ssi $e \notin F$.

Une boucle auto-évitante (courbe de Jordan !) de longueur n est un chemin $(x_0, x_1, \ldots, x_n)$ tel que le chemin $(x_0, x_1, \ldots, x_{n-1})$ est auto-évitant et $x_n = x_0$. Son complémentaire est formé de deux composantes connexes, l'une bornée (appelée *intérieur*) et l'autre pas.

Soit $\Lambda(m) := [-m, m]^2$. On introduit les évènements suivants :

— $F_m = \{\Lambda(m)$ est à l'intérieur d'une boucle auto-évitante de $\mathcal{E}_2^*$ à arêtes dans $F^*\}$,

— $G_m = \{$les arêtes reliant les sommets de $\Lambda(m)$ sont dans $F\}$.

Il y a percolation en 0 sur l'événement $F_m^c \cap G_m$, et comme F_m et G_m concernent des ensembles d'arêtes disjoints, ils sont indépendants sur $\mathbb{P}_p$. Considérons à présent une boucle γ dont l'intérieur contient $\Lambda(m)$. Sa longueur n est nécessairement supérieure à $4m$. De plus, elle contient nécessairement au moins un sommet de la forme $(1/2+k, 1/2)$ avec $0 \leqslant k \leqslant n$. Notons x_0 le point de cette forme le plus à droite et numérotons les éléments de γ en parcourant la courbe dans le sens trigonométrique de sorte que $\gamma = (x_0, x_1, \ldots, x_{n-1})$. Ce chemin est auto-évitant. Ainsi, il y a au plus $n\kappa_n(2)$ boucles auto-évitantes de longueur n entourant $\Lambda(m)$. Donc si M_n est le nombre de boucles auto-évitantes de longueur n entourant $\Lambda(m)$ et dont toutes les arêtes sont dans F, on a

$$\mathbb{P}_p(F_m) \leqslant \mathbb{P}_p\left(\sum_{n \geqslant 4m} M_n \geqslant 1\right) = \mathbb{E}_p(\mathbb{1}_{\{\sum_{n \geqslant 4m} M_n \geqslant 1\}}) \leqslant \sum_{n \geqslant 4m} \mathbb{E}_p(M_n) \leqslant \sum_{n \geqslant 4m} n\kappa_n(2)(1-p)^n.$$

Puisque $\kappa_n(2)^{1/n} = \kappa(2) + o_n(1)$, la série ci-dessus (membre de droite droite) est convergente dès que $\kappa(2)(1-p) < 1$ et par suite, pour m assez grand, $\mathbb{P}_p(F_m) < 1/2$. Pour un tel m on a

$$\theta_2(p) \geqslant \mathbb{P}_p(F_m^c \cap G_m) = \mathbb{P}_p(F_m^c)\mathbb{P}_p(G_m) \geqslant \frac{1}{2}p^{4m^2} > 0.$$

Nous avons démontré que si $(1-p)\kappa(2) < 1$ alors $\theta_2(p) > 0$, d'où $p_c(2) \leqslant 1 - 1/\kappa(2)$. □

Dans le régime sur-critique $p > p_c(d)$, la classe $C(0)$ de l'origine 0 peut être infinie. Le théorème suivant complète ce résultat.

Théorème 6.12 (Existence de la classe infinie)**.** *Si $p > p_c(d)$, alors $\mathbb{P}_p$-p.s. il existe au moins une classe infinie.*

Démonstration. L'événement A :=«Il existe au moins une classe infinie» est dans la tribu terminale des événements $(A_e)_{e \in E}$ où $A_e := \{e \in F\}$. Or sous $\mathbb{P}_p$ ces événements sont indépendants, et donc $\mathbb{P}_p(A) = 0$ ou $\mathbb{P}_p(A) = 1$ par la loi du 0-1 de Kolmogorov. À présent si $p > p_c(d)$ alors $\mathbb{P}_p(|C(0)| = \infty) > 0$ et $\{|C(0)| = \infty\} \subset A$, d'où $\mathbb{P}_p(A) = 1$. □

Toutefois, la situation est bien différente de celle du graphe de Bethe : la classe infinie est unique ! Cependant, la démonstration dépasse le cadre de ce chapitre. Nous nous contentons d'une démonstration sur $\mathcal{E}_2$ en supposant connu le fait que $\theta_2(1/2) = 0$.

Recueil de modèles stochastiques – Page 66/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

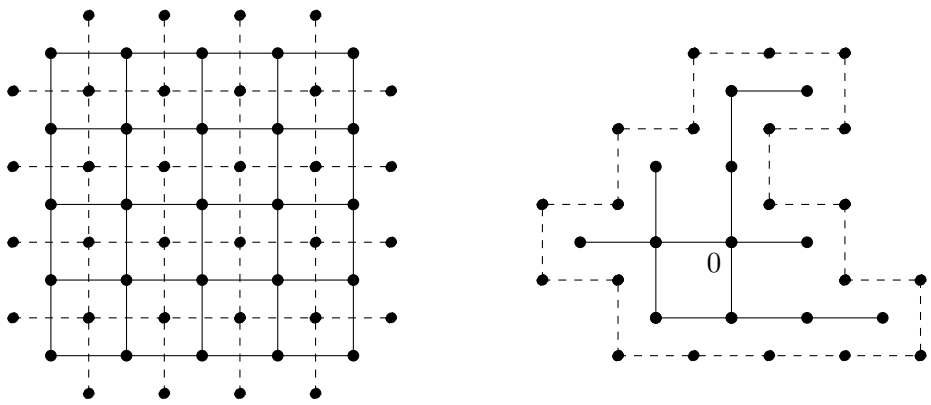


FIGURE 6.4 – À gauche, le graphe euclidien $\mathcal{E}_2$ (trait plein) et son dual $\mathcal{E}_2^*$ (trait pointillé). À droite, une composante finie et sa boucle ceinturante.

Théorème 6.13 (Unicité de la classe infinie)**.** *Supposons que $d = 2$ et que $\theta_2(1/2) = 0$ sur le graphe $\mathcal{E}_2$. Alors, pour tout $p > p_c(2)$, il existe $\mathbb{P}_p$-p.s. une unique classe infinie.*

Démonstration. Reprenons les notations de la preuve du lemme 6.11. Pour $p = 1/2$, la percolation sur $\mathcal{E}_2$ et sur son dual $\mathcal{E}_2^*$ sont de même nature : $\mathbb{P}_p$-p.s. il n'y a pas de composante infinie car $\theta_2(1/2) = 0$. Soit $\Lambda(m) = [-m, m]^2$. Il existe une boucle γ_* auto-évitante dans $\mathcal{E}_2^*$ entourant $\Lambda(m)$ et dont les arêtes sont dans F^* (donc les arêtes duales dans $\mathcal{E}_2$ ne sont pas dans F). Par le même argument, il existe une boucle γ dans $\mathcal{E}_2$ qui contient γ_* et donc $\Lambda(m)$. Par couplage, ceci est encore vrai pour tout $p > 1/2$. Si deux classes d'équivalence infinies intersectent $\Lambda(m)$ alors elles intersectent également γ et sont donc confondues ! Puisque ceci est vrai pour tout m, il ne peut y avoir qu'une classe infinie. □

6.4 Graphe complet et modèle de Erdős-Rényi

On considère dans cette section le cas où (V, E) est le *graphe complet infini* $\mathcal{K}_\infty$, c'est-à-dire que V est infini dénombrable et $E = \mathcal{P}_2(V)$. Ainsi $x \sim y$ pour tous $x \neq y$ dans V. Les sommets jouent tous le même rôle. Comme tous les sommets sont voisins dans $\mathcal{K}_\infty$, il n'y a donc pas de géométrie comme dans $\mathcal{B}_r$ ou dans $\mathcal{E}_d$. Sous $\mathbb{P}_p$ le graphe aléatoire (V, F) est appelé modèle de Erdős-Rényi infini. Pour tous $p \in]0, 1]$ et $x \in V$, le sommet x a un nombre infini de voisins dans (V, F) et $\mathbb{P}_p(|C(x)| = \infty) = 1$, d'où $\theta = \mathbb{1}_{]0,1]}$ et $p_c = 0$.

Pour rendre le modèle plus passionnant, on peut considérer le phénomène de la percolation dans le graphe complet fini $\mathcal{K}_n$ à n sommets $V = \{1, \ldots, n\}$ et faire dépendre de n le paramètre p de $\mathbb{P}_p$. On note $\mathcal{G}(n, p)$ la loi du graphe aléatoire (V, F) sous $\mathbb{P}_p$. Dans ce modèle de *Erdős-Rényi* fini $\mathcal{G}(n, p)$, chaque sommet $x \in V$ possède un nombre aléatoire de voisins, qui suit la loi binomiale $\mathrm{Bin}(n-1, p)$ de moyenne $(n-1)p \sim np$. Lorsque $n \to \infty$ avec $np \to \lambda > 0$ alors $p_n \sim \lambda/n \to 0$ et le nombre de voisins de chaque site converge en loi vers la loi de Poisson $\mathrm{Poi}(\lambda)$.

Théorème 6.14 (Phénomène de seuil et composante connexe géante)**.** *Soit $\lambda > 0$ un paramètre réel fixé, et $\alpha := \lambda - 1 - \log(\lambda) > 0$. Soit $(G_n)_{n \geqslant 1}$ une suite de graphes de Erdős-Rényi définis sur un même espace de probabilité, avec $G_n = (V_n, F_n)$ de loi de Erdős-Rényi $\mathcal{G}(n, p)$ avec $p = \lambda/n$.*

1. *Si $\lambda < 1$ alors pour tout $c > 1/\alpha$, en notant $G_n(v)$ la composante connexe du sommet v,*

$$\lim_{n \to \infty} \mathbb{P}\left(\max_{v \in V_n} |G_n(v)| \geqslant c \log(n) \right) = 0.$$

Recueil de modèles stochastiques – Page 67/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

2. *Si* $\lambda > 1$ *alors*
$$\frac{\max_{v\in V_n}|G(v)|}{n} \xrightarrow[n\to\infty]{p.s.} 1-\rho$$
où $\rho \in]0,1[$ *est la probabilité d'extinction d'un* processus de branchement de Galton-Watson *de loi de reproduction* $\mathrm{Poi}(\lambda)$. *De plus, il existe un réel* $c > 0$ *tel que presque sûrement la seconde composante connexe (en taille) a une taille au plus* $c\log(n)$.

En substance, si $\lambda < 1$ alors la connectivité du graphe aléatoire est si faible qu'il ne comporte pas de composante connexe plus grande que $\mathcal{O}(\log(n))$. Le graphe est émietté. En revanche, si $\lambda > 1$ alors la connectivité du graphe aléatoire est si forte qu'il contient une unique composante connexe contenant une fraction > 0 des sommets – on parle de composante connexe géante – tandis que les autres composantes connexes ont une taille qui n'excède pas $\mathcal{O}(\log(n))$.

Démonstration. Par souci de simplicité, nous nous contentons d'établir la première propriété seulement. Pour ce faire, on explore $G(v)$ en utilisant un *algorithme de parcours en largeur*[2], qui consiste à explorer les voisins de v, qui constituent la première génération, puis leurs voisins qui n'ont pas déjà été visités, qui constituent la seconde génération, etc. Cela donne un arbre fini couvrant[3] $G(v)$, similaire à celui de la figure 5.2. En forçant l'absence de cycles, cette approche fournit également un couplage avec un arbre aléatoire de Galton-Watson de loi de reproduction $\mathrm{Bin}(n-1,p)$, sous-critique car $(n-1)p \leqslant \lambda < 1$, dont la taille totale N vérifie $N \geqslant |G(v)|$ (le fait que l'arbre est plus gros vient de l'absence de cycles). Or le théorème 5.15 affirme que
$$N \overset{\text{loi}}{=} T \quad \text{où} \quad T := \inf\{k \geqslant 1 : S_k = -1\}$$
est le temps d'atteinte de -1 d'une *marche aléatoire* $(S_k)_{k\geqslant 0}$ sur $\mathbb{Z}$ issue de $S_0 = 0$ et d'incréments $(U_i)_{i\geqslant 1}$ i.i.d. tels que $1 + U_i \sim \mathrm{Bin}(n-1,p)$. Fixons $\theta > 0$ quelconque et posons
$$M_k = e^{\theta S_k}\varphi(\theta)^{-k} \quad \text{où} \quad \varphi(\theta) = \mathbb{E}(e^{\theta U_1})$$
est la transformée de Laplace de la loi des incréments de $(S_k)_{k\geqslant 0}$. Or $(M_k)_{k\geqslant 0}$ est une martingale positive[4] pour la filtration naturelle de $(S_k)_{k\geqslant 0}$, et T est un temps d'arrêt fini p.s. On a $\lim_{n\to\infty} M_{T\wedge n} = M_T$ p.s. de sorte que grâce au lemme de Fatou et au théorème d'arrêt de Doob,
$$e^{-\theta}\mathbb{E}(\varphi(\theta)^{-T}) = \mathbb{E}(M_T) = \mathbb{E}(\varliminf_{n\to\infty} M_{T\wedge n}) \leqslant \varliminf_{n\to\infty} \mathbb{E}(M_{T\wedge n}) = \mathbb{E}(M_0) = 1.$$
Choisissons $\theta > 0$ tel que $\varphi(\theta) < 1$, de sorte que pour tout $r > 0$, par l'inégalité de Markov,
$$\mathbb{P}(T \geqslant r) = \mathbb{P}(\varphi(\theta)^{-T} \geqslant \varphi(\theta)^{-r}) \leqslant \varphi(\theta)^r\mathbb{E}(\varphi(\theta)^{-T}) = \varphi(\theta)^r e^{\theta}.$$
L'inégalité de convexité $1 + x \leqslant e^x$ valable pour tout $x \in \mathbb{R}$ donne
$$\varphi(\theta) = e^{-\theta}(1-p+pe^{\theta})^{n-1} = e^{-\theta}\left(1 + \frac{\lambda(e^{\theta}-1)}{n}\right)^{n-1} \leqslant e^{\lambda(e^{\theta}-1)-\theta}$$
(notons que $e^{\lambda(e^{\theta}-1)}$ est la transformée de Laplace de $\mathrm{Poi}(\lambda)$). La fonction $\theta \in \mathbb{R} \mapsto \lambda(e^{\theta}-1)-\theta$ atteint son minimum en $\theta_* = -\log(\lambda) > 0$, et $\varphi(\theta_*) = e^{-\alpha}$ où $\alpha := 1 - \lambda + \log(\lambda) > 0$, d'où
$$\mathbb{P}(T \geqslant r) \leqslant \varphi(\theta_*)^r e^{\theta_*} = e^{-r\alpha+\theta_*} = \lambda^{-1}e^{-\alpha r}.$$

2. «Breadth-First Search» en anglais (BFS).
3. «Spanning tree» en anglais et on parle de «Minimal Spanning Tree» (MST).
4. Le théorème de convergence des sous-martingales bornées dans L^1 implique que toute martingale positive converge p.s. vers une v.a.r. de L^1 (la convergence a lieu dans L^1 si la martingale est uniformément intégrable).

Recueil de modèles stochastiques – Page 68/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Par conséquent, pour tout $c > 0$,

$$\mathbb{P}(|G(v)| \geqslant c\log(n)) \leqslant \mathbb{P}(N \geqslant c\log(n)) = \mathbb{P}(T \geqslant c\log(n)) \leqslant \lambda^{-1}n^{-\alpha c}.$$

À présent, comme les variables aléatoires $(G(v))_{v\in V}$ sont identiquement distribuées, on a

$$\mathbb{P}\left(\max_{v\in V}|G(v)| \geqslant c\log(n)\right) \leqslant \sum_{v\in V}\mathbb{P}(|G(v)| \geqslant c\log(n)) = n\mathbb{P}(|G(1)| \geqslant c\log(n)) \leqslant \lambda^{-1}n^{1-\alpha c},$$

d'où le résultat en prenant $c > 1/\alpha$. □

6.5 Notes et commentaires

La géométrie des arbres réguliers (graphe de Bethe) et des graphes complets est plus simple que celle du graphe euclidien (grille), ce qui facile en principe l'étude de propriétés probabilistes. En informatique, en télécommunication, mais aussi en électricité, un *réseau*[5] est un graphe muni d'une marque ou d'un poids sur chaque arrête, représentant une conductance, une résistance, une longueur, un coût, etc, dans notre cas 0 (arrête indisponible) ou 1 (arrête disponible). En ce sens, nos graphes aléatoires sont des réseaux aléatoires[6]. D'autre part, dans le cas euclidien, il se trouve que notre graphe est également un réseau au sens de la géométrie[7], c'est-à-dire un sous-groupe discret de l'espace vectoriel euclidien.

Le phénomène de la percolation est un classique de la physique statistique, qui peut être étudié sur tout graphe aléatoire. Le mécanisme de la percolation possède par ailleurs de nombreuses variantes : percolation orientée (les arêtes sont orientées), percolation de premier passage, de dernier passage, percolation par sommets (ce sont les sommets et non les arêtes qui sont déclarés ouverts ou fermés), etc. Une bonne partie de ce chapitre est inspirée des notes de cours de Thierry Lévy [Lé08]. La théorie mathématique de la percolation, développée depuis le milieu du vingtième siècle par Hammersley et Kesten notamment, constitue depuis lors un domaine très actif des probabilités. Un panorama sur le sujet se trouve dans les livres [Gri99, Gri10] de Geoffrey Grimmett, dont la thèse soutenue en 1974 sous la direction de John Hammersley et Dominic Welsh (ancien élève de Hammersley également) portait déjà sur les champs et graphes aléatoires ! Le théorème 6.14 date des travaux de Paul Erdős et Alfréd Rényi du milieu du vingtième siècle [ER61]. Le théorème 6.14 est inspiré des notes de cours de Charles Bordenave [Bor14a]. Plus généralement, on peut établir le résultat suivant :

Condition sur (n,p)	Comportement p.s. du graphe aléatoire $G(n,p)$ quand $n \to \infty$
$np < (1-\epsilon)\log(n)$	existence de sommets isolés (donc graphe non connexe)
$np > (1+\epsilon)\log(n)$	graphe connexe (donc $np \sim \log(n)$ est un seuil de connectivité)
$np < 1$	pas de composante connexe de taille $> \mathcal{O}(\log(n))$
$np = 1$	plus grande composante connexe de taille $\approx n^{2/3}$
$np \to \lambda > 1$	unique composante connexe géante (fraction > 0 des sommets)
	aucune autre composante connexe n'est $> \mathcal{O}(\log(n))$.

On renvoie à ce sujet à [Bor14a] ainsi qu'au livre de Remco van der Hofstad [vdH14] notamment.

Le lemme 6.8 donnant un encadrement de la constante de connectivité du graphe euclidien est l'une des clés de la démonstration du théorème 6.7. Toutefois, la valeur exacte de cette constante $\kappa(d)$ est inconnue pour tout $d \geqslant 2$. Hugo Duminil-Copin et Stanislav Smirnov ont montré qu'elle vaut $\sqrt{2+\sqrt{2}}$ pour le graphe hexagonal sur $\mathbb{Z}^2$. La démonstration du théorème 6.13, tirée de l'article de synthèse [HJ06], est due à Theodore Harris [Har60]. Dans

5. «Network» en anglais.
6. «Random networks» en anglais
7. «Lattice» en anglais.

Recueil de modèles stochastiques – Page 69/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

ce travail, Harris établit également que $\theta_2(1/2) = 0$ c'est-à-dire que $p_c(2) \geqslant 1/2$. Il faudra attendre le travail de Kesten [Kes80] pour que l'égalité $p_c(2) = 1/2$ soit démontrée. La question de la continuité de θ_2 au point $1/2$ est encore ouverte. Pour les autres valeurs de d, la valeur de la probabilité critique n'est même pas connue. On peut toutefois établir que θ_d est de classe $\mathcal{C}^\infty$ sur l'intervalle $]p_c(d), 1[$.

Recueil de modèles stochastiques – Page 70/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Chapitre 7

Renforcement

Recueil de modèles stochastiques – Page 71/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Mots-clés. Renforcement ; urne ; chaîne de Markov ; martingale ; échangeabilité ; graphe aléatoire ; marche aléatoire ; lois exponentielles.

Le phénomène du *renforcement* est très présent dans la nature, notamment en génétique, en physique statistique, en sociologie, en psychologie, et en neurosciences. Ce chapitre présente trois modèles emblématiques du phénomène : l'*urne de Pólya*, le *graphe aléatoire à attachement préférentiel de Barabási-Albert*, et une *marche aléatoire renforcée*. Il se termine par un *théorème de Rubin* sur les urnes de Pólya généralisées. D'autres instances se trouvent dans les modèles de Fisher-Wright et de Moran (chapitre 4), de Ewens (chapitre 9), dans certains modèles de croissance (chapitre 11), et dans le modèle d'Ehrenfest (chapitre 3). Dans les modèles markoviens, le phénomène du renforcement apparaît souvent en liaison avec une propriété de monotonie partielle dans une récurrence aléatoire.

7.1 Urne de Pólya

L'*urne de Pólya* est un modèle remarquablement simple de renforcement, qui modélise bien par exemple le fait que le succès ou la richesse s'auto-amplifie au cours du temps.

Au temps $n = 0$, on prépare une urne contenant $a > 0$ boules argentées et $b > 0$ boules blanches. Pour fabriquer la configuration de l'urne au temps $n = 1$, on tire au hasard une boule dans l'urne, puis on remet la boule tirée dans l'urne, ainsi qu'une nouvelle boule de même couleur. On répète ce mécanisme de manière indépendante pour fabriquer la configuration de l'urne en tout temps $n \in \mathbb{N}$. À l'instant $n \in \mathbb{N}$, l'urne contient $a + b + n$ boules. Si $M_n \in [0, 1]$ est la proportion de boules argentées à l'instant n, alors $M_0 = a/(a + b)$, et

$$M_{n+1} = \frac{(a + b + n)M_n + \mathbb{1}_{U_{n+1} \leqslant M_n}}{a + b + n + 1}$$

où $(U_n)_{n \geqslant 1}$ est une suite de variables aléatoires indépendantes et identiquement distribuées de loi uniforme sur $[0, 1]$. La suite récurrente aléatoire $(M_n)_{n \in \mathbb{N}}$ est une chaîne de Markov non homogène d'espace d'états $\mathbb{R}_+$, mais aussi une martingale. On pose $U_0 = 1$. Codons le résultat du n^{e} tirage par une variable aléatoire X_n à valeurs dans $\{\alpha, \beta\}$ où α et β indiquent que la boule tirée est argentée ou blanche. Alors à l'instant n l'urne contient $Y_n = \sum_{k=1}^{n} \mathbb{1}_{\{X_k = \alpha\}} = (a + b + n)M_n$ boules argentées et $Z_n = (a + b + n - Y_n)$ boules blanches, et $(X_n, Y_n, Z_n)_{n \geqslant 1}$ constitue une suite récurrente aléatoire :

$$(X_{n+1}, Y_{n+1}, Z_{n+1}) = \begin{cases} (\alpha, Y_n + 1, Z_n) & \text{si } U_{n+1} \leqslant \frac{X_n}{X_n + Y_n} \\ (\beta, Y_n, Z_n + 1) & \text{si } U_{n+1} > \frac{X_n}{X_n + Y_n} \end{cases}.$$

Théorème 7.1 (Martingale)**.** *La suite* $(M_n)_{n\in\mathbb{N}}$ *est une martingale à valeurs dans* $[0,1]$ *pour la filtration* $(\mathcal{F}_n)_{n\in\mathbb{N}}$ *définie par* $\mathcal{F}_n = \sigma(U_0,\dots,U_n)$*, et en particulier la proportion moyenne de boules argentées est conservée au cours du temps : pour tout* $n \in \mathbb{N}$*,*

$$\mathbb{E}(M_n) = \mathbb{E}(M_0) = \frac{a}{a+b}.$$

De plus, il existe une variable aléatoire M_∞ *sur* $[0,1]$ *telle que*

$$\lim_{n\to\infty} M_n = M_\infty$$

presque sûrement et dans L^p *pour tout* $p \geqslant 1$*. En particulier* $\mathbb{E}(M_\infty) = \mathbb{E}(M_0) = \frac{a}{a+b}$*.*

Démonstration. La suite $(M_n)_{n\in\mathbb{N}}$ est une martingale pour $(\mathcal{F}_n)_{n\geqslant 0}$ car pour tout $n \in \mathbb{N}$, M_n est $\mathcal{F}_n$-mesurable, à valeurs dans $[0,1]$ donc bornée donc intégrable, et

$$\mathbb{E}(M_{n+1} \,|\, \mathcal{F}_n) = \mathbb{E}(M_{n+1} \,|\, M_n) = \frac{(a+b+n)M_n + M_n}{a+b+n+1} = M_n.$$

Le théorème de convergence des martingales uniformément bornées donne $M_n \to M_\infty$ presque sûrement et dans L^p où M_∞ est une variable aléatoire à valeurs dans $[0,1]$ [1]. □

Théorème 7.2 (Équilibre)**.** *La v.a.r.* M_∞ *suit la loi Beta de densité*

$$u \in [0,1] \mapsto \frac{1}{\mathrm{Beta}(a,b)} u^{a-1}(1-u)^{b-1} \quad \textit{où} \quad \mathrm{Beta}(a,b) = \frac{(a-1)!(b-1)!}{(a+b-1)!}.$$

En particulier, si $a = b = 1$ *alors* M_∞ *suit la loi uniforme sur* $[0,1]$*.*

Démonstration. Pour tous c et k on note

$$c^{(k)} = c(c+1)\cdots(c+k-1) = \frac{(c+k-1)!}{(c-1)!}.$$

Soit X_n la variable aléatoire de Bernoulli valant 1 si la boule du n-ième tirage est argentée, et 0 si elle est blanche. Pour tous $x_1,\dots,x_n$ dans $\{0,1\}$, on a, en notant $k = x_1 + \cdots + x_n$,

$$\mathbb{P}(X_1 = x_1,\dots,X_n = x_n) = \frac{a^{(k)} b^{(n-k)}}{(a+b)^{(n)}}.$$

Cette probabilité est invariante par permutation des $x_1,\dots,x_n$: la loi du vecteur aléatoire $(X_1,\dots,X_n)$ est *échangeable*. Ainsi le nombre (aléatoire) $Y_n = X_1 + \cdots + X_n$ de boules argentées tirées au cours des n premiers tirages vérifie, pour tout $k \in \{0,1,\dots,n\}$,

$$\begin{aligned}
\mathbb{P}(Y_n = k) &= \binom{n}{k} \frac{a^{(k)} b^{(n-k)}}{(a+b)^{(n)}} \\
&= \binom{n}{k} \frac{\Gamma(a+k)\Gamma(b+n-k)\Gamma(a+b)}{\Gamma(a)\Gamma(b)\Gamma(a+b+n)} \\
&= \binom{n}{k} \frac{\mathrm{Beta}(a+k, b+n-k)}{\mathrm{Beta}(a,b)}
\end{aligned}$$

où $\mathrm{Beta}(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ et Γ sont les fonctions Beta et Gamma d'Euler. On dit que Y_n suit la loi Beta-Binomiale. On a $M_n = (a+Y_n)/(a+b+n)$ avec $Y_0 = 0$. Lorsque $a = b = 1$, la formule pour la loi de Y_n indique que Y_n suit la loi uniforme sur $\{0,1,\dots,n\}$, et donc M_n

1. La martingale est uniformément intégrable. Alternativement, on peut utiliser le théorème de convergence des martingales positives pour obtenir la convergence presque sûre puis le théorème de convergence dominée.

Recueil de modèles stochastiques – Page 72/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

suit la loi uniforme sur $\{1/(n+2), \ldots, (n+1)/(n+2)\}$, ce qui entraîne que M_∞ suit la loi uniforme sur $[0,1]$. Dans le cas général, on peut écrire, pour tout $t \in [0,1]$,

$$\begin{aligned}
\mathbb{P}(M_\infty \leqslant t) &= \lim_{n\to\infty} \mathbb{P}(Y_n \leqslant (a+b+n)t) \\
&= \lim_{n\to\infty} \sum_{k=0}^{\lfloor t(a+b+n)\rfloor} \binom{n}{k} \frac{\mathrm{Beta}(a+k, b+n-k)}{\mathrm{Beta}(a,b)} \\
&= \frac{1}{\mathrm{Beta}(a,b)} \int_0^t u^{a-1}(1-u)^{b-1}\, du
\end{aligned}$$

(le détail du calcul asymptotique est omis). □

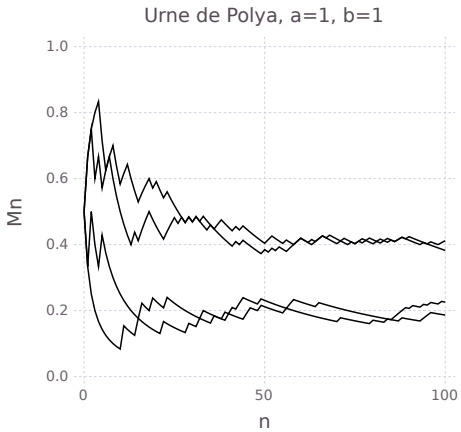


FIGURE 7.1 – Quelques trajectoires de l'urne de Pólya.

L'instant $T = \inf\{n \geqslant 1 : X_n = 1\} = \inf\{n \geqslant 0 : U_n \leqslant M_n\}$ du premier tirage produisant une boule argentée est un temps d'arrêt pour la filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$ (note : $U_0 = 1$). Le théorème d'arrêt de Doob donne $a/(a+b) = \mathbb{E}(M_0) = \mathbb{E}(M_{n\wedge T}) = \mathbb{E}((a+1)/(a+b+n\wedge T))$.

Remarque 7.3 (Urne de Pólya généralisée)**.** *Généralisons le mécanisme de renforcement comme suit : on se donne un entier $r \geqslant -1$, et, à chaque tirage, on remet dans l'urne $1+r$ boules de la couleur tirée. Si $r = 1$ on retrouve l'urne de Pólya que nous avons étudiée, pour laquelle Y_n suit la loi Beta-Binomiale. Si $r = 0$, alors on obtient des tirages avec remise et Y_n suit la loi binomiale, tandis que si $r = -1$ alors on obtient des tirages sans remise et Y_n suit la loi hypergéométrique. De manière générale, pour tout $r \geqslant -1$, la loi de Y_n est donnée par*

$$\mathbb{P}(Y_n = k) = \frac{a^{(r,k)} b^{(r,k)}}{(a+b)^{(r,k)}} \quad \text{où} \quad c^{(r,k)} := c(c+r)\cdots(c+(k-1)r).$$

Il est possible de considérer un nombre arbitraire de couleurs, ce qui fournit un modèle incluant le modèle d'échantillonnage de la loi hypergéométrique multitypes. De nombreux modèles de renforcement peuvent être obtenus comme une version généralisée de l'urne de Pólya.

Remarque 7.4 (Matrice de remise)**.** *Soit $A = (a_{i,j})_{1\leqslant i,j\leqslant k}$ une matrice $k \times k$ de nombres entiers. Considérons une urne de Pólya généralisée à k couleurs qui évolue comme suit : on tire une boule au hasard dans l'urne, on repère sa couleur, notée i, puis on remet dans l'urne $a_{i,j}$ boules de couleur j pour tout $1 \leqslant j \leqslant k$. On dit que A est la* matrice de remise *de l'urne. Pour l'urne de Pólya standard que nous avons étudiée on a $k = 2$ et $A = 2I_2$.*

Recueil de modèles stochastiques – Page 73/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

7.2 Graphe à attachement préférentiel de Barabási-Albert

Les graphes aléatoires permettent de modéliser un certain nombre de phénomènes naturels, comme les structures d'amitié dans les réseaux sociaux, les structures des liens entre pages dans le World Wide Web, les structures de collaborations dans les productions artistiques et scientifiques, les structures de régulation entre protéines, les liaisons entre machines dans le réseau Internet, etc. Les arbres de type Galton-Watson du chapitre 5 constituent un modèle de graphe aléatoire adapté aux structures de filiations. Le modèle le plus célèbre et le plus simple de graphe aléatoire est sans doute celui de Erdős-Rényi, évoqué dans le chapitre 6 : il se construit récursivement en ajoutant un nouveau site puis en tirant à pile ou face de manière indépendante sa connexion avec chacun des sites existants. Ce modèle ne colle pas avec la réalité de graphes aléatoires sociaux, pour lesquels les nouveaux sites se connectent préférentiellement aux sites existants les plus importants au sens de la connectivité (degré). Il y a là une instance du phénomène de renforcement dont il faut tenir compte spécifiquement.

Le *graphe aléatoire à attachement préférentiel de Barabási-Albert* est défini de la manière suivante : au temps $n \geqslant 1$, le graphe contient n sites (sommets) et un certain nombre de liens non orientés (arêtes) entre ces sites. Au temps $n = 1$, le site 1 est relié à lui même. Pour faire évoluer récursivement le graphe, du temps n au temps $n+1$, on considère les degrés $d_{n,1}, \ldots, d_{n,n}$ des n sites du graphe au temps n, et la loi de probabilité associée

$$p_{n,k} = \frac{d_{n,k}}{d_{n,1} + \cdots + d_{n,n}},$$

puis on connecte le nouveau site $n+1$ à un site choisi aléatoirement et indépendamment parmi les n sites existants, avec la loi de probabilité $p_{n,\cdot}$. Avec ce mécanisme, $d_{1,1} = 2$, $d_{2,1} = 3$, $d_{2,2} = 1$, et pour tout $n \geqslant 1$, $d_{n,1} + \cdots + d_{n,n} = 2n$ (soit n arêtes). On peut réaliser cette construction comme un modèle d'urne de Pólya généralisée : au temps $n \geqslant 1$ l'urne contient $2n$ boules dont les couleurs peuvent aller de 1 à n, on tire alors une boule au hasard, et si k est sa couleur, on remet dans l'urne 2 boules de couleur k ainsi qu'une boule nouvelle de couleur $n+1$ ce qui correspond à renforcer le site de couleur k et à introduire un nouveau site de couleur $n+1$ (le graphe gagne une arête : $k \leftrightarrow n+1$).

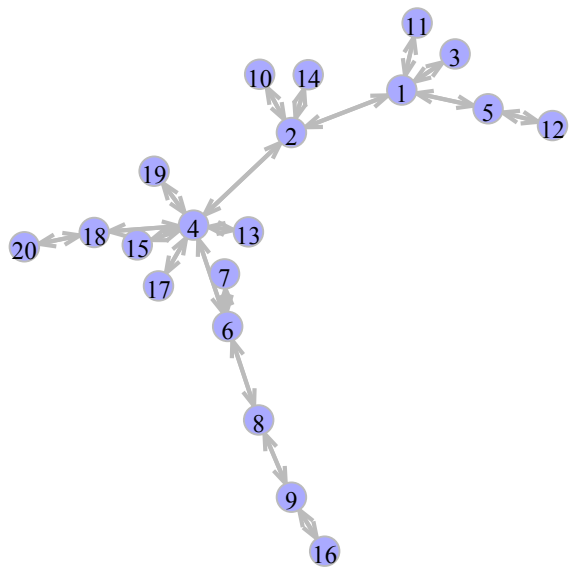


Figure 7.2 – Réalisation d'un graphe à attachement préférentiel de Barabási-Albert.

Le graphe à attachement préférentiel de Barabási-Albert n'a pas de cycles : c'est un arbre.

Recueil de modèles stochastiques – Page 74/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Théorème 7.5 (Loi de puissance)**.** *Fixons $k \geqslant 1$, et soit $d_{n,k}$ le degré du site k dans le graphe aléatoire à attachement préférentiel de Barabási-Albert à n sites avec $n \geqslant k$. Alors la proportion $d_{n,k}/n$ de sites connectés au site k converge presque sûrement quand $n \to \infty$ vers la loi Beta* $\mathrm{Beta}(1, 2k-1)$ *sur* $[0,1]$ *de densité* $u \in [0,1] \mapsto (2k-1)(1-x)^{2(k-1)}$.

La loi de puissance qui apparaît dans ce modèle à attachement préférentiel correspond bien aux réseaux sociaux réels, et diffère du comportement sous-exponentiel de la loi de Poisson des modèles de graphes aléatoires de Erdős-Rényi à attachement non préférentiel.

Démonstration. Au temps k, on décide d'une coloration parallèle : les $2k-1$ boules de couleur inférieure à k sont blanches tandis que la $2k$-ième est argentée. Au temps $n > k$, on ne suit que l'évolution des boules de ces couleurs là. Elle correspond à une urne de Pólya avec une composition initiale de $a = 1$ et $b = 2k-1$. On peut utiliser alors le théorème 7.2. □

Théorème 7.6 (Loi de puissance pour le nombre moyen de sites de degré donné)**.** *Pour tous $n, d \geqslant 1$, si $N(n,d)$ désigne le nombre moyen de sites de degré d au temps n dans le graphe aléatoire à attachement préférentiel de Barabási-Albert à n sites, alors*

$$\lim_{n\to\infty} \mathbb{E}\left(\frac{N(n,d)}{n}\right) = \frac{4}{d(d+1)(d+2)}.$$

Ainsi, dans un très grand graphe à attachement préférentiel ($n \gg 1$), la probabilité qu'une arête soit de degré d a une décroissance polynomiale en d^{-3} quand $d \to \infty$. La queue lourde de la loi du degré moyen est liée à la présence, due au renforcement, de sites fortement connectés.

Démonstration. L'heuristique des physiciens Barabási et Albert pour le comportement en d^{-3} est la suivante : si les degrés et le temps étaient continus, on aurait l'équation d'évolution

$$\partial_t d_{t,k} = \frac{d_{t,k}}{\sum_k d_{t,k}} = \frac{d_{t,k}}{2t},$$

qui donne $d_{t,k} = (t/t_k)^{1/2}$ où t_k est le temps d'apparition du site k, et comme les sites sont ajoutés uniformément sur $[0,t]$, on obtient $\mathbb{P}(d_{t,k} > d) = \mathbb{P}(t_k < t/d^2) = 1/d^2$, d'où en dérivant, $\mathbb{P}(d_{t,k} = d) = 2/d^3$, qui est le comportement polynomial suggéré par les simulations.

En fait, la suite $(N(n,\cdot))_{n\geqslant 1}$ est une chaîne de Markov. Conditionnellement à $N(n,\cdot)$, pour créer un nouveau site de degré d, il faut ajouter une arête à un site de degré $d-1$, ce qui se produit avec probabilité $(d-1)N(n,d-1)/(2n)$, tandis qu'un site de degré d disparaît si on lui ajoute une arête, ce qui se produit avec probabilité $dN(n,d)/(2n)$. Enfin, un site de degré 1 est créé à chaque étape par construction. Aussi, le nombre moyen $m_n(d) := \mathbb{E}N(n,d)$ de sites de degré d au temps n vérifie une équation de récurrence linéaire, qualifiée d'*équation maître* par Dorogovstev, Mendes, et Samukhin : pour tout $n, d \geqslant 1$:

$$m_{n+1}(d) - m_n(d) = -\frac{d}{2n} m_n(d) + \frac{d-1}{2n} m_n(d-1) + \mathbb{1}_{d=1},$$

avec pour condition initiale $m_1 = \mathbb{1}_{d=2}$. Pour $d = 1$, l'équation s'écrit

$$m_{n+1}(1) = c + \left(1 - \frac{b}{n}\right) m_n(1) \quad \text{avec } c = 1 \text{ et } b = 1/2,$$

ce qui donne

$$\begin{aligned} m_{n+1}(d) &= c + \left(1 - \frac{b}{n}\right)c + \left(1 - \frac{b}{n}\right)\left(1 - \frac{b}{n-1}\right) m_{n-1}(d) \\ &= c \sum_{k=1}^{n} \prod_{j=k+1}^{n} \left(1 - \frac{b}{j}\right) + \underbrace{m_1(1)}_{=0} \prod_{k=1}^{n} \left(1 - \frac{b}{k}\right). \end{aligned}$$

Recueil de modèles stochastiques – Page 75/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Or $\prod_{j=k+1}^{n}\left(1-\frac{b}{j}\right) \approx \exp(-\sum_{k=k+1}^{n} b/j) \approx \exp(-b(\log(n)-\log(k))) = (k/n)^b$, d'où

$$m_n(1) \approx cn^{-b}\int_0^n s^b\,ds = cn^{-b}\frac{n^{b+1}}{b+1} = \frac{cn}{b+1},$$

de sorte que $\lim_{n\to\infty} m_n(1)/n = c/(b+1) = 2/3$. Plus généralement, pour tout $d > 1$, on a

$$m_{n+1}(d) = c_n + \left(1-\frac{b}{n}\right)m_n(d)$$

où $b = d/2$ et $c_n = ((d-1)/2)m_n(d-1)/n$, et on montre par récurrence sur d que $c = c(d) = \lim_{n\to\infty} c_n$ existe et

$$\frac{m_n(d)}{n} \underset{n\to\infty}{\longrightarrow} \frac{c}{b+1}.$$

Ensuite, en posant $\ell_d := \lim_{n\to\infty} m_n(d)/n$ on obtient $p_1 = 2/3$, tandis que pour tout $d > 1$,

$$\ell_d = \frac{\frac{(d-1)}{2}\ell_{d-1}}{\frac{d}{2}+1} = \ell_{d-1}\frac{d-1}{d-2}.$$

Enfin, la solution de cette récurrence en d est (facile à vérifier !)

$$\ell_d = \frac{4}{d(d+1)(d+2)}.$$

□

7.3 Marche aléatoire renforcée

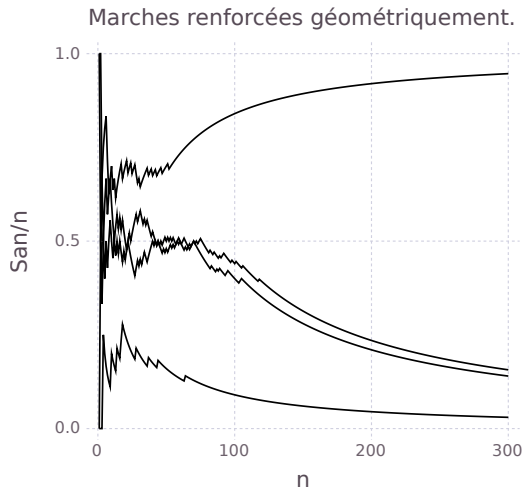


Figure 7.3 – Occupation d'un chemin pour la marche renforcée géométriquement.

Le phénomène du renforcement est à l'œuvre dans l'apparition des chemins empruntés par les passants dans les montagnes ou par les fourmis sur le sol. Considérons deux chemins distincts α et β reliant les mêmes points de départ et d'arrivée, empruntés par des passants successifs. Pour tout $n \geqslant 1$, on code par une variable aléatoire X_n à valeur dans $\{\alpha, \beta\}$ le chemin emprunté lors du n-ième passage. On code par des v.a. A_n et B_n l'attractivité des

Recueil de modèles stochastiques – Page 76/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

chemins α et β au moment du $(n+1)^{\text{e}}$ passage : pour les humains, il peut s'agir par exemple de la raréfaction de l'herbe, tandis que pour les fourmis, il peut s'agir de la quantité de phéromone. On se donne (A_0, B_0), ainsi qu'une fonction $r :]0,\infty[\to]0,\infty[$ appelée *fonction de renforcement* telle que $r(x) \geqslant x$ pour tout $x > 0$, et on modélise $(X_n)_{n\geqslant 1}$ par

$$(X_{n+1}, A_{n+1}, B_{n+1}) = \begin{cases} (\alpha, r(A_n), B_n) & \text{si } U_{n+1} \leqslant \frac{A_n}{A_n+B_n}; \\ (\beta, A_n, r(B_n)) & \text{si } U_{n+1} > \frac{A_n}{A_n+B_n} \end{cases}$$

où $(U_n)_{n\geqslant 1}$ est une suite de variables aléatoires indépendantes et identiquement distribuées de loi uniforme sur $[0,1]$. La suite récurrente aléatoire $((X_n, A_n, B_n))_{n\in\mathbb{N}}$ est une chaîne de Markov non homogène (la valeur de X_0 ne joue aucun rôle dans la récurrence). Lorsque A_0, B_0, et r prennent des valeurs entières, tout se passe comme si nous avions une urne contenant, à tout instant n, A_n boules argentées et B_n boules blanches : on tire une boule au hasard dans cette urne, puis on introduit dans l'urne $r(A_n)$ boules de la même couleur, de sorte qu'on obtient le cas des tirages sans remise si $r(x) = x$, l'urne de Pólya si $r(x) = x+1$, et une sorte d'urne de Pólya non-linéaire dans le cas général.

La suite $(X_n)_{n\geqslant 1}$ constitue une marche aléatoire non markovienne sur l'ensemble à deux points $\{\alpha, \beta\}$. Ses transitions dépendent les unes des autres via le mécanisme de renforcement lié au temps passé sur chaque site. On parle de *marche aléatoire renforcée par sites*. Le mécanisme de renforcement peut également être vu comme une sorte d'algorithme stochastique.

Pour simplifier, on suppose que $A_0 = B_0 = 1$ (X_0 est inutile). Le nombre de passages par le chemin α et par le chemin β à l'instant n sont donnés par

$$S_n(\alpha) = \sum_{k=1}^{n} \mathbb{1}_{\{X_k=\alpha\}} \quad \text{et} \quad S_n(\beta) = \sum_{k=1}^{n} \mathbb{1}_{\{X_k=\beta\}}.$$

On pose par commodité $S_0(\alpha) = 0$ et $S_0(\beta) = 0$ de sorte que $S_n(\alpha) + S_n(\beta) = n$ pour tout $n \in \mathbb{N}$. Voici le comportement de ces deux suites pour trois cas de fonction de renforcement r.

Théorème 7.7 (Comportement asymptotique presque sûr quand $n \to \infty$)**.**

— Absence de renforcement. *Si $r(x) = x$ pour tout $x > 0$, alors presque sûrement*

$$S_n(\alpha) \sim \frac{n}{2}, \quad S_n(\beta) \sim \frac{n}{2}, \quad \liminf |S_n(\alpha) - S_n(\beta)| = 0, \quad \limsup \frac{|S_n(\alpha) - S_n(\beta)|}{\sqrt{2n\log(\log(n))}} = 1.$$

— Renforcement linéaire. *Si $r(x) = x+1$ pour tout $x > 0$ alors il existe une variable aléatoire U uniforme sur $[0,1]$ telle que presque sûrement,*

$$\frac{S_n(\alpha)}{n} \to U \quad \textit{et} \quad \frac{S_n(\beta)}{n} \to 1 - U.$$

— Renforcement géométrique. *Si $r(x) = \rho x$ pour tout $x > 0$ où $\rho > 1$ est une constante alors presque sûrement la suite aléatoire $(X_n)_{n\in\mathbb{N}}$ est constante à partir d'un certain rang sur n, et sa valeur limite suit la loi de Bernoulli symétrique sur $\{\alpha, \beta\}$.*

En l'absence de renforcement ou dans le cas du renforcement linéaire, la fréquence d'emprunt de chacun des deux chemins converge au fil du temps vers $1/2$ dans le premier cas et vers un nombre aléatoire uniforme sur $[0,1]$ dans le second cas. Intuitivement, un renforcement sur-linéaire pourrait forcer la fréquence d'emprunt des chemins à converger vers les valeurs extrêmes 0 ou 1. Cette intuition est confirmée par le cas du renforcement géométrique, pour lequel au bout d'un certain temps, l'un des deux chemins est emprunté systématiquement. Le modèle du renforcement linéaire coïncide avec l'urne de Pólya étudiée précédemment. Le modèle du renforcement géométrique est particulièrement attrayant, car il fait apparaître un chemin privilégié, choisi aléatoirement au fil du renforcement. Il s'agit en quelque sorte d'une urne de Pólya généralisée non-linéaire, plus précisément sur-linéaire. Le théorème 7.8 de Rubin ci-après fournit un critère sur le renforcement pour que ce phénomène apparaissent.

Recueil de modèles stochastiques – Page 77/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Démonstration.

— Absence de renforcement. Dans ce cas $(A_n)_{n\in\mathbb{N}}$ et $(B_n)_{n\in\mathbb{N}}$ sont constantes et égales à 1. On retrouve les tirages avec remise, le processus de Bernoulli (jeu de pile ou face). La suite $(S_n(\alpha))_{n\in\mathbb{N}}$ est un processus de Bernoulli sur $\mathbb{N}$ issu de 0 dont les incréments sont de loi de Bernoulli sur $\{0,1\}$ de paramètre $A_0/(A_0+B_0)=1/2$. D'après la loi forte des grands nombres, presque sûrement, $S_n(\alpha)/n \to 1/2$ et $S_n(\beta)/n = 1 - S_n(\alpha)/n \to 1/2$ quand $n\to\infty$. La suite $(S_n(\alpha)-S_n(\beta))_{n\in\mathbb{N}}$ est une marche aléatoire sur $\mathbb{Z}$ issue de 0 dont les incréments sont de loi de Rademacher sur $\{-1,1\}$ de paramètre 1/2. C'est une chaîne de Markov irréductible récurrente : presque sûrement chaque état est visité une infinité de fois, et en particulier l'état 0, d'où $\liminf |S_n(\alpha)-S_n(\beta)| = 0$ presque sûrement. Le résultat sur $\limsup$ provient de la loi du logarithme itéré de Strassen ;

— Renforcement linéaire. Dans ce cas $A_n = 1+S_n(\alpha)$ et $B_n = 1+S_n(\beta)$ pour tout $n\in\mathbb{N}$. On retrouve l'urne de Pólya et le résultat attendu découle alors directement du cas uniforme dans le théorème 7.2. Effectuons malgré tout le raisonnement allégé avec les notations actuelles. La relation $S_n(\alpha)+S_n(\beta)=n$ réduit le problème à l'étude de $S_n(\alpha)$. Soit $(\mathcal{F}_n)_{n\in\mathbb{N}}$ la filtration définie par $\mathcal{F}_n=\sigma(U_0,\dots,U_n)$ avec $U_0=0$. On a

$$\begin{aligned}\mathbb{E}(1+S_{n+1}(\alpha)\,|\,\mathcal{F}_n) &= 1+S_n(\alpha)+\mathbb{E}(\mathbb{1}_{\{X_{n+1}=a\}}\,|\,\mathcal{F}_n)\\ &= 1+S_n(\alpha)+\mathbb{E}(\mathbb{1}_{\{U_{n+1}\leqslant\frac{1+S_n(\alpha)}{n+2}\}})\\ &= 1+S_n(\alpha)+\frac{1+S_n(\alpha)}{n+2} = (2+(n+1))\frac{1+S_n(\alpha)}{2+n},\end{aligned}$$

et donc $((1+S_n(\alpha))/(n+2))_{n\in\mathbb{N}}$ est une martingale pour $(\mathcal{F}_n)_{n\in\mathbb{N}}$. À valeurs dans $[0,1]$, elle est uniformément bornée et converge donc presque sûrement en moyenne vers une variable aléatoire U à valeurs dans $[0,1]$. On montre enfin par récurrence sur n que $(1+S_n)/(n+2)$ suit la loi uniforme sur $\{1/(n+2),\dots,(n+1)/(n+2)\}$;

— Renforcement géométrique. Dans ce cas $A_n=\rho^{S_n(\alpha)}$ et $B_n=\rho^{S_n(\beta)}$ pour tout $n\in\mathbb{N}$. La variable aléatoire $\Delta_n := S_n(\alpha)-S_n(\beta)$ vérifie $\Delta_0=0$ et

$$\Delta_{n+1}=\Delta_n+\mathbb{1}_{\{U_n\leqslant 1/(1+\rho^{-\Delta_n})\}}-\mathbb{1}_{\{U_n>1/(1+\rho^{-\Delta_n})\}}.$$

Posons $\mathcal{F}_n=\sigma(U_0,\dots,U_n)$, avec $U_0=0$. Il en découle que

$$\begin{aligned}\mathbb{P}(|\Delta_{n+1}|=|\Delta_n|+1\,|\,\mathcal{F}_n) &= \frac{\rho^{|\Delta_n|}}{1+\rho^{|\Delta_n|}}\mathbb{1}_{\{\Delta_n\neq 0\}}+\mathbb{1}_{\{\Delta_n=0\}}\\ \mathbb{P}(|\Delta_{n+1}|=|\Delta_n|-1\,|\,\mathcal{F}_n) &= \frac{1}{1+\rho^{|\Delta_n|}}\mathbb{1}_{\{\Delta_n\neq 0\}}.\end{aligned}$$

Si $f:\mathbb{N}\to\mathbb{R}$ vérifie $f(0)=0$ et pour tout $n\geqslant 1$,

$$\rho^n(f(n+1)-f(n))=f(n)-f(n-1),$$

alors

$$\mathbb{E}(f(|\Delta_{n+1}|)\,|\,\mathcal{F}_n)=f(|\Delta_n|)+\mathbb{1}_{\{\Delta_n=0\}}\geqslant f(|\Delta_n|),$$

et $(f(|\Delta_n|))_{n\in\mathbb{N}}$ est une sous-martingale. Considérons le cas où pour tout $n\geqslant 1$,

$$f(n)=\sum_{k=0}^{n-1}\rho^{-\frac{k(k+1)}{2}}.$$

Comme f est bornée, il en découle que la sous-martingale $f(|\Delta|)$ est uniformément bornée : elle converge donc presque sûrement (et en moyenne). Comme f est injective et comme $|\Delta|$ prend ses valeurs dans $\mathbb{N}$, les valeurs prises par $f(|\Delta|)$ sont discrètes.

Recueil de modèles stochastiques – Page 78/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Comme $f(|\Delta|)$ converge presque sûrement, elle ne peut converger que vers le seul point d'accumulation de $f(\mathbb{N})$, qui est $\lim_{n\to\infty} f(n) = \sum_{k=0}^{\infty} \rho^{-\frac{k(k+1)}{2}}$. Ainsi presque sûrement $\lim_{n\to\infty} |\Delta_n| = \infty$. À présent, pour tous $N, d \geqslant 0$, la probabilité que la suite $|\Delta|$ soit croissante à partir du rang N sachant que $|\Delta_N| = d$ vaut

$$\mathbb{P}(\forall n \geqslant N : |\Delta_{n+1}| = |\Delta_n| + 1 \,|\, |\Delta_N| = d) = \prod_{k=d}^{\infty} \frac{1}{1+\rho^{-k}} =: p_d.$$

La probabilité p_d ne dépend pas de N et vérifie $p_d \nearrow 1$ quand $d \to \infty$. Ainsi la probabilité que $|\Delta|$ soit croissante à partir d'un certain rang (aléatoire) s'écrit

$$\begin{aligned}\mathbb{P}(\exists N \geqslant 0, \forall n \geqslant N : |\Delta_{n+1}| = |\Delta_n| + 1) &\geqslant \sup_{N\geqslant 0} \mathbb{P}(\forall n \geqslant N, |\Delta_{n+1}| = |\Delta_n| + 1) \\ &= \sup_{N\geqslant 0} \sum_{d=0}^{\infty} p_d \mathbb{P}(|\Delta_N| = d).\end{aligned}$$

Or pour tout $d \geqslant 0$, on a $\lim_{N\to\infty} \mathbb{P}(|\Delta_N| = d) = 0$ car nous savons que presque sûrement $\lim_{N\to\infty} |\Delta_N| = \infty$. Par conséquent, pour tout $D \geqslant 0$,

$$\sup_{N\geqslant 0} \sum_{d=0}^{\infty} p_d \mathbb{P}(|\Delta_N| = d) \geqslant \sup_{N\geqslant 0} \sum_{d=D}^{\infty} p_d \mathbb{P}(|\Delta_N| = d) \geqslant p_D.$$

Comme $P_D \to 1$ quand $D \to \infty$, il en découle que presque sûrement $|\Delta|$ est croissante à partir d'un certain rang, et donc la suite X est absorbée par α ou par β. La loi de la limite de X est symétrique par symétrie du modèle en α et β quand $A_0 = B_0$.

□

7.4 Théorème de Rubin sur les urnes de Pólya généralisées

On se donne deux fonctions $s_\alpha, s_\beta : \mathbb{R}\ + \to \mathbb{R}\ +$ telles que $s_\alpha(0) > 0$ et $s_\beta(0) > 0$. On pose $S_\alpha(n) = s_\alpha(0) + \cdots + s_\alpha(n)$ et $S_\beta(n) = s_\beta(0) + \cdots + s_\beta(n)$. On construit une suite récurrente aléatoire $(X_n)_{n\geqslant 1}$ à valeurs dans $\{\alpha, \beta\}$ comme suit : pour tout $n \in \mathbb{N}$, conditionnellement à $X_1, \ldots, X_n$, en notant $Y_n = \sum_{k=1}^{n} \mathbb{1}_{\{X_k=\alpha\}}$ le nombre de α et $Z_n = n - Y_n$ le nombre de β,

$$(X_{n+1}, Y_{n+1}, Z_{n+1}) = \begin{cases} (\alpha, Y_n + 1, Z_n) & \text{si } U_{n+1} \leqslant \frac{S_\alpha(Y_n)}{S_\alpha(Y_n)+S_\beta(Z_n)} \\ (\beta, Y_n, Z_n + 1) & \text{si } U_{n+1} > \frac{S_\alpha(Z_n)}{S_\alpha(Y_n)+S_\beta(Z_n)}, \end{cases}$$

où $(U_n)_{n\geqslant 1}$ est une suite de v.a. i.i.d. de loi uniforme sur $[0,1]$. La suite $(X_n)_{n\geqslant 1}$ ainsi construite code les tirages successifs d'une *urne de Pólya généralisée*. Le modèle du théorème 7.7 dans le cas où (A_0, B_0) est déterministe s'obtient avec $(s_\alpha(0), s_\beta(0)) = (A_0, B_0)$, et $S_\alpha(x) = S_\beta(x) = r(x)$ pour tout $x > 0$. Si r est linéaire alors on retrouve l'absence de renforcement (tirages avec remise). Si r est affine alors on retrouve le renforcement linéaire (urne de Pólya standard). Enfin, si r est une fonction puissance, alors on retrouve le renforcement géométrique.

Pour étudier le cas général, on introduit la probabilité p_α (respectivement p_β) que la suite $(X_n)_{n\geqslant 1}$ ne comporte qu'un nombre fini de β (respectivement de α) c'est-à-dire que des α (respectivement que des β) à partir d'un certain rang sur n :

$$p_\alpha = \mathbb{P}(\cup_n \cap_{k\geqslant n} \{X_k = \alpha\}) \quad \text{et} \quad p_\beta = \mathbb{P}(\cup_n \cap_{k\geqslant n} \{X_k = \beta\}).$$

On a $p_\alpha + p_\beta \leqslant 1$. Les nombres suivants dans $[0, \infty]$ sont essentiels au théorème ci-dessous :

$$\varphi_\alpha = \sum_{n=0}^{\infty} \frac{1}{S_\alpha(n)} \quad \text{et} \quad \varphi_\beta = \sum_{n=0}^{\infty} \frac{1}{S_\beta(n)}.$$

Recueil de modèles stochastiques – Page 79/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Théorème 7.8 (Théorème de Rubin)**.** *Les valeurs de p_α, p_β sont liées à $\varphi_\alpha, \varphi_\beta$:*

	$\boldsymbol{\varphi_\beta < \infty}$	$\boldsymbol{\varphi_\beta = \infty}$
$\boldsymbol{\varphi_\alpha < \infty}$	$p_\alpha > 0,\ p_\beta > 0,\ p_\alpha + p_\beta = 1$	$p_\alpha = 1$ *(donc* $p_\beta = 0$*)*
$\boldsymbol{\varphi_\alpha = \infty}$	$p_\beta = 1$ *(donc* $p_\alpha = 0$*)*	$p_\alpha = 0,\ p_\beta = 0$

Ainsi, la trivialité des probabilités p_α et p_β ne dépend que du comportement asymptotique des fonctions de renforcement s_α et s_β, et les cas critiques sont liés à la condition de Riemann. En absence de renforcement (S_α et S_β sont linéaires) ainsi que dans le cas d'un renforcement linéaire (S_α et S_β sont affines) on a $\varphi_\alpha = \varphi_\beta = \infty$ car la série harmonique diverge. Dès que le renforcement est sur-linéaire, alors $\varphi_\alpha < \infty$ et $\varphi_\beta < \infty$, en particulier pour le cas du renforcement quadratique, et bien sûr le cas du renforcement géométrique.

Démonstration. Soient $(E_n^\alpha)_{n\geqslant 1}$ et $(E_n^\beta)_{n\geqslant 1}$ deux suites indépendantes de v.a.r. indépendantes de loi exponentielle avec $\mathbb{E}(E_n^\alpha) = 1/S_\alpha(n)$ et $\mathbb{E}(E_n^\beta) = 1/S_\beta(n)$ pour tout $n \geqslant 1$. On définit les ensembles aléatoires $\mathcal{A} = \{\sum_{k=1}^n E_k^\alpha : n \geqslant 1\}$ et $\mathcal{B} = \{\sum_{k=1}^n E_k^\beta : n \geqslant 1\}$ et $\mathcal{G} = \mathcal{A} \cup \mathcal{B}$. Soit $\xi_1 < \xi_2 < \cdots$ les éléments de $\mathcal{G}$ rangés par ordre croissant. On considère à présent la suite aléatoire $(X'_n)_{n\geqslant 1}$ à valeurs dans $\{\alpha, \beta\}$ définie par $X_n = \alpha$ si $\xi_n \in \mathcal{A}$ et $X_n = \beta$ si $\xi_n \in \mathcal{B}$.

Les suites $(X'_n)_{n\geqslant 1}$ et $(X_n)_{n\geqslant 1}$ ont même loi, et cela découle des propriétés des lois exponentielles dont l'*absence de mémoire*. Examinons l'égalité en loi de X_1 et X'_1. Si U et V sont deux v.a.r. indépendantes de loi exponentielles de moyennes u et v respectivement alors $\mathbb{P}(U < V) = u^{-1}/(u^{-1} + v^{-1})$ et $\mathbb{P}(V < U) = v^{-1}/(u^{-1} + v^{-1})$, ce qui fait que

$$\mathbb{P}(X'_1 = \alpha) = \mathbb{P}(\xi_1 \in \mathcal{A}) = \mathbb{P}(E_1^\alpha \leqslant E_1^\beta) = \frac{S_\alpha(1)}{S_\alpha(1) + S_\beta(1)} = \mathbb{P}(X_1 = \alpha).$$

La même idée fournit avec un peu de labeur l'égalité en loi de $(X_n)_{n\geqslant 1}$ et $(X'_n)_{n\geqslant 1}$.

À présent, la *loi du zéro-un pour les sommes de v.a.r. exponentielles indépendantes* affirme que l'événement $\{\sum_{n=0}^\infty E_n^\alpha < \infty\}$ est de probabilité 0 ou 1, et est de probabilité 1 si et seulement si $\sum_{n=0}^\infty 1/S_\alpha(n) < \infty$ (pour le voir on peut utiliser le théorème de convergence monotone et la transformée de Laplace). La même propriété a lieu pour la suite $(E_n^\beta)_{n\geqslant 1}$ avec G. Par ailleurs, on a $p_\alpha = \mathbb{P}(\sum_{k=0}^n E_n^\alpha < \sum_{k=0}^n E_k^\beta)$ et $p_\beta = \mathbb{P}(\sum_{k=0}^n E_n^\beta < \sum_{k=0}^n E_k^\alpha)$. □

7.5 Notes et commentaires

Les urnes de Pólya sont étudiées notamment dans le livre de Hosam Mahmoud [Mah09], dans celui de Norman Johnson et Samuel Kotz [JK77], et dans les articles de survol de Samuel Kotz et Narayanaswamy Balakrishnan [KB97] et de Robin Pemantle [Pem07]. Les urnes de Pólya portent le nom du mathématicien hongrois George Pólya. L'école hongroise de mathématiques a beaucoup développé les mathématiques discrètes aléatoires, avec notamment les travaux fondateurs de Paul Erdős et Alfréd Rényi des années 1950 sur les graphes aléatoires. Le modèle de graphe aléatoire à attachement préférentiel a été introduit par les physiciens hongrois Albert-László Barabási et Réka Albert [AB99, AB02] puis étudié par Béla Bollobás, Oliver Riordan, Joel Spencer, et Gábor Tusnády [BRST01]. Il peut être vu comme une instance du phénomène d'auto-organisation. Le sujet fait toujours l'objet de recherches à l'heure où nous écrivons ces lignes, comme en témoigne par exemple un article de Bubeck, Mossel, et Rácz [BMR14]. La partie de ce chapitre sur les graphes aléatoires à attachement préférentiel est inspirée du cours [Bod14] de Thierry Bodineau à l'École Polytechnique, ainsi que des livres de Remco van der Hofstad [vdH14] et de Rick Durrett [Dur10]. Il est possible d'utiliser l'inégalité de concentration de Azuma-Hoeffding (chapitre 17) pour établir que $N(n, d)$ (sans espérance) a un comportement polynomial en d quand $n \to \infty$. Il est également possible de modifier la règle d'attachement préférentiel afin d'obtenir un comportement polynomial de degré différent

Recueil de modèles stochastiques – Page 80/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

de 3, ce qui correspond à considérer une urne de Pólya modifiée (non-linéaire). Le mécanisme d'attachement préférentiel est également présent dans le processus de branchement de Herbert Simon et Udny Yule. La loi discrète à décroissance polynomiale, version discrète de la loi de (Vilfredo) Pareto, est connue sous le nom de loi de (George Kingsley) Zipf.

La partie sur la marche aléatoire renforcée est inspirée d'un texte de Thierry Lévy. Nous renvoyons plus généralement au survol [Pem07] de Robin Pemantle sur les processus aléatoires renforcés, ainsi qu'à celui de Pierre Tarrès [Tar11] sur les marches aléaoires renforcées. Un résultat récent sur le renforcement se trouve par exemple dans l'article de Margherita Disertori, Christophe Sabot, et Pierre Tarrès [DST14]. Un modèle continu de renforcement est étudié par Michel Benaïm et Olivier Raimond dans [BR11]. Le théorème 7.8 de Herman Rubin figure dans un appendice d'un article de Burgess Davis [Dav90].

Au delà des modèles évoqués, le phénomène de renforcement joue un rôle important dans la théorie de l'apprentissage. Il est également relié à une gamme d'algorithmes stochastiques comme l'algorithme d'approximation stochastique de Robbins-Monro, lié à la théorie des systèmes dynamiques et ses attracteurs. Les martingales constituent un outil privilégié de l'approche stochastique de la théorie des jeux et stratégies, présentée brièvement dans le livre de Michel Benaïm et Nicole El Karoui [BK05].

Recueil de modèles stochastiques – Page 81/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Chapitre 8

Simulation de lois discrètes

Recueil de modèles stochastiques – Page 83/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Mots-clés. Loi discrète ; simulation ; chaîne de Markov ; groupe symétrique ; partition aléatoire ; permutation aléatoire ; graphe aléatoire ; modèle des configurations ; algorithme de Metropolis-Hastings ; algorithme du recuit simulé ; algorithme de Propp-Wilson.

Ce chapitre est consacré à quelques algorithmes remarquables de simulation de lois 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.

8.1 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 fini ou dénombrable. Nous pouvons numéroter en utilisant une bijection $\varphi : E \to \{1, 2, \ldots\}$ et en posant $a_k := \varphi^{-1}(k)$. À présent, si on partitionne l'intervalle réel $[0, 1]$ en blocs de mesures de Lebesgue respectives $\mu(a_1), \mu(a_2)$, etc, par exemple en utilisant les intervalles

$$I_1 = [0, \mu(a_1)[\,,\ I_2 = [\mu(a_1), \mu(a_2) + \mu(a_2)[\,,\ \ldots,$$

et si U est une variable aléatoire de loi uniforme sur l'intervalle $[0, 1]$, alors $\mathbb{P}(U \in I_{\varphi(a)}) = \mu(a)$ pour tout $a \in E$. L'algorithme basique de simulation de μ est alors le suivant : on génère une réalisation u de U, ensuite si $u \leqslant \mu(a_1)$ alors on décide a_1 ; sinon, si $u \leqslant \mu(a_1) + \mu(a_2)$, alors on décide a_2, etc. Si F est la fonction de répartition de la loi $\mu \circ \varphi^{-1}$ sur $\{1, 2, \ldots\} \subset \mathbb{R}$, d'inverse généralisé F^{-1}, alors $(\varphi^{-1} \circ F^{-1})(U) \sim \mu$. 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) ! La numérotation φ qui minimise le coût moyen $\mathbb{E}(N)$ vérifie

$$\mu(a_1) \geqslant \mu(a_2) \geqslant \cdots .$$

Pour la loi géométrique (de moyenne quelconque) et pour la loi de Poisson (de moyenne 1), la numérotation naturelle est à poids décroissants. D'autre part, si $\mathrm{card}(E)$ est petit, alors on peut déterminer l'ordre à poids décroissants en utilisant un algorithme de tri (qui a un coût).

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 amincissement [1]. Plus précisément, on simule $\lceil\lambda\rceil$ variables aléatoires $X_1, \ldots, X_{\lceil\lambda\rceil}$ i.i.d. de loi Poi(1), puis, conditionnellement à leur somme $S = X_1 + \cdots + X_{\lceil\lambda\rceil}$, 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 $B_1 + \cdots + B_S \sim \mathrm{Poi}(\lambda)$.

Le cas de la loi uniforme sur un ensemble E fini est justiciable d'une approche directe : $\varphi^{-1}(\lceil \mathrm{card}(E)U\rceil)$ suit cette loi ! Cependant, l'algorithme basique est impraticable lorsque E est difficile à énumérer et donc φ est difficile d'accès, ou lorsque $\mathrm{card}(E)$ est très grand et μ est peu concentrée. Nous étudions par la suite deux exemples de ce type : les ensembles des permutations de $\{1, \ldots, n\}$ et des partitions de $\{1, \ldots, n\}$, 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, on dispose d'algorithmes à base de chaîne de Markov comme l'algorithme de Metropolis-Hastings (approché) ou celui de Propp-Wilson (exact) brièvement abordés plus loin sans le chapitre.

8.2 Permutations et partitions 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 $\sum_{\sigma\in\mathcal{S}_n} \mathrm{card}(\mathcal{S}_n)^{-1}\delta_\sigma$ sur l'ensemble fini $\mathcal{S}_n$ des permutations de $\{1, \ldots, n\}$ (groupe symétrique). Or $\mathrm{card}(\mathcal{S}_n) = n! \sim \sqrt{2\pi n}(n/e)^n$ 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 $U_1, \ldots, U_n$ sont des variables aléatoires i.i.d. de loi uniforme sur $[0, 1]$ et si σ est une permutation aléatoire telle que $U_{\sigma(1)} \leqslant \cdots \leqslant U_{\sigma(n)}$, alors σ suit la loi uniforme sur $\mathcal{S}_n$. La complexité est celle de l'algorithme de tri utilisé [2]. Un algorithme plus naïf pour simuler une permutation σ uniforme consiste à tirer uniformément et sans remise les valeurs de $\sigma(1), \ldots, \sigma(n)$ dans $\{1, \ldots, n\}$. 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. Un bon algorithme de simulation (exacte !) de la loi uniforme sur $\mathcal{S}_n$ est connu sous le nom de *Fisher-Yates shuffle* ou de *Knuth shuffle*. Sa complexité est de $n - 1$.

Théorème 8.1 (Algorithme de Fisher-Yates-Knuth)**.** *Si $U_1, \ldots, U_n$ sont des variables aléatoires indépendantes avec U_i de loi uniforme sur $\{1, \ldots, i\}$ pour tout $1 \leqslant i \leqslant n$ alors le produit de transpositions aléatoires $(1, U_1) \cdots (n, U_n)$ suit la loi uniforme sur $\mathcal{S}_n$.*

L'inversion dans $\mathcal{S}_n$ 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 $\mathcal{S}_n$. La transposition $(1, U_1)$ est triviale (élément neutre de $\mathcal{S}_n$) et il n'est pas nécessaire d'en tenir compte si $n \geqslant 2$ (elle rend cependant la formule valable pour $n = 1$). L'algorithme de Fisher-Yates-Knuth pour permuter un vecteur `v` s'écrit en pseudo-code :

```
for k from length(v) downto 2 do swap(v[ceil(k*rand)],v[k])
```

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

1. Si le couple (X, Y) vérifie $\mathrm{Loi}(X) = \mathrm{Poi}(\lambda)$ et $\mathrm{Loi}(Y|X = n) = \mathrm{Bin}(n, p)$ pour tout n alors $Y \sim \mathrm{Poi}(p\lambda)$.
2. De l'ordre de $n\log(n)$ avec grande probabilité et n^2 au pire pour l'*algorithme de tri rapide* (« *quick sort* »).

Recueil de modèles stochastiques – Page 84/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

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

$$\mathbb{P}(\sigma_{n+1} = \sigma) = \sum_{i=1}^{n+1} \mathbb{P}(\sigma_n = \sigma(n+1,i))\mathbb{P}(U_{n+1} = i) = \frac{1}{n+1}\sum_{i=1}^{n+1} \mathbb{P}(\sigma_n = \sigma(n+1,i)).$$

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

$$\mathbb{P}(\sigma_{n+1} = \sigma) = \frac{1}{n+1}\mathbb{P}(\sigma_n = \sigma(n+1, i_\sigma)) = \frac{1}{n+1}\frac{1}{n!} = \frac{1}{(n+1)!}.$$

□

Remarque 8.2 (Loi uniforme sur les dérangements)**.** *Une permutation $\sigma \in \mathcal{S}_n$ sans points fixes est appelée* dérangement [3] *: $\sigma(i) \neq i$ pour tout $1 \leqslant i \leqslant n$. Soit $\mathcal{D}_n \subset \mathcal{S}_n$ l'ensemble des dérangements. À présent, si σ est une permutation aléatoire de loi uniforme sur $\mathcal{S}_n$, alors il vient $\{\sigma \notin \mathcal{D}_n\} = \cup_{i=1}^n A_i$ où $A_i = \{\sigma(i) = i\}$, et donc, grâce au* principe d'inclusion-exclusion [4]*,*

$$\mathbb{P}(\sigma \in \mathcal{D}_n) = 1 - \mathbb{P}(\cup_{1\leqslant i\leqslant n} A_i) = 1 - \sum_{p=1}^{n}(-1)^{p+1}\sum_{1\leq i_1<\cdots<i_p\leq n}\mathbb{P}(A_{i_1}\cap\cdots\cap A_{i_p}).$$

Or pour tout $1 \leqslant p \leqslant n$,

$$\sum_{1\leqslant i_1<\cdots i_p\leqslant n}\mathbb{P}(A_{i_1}\cap\cdots\cap A_{i_p}) = \sum_{1\leqslant i_1<\cdots<i_p\leqslant n}\frac{(n-p)!}{n!} = \binom{n}{p}\frac{(n-p)!}{n!} = \frac{1}{p!}$$

d'où $\mathbb{P}(\sigma \in \mathcal{D}_n) = 1 - \sum_{p=1}^n \frac{(-1)^{p+1}}{p!} \to e^{-1} \approx 0.37$. Donc [5] $\mathrm{card}(\mathcal{D}_n)/\mathrm{card}(\mathcal{S}_n) \approx 0.37$ *quand $n \gg 1$. La loi uniforme sur $\mathcal{D}_n$ peut être simulée par la méthode du rejet en utilisant des permutations aléatoires i.i.d. de loi uniforme sur $\mathcal{S}_n$ obtenues par l'algorithme de Fisher-Yates-Knuth. La complexité suit une loi géométrique, de moyenne $e \times (n-1) \approx 2.72 \times n$ quand $n \gg 1$. On peut améliorer la performance en stoppant à chaque étape de proposition l'algorithme de Fisher-Yates-Knuth dès qu'un point fixe apparaît.*

Remarque 8.3 (Loi uniforme sur les appariements)**.** *Un* appariement [6] *de $2n$ points est une partition de $\{1,\dots,2n\}$ en n parties de cardinal 2, chacune constituant un «couple de points appariés». L'ensemble $\mathcal{A}_{2n}$ des appariements de $2n$ points a pour cardinal*

$$\mathrm{card}(\mathcal{A}_{2n}) = (2n-1)!! = \prod_{k=1}^{n-1}(2k+1) = \frac{(2n)!}{2n!}.$$

L'algorithme suivant permet de simuler la loi uniforme sur $\mathcal{A}_{2n}$: on partitionne aléatoirement $(1,\dots,2n)$ en deux vecteurs V et V' de taille n, par exemple en utilisant $2n$ réalisations i.i.d. de loi de Bernoulli de paramètre $1/2$, ensuite on permute aléatoirement et uniformément les éléments de V avec l'algorithme de Fisher-Yates-Knuth, et enfin on apparie V_i et V'_i pour tout $1 \leqslant i \leqslant n$. L'ensemble $\mathcal{A}_{2n}$ est également en bijection avec l'ensemble des dérangements involutifs de $\{1,\dots,2n\}$ (permutations sans points fixes et qui sont leur propre inverse).

3. Un point fixe de la permutation σ est un cycle de longueur 1.
4. ou *crible de Poincaré*.
5. On note parfois $!n = \mathrm{card}(\mathcal{D}_n)$, et on a $!(n+1) = (n+1)\times !n + (-1)^{n+1}$, analogue de $(n+1)! = (n+1)\times n!$.
6. «*Matchings*» en anglais.

Recueil de modèles stochastiques – Page 85/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

La décomposition en cycles d'une permutation aléatoire de loi uniforme sur $\mathcal{S}_n$ fournit une partition aléatoire de $\{1,\dots,n\}$. La loi de cette partition n'est pas uniforme sur l'ensemble des partitions Π_n de $\{1,\dots,n\}$ (remarque 9.6). Intéressons-nous à la simulation de la loi uniforme sur Π_n. Cette loi affecte le même poids $1/B_n$ à chaque élément de Π_n, où $B_n = \mathrm{card}(\Pi_n)$. En combinatoire, la suite $(B_n)_{n\geqslant 1}$ constitue les *nombres 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 la formule $G(X) = \exp(\exp(X) - 1)$. 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 :

$$B_n = \frac{1}{e} \sum_{k=1}^{\infty} \frac{k^n}{k!}.$$

Cette formule intervient dans un algorithme de simulation de la loi uniforme sur Π_n.

Théorème 8.4 (Algorithme de Stam)**.** *Soit K un entier aléatoire valant k avec probabilité $k^n/(k!eB_n)$ pour tout $k \geqslant 0$. 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,

$$\mathbb{P}(P = p) = \sum_{k=b}^{\infty} \mathbb{P}(P = p | K = k)\mathbb{P}(K = k) = \sum_{k=b}^{\infty} \frac{k(k-1)\cdots(k-b+1)}{k^n} \frac{k^n}{k!eB_n} = \frac{1}{B_n}.$$

□

8.3 Graphes aléatoires

Un *graphe fini* $G = (V, E)$ est constitué d'un ensemble fini et non vide de sommets[7] V, et d'un ensemble d'arêtes[8] $E \subset \{\{i,j\} : i,j \in V, i \neq j\}$. On note parfois $i \sim j$ pour signifier que $\{i,j\} \in E$, et $i \not\sim j$ lorsque $\{i,j\} \notin E$. Il existe au plus une arête entre deux sommets distincts, et aucune entre un sommet et lui même (absence de boucles). Les arêtes ne sont pas orientées. On peut toujours numéroter les sommets et poser $V = \{1,\dots,n\}$ où $n = \mathrm{card}(V)$. La *matrice d'adjacence* A de G est alors la matrice symétrique $n \times n$ définie par $A_{j,k} = \mathbb{1}_{\{j,k\}\in E}$.

Théorème 8.5 (Loi uniforme sur les graphes finis)**.** *Pour tout entier $n \geqslant 1$, la loi uniforme sur l'ensemble $\mathcal{G}_n$ des graphes à n sommets est obtenue en rendant les $\binom{n}{2} = \frac{1}{2}n(n-1)$ arêtes indépendantes et identiquement distribuées de loi de Bernoulli $(1-p)\delta_0 + p\delta_1$ avec $p = \frac{1}{2}$.*

Il s'agit d'un *graphe aléatoire de Erdős-Rényi* de taille n et de paramètre $p = \frac{1}{2}$.

7. On dit aussi *sites* en français, et en anglais «*vertices*», pluriel de «*vertex*», d'où la notation V.
8. On dit aussi *liens* en français, et en anglais «*edges*», d'où la notation E.

Recueil de modèles stochastiques – Page 86/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Démonstration. L'ensemble $\mathcal{G}_n$ est en bijection avec l'ensemble des matrices $n \times n$ symétriques à coefficients dans $\{0,1\}$ et à diagonale nulle, lui même en bijection avec l'ensemble produit $\{0,1\}^{n(n-1)/2}$. Or la loi uniforme sur un ensemble produit est le produit des lois uniformes sur les facteurs, et la loi uniforme sur $\{0,1\}$ est la loi de Bernoulli $(1-p)\delta_0 + p\delta_1$ avec $p = \frac{1}{2}$. □

Dans un graphe fini $G = (V, E)$, le *degré d'un sommet* $i \in V$ est le nombre noté $d(i)$ de sommets reliés à i directement par une arête : $d(i) = \mathrm{card}\{j \in V : \{i,j\} \in E\}$. On dit que G est un *graphe d-régulier*, où $d \geqslant 0$ est un entier fixé, lorsque $d(i) = d$ pour tout $i \in V$. Les graphes 0-réguliers sont constitués de sommets isolés et ne comportent aucune arête. Les graphes 1-réguliers sont constitués d'arêtes déconnectées les unes des autres. Les graphes finis 2-réguliers sont constitués de cycles déconnectés les uns des autres.

Si G est un graphe à n sommets, numérotés de 1 à n de sorte que $d_1 \geqslant \cdots \geqslant d_n$ où $d_i = d(i)$ pour tout $1 \leqslant i \leqslant n$, alors les deux propriétés suivantes ont lieu :

1. l'entier $d_1 + \cdots + d_n$ est pair ;
2. pour tout $1 \leqslant k \leqslant n$, $d_1 + \cdots + d_k \leqslant k(k-1) + \min(k, d_{k+1}) + \cdots + \min(k, d_n)$.

La parité de la somme vient du fait que chaque arête compte deux fois, tandis que la quantité $k(k-1) + \min(d_{k+1}, k) + \cdots + \min(d_n, k)$ est la contribution maximale à $d_1 + \cdots + d_k$ des arêtes liées aux sommets 1 à k : au plus $\binom{k}{2} = k(k-1)/2$ arêtes (comptent double) entre les sommets 1 à k, et au plus $\min(k, d_{k+i})$ arêtes entre le sommet $i > k$ et les sommets 1 à k. Un *théorème de Erdős-Gallai* affirme que pour tous $d_1 \geqslant \cdots \geqslant d_n \geqslant 0$, il existe un graphe à $n \geqslant 1$ sommets de degrés $d_1, \ldots, d_n$ si et seulement si les deux conditions ci-dessus sont vérifiés. On dit que $d_1, \ldots, d_n$ est la *suite de degrés*[9] du graphe.

Les *multigraphes* sont obtenus à partir de la définition des graphes en relaxant deux contraintes : on accepte les arêtes multiples entre sommets ainsi que les boucles.

Théorème 8.6 (Algorithme des configurations de Bollobás)**.** *Soient $d_1 \geqslant \cdots \geqslant d_n \geqslant 0$ des entiers vérifiant les conditions de Erdős-Gallai, et $2r := d_1 + \cdots + d_n$. Soit C_n le multigraphe aléatoire à n sommets de degrés prescrits $d_1, \ldots, d_n$ construit comme suit :*

— *pour tout $1 \leqslant k \leqslant n$, on dispose d_k demi-arêtes sur le sommet k, soit au total $2r$ demi-arêtes numérotées arbitrairement de 1 à $2r$;*

— *on associe aléatoirement et uniformément la première demi-arête à l'une des $2r-1$ autres demi-arêtes, ce qui produit une arête, et on itère le procédé jusqu'à épuisement.*

On utilise alors l'algorithme du rejet pour obtenir un graphe à partir du multigraphe C_n : on répète de manière indépendante et identiquement distribuée l'algorithme précédent jusqu'à ce que C_n soit un graphe (ni arêtes multiples ni boucles). Cela se produit au bout d'un temps aléatoire de loi géométrique fini presque sûrement, et le graphe aléatoire ainsi obtenu suit la loi uniforme sur l'ensemble $\mathcal{G}_{n,(d_1,\ldots,d_n)}$ des graphes à n sommets de degrés prescrits $d_1, \ldots, d_n$.

On dit que le multigraphe aléatoire C_n est le *modèle des configurations* à degrés prescrits.

Démonstration. Tout d'abord, la probabilité que C_n soit un graphe n'est pas nulle car d'une part tout multigraphe et donc en particulier tout graphe à n sommets de degré $d_1, \ldots, d_n$ est probable, et car d'autre part $d_1, \ldots, d_n$ vérifie les conditions de Erdős-Gallai. Ceci montre que l'algorithme du rejet stoppera au bout d'un temps géométrique fini presque sûrement.

Ensuite, les graphes sont tous équiprobables pour la loi de C_n et donc si on conditionne la loi de C_n par le fait que le multigraphe est un graphe, alors on obtient la loi uniforme sur les graphes de suite de degrés $d_1, \ldots, d_n$. On peut alors appliquer la méthode du rejet. Un algorithme pour l'étape d'*appariement* des demi-arêtes figure dans la remarque 8.3.

Notons que les multigraphes qui ne sont pas des graphes ne sont pas équiprobables pour la loi de C_n : deux multigraphes ne différant que par le nombre d'arêtes multiples entre deux sommets précis n'ont pas la même probabilité d'apparaître car ces arêtes sont indistinguables (facteur $j!$ si arête multiple de multiplicité j). Idem pour les boucles. □

9. «*degree sequence*» en anglais.

Recueil de modèles stochastiques – Page 87/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

8.4 Algorithme de Metropolis-Hastings

Considérons le problème de la simulation d'une loi μ sur un ensemble au plus dénombrable E chargeant tous les éléments. C'est typiquement le cas lorsque μ est une mesure de Boltzmann [10], c'est-à-dire que pour tout $x \in E$,

$$\mu(x) = \frac{e^{-\beta H(x)}}{Z} \quad \text{où} \quad Z = \sum_{x \in E} e^{-\beta H(x)},$$

où $\beta > 0$ est un paramètre réel et où $H : E \to \mathbb{R}$ est une fonction qui attribue à chaque état une « énergie ». En pratique, il est souvent difficile voire impossible d'évaluer la constante de normalisation Z (penser au cas où E est le groupe symétrique !). En revanche, le rapport

$$\frac{\mu(x)}{\mu(y)} = e^{\beta H(y) - \beta H(x)}$$

ne dépend que de la différence d'énergie $H(y) - H(x)$, dont l'évaluation en pratique est même souvent moins coûteuse que l'évaluation de $H(x)$ et $H(y)$ séparément. L'algorithme de Metropolis-Hastings permet de construire et de simuler une chaîne de Markov récurrente apériodique $(X_t)_{t \in \mathbb{N}}$ sur E de loi invariante μ, qui ne fait intervenir μ qu'à travers les rapports $\mu(x)/\mu(y)$. On dispose alors d'un générateur approché [11] de μ, car comme $X_t \to \mu$ en loi, on a

$$\mathrm{Loi}(X_t) \approx \mu \text{ pour } t \gg 1.$$

Pour construire le noyau $\mathbf{P}$ de $(X_t)_{t \in \mathbb{N}}$, on se donne tout d'abord un noyau de transition irréductible $\mathbf{Q}$ sur E vérifiant $\mathbf{Q}(x, y) = 0$ ssi $\mathbf{Q}(y, x) = 0$. Si μ est réversible pour $\mathbf{Q}$, c'est-à-dire que $\mu(x)\mathbf{Q}(x, y) = \mu(y)\mathbf{Q}(y, x)$ pour tous $x, y \in E$, alors on peut prendre $\mathbf{P} = (1 - \varepsilon)\mathbf{Q} + \varepsilon \mathbf{I}$ avec $\varepsilon \in [0, 1]$, le paramètre ε permettant de forcer l'apériodicité si nécessaire. Si μ n'est pas réversible pour $\mathbf{Q}$, alors on construit le noyau $\mathbf{P}$ en posant, pour tous $x, y \in E$,

$$\mathbf{P}(x, y) = \begin{cases} \mathbf{Q}(x, y)\rho(x, y) & \text{si } x \neq y, \\ 1 - \sum_{z \neq x} \mathbf{P}(x, z) & \text{si } x = y, \end{cases} \quad \text{où} \quad \rho(x, y) = \alpha\left(\frac{\mu(y)\mathbf{Q}(y, x)}{\mu(x)\mathbf{Q}(x, y)}\right) \mathbb{1}_{\mathbf{Q}(x,y) > 0},$$

et où $\alpha : \mathbb{R} \to]0, 1]$ est une fonction fixée d'avance qui vérifie $\alpha(u) = u\alpha(1/u)$ pour tout réel $u \in \mathbb{R}$, comme par exemple $\alpha(u) = \min(1, u)$ ou encore $\alpha(u) = u/(1 + u)$.

Ainsi $\mathbf{P} = \mathbf{Q}$ si μ est réversible pour $\mathbf{Q}$. Notons que si μ est la loi uniforme, ce qui impose que E est fini, et si $\mathbf{Q}(x, y) = \mathbf{Q}(y, x)$ pour tous $x, y \in E$, alors μ est réversible pour $\mathbf{Q}$.

Théorème 8.7 (Metropolis-Hastings)**.** $\mathbf{P}$ *est un noyau de transition irréductible récurrent positif, de loi invariante réversible* μ*, apériodique si* $h < 1$ *ou si* $\mathbf{Q}$ *est apériodique.*

Démonstration. On a $0 \leqslant \mathbf{P}(x, y) \leqslant \mathbf{Q}(x, y)$ pour tous $x \neq y$ donc $\mathbf{P}$ est un noyau de transition. Comme il hérite du squelette de $\mathbf{Q}$, il est irréductible, et apériodique si $\mathbf{Q}$ l'est. Si $h < 1$ alors $\mathbf{P}(x, x) > 0$ pour tout $x \in E$ et donc $\mathbf{P}$ est apériodique. La propriété de α donne $\mu(x)\mathbf{P}(x, y) = \mu(y)\mathbf{P}(y, x)$ si $x \neq y \in E$ et donc μ est réversible, donc invariante. Comme $\mathbf{P}$ est irréductible, il est récurrent positif car il possède une loi invariante, qui est unique. □

L'algorithme de Metropolis-Hastings consiste à simuler les trajectoires de la chaîne $(X_t)_{t \in \mathbb{N}}$ de noyau $\mathbf{P}$. Cela se ramène au problème de la simulation de la loi discrète $\mathbf{P}(x, \cdot)$ pour un x quelconque. Pour ce faire, soit Y une variable aléatoire sur E de loi $\mathbf{Q}(x, \cdot)$, et U une variable

10. Maximise l'entropie de Boltzmann parmi les lois de même énergie moyenne.

11. Lorsque E est fini, la condition de Doeblin fournit un contrôle de l'erreur. Cependant, on ne dispose pas d'un contrôle de l'erreur utile en toute généralité, sauf dans quelques catégories d'exemples spéciales.

Recueil de modèles stochastiques – Page 88/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

aléatoire de loi uniforme sur $[0,1]$, indépendante de Y. Soit Z la variable aléatoire définie par $Z = Y$ si $U < \rho(x,Y)$ et $Z = x$ sinon. Alors $Z \sim \mathbf{P}(x,\cdot)$ car pour tous $y \neq x$,

$$\mathbb{P}(Z = y) = \mathbb{P}(U < \rho(x,Y), Y = y) = \rho(x,y)\mathbf{Q}(x,y) = \mathbf{P}(x,y).$$

Autrement dit la proposition Y de loi $\mathbf{Q}(x,\cdot)$ est acceptée ou rejetée selon que $U < \rho(x,Y)$ ou pas. L'évaluation de $\mathbf{P}(x,x)$ n'est pas requise. On dit que ρ est la *fonction d'acceptation-rejet* tandis que $\mathbf{Q}$ le *noyau de proposition ou d'exploration*. Le noyau $\mathbf{Q}$ doit être facile à simuler. En pratique, on utilise souvent le noyau de la marche aléatoire simple aux plus proches voisins sur le graphe associé à une distance naturelle sur E :

$$\mathbf{Q}(x,y) = \frac{\mathbb{1}_{y \in V_x}}{|V_x|} \quad \text{où} \quad V_x := \{y \in E : \mathrm{dist}(x,y) = 1\}.$$

Si le graphe est régulier ($|V_x| = |V_y|$ pour tous $x, y \in E$) et si μ est uniforme, alors $\mathbf{P} = \mathbf{Q}$.

Exemple 8.8 (Échantillonneur de Gibbs)**.** *Soit F une partie finie de $\mathbb{Z}$ et μ une loi sur l'espace produit $E = F^r$ chargeant tous les états. L'*échantillonneur de Gibbs *est le nom donné à l'algorithme de Metropolis-Hastings associé à la loi μ avec pour noyau de proposition*

$$\mathbf{Q}(x,y) = \sum_{i=1}^{d} q(i)\mu_{i,x_{-i}}(y_i)\mathbb{1}_{\{x_i = y_i\}},$$

où q est une loi sur $\{1,\ldots,r\}$ chargeant tous les états comme par exemple la loi uniforme, où $x_{-i} = (x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_r)$ et où $\mu_{i,x_{-i}}$ est la loi de x_i sachant x_{-i} sous μ. Si i est une réalisation de loi q et si y_i est une réalisation de loi $\mu_{i,x_{-i}}$ alors $(x_1,\ldots,x_{i-1},y_i,x_{i+1},\ldots,x_r)$ est une réalisation de loi $\mathbf{Q}(x,\cdot)$. L'échantillonneur de Gibbs ramène la simulation de μ à celle des lois conditionnelles $\mu_{i,x(-i)}$. Un exemple éponyme est fourni par le modèle de magnétisme du physicien Ising : $E = \{-1,+1\}^\Lambda$ où Λ est une partie finie de $\mathbb{Z}^d$, identifié à $\{-1,1\}^r$ avec $r = \mathrm{card}(\Lambda)$, et μ est une loi de Boltzmann d'énergie $H(x) = \sum_{|i-j|_1=1} \varepsilon x_i x_j + m\sum_i x_i$. Ce modèle issu de la physique statistique est également utilisé en imagerie.

8.5 Algorithme du recuit simulé

Soit μ la loi de Boltzmann sur E au plus dénombrable de la forme $\mu(x) = Z^{-1}e^{-\beta H(x)}$ comme précédemment. Il est commode d'interpréter le paramètre $1/\beta$ comme une température ou une variance par analogie gaussienne. La loi μ favorise les configurations de faible énergie, d'autant plus que la température $1/\beta$ est basse. On dispose en fait du *principe de Laplace* :

$$\mu_*(x) := \lim_{\beta\to\infty} \mu(x) = \frac{\mathbb{1}_{x \in M}}{|M|} \quad \text{où} \quad M := \left\{y \in E : \mu(y) = \inf_E H\right\}.$$

La loi μ_* est uniforme sur l'ensemble M des minima globaux de H. Cela conduit à un algorithme stochastique pour simuler μ_* appelé *algorithme du recuit simulé*[12], qui consiste à utiliser l'algorithme de Metropolis-Hastings mais avec une température $1/\beta = 1/\beta_t$ décroissante au cours du temps t. La chaîne de Markov de Metropolis-Hastings $(X_t)_{t\in\mathbb{N}}$ ainsi modifiée n'est plus homogène en temps. Considérons par exemple le cas où $\alpha(u) = \min(1,u)$ pour tout $u \in \mathbb{R}$, et où $\mathbf{Q}$ est tel que $\mathbf{Q}(x,y) = \mathbf{Q}(y,x)$ pour tous $x, y \in E$. On a alors

$$\rho(x,y) = \alpha\left(\frac{\mu(y)}{\mu(x)}\right) = \alpha\left(e^{-\beta_n(H(y)-H(x))}\right) = e^{-\beta_n \max(0,H(y)-H(x))}.$$

12. «*Simulated annealing*» en anglais.

Recueil de modèles stochastiques – Page 89/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Sachant qu'on se trouve en l'état x au temps t, on simule une proposition y de loi $\mathbf{Q}(x,\cdot)$, qu'on accepte ssi $U < \rho(x,y)$ où U est une variable aléatoire uniforme indépendante, c'est-à-dire avec probabilité 1 si $H(y) \leqslant H(x)$ et probabilité $e^{-\beta_t(H(y)-H(x))}$ sinon (vaut 0 si $1/\beta_t = 0$). L'algorithme explore le graphe de H, accepte toujours de descendre, et accepte de monter avec une probabilité décroissant vers zéro au fil du temps. Ce mécanisme subtil permet d'échapper, au début, aux minima locaux. Le réglage de la décroissance de la température $1/\beta_t$ est délicat en pratique ! La performance de l'algorithme dépend beaucoup de la régularité du graphe de H par rapport à la structure de voisinage de l'exploration.

Considérons à présent le *problème du voyageur de commerce*[13] pour n villes positionnées dans l'espace en $v_1, \ldots, v_n$, qui consiste à trouver une tournée de longueur minimale, c'est-à-dire une permutation $x \in \mathcal{S}_n$ minimisant (la longueur d'une tournée joue le rôle d'énergie ici)

$$H(x) := \sum_{i=1}^{n} \operatorname{dist}\bigl(v_{x(i)}, v_{x(i+1)}\bigr) \quad \text{où} \quad x(n+1) := x(1).$$

Le cardinal de $E = \mathcal{S}_n$ est énorme : $n!$. À coût total fixé, on ne pourra qu'explorer une partie restreinte $F \subset E$ et retenir le minimum de H sur F comme approximation du minimum de H sur E. La génération d'un $F \subset E$ peut se faire de manière déterministe[14], ou alternativement de manière stochastique, par exemple avec l'algorithme du recuit simulé, avec pour noyau d'exploration $\mathbf{Q}$ celui de la marche aléatoire simple sur $\mathcal{S}_n$ associée aux transpositions :

$$\mathbf{Q}(x,y) = \begin{cases} \frac{2}{n(n-1)} & \text{si } y = x\tau \text{ pour une transposition } \tau, \\ 0 & \text{sinon.} \end{cases}$$

8.6 Algorithme de Propp-Wilson

L'algorithme de Metropolis-Hasting a l'inconvénient de ne pas être exact. Adoptons l'interprétation des chaînes de Markov sous forme de suites récurrentes aléatoires. Soit donc $g : E \times [0,1] \to E$ une fonction telle que $g(x,U) \sim \mathbf{P}(x,\cdot)$ pour tout $x \in E$, où U est une variable aléatoire de loi uniforme sur $[0,1]$. On note $G : E \to E$ la fonction aléatoire définie par $G(x) = g(x,U)$. Soit $(G_n)_{n\geqslant 0}$ 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 $(U_n)_{n\geqslant 0}$ de variables aléatoires i.i.d. de loi uniforme sur $[0,1]$. Pour tout n, on construit les applications aléatoires $A_n : E \to E$ et $B_n : E \to E$ par composition de la manière suivante :

$$A_n = G_n \circ \cdots \circ G_1 \quad \text{et} \quad B_n = G_1 \circ \cdots \circ G_n.$$

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 $(A_n(x))_{n\geqslant 0}$ et $(B_n(x))_{n\geqslant 0}$ ont les mêmes lois marginales de dimension 1, mais n'ont pas la même loi en général. La suite $(A_n(x))_{n\geqslant 0}$ est une chaîne de Markov sur E de noyau $\mathbf{P}$ et de loi initiale δ_x. En revanche, $(B_n(x))_{n\geqslant 0}$ 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

$$T_A = \inf\{n \geqslant 0 : \operatorname{card}(A_n(E)) = 1\} \quad \text{et} \quad T_B = \inf\{n \geqslant 0; \operatorname{card}(B_n(E)) = 1\}.$$

On dit que G est *contractante* lorsque $p = \mathbb{P}(\operatorname{card}(G(E)) = 1) > 0$.

13. L'étude de l'ordre de grandeur de $H(x)$ lorsque $v_1, \ldots, v_n$ sont aléatoires i.i.d. fait l'objet du chapitre 17.

14. L'algorithme (déterministe) de Steinhaus-Johnson-Trotter permet de lister les éléments de $\mathcal{S}_n$ sans comparaisons aux éléments déjà produits. Algébriquement, il correspond à parcourir un graphe de Cayley de $\mathcal{S}_n$, tandis que géométriquement, il correspond à parcourir les sommets adjacents du polytope appelé permutaèdre.

Recueil de modèles stochastiques – Page 90/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

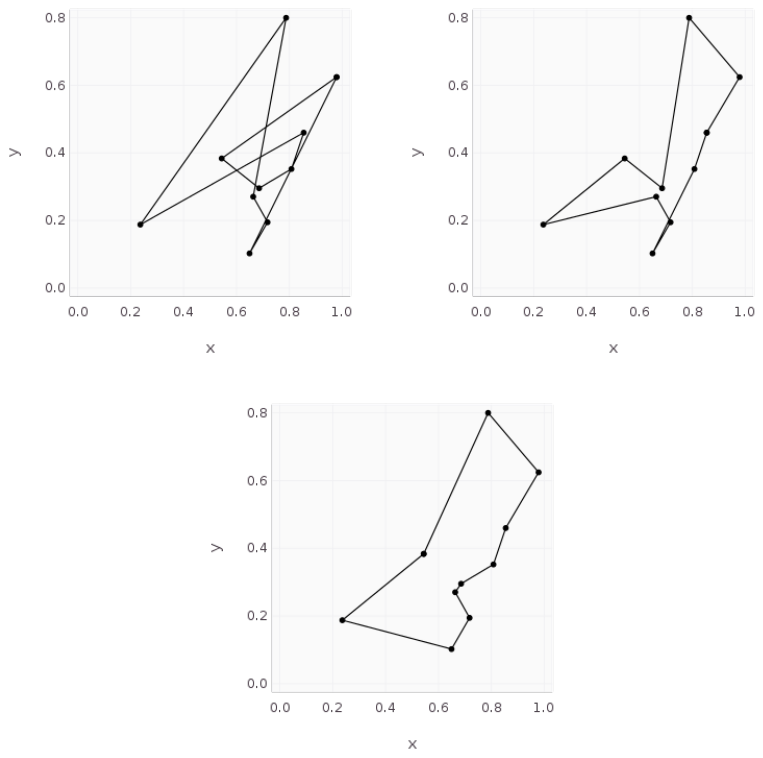


FIGURE 8.1 – Quelques étapes de l'algorithme du recuit pour le voyageur de commerce ($n = 10$).

Théorème 8.9 (Convergence vers μ à la coalescence pour la suite renversée)**.** *Si G est contractante alors $\mathbb{P}(T_B < \infty) = 1$, pour tout $x \in E$, $B_{T_B}(x) \sim \mu$, et $B_n(x) = B_{T_B}(x)$ pour tout $n \geqslant T_B$.*

Notons que $(A_n(x))_{n\geqslant 0}$ converge en loi vers μ tandis que $(B_n(x))_{n\geqslant 0}$ converge p.s. vers μ.

Démonstration. Les événements $C_n = \{\mathrm{card}(G_n(E)) = 1\}$ sont indépendants et de même probabilité $p = \mathbb{P}(\mathrm{card}(G(E)) = 1) > 0$. Par le *lemme de Borel-Cantelli*, presque sûrement, $\mathrm{card}(G_n(E)) = 1$ 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 $n \geqslant T_B$ car

$$B_n(x) = B_{T_B}((G_{T_B+1} \circ \cdots \circ G_n)(x)) = x_B.$$

Il y a *coalescence* des trajectoires de $(B_n(x))_{n\geqslant 0}$ quel que soit l'état initial x. Soit μ_B la loi de x_B. Par convergence dominée, pour tout $x \in E$ et toute fonction bornée $f : E \to \mathbb{R}$,

$$\lim_{n\to\infty} \mathbb{E}[f(B_n(x)] = \lim_{n\to\infty} \mathbb{E}[f(B_n(x))\mathbb{1}_{\{T_B\leqslant n\}}] = \mathbb{E}[f(x_B)] = \mu_B f.$$

Or $\mathbb{E}[f(B_n(x))] = \mathbf{P}^n(x,\cdot)f$, donc $\lim_{n\to\infty} \mathbf{P}^n(x,y) = \mu_B(y)$ pour tous x et y, d'où $\mu_B = \mu$. □

Si G est contractante alors on a aussi $\mathbb{P}(T_A < \infty) = 1$ et T_A définit p.s. un singleton aléatoire $\{x_A\}$, vérifiant $A_{T_A}(x) = x_A$ pour tout $x \in E$, mais rien n'assure que $A_n(x) = x_A$ pour $n > T_A$ et tout $x \in E$, ni que x_A suit la loi μ. Pour tous $x, y \in E$, la suite $((A_n(x), A_n(y)))_{n\geqslant 0}$ est un couplage coalescent : les deux composantes sont des chaînes de

Recueil de modèles stochastiques – Page 91/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

même noyau et qui restent collées après leur première rencontre. On dit que $((B_n(x), B_n(y)))_{n\geqslant 0}$ est un *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 8.10 (Contrôle géométrique du temps de coalescence par monotonie)**.** *Si*
— *E est ordonné (éventuellement par un ordre partiel) ;*
— *E possède un plus petit élément x_* et $\varepsilon = \inf_{x\in E} \mathbf{P}(x, x_*) > 0$;*
— *pour tout $x \in E$ la fonction $g(x, \cdot) : [0, 1] \to E$ est croissante et constante par morceaux ;*
alors $B_n(x) = B_{\sigma-1}(x_*) \sim \mu$ pour tous $x \in E$ et $n \geqslant \sigma$, où $\sigma = \inf\{n \geqslant 1; U_n < \varepsilon\} \sim \mathrm{Geo}(\varepsilon)$.

On dispose donc d'un algorithme de simulation exacte de μ de complexité géométrique. La condition $\inf_{x\in E} \mathbf{P}(x, x_*) > 0$ peut être relaxée en $\inf_{x\in E} \mathbf{P}^r(x, x_*) > 0$ pour un entier $r \geqslant 1$, en considérant la chaîne $(X_{nr})_{n\geqslant 0}$ de noyau $\mathbf{P}^r$, également de loi invariante μ.

Démonstration. On a $\mathbb{P}(G(E) = \{x_*\}) \geqslant \varepsilon > 0$ et donc G est contractante. La mesure de Lebesgue du premier morceau de $g(x, \cdot)$ vaut $\mathbf{P}(x, x_*)$. Par conséquent, $g(x, u) = x_*$ pour $u < \varepsilon$ et tout $x \in E$. 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 $n \geqslant \sigma$. On procède alors comme dans la preuve du théorème 8.9. □

Exemple 8.11 (Propp-Wilson pour une mesure de Boltzmann)**.** *On modélise un substrat avec $\Lambda := \{0, 1, \ldots, L\}^2 \subset \mathbb{Z}^2$, où chaque site $i \in \Lambda$ est soit vierge, soit occupé par un composé chimique de type $+1$ ou -1. Les deux composés se repoussent, et cela conduit à considérer*

$$E = \{x \in \{-1, 0, +1\}^\Lambda : \forall i, j \in \Lambda, |i - j|_1 = 1 \Rightarrow x_i x_j \neq -1\}.$$

On considère la loi de Boltzmann μ sur E définie par $\mu(x) = Z^{-1} \exp(-\beta H(x))$ pour tout $x \in E$, où $\beta \in \mathbb{R}$, où $Z = \sum_{x\in E} e^{-\beta H(x)}$, et où $H(x) = \mathrm{card}\{i \in \Lambda : x_i = +1\}$. Cette loi favorise le composé $+1$ si $\beta > 0$ et le composé -1 si $\beta < 0$. Lorsque $\beta = 0$, la loi μ est uniforme sur E. Bien que E soit fini, la simulation de μ par la méthode du rejet est coûteuse voire impraticable. L'algorithme de Metropolis-Hasting est en revanche praticable, avec par exemple un noyau d'exploration $\mathbf{Q}$ vérifiant $\mathbf{Q}(x, y) \neq 0$ ssi x et y ne diffèrent qu'en un seul site. Examinons comment on pourrait mettre en œuvre l'algorithme de Propp-Wilson. Pour tous $u \in [-1, 1]$, $i \in \{-1, 0, +1\}$, $x \in E$, on note $h_u^i(x) \in E$ la configuration obtenue à partir de x en forçant le site i à -1 si $u < -1/(1 + 2e^{-\beta})$, à $+1$ si $u > 1/(1 + 2e^{-\beta})$, et à 0 sinon. Soient $(U_n)_{n\geqslant 1}$ et $(I_n)_{n\geqslant 1}$ des suites indépendantes de v.a. i.i.d. de loi uniforme sur $[-1, 1]$ et sur Λ respectivement, et $G_n := h_{U_{n+1}}^{I_{n+1}}$. La suite récurrente aléatoire $(X_n)_{n\in\mathbb{N}}$ définie par

$$X_{n+1} = G_n(X_n)$$

est alors une chaîne de Markov sur E, irréductible, de loi invariante μ. De plus, pour tout $u \in [-1, 1]$ et tout $i \in \Lambda$, la fonction $x \in E \mapsto h_u^i(x)$ est croissante pour l'ordre (partiel) sur E défini par $x \leqslant y$ ssi $x_i \leqslant y_i$, ce qui permet d'utiliser le théorème 8.10 avec $x_* = -1$. En fait, on dispose également d'un plus grand élément $x^* = 1$, et la monotonie fournit l'encadrement

$$G_n(-1) \leqslant G_n(x) \leqslant G_n(1),$$

indiquant que la coalescence partant de x quelconque a lieu avant la coalescence partant de ± 1.

8.7 Notes et commentaires

Une analyse probabiliste de la complexité de l'*algorithme de tri rapide randomisé* se trouve par exemple dans le livre de Rajeev Motwani et Prabhakar Raghavan [MR95]. Le livre

Recueil de modèles stochastiques – Page 92/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

monumental de Donald 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 Ronald Fisher et Frank Yates [FY48]. Il est implémenté en standard dans les logiciels de calcul. Cet algorithme correspond exactement au processus des restaurants chinois du chapitre 9, et la loi uniforme sur $\mathcal{S}_n$ coïncide avec la loi d'Ewens de paramètre $\theta = 1$.

Pour tout $n \geqslant 1$ fixé, la marche aléatoire sur le groupe symétrique $\mathcal{S}_n$ dont les pas sont i.i.d. de loi uniforme sur l'ensemble des transpositions, étudiée par Persi Diaconis et Mehrdad 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 2. 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 des 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 un article de Persi Diaconis, Eddy Mayer-Wolf, Ofer Zeitouni, et Martin Zerner [DMWZZ04]. Plus généralement, au delà des transpositions aléatoires, Nathanaël Berestycki, Oded Schramm, et Ofer Zeitouni ont établi dans [BSZ11] que pour tout pour tous $n \geqslant k \geqslant 2$, 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 Aart Stam de simulation de la loi uniforme sur les partitions se trouve dans [Sta83] et dans le livre de Knuth [Knu05, Volume 4A]. 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 Alfred Young ou de Norman Ferrers. Le théorème de Paul Erdős et Tibor Gallai figure dans [EG60], et a été redémontré par plusieurs auteurs, dont Claude Berge [Ber76]. Une preuve courte et constructive se trouve par exemple dans un article de Amitabha Tripathi, Sushmita Venugopalan, et Douglas West [TVW10]. L'algorithme des configurations remonte à Béla Bollobás [Bol80], et a été raffiné et étendu notamment par Brendan McKay et Nicholas Wormald [MW90]. Il est abordé dans les cours de Charles Bordenave [Bor14a] et de Remco van der Hofstad [vdH14]. À ce sujet, soit C_n est le multigraphe aléatoire du modèle des configurations à n sommets pour la suite de degrés $d_{n,1},\dots,d_{n,n}$, et $p_{n,k} = (1/n)\sum_{i=1}^n \mathbb{1}_{\{d_{n,i}=k\}}$ la proportion de sommets de degré k, et supposons qu'il existe une loi $(p_k)_{k\geqslant 1}$ telle que $\lim_{n\to\infty} p_{n,k} = p_k$, et telle que les deux premiers moments convergent :

$$\mu = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n d_{n,i} = \sum_{k=1}^\infty k p_k < \infty \quad \text{et} \quad \nu = \lim_{n\to\infty} \sum_{i=1}^n \frac{d_{n,i}(d_{n,i}-1)}{\sum_{j=1}^n d_{n,j}} = \frac{1}{\mu}\sum_{k=1}^\infty k(k-1)p_k < \infty,$$

alors on peut établir que la probabilité que C_n soit un graphe tend vers $e^{-\frac{1}{2}\nu-\frac{1}{4}\nu^2}$ quand $n \to \infty$, ce qui signifie que la méthode du rejet du théorème 8.6 reste raisonnable si $n \gg 1$.

La partie sur les algorithmes de Metropolis-Hastings, du recuit simulé, et de Propp-Wilson est inspirée d'un livre précédent [BC07, Chapitre 4] et du cours de Thierry Bodineau à l'École Polytechnique [Bod14]. L'algorithme de Metropolis-Hastings a été introduit par Nicholas Metropolis [MRR+53] puis généralisé par W. Keith 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. Les méthodes MCMC sont au cœur des implémentations quantitatives de la statistique bayésienne abordée par

Recueil de modèles stochastiques – Page 93/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

exemple dans le livre de Christian Robert et George Casella [RC04]. Des version à particules permettent d'améliorer les performances pratiques notamment lorsque la loi μ est multimodale.

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 s'en inspire, ce qui explique son nom. Le recuit simulé converge théoriquement en temps infini lorsque la température suit un schéma de décroissance par paliers logarithmiques, mais ce résultat d'analyse asymptotique n'est pas vraiment pertinent en pratique. Une analyse de l'algorithme se trouve par exemple dans le livre de Étienne Pardoux [Par07] et dans le panorama de Olivier Catoni [Cat99]. Des versions améliorées du recuit simulé, comme l'algorithme de Wang-Landau introduit par Fugao Wang et David Landau [WL01], constituent un sujet de recherche actuel. Les algorithmes de Metropolis-Hastings, du recuit simulé, et de Wang-Landau sont tous disponibles dans des cadres à temps et espace continus. L'algorithme de Propp-Wilson, proposé par James Propp et David Wilson [PW96, PW98], est présenté dans le livre de David Levin, Yuval Peres, et Elizabeth Wilmer [LPW09], et dans celui de Olle Häggström [Häg02]. L'exemple 8.11 est inspiré d'un texte de Christophe Sabot. D'autres algorithmes de simulation exacte de la loi invariante ont été développés, comme par exemple l'algorithme de James Fill [Fil98].

Recueil de modèles stochastiques – Page 94/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Chapitre 9

Restaurants chinois

Mots-clés. Partition aléatoire ; permutation aléatoire ; groupe symétrique ; combinatoire.

Pour tout entier $n \geqslant 1$, on note $\mathcal{S}_n$ l'ensemble des permutations de $\{1, \ldots, n\}$ (groupe symétrique). Le processus des restaurants chinois est une chaîne de Markov inhomogène $(\sigma_n)_{n\geqslant 1}$, d'espace d'états $\cup_{n\geqslant 1}\mathcal{S}_n$, où σ_n est à valeurs dans $\mathcal{S}_n$ pour tout $n \geqslant 1$, 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

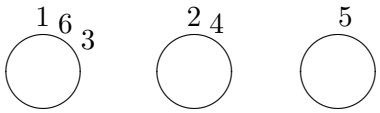
$$\mathbb{P}(\sigma_{n+1} = \sigma' \,|\, \sigma_n = \sigma) = \begin{cases} \dfrac{1}{\theta + n} & \text{si } \sigma' \text{ s'obtient en insérant } n+1 \text{ dans l'un des cycles de } \sigma; \\[2ex] \dfrac{\theta}{\theta + n} & \text{si } \sigma' \text{ s'obtient en ajoutant le cycle } (n+1) \text{ à } \sigma; \\[2ex] 0 & \text{sinon;} \end{cases}$$

où $\theta \geqslant 0$ 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 $\sigma_1, \ldots, \sigma_n$, 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 $\{1, \ldots, n\}$ 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 $n \geqslant 1$, π_n est à valeurs dans l'ensemble Π_n des partitions de $\{1, \ldots, n\}$. En général, l'image d'une chaîne de Markov par une fonction n'est pas une chaîne de Markov [1]. Cependant, il se trouve ici que $(\pi_n)_{n\geqslant 1}$ est une chaîne de Markov inhomogène sur $\cup_{n\geqslant 1}\Pi_n$, de

1. Un critère dû à Dynkin fournit une condition suffisante sur la fonction.

Recueil de modèles stochastiques – Page 95/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.



$$\sigma_6 = (1,6,3)(2,4)(5) = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 6 & 4 & 1 & 2 & 5 & 3 \end{pmatrix} \text{ et } \pi_6 = \{\{1,6,3\},\{2,4\},\{5\}\}$$

Figure 9.1 – Configuration avec $n = 6$ clients répartis sur 3 tables dans le restaurant chinois.

Recueil de modèles stochastiques – Page 96/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

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

$$\mathbb{P}(\pi_{n+1} = \pi' \,|\, \pi_n = \pi) = \begin{cases} \dfrac{|b|}{\theta + n} & \text{si } \pi' \text{ s'obtient en insérant } n+1 \text{ dans le bloc } b \text{ de } \pi; \\[2ex] \dfrac{\theta}{\theta + n} & \text{si } \pi' \text{ s'obtient en ajoutant le bloc singleton } \{n+1\} \text{ à } \pi; \\[2ex] 0 & \text{sinon;} \end{cases}$$

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 9.1 (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 9.2 (Remonter le temps)**.** *Voici un exemple de début de trajectoire $(\pi_n)_{n \geqslant 1}$:*

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

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 $n \geqslant 1$ et tout $\pi' \in \Pi_{n+1}$, il existe un unique $\pi \in \Pi_n$ tel que $\mathbb{P}(\pi_{n+1} = \pi' \,|\, \pi_n = \pi) > 0$. 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).

9.1 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 $n \geqslant 1$.

Théorème 9.3 (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 $\mathbb{P}(\sigma_{n+1} = \sigma' | \sigma_n = \sigma) > 0$. Si $|\sigma'| = |\sigma| + 1$ alors $\mathbb{P}(\sigma_{n+1} = \sigma' | \sigma_n = \sigma) = \theta/(\theta + n)$ et si $|\sigma'| = |\sigma|$ alors $\mathbb{P}(\sigma_{n+1} = \sigma' | \sigma_n = \sigma) = 1/(\theta + n)$. Dans les deux cas, on a bien

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

□

Remarque 9.4 (Loi d'Ewens, marche aléatoire, transpositions, et loi uniforme sur $\mathcal{S}_n$)**.** *Conditionnellement à l'événement $\{\sigma_n = \sigma\}$, on peut construire σ_{n+1} à partir de σ en choisissant d'abord un élément aléatoire K de $\{1, \ldots, n+1\}$ valant $n+1$ avec probabilité $\theta/(\theta+n)$ et $1, \ldots, n$ avec probabilité $1/(\theta+n)$ (si $\theta = 1$ alors K est uniforme sur $\{1, \ldots, n+1\}$), puis en ajoutant $n+1$ au cycle de σ contenant K si $K \leqslant n$ ou en créant un nouveau cycle $(n+1)$ si $K = n+1$. Ceci revient à fabriquer à partir de σ un élément σ' de $\mathcal{S}_{n+1}$ en ajoutant à σ le cycle $(n+1)$, puis à calculer le produit $\sigma'\tau$ dans $\mathcal{S}_{n+1}$ où τ est la transposition aléatoire $(n+1, K)$. Ainsi*

$$(\sigma_n)_{n\geqslant 1} \overset{\text{loi}}{=} ((1, K_1) \cdots (n, K_n))_{n\geqslant 1}$$

où $(K_n)_{n\geqslant 1}$ est une suite de variables aléatoires indépendantes avec $K_1 = 1$ et pour tout $n \geqslant 1$, K_{n+1} de loi $\frac{1}{\theta+n}\delta_1 + \cdots + \frac{1}{\theta+n}\delta_n + \frac{\theta}{\theta+n}\delta_{n+1}$, et $(\sigma_n)_{n\geqslant 0}$ est une marche aléatoire sur $\mathcal{S}_n$, non homogène en temps, sur le graphe de Cayley de $\mathcal{S}_n$ engendré par les transpositions. Si $\theta = 1$ alors pour tout $n \geqslant 1$, σ_n suit la loi uniforme sur $\mathcal{S}_n$ (algorithme de Fisher-Yates-Knuth, chapitre 8).

Théorème 9.5 (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 9.3 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 $\sum_{\sigma\in E_\pi} 1 = \prod_{b\in\pi}(|b| - 1)!$. □

Remarque 9.6 (Loi d'Ewens et loi uniforme sur Π_n)**.** *Pour $n \geqslant 3$, quelque soit θ, la loi d'Ewens sur Π_n n'est jamais la loi uniforme étudiée dans le chapitre 8. En effet, notons*

$$\pi = \{\{1\}, \{2\}, \ldots, \{n\}\}, \quad \pi' = \{\{1, 2\}, \{3\}, \ldots, \{n\}\}, \quad \textit{et} \quad \pi'' = \{\{1, 2, \ldots, n\}\}.$$

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 $n \geqslant 1$ et tout $1 \leqslant k \leqslant n$, 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 9.7 (Loi d'Ewens sur les blocs)**.** *Pour tout $n \in \mathbb{N}^*$ et tout $(a_1, \ldots, a_n) \in \mathbb{N}$ vérifiant $a_1 + 2a_2 + \cdots + na_n = n$, on a*

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

Recueil de modèles stochastiques – Page 97/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Démonstration. Le théorème 9.5 donne

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

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

Remarque 9.8. *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 !*

9.2 Nombre de tables

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

Théorème 9.9 (Nombre de tables)**.** *Pour tout $n \geqslant 1$ on a*

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

et

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

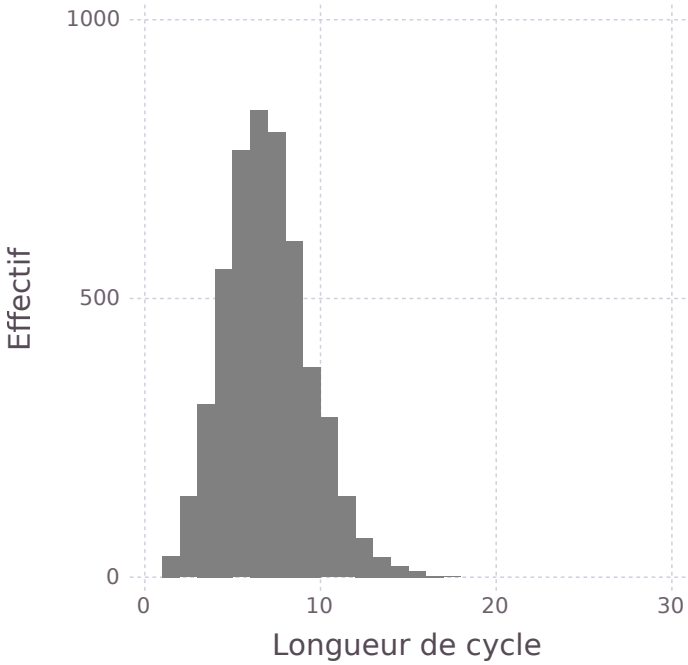


Figure 9.2 – Histogramme d'un échantillon de taille 5000 de la loi d'Ewens de taille $n = 1000$ et de paramètre $\theta = 1$.

Démonstration. Comme $(\pi_n)_{n\geqslant 1}$ 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 $(\varepsilon_n)_{n\geqslant 1}$ sont i.i.d. de loi uniforme sur $[0,1]$. La structure de la fonction f entraîne alors qu'il existe une suite $(\xi_k)_{k\geqslant 1}$ de variables aléatoires indépendantes de lois de Bernoulli, vérifiant pour tout $k \geqslant 1$

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

Recueil de modèles stochastiques – Page 98/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

et telle que que la variable aléatoire $|\pi_n|$ a même loi que $\xi_1 + \cdots + \xi_n$ pour tout $n \geqslant 1$. 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 $\mathbb{V}(|\pi_n|)$ les formules en sommes. Une comparaison série-intégrale[2] montre que ces sommes sont des $\theta \log(n) + \mathcal{O}_{n\to\infty}(1) \sim_{n\to\infty} \theta \log(n)$. □

Remarque 9.10 (Un jeu de pile ou face inhomogène)**.** *Par construction,* $(|\pi_n|)_{n\geqslant 1}$ *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* $(\pi_n)_{n\geqslant 1}$ *soit croissante, elle est cependant de moins en moins croissante en quelque sorte puisque la quantité*

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

décroît 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 9.11 (Comportement asymptotique du nombre de tables en $\mathbb{P}$)**.** *On a*

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

Démonstration. Par l'inégalité de Markov et le théorème 9.9, pour tous $\varepsilon > 0$ et $n \geqslant 2$,

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

□

Le résultat suivant affirme que la convergence a lieu presque sûrement. La borne $\mathcal{O}(1/\log(n))$ obtenue par la méthode du second moment ci-dessus n'est pas sommable : la convergence presque sûre ne peut donc pas être déduite d'une simple application du *lemme de Borel-Cantelli*.

Théorème 9.12 (Comportement asymptotique presque sûr du nombre de tables)**.** *On a*

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

Démonstration. Soient $(\xi_n)_{n\geqslant 1}$ comme dans la preuve du théorème 9.9 de sorte que

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

Notons que $\xi_1 = 1$ p.s. et donc $\xi_1 - \mathbb{E}(\xi_1) = 0$. Pour tout $n \geqslant 2$, 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 $n \to \infty$ (théorème 9.9). Il suffit donc d'établir que le premier terme du membre de droite converge p.s. vers 0. Or si

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

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

Recueil de modèles stochastiques – Page 99/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

alors $(S_n)_{n\geqslant 2}$ est une martingale de carré intégrable bornée dans L^2 car (théorème 9.9)

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

(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 $(S_n)_{n\geqslant 2}$ converge p.s. (et dans L^2) vers une v.a.r. finie p.s. (et dans L^2). Le lemme de Kronecker [3] assure alors la convergence p.s. vers 0 quand $n \to \infty$ 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 9.13 (Comportement asymptotique en loi du nombre de tables)**.** *On a*

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

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

Démonstration. Tout d'abord, on observe que $\mathbb{E}(|\pi_n|)$ et $\mathbb{V}(|\pi_n|)$ peuvent être remplacés par $\theta \log(n)$ en utilisant le *lemme de Slutsky* et le théorème 9.9.

Le résultat découle du *théorème limite central quantitatif de Berry-Esseen* inhomogène de la section 1.6 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 de la section 1.6, avec $p_k = \theta/(\theta+k-1)$. En effet, cela donne $d_V(\mu_n, \nu_n) = o_{n\to\infty}(1)$, et par conséquent, pour toute fonction continue et bornée $f : \mathbb{R} \to \mathbb{R}$, on a, avec $\sigma_n^2 = \mathbb{V}(|\pi_n|)$,

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

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

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

par le *théorème limite central*, le théorème 9.9 ($\sqrt{\lambda_n}/\sigma_n \to 1$), et le *lemme de Slutsky*. □

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

9.3 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 9.14 (Clients solitaires)**.** *Pour tout temps* $n \geqslant 1$*, le nombre* $A_{n,1}$ *de tables réduites à un seul client vérifie*

$$\mathbb{E}(A_{n,1}) = \frac{n\theta}{n+\theta-1} \qquad \textit{et} \qquad \mathbb{V}(A_{n,1}) = \frac{n(n-1)(n-2+2\theta)\theta}{(n+\theta-2)(n+\theta-1)^2}.$$

En particulier,

$$\lim_{n\to\infty} \mathbb{E}(A_{n,1}) = \theta \qquad \textit{et} \qquad \lim_{n\to\infty} \mathbb{V}(A_{n,1}) = \theta.$$

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

Recueil de modèles stochastiques – Page 100/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

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 $n \to \infty$.

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 $0 \leqslant A_{n,1} \leqslant n$ pour tout $n \in \mathbb{N}^*$. De plus, pour tous entiers $1 \leqslant a_1, \ldots, a_n \leqslant n$ et $0 \leqslant a_{n+1} \leqslant n+1$ on a

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

qui vaut

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

En particulier, $(A_{n,1})_{n\geqslant 1}$ 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 !)

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

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

$$\mathbb{V}(X) = \mathbb{E}(\mathbb{V}(X \,|\, Y)) + \mathbb{V}(\mathbb{E}(X \,|\, Y)).$$

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 9.15 (Table unique)**.** *On a $A_{1,1} = 1$ et $A_{n,n} \in \{0,1\}$ pour tout $n \geqslant 1$. Pour tout $n \geqslant 2$, 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 $(A_{n,n})_{n\geqslant 1}$ 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 $\mathbb{P}(A_{n,n} = 1) = \mathbb{P}(\xi_2 = 0, \ldots, \xi_n = 0) = \mathbb{P}(\xi_2 = 0) \cdots \mathbb{P}(\xi_n = 0)$. 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

Recueil de modèles stochastiques – Page 101/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

série harmonique diverge, et par conséquent $\lim_{n\to\infty} \mathbb{P}(A_{n,n} = 1) = 0$. D'autre part, la suite d'événements $(E_n)_{n\geqslant 1}$ définie par $E_n = \{A_{n,n} = 0\}$ est croissante, et donc

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

En fait, $A_{n,n} = \mathbb{1}_{\{n<T\}}$ où $T := \inf\{n \geqslant 1 : \xi_n = 1\}$ 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 $\sum_{n\geqslant 1} \mathbb{P}(F_n = 1) = \sum_{n\geqslant 1} \frac{\theta}{\theta+n} = \infty$ et $\overline{\lim}\, F_n = \{\sum_{n\geqslant 1} \mathbb{1}_{\{\xi_n=1\}} = \infty\} \subset \{T < \infty\}$. □

9.4 Compléments de combinatoire

D'après le chapitre 8, le cardinal de B_n est le n-ème *nombre de Bell* B_n. On a

$$B_n = \sum_{k=1}^{n} \left\{ {n \atop k} \right\}$$

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

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

car pour choisir une partition de $\{1, \ldots, n\}$ ayant k blocs il faut et il suffit soit de choisir une partition de $\{1, \ldots, n-1\}$ 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 partition de $\{1, \ldots, n-1\}$ ayant k blocs. Si X est une variable aléatoire de loi de Poisson de paramètre λ alors

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

On dispose également de la formule explicite suivante :

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

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

Une partition de $\{1, \ldots, n\}$ 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 $\{1, \ldots, n\}$ qui possèdent exactement k cycles est donné par le *nombre de Stirling de première espèce* non signé, noté

$$\left[{n \atop k} \right],$$

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

Recueil de modèles stochastiques – Page 102/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

9.5 Notes et commentaires

L'essentiel de ce chapitre est tiré de l'article didactique [CDM13]. 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 4. Le paramètre θ apparaît comme un taux de mutation des allèles, comme expliqué dans le chapitre 10. 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
$$\mathcal{L}(Z_1, \dots, Z_n \mid Z_1 + 2Z_2 + \cdots + nZ_n = n) \sim \mathrm{Ewens}(n, \theta);$$
2. Soit $(P_r)_{r \geqslant 1}$ 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 $(W_r)_{r \geqslant 1}$ sont i.i.d. de loi Beta $w \mapsto \theta(1-w)^{\theta-1}\mathbb{1}_{[0,1]}(w)$. Conditionnellement à $(P_r)_{r \geqslant 1}$, soient $X_1, \dots, X_n$ des v.a. i.i.d. sur $\mathbb{N}^*$ de loi $(P_r)_{r \geqslant 1}$. La suite $X_1, \dots, X_n$ fait apparaître au plus n entiers différents, et leurs effectifs suit la loi $\mathrm{Ewens}(n, \theta)$.

Signalons qu'il est possible d'établir que pour tous $0 \leqslant k \leqslant n$,
$$\mathbb{P}(|\pi_n| = k) = \begin{bmatrix} n \\ k \end{bmatrix} \frac{\theta^k}{\theta(\theta+1)\cdots(\theta+n-1)}$$
et d'obtenir une formule du même genre pour la loi du couple $(A_n, |\pi_n|)$, cf. [JKB97].

Recueil de modèles stochastiques – Page 103/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Chapitre **10**

Généalogies et coalescence

Recueil de modèles stochastiques – Page 105/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Mots-clés. Arbre ; branchement ; généalogie ; loi multinomiale ; loi de Poisson ; processus de Galton-Watson ; modèle de Fisher-Wright ; loi géométrique ; loi exponentielle ; loi de Ewens ; coalescent de Kingman.

On s'intéresse à l'évolution d'une population de taille finie N constante au fils des générations. On suppose que les générations ne se chevauchent pas, et que pour tout $n \geqslant 0$, 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 $n \geqslant 0$, 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 $(\nu_i^n)_{1 \leqslant i \leqslant N}$ ne sont pas indépendantes puisqu'elles doivent vérifier la relation $\nu_1^n + \cdots + \nu_N^n = N$. On a

$$\nu_i^n = \sum_{j=1}^{N} \mathbb{1}_{\{a_j^{n+1} = i\}}.$$

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

$$\mathbb{P}(\nu_1^n = n_1, \dots, \nu_N^n = n_N) = \frac{N!}{n_1! \dots n_N!} \left(\frac{1}{N}\right)^N \mathbb{1}_{\{n_1 + \cdots + n_N = N\}}.$$

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 $\mathrm{Bin}(N, x/N)$. 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 10.1 (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*

$$\mathcal{L}((Y_1, \dots, Y_N) \,|\, Y_1 + \cdots + Y_N = N) = \mathrm{Mul}(N, (p_1, \dots, p_N))$$

où $p_k = \theta_k/(\theta_1 + \cdots + \theta_N)$ pour tout $1 \leqslant k \leqslant N$. En particulier, si $\theta_1 = \cdots = \theta_N$ alors $p_1 = \cdots = p_N = 1/N$. 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 5).

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.

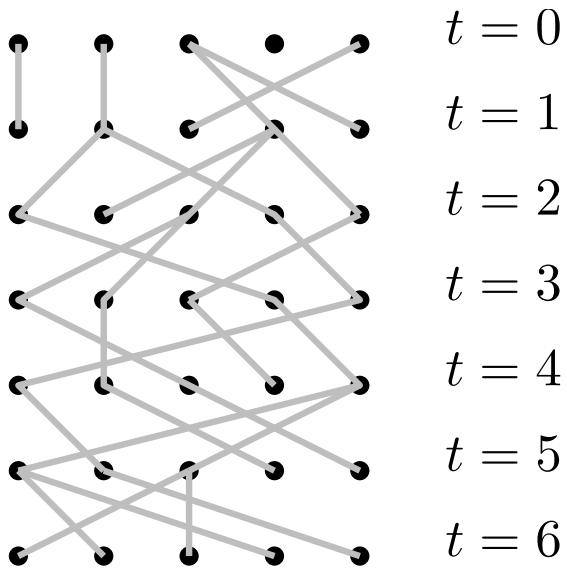


Figure 10.1 – Dans le modèle de Fisher-Wright, tout se passe comme si pour chaque génération, chaque individu, indépendemment de tous les autres, choisissait uniformément un père dans la génération précédente et en héritait. Le graphique ci-dessus illustre par exemple l'évolution de ces relations de filiation sur quelques générations pour une population de taille $N = 5$. Ce point de vue généalogique est cœur de ce chapitre.

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 $k \geqslant 1$,

$$\mathbb{P}(T_3' > k) = \left(1 - \frac{3}{N} + \frac{2}{N^2}\right)^k.$$

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

Recueil de modèles stochastiques – Page 106/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

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 $\mathrm{Bern}(1 - 1/(3N-2))$ et $\mathrm{Geom}_{\mathbb{N}^*}(p_2)$.

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 espèce [1]. 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

$$G_N = \left(g_{j,k}^{(N)} \mathbb{1}_{\{j \leqslant k\}}\right)_{1 \leqslant j,k \leqslant n}.$$

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.

10.1 Modèle généalogique à temps continu

Considérons la chaîne de Markov à temps continu $(A(t))_{t\geqslant 0}$ sur $\mathbb{N}^*$ de générateur

$$L(j,k) = \begin{cases} \binom{j}{2} & \text{si } j \geqslant 2 \text{ et } k = j-1, \\ -\binom{j}{2} & \text{si } j \geqslant 2 \text{ et } k = j, \\ 0 & \text{sinon.} \end{cases}$$

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 $\mathrm{Exp}(1)$ est la loi $\mathrm{Exp}(\binom{n}{2})$. On note $(A_n(t))_{t\geqslant 0}$ 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 $p_2 = 1/N$.

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

Recueil de modèles stochastiques – Page 107/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Lemme 10.2 (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* $n \to \infty$*. 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 $V \sim \mathrm{Exp}(\mu)$, alors

$$\varphi_{V_n/n}(t) = \frac{\mu_n e^{it/n}}{1-(1-\mu_n)e^{it/n}} = \frac{n\mu_n}{n\mu_n - n(1-e^{it/n})} \underset{n\to\infty}{\longrightarrow} \frac{\mu}{\mu - it} = \varphi_V(t),$$

Alternativement, pour tout $x \geqslant 0$, $\mathbb{P}(n^{-1}V_n \geqslant x) = \mathbb{P}(V_n \geqslant nx) = e^{-\mu_n \lfloor nx \rfloor} \to e^{-\mu x}$. □

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

$$(A_2^N([Nt]))_{t\geqslant 0} \underset{N\to\infty}{\overset{\text{loi}}{\longrightarrow}} (A_2(t))_{t\geqslant 0}.$$

Considérons l'arbre généalogique de trois individus. Quand N tend vers l'infini, T_2'/N converge en loi vers $\mathrm{Exp}(1)$. 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 $\mathrm{Exp}(1)$ et $\mathrm{Exp}(3)$:

$$\begin{aligned}
\varphi_{(T_2'/N, T_3'/N)}(s,t) &= \mathbb{E}\big(\exp(isT_2'/N + itT_3'/N)\big) \\
&= \mathbb{E}\big[\mathbb{E}\big(\exp(isT_2'/N)|T_3'\big)\exp(itT_3'/N)\big] \\
&= \frac{1}{3N-2}\mathbb{E}\Big[e^{itT_3'/N}\Big] + \frac{3N-3}{3N-2}\mathbb{E}\Big[e^{itV/N}\Big]\mathbb{E}\Big[e^{itT_3'/N}\Big]
\end{aligned}$$

et le lemme 10.2 permet de conclure que $(A_3^N([Nt]))_{t\geqslant 0}$ converge en loi vers le processus $(A_3(t))_{t\geqslant 0}$ 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 $T_3 \sim \mathrm{Exp}(3)$, un ancêtre commun à 2 individus apparaît,
— l'ancêtre commun apparaît alors après un temps $T_2 \sim \mathrm{Exp}(1)$,
— 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 10.3 (Du discret au continu)**.** *Pour tout* $n \geqslant 2$*, au sens des lois marginales*

$$(A_n^N([Nt]))_{t\geqslant 0} \underset{N\to\infty}{\overset{\text{loi}}{\longrightarrow}} (A_n(t))_{t\geqslant 0}.$$

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 $\mathcal{S}_{k,k-1} = \binom{k}{2}$, pour tout $2 \leqslant k \leqslant n$,

$$g_{k,k-1}^{(N)} = S_{k,k-1}\frac{N(N-1)\cdots(N-k+1)}{N^k} = \binom{k}{2}\frac{1}{N} + O(N^{-2})$$

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 $(G_N)^{[Nt]}$... □

Recueil de modèles stochastiques – Page 108/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

10.2 Longueur de l'arbre généalogique

Le processus à temps continu $(A_n(t))_{t\geqslant 0}$ 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 $k\geqslant 2$, 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 $\mathrm{Exp}(\binom{k}{2})$. De plus, les variables aléatoires $(T_k)_{2\leqslant k\leqslant n}$ sont indépendantes. Si W_n est le temps d'apparition de l'ACPR alors

$$W_n = T_n + \cdots + T_2,$$

où les $(T_k)_{2\leqslant k\leqslant n}$ sont les temps de séjours dans les états $2,\dots,n$. En particulier, on a donc

$$\mathbb{E}(W_n) = 2\sum_{k=2}^{n}\frac{1}{k(k-1)} = 2\sum_{k=2}^{n}\left(\frac{1}{k-1}-\frac{1}{k}\right) = 2\left(1-\frac{1}{n}\right)\to 2$$

et les $T_2,\dots,T_n$ étant indépendantes, on a de plus

$$\begin{aligned}\mathbb{V}(W_n) &= \sum_{k=2}^{n}\mathbb{V}(T_k)\\ &= \sum_{k=2}^{n}\frac{4}{k^2(k-1)^2}\\ &= 4\sum_{k=2}^{n}\left(\frac{1}{k^2}+\frac{1}{(k-1)^2}-\frac{2}{k(k-1)}\right)\\ &= 8\sum_{k=1}^{n-1}\frac{1}{k^2}+\frac{4}{n^2}-4-8\left(1-\frac{1}{n}\right)\end{aligned}$$

On a $\mathbb{V}(W_n)\nearrow \frac{4\pi^2}{3}-12\simeq 1.16$ (notons que $\mathbb{V}(T_2)=1$). Il est possible d'obtenir une expression explicite de la densité de W_n en remarquant que $\mathbb{P}(W_n\leqslant t)=\mathbb{P}(A_n(t)=1)$ 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 :

$$L_n = 2T_2 + \cdots + nT_n.$$

En particulier, comme $T_2,\dots,T_n$ sont indépendantes avec $T_k\sim\mathrm{Exp}(\binom{k}{2})$,

$$\mathbb{E}(L_n)\sim_{n\to\infty} 2\log(n) \quad\text{et}\quad \mathbb{V}(L_n)\sim_{n\to\infty}\frac{2\pi^2}{3}.$$

Théorème 10.4 (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*

$$\frac{L_n}{\log(n)}\xrightarrow[n\to\infty]{p.s.} 2 \quad \textit{et} \quad \frac{L_n}{2}-\log(n)\xrightarrow[n\to\infty]{\text{loi}} G$$

où G est la loi de Gumbel de fonction de répartition $t\in\mathbb{R}\mapsto e^{-e^{-t}}$.

Recueil de modèles stochastiques – Page 109/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Démonstration. La v.a.r. L_n est la somme des v.a.r. indépendantes $2T_2, \dots, nT_n$ qui suivent les lois $\mathrm{Exp}(1/2), \dots, \mathrm{Exp}((n-1)/2)$. Le premier résultat découle alors de la propriété suivante : si $E_1, \dots, E_n$ sont indépendantes de lois $\mathrm{Exp}(\lambda), \mathrm{Exp}(2\lambda), \dots, \mathrm{Exp}(n\lambda)$ alors $S_n := E_1 + \cdots + E_n$ 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 $\mathrm{Exp}(\lambda)$. La densité de M_n est

$$f_n(x) := (\mathbb{P}(M_n \leqslant x))' = ((1 - e^{-\lambda x})^n \mathbb{1}_{\mathbb{R}_+}(x))' = n\lambda(1 - e^{-\lambda x})^{n-1} e^{-\lambda x} \mathbb{1}_{\mathbb{R}_+}(x).$$

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 $\mathrm{Exp}(\lambda)$,

$$\begin{aligned} f_n * g_{(n+1)\lambda}(y) &= (n+1)\lambda \int_{-\infty}^{y} f_n(x) e^{-\lambda(n+1)(y-x)}\, dx \\ &= \lambda(n+1) n\lambda e^{-\lambda(n+1)y} \int_0^y (e^{\lambda x} - 1)^{n-1} e^{\lambda x}\, dx \\ &= \lambda(n+1) e^{-\lambda(n+1)y} (e^{\lambda y} - 1)^n \\ &= \lambda(n+1) e^{-\lambda y} (1 - e^{-n\lambda y})^n = f_{n+1}(y). \end{aligned}$$

Alternativement, on peut remarquer que si $F_1, \dots, F_n$ sont des v.a.r. i.i.d. de loi exponentielle $\mathrm{Exp}(\lambda)$, et si $\min(F_1, \dots, F_n) = F_{(1)} < \cdots < F_{(n)} = \max(F_1, \dots, F_n)$ est le réordonnement croissant (*statistique d'ordre*) de $F_1, \dots, F_n$, alors l'identité en loi suivante a lieu [2] :

$$(F_{(1)}, \dots, F_{(n)}) \stackrel{\text{loi}}{=} \left(\frac{F_1}{n}, \cdots, \frac{F_1}{n} + \cdots + \frac{F_n}{1} \right).$$

En particulier $F_{(k)} \stackrel{\text{loi}}{=} \frac{F_1}{n} + \cdots + \frac{F_k}{n-k+1}$ pour tout $1 \leqslant k \leqslant n$. Pour le voir, on écrit, pour toute fonction mesurable et bornée $\varphi : \mathbb{R}^n \to \mathbb{R}$, en utilisant des changement de variables,

$$\begin{aligned} \mathbb{E}(\varphi(F_{(1)}, \dots, F_{(n)})) &= \int_{[0,\infty)^n} \varphi(x_{(1)}, \dots, x_{(n)})\, \lambda^n e^{-\lambda(x_1 + \cdots + x_n)}\, dx_1 \cdots dx_n \\ &= n! \int_{0 \leqslant y_1 < \cdots < y_n} \varphi(y_1, \dots, y_n)\, \lambda^n e^{-\lambda(y_1 + \cdots + y_n)}\, dy_1 \cdots dy_n \\ &= \int_{[0,\infty)^n} \varphi(z_1, \dots, z_1 + \cdots + z_n)\, n! \lambda^n e^{-\lambda(nz_1 + \cdots + 1z_n)}\, dz_1 \cdots dz_n, \end{aligned}$$

ce qui signifie que les v.a.r. $F_{(1)}, F_{(2)} - F_{(1)}, \dots, F_{(n)} - F_{(n-1)}$ sont indépendantes avec pour tout $1 \leqslant k \leqslant n$, $F_{(k)} - F_{(k-1)}$ de loi exponentielle de paramètre $(n-k+1)\lambda$.

Pour établir le second résultat du théorème (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

$$\sum_{n \geqslant 2} \mathbb{E}(X_n^2) = \sum_{n \geqslant 2} \mathbb{V}(X_n) = \sum_{n \geqslant 2} \frac{4}{(n-1)^2} < \infty$$

2. On parle de théorème de Rényi. L'identité $F_{(1)} \stackrel{d}{=} \frac{1}{n} F_1$ est un cas spécial de l'identité en loi plus générale

$$\min\left(\frac{F_1}{\lambda_1}, \dots, \frac{F_n}{\lambda_n} \right) \stackrel{\text{loi}}{=} \frac{E_1}{\lambda_1 + \cdots + \lambda_n},$$

valide pour tout for tous $\lambda_1 > 0, \dots, \lambda_n > 0$, qui joue un rôle important dans la structure des processus de Markov. Presque sûrement, le min est atteint en un unique entier aléatoire N dans $\{1, \dots, n\}$, indépendant de la valeur du min, de loi discrète $\mathbb{P}(N = k) = \lambda_k / (\lambda_1 + \cdots + \lambda_n)$ pour tout $1 \leqslant k \leqslant n$. L'identité en loi $F_{(n)} \stackrel{\text{loi}}{=} \frac{F_1}{n} + \cdots + \frac{F_n}{1}$ apparaît dans la structure du *processus de Yule* du chapitre 21.

Recueil de modèles stochastiques – Page 110/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

donc $\sum_n X_n$ converge p.s. et donc

$$\frac{L_n}{\log(n)} = \frac{1}{\log(n)} \sum_{k=2}^{n} X_k + \frac{1}{\log(n)} \sum_{k=2}^{n} \mathbb{E}(kT_k) = \frac{1}{\log(n)} \sum_{k=2}^{n} X_k + \frac{2}{\log(n)} \sum_{k=1}^{n-1} \frac{1}{k} \xrightarrow[n\to\infty]{\text{p.s.}} 2.$$

Le troisième résultat (fluctuation Gumbel) vient de $L_n \stackrel{d}{=} M_n$ car pour tout $t \in \mathbb{R}$,

$$\mathbb{P}(L_n - 2\log(n) \leqslant 2t) = \left(1 - e^{-\frac{1}{2}(2t+2\log(n))}\right)^{n-1} = \left(1 - \frac{e^{-t}}{n}\right)^{n} \underset{n\to\infty}{\longrightarrow} e^{-e^{-t}}.$$

□

10.3 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 diversifications 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 10.5 (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 $\eta_1 + \cdots + \eta_n$ où $(\eta_k)_{k\geqslant 1}$ sont indépendantes de loi de Bernoulli de paramètres $(\theta/(k-1+\theta))_{k\geqslant 1}$.*

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 9). En particulier, on sait calculer les deux premiers moments, et on connaît le comportement asymptotique (convergence et fluctuation).

Lemme 10.6 (Horloges exponentielles en compétition)**.** *Si $E_1, \ldots, E_n$ sont des v.a.r. exponentielles indépendantes de paramètres respectifs $\lambda_1, \ldots, \lambda_n$ alors $M := \min(E_1, \ldots, E_n)$ suit la loi exponentielle de paramètre $\lambda_1 + \ldots + \lambda_n$. De plus, presque sûrement le minimum dans la définition de M est atteint en un unique entier I aléatoire de $\{1, \ldots, n\}$, indépendant de M, de loi $\mathbb{P}(I = i) = \mathbb{P}(M = E_i) = \lambda_i/(\lambda_1 + \cdots + \lambda_n)$ pour tout $i \in \{1, \ldots, n\}$.*

Preuve du lemme 10.6. Pour tout $i \in \{1, \ldots, n\}$ et tout $x \in \mathbb{R}_+$,

$$\mathbb{P}(\min_{k\neq i}(E_k) > E_i \geqslant x) = \lambda_i \int_x^\infty \prod_{k\neq i} \mathbb{P}(E_k > y)\, e^{-\lambda_i y}\, dy = \frac{\lambda_i}{\lambda_1 + \cdots + \lambda_n} e^{-(\lambda_1+\cdots+\lambda_n)x}$$

C'est la formule pour $\mathbb{P}(I = i, M \geqslant x)$, qui donne la loi du couple (I, M). □

Recueil de modèles stochastiques – Page 111/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Preuve du théorème 10.5. 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 $\mathrm{Exp}(\binom{n}{2})$) et $R_1, \ldots, R_n$ premiers temps de mutation des individus $1, \ldots, n$ de même loi $\mathrm{Exp}(\theta/2)$. 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 10.6)

$$\frac{\binom{n}{2}}{\frac{n(n-1+\theta)}{2}} = \frac{n-1}{n-1+\theta}.$$

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 $(\eta_k)_k$. □

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 $\boldsymbol{c} = (c_1, \ldots, c_n)$ où c_i est le nombre d'allèles représentés i fois et $|\boldsymbol{c}| = c_1 + 2c_2 + \cdots + nc_n = n$. On notera $\boldsymbol{e}_i = (0, 0, \ldots, 0, 1, 0, \ldots, 0)$ le i^{ème} vecteur unitaire. Il s'agit à présent d'établir une équation satisfaite par les probabilités $(q(\boldsymbol{c}))$ où $q(\boldsymbol{c})$ est la probabilité qu'un échantillon de taille $|\boldsymbol{c}|$ tiré sous la probabilité stationnaire ait la configuration $\boldsymbol{c}$. On pose $q(\boldsymbol{e}_1) = 1$. Supposons que la configuration soit $\boldsymbol{c}$. 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 10.6)

$$\frac{n\theta/2}{n\theta/2 + n(n-1)/2} = \frac{\theta}{\theta + n - 1}.$$

La configuration $\boldsymbol{b}$ qui a conduit à la $\boldsymbol{c}$ est

— $\boldsymbol{b} = \boldsymbol{c}$ si la mutation a eu lieu sur l'un des c_1 singletons (probabilité c_1/n),

— $\boldsymbol{b} = \boldsymbol{c} - 2\boldsymbol{e}_1 + \boldsymbol{e}_2$ si la mutation a eu lieu sur un des allèles représenté 2 fois (probabilité $2b_2/n = 2(c_2+1)/n$),

— $\boldsymbol{b} = \boldsymbol{c} - \boldsymbol{e}_1 - \boldsymbol{e}_{j-1} + \boldsymbol{e}_j$ si la mutation a eu lieu sur un des allèles représenté j fois (probabilité $jb_j/n = j(c_j+1)/n$)

Recueil de modèles stochastiques – Page 112/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

— 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 $\boldsymbol{b} = \boldsymbol{c} + \boldsymbol{e}_j - \boldsymbol{e}_{j+1}$: 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 9, qui nous enseigne que $\boldsymbol{c}$ suit la loi d'Ewens (on parle de formule d'échantillonnage d'Ewens en génétique[3] : pour tous entiers $a_1, \dots, a_n$ vérifiant $a_1 + 2a_2 + \cdots + na_n = n$,

$$q(\boldsymbol{c}) = \mathbb{P}(\boldsymbol{c}_1(n) = a_1, \dots, \boldsymbol{c}_n(n) = a_n) = \frac{n!}{\theta(\theta+1)\cdots(\theta+n-1)} \prod_{j=1}^{n} \left(\frac{\theta}{j}\right)^{c_j} \frac{1}{c_j}.$$

10.4 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 $(C(t))_{t\geqslant 0}$? 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 conséquence 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 $(C(t))_{t\geqslant 0}$ 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 rélié à personne), et de générateur Q

$$Q(\alpha, \beta) = \begin{cases} -\binom{k}{2} & \text{si } \alpha = \beta \text{ et } |\alpha| = k, \\ 1 & \text{si } \alpha \sim \beta, \\ 0 & \text{sinon,} \end{cases}$$

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 :

$$\mathbb{P}(C_{k-1}^{\text{inc}} = \beta \,|\, C_k^{\text{inc}} = \alpha) = \begin{cases} \frac{1}{\binom{k}{2}} & \text{si } \alpha \sim \beta \text{ et } |\alpha| = k, \\ 0 & \text{sinon.} \end{cases}$$

Le processus C évolue donc selon une suite $\Delta = C_n^{\text{inc}} \sim C_{n-1}^{\text{inc}} \sim \cdots \sim C_1^{\text{inc}} = \Theta$, 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

$$|C(t)| = A_n(t),$$

3. ESF : Ewens Sampling Formula.

Recueil de modèles stochastiques – Page 113/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

puisque les classes de $C(t)$ sont en bijection avec les ancêtres au temps t. Ainsi, les processus $(A_n(t))_{t\geqslant 0}$ et $(C_k^{\mathrm{inc}})_k$ sont indépendants et $C(t) = C_{A_n(t)}^{\mathrm{inc}}$ pour tout $t \geqslant 0$. Plus précisément,

$$\mathbb{P}(C(t) = \alpha) = \mathbb{P}(A_n(t) = |\alpha|)\mathbb{P}(C_{|\alpha|}^{\mathrm{inc}} = \alpha).$$

Théorème 10.7 (Loi du coalescent de Kingman)**.** *Soit $1 \leqslant j \leqslant n$ et α une partition de $[n]$ dont les classes d'équivalence admettent les cardinaux $\lambda_1, \dots, \lambda_j$. Alors,*

$$\mathbb{P}(C_j^{\mathrm{inc}} = \alpha) = \frac{(n-j)!j!(j-1)!}{n!(n-1)!}\lambda_1! \cdots \lambda_j!.$$

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 $j \geqslant 2$. Alors

$$p_{j-1}(\beta) := \mathbb{P}(C_{j-1}^{\mathrm{inc}} = \beta) = \sum_{\alpha \in \mathcal{E}_n} p_j(\alpha)\mathbb{P}(C_{j-1}^{\mathrm{inc}} = \beta \,|\, C_j^{\mathrm{inc}} = \alpha) = \sum_{\alpha \sim \beta} p_j(\alpha)\frac{2}{j(j-1)}.$$

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 $1 \leqslant l \leqslant j-1$ et $1 \leqslant m \leqslant \lambda_l - 1$. Grâce à l'hypothèse de récurrence [4],

$$\begin{aligned}
p_{j-1}(\beta) &= \frac{1}{2}\sum_{l=1}^{j-1}\sum_{m=1}^{\lambda_l - 1}\frac{2(n-j)!j!(j-1)!}{j(j-1)n!(n-1)!}\lambda_1! \cdots \lambda_{l-1}!m!(\lambda_l - m)!\lambda_{l+1}! \cdots \lambda_{j-1}!\binom{\lambda_l}{m} \\
&= \frac{(n-j)!(j-1)!(j-2)!}{n!(n-1)!}\lambda_1! \cdots \lambda_{j-1}!\sum_{l=1}^{j-1}\sum_{m=1}^{\lambda_l - 1} 1 \\
&= \frac{(n-j+1)!(j-1)!(j-2)!}{n!(n-1)!}\lambda_1! \cdots \lambda_{j-1}!
\end{aligned}$$

car $\sum_{l=1}^{j-1}\sum_{m=1}^{\lambda_l - 1} 1 = \lambda_1 + \cdots + \lambda_{j-1} - j + 1 = n - j + 1$. □

10.5 Notes et commentaires

En biologie, la *phylogénie* est l'étude des relations de parenté entre êtres vivants. La modélisation stochastique en phylogénie est abordée dans les livres de Warren Ewens [Ewe04], de Rick Durrett [Dur08], et de Benjamin Jourdain et Jean-François Delmas [DJ06]. Les aspects statistiques sont abordés dans le cours de Saint-Flour de [Tav04]. Sir John Frank Charles Kingman étudie le processus qui porte son nom dans [Kin93]. Son travail a donné lieu a de nombreux développement, notamment une théorie de la fragmentation et de la coalescence autonome, en liaison avec les processus de branchement, les processus de Lévy, et les partitions aléatoires. On pourra consulter par exemple les livres de Nathanaël Berestycki [Ber09] et de Jean Bertoin [Ber06], ainsi que le cours de Saint-Flour de Jim Pitman [Pit06].

4. Le facteur 1/2 est dû à la permutation de m et $\lambda_l - m$. Considérer l'exemple $\lambda_l = 2$, puis $\lambda_l = 3$.

Recueil de modèles stochastiques – Page 114/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Chapitre 11

Agrégation limitée par diffusion interne

Mots-clés. Chaîne de Markov inhomogène ; martingale ; convergence en loi ; couplage.

11.1 Modèle d'agrégation par diffusion

Un fût de déchets radioactifs est enterré secrètement dans le Cantal. Après quelques années, il devient poreux et laisse échapper son contenu. Pour éviter une contamination excessive, on a disposé des pièges à particules qui capturent la première particule qui passe mais deviennent ensuite inertes (une nouvelle particule passe sans être arrêtée). Le milieu est isotrope : une particule radioactive se déplace de la même manière dans toutes les directions. Cette particule continuera à se déplacer tant qu'elle passe sur des pièges qui ont déjà été activés et est capturée par le premier piège libre qu'elle rencontre. On souhaite connaître la forme typique des zones qui seront contaminées par cette fuite.

Pour modéliser l'évolution de la région polluée, on introduit le modèle de croissance suivant. On assimile l'espace à $\mathbb{Z}^d$ où d vaut 1, 2 ou 3 selon le nombre de dimensions que l'on souhaite prendre en compte. Deux éléments x et y de $\mathbb{Z}^d$ sont dits voisins, et nous noterons $x \sim y$, si

$$|x-y|_1 = |x_1 - y_1| + \cdots + |x_d - y_d| = 1.$$

On place le fût à l'origine $0 \in \mathbb{Z}^d$. Sur chaque point du réseau est disposé un piège. Notons $A(0)$ le singleton $\{0\}$. Le premier piège activé sera l'un des $2d$ emplacements voisins de l'origine. On modélise la trajectoire d'une particule radioactive sortant du fût par une marche aléatoire simple symétrique sur $\mathbb{Z}^d$ issue de 0 et arrêtée lorsqu'elle sort de $A(0)$. On note $A(1)$ l'ensemble aléatoire composé de 0 et du point où la particule est sortie. Pour tout x voisin de 0, $A(1)$ est égal à $\{0, x\}$ avec probabilité $1/2d$. On itère ensuite le procédé. Étant donné un ensemble $A(n) \subset \mathbb{Z}^d$, on considère une marche aléatoire symétrique $(S_k)_{k\geqslant 0}$ issue de 0 arrêtée lorsqu'elle sort de $A(n)$. On définit alors $A(n+1)$ comme l'ensemble des éléments de $A(n)$ et du point où est sortie la marche S. Ce modèle est connu sous le nom de *agrégation limitée par diffusion interne*[1].

On obtient ainsi une suite $(A(n))_{n\geqslant 0}$ d'ensembles aléatoires. La résultat suivant regroupe quelques propriétés immédiate de cette suite.

Théorème 11.1. *La suite $(A(n))_{n\geqslant 0}$ est croissante au sens de l'inclusion. Le cardinal de $A(n)$ vaut exactement $n+1$. Pour tout $n \in \mathbb{N}$, soit x et y deux éléments de $A(n)$, alors il*

1. «*Internal Diffusion Limited Agregation* (IDLA) en anglais

Recueil de modèles stochastiques – Page 115/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

existe $x_0, \ldots, x_m$ *éléments de* $A(n)$ *tels que* $x_0 = x$, $x_m = y$ *et* $x_i \sim x_{i+1}$ *pour* $i = 0, \ldots, m-1$. *On pourrait dire que l'ensemble* $A(n)$ *est* connexe par arcs *dans* $\mathbb{Z}^d$.

On souhaite à présent étudier le comportement asymptotique de la suite $(A(n))_n$. On étudie en détail le cas de la dimension 1. L'étude en dimensions supérieures est beaucoup plus délicate. Elle est évoquée dans les sections 11.5 et Notes et commentaires.

11.2 Modèle unidimensionnel

On suppose que la propagation se fait selon un axe horizontal. On se place donc dans le cas où $d = 1$. L'ensemble initial A_0 est le singleton $\{0\}$ puis A_1 est égal à $\{0, 1\}$ avec probabilité $1/2$ et $\{-1, 0\}$ avec probabilité $1/2$, etc. Notons $G_n = \min A_n$ et $D_n = \max A_n$. L'ensemble A_n est de la forme $A_n = \{G_n, G_n + 1, \ldots, D_n - 1, D_n\}$. Puisque le cardinal de A_n est $n + 1$, on a $D_n - G_n = n$. Ainsi, A_n est caractérisé par $X_n = D_n + G_n$ et, en particulier,

$$D_n = \frac{X_n + n}{2} \quad \text{et} \quad G_n = \frac{X_n - n}{2}.$$

Les accroissements $(X_{n+1} - X_n)_{n \geqslant 0}$ ne peuvent prendre que les valeurs -1 ou 1. Plus précisément, $X_{n+1} = X_n - 1$ si la marche issue de 0 atteint $G_n - 1$ avant $D_n + 1$, et $X_{n+1} = X_n + 1$ si la marche issue de 0 atteint $D_n + 1$ avant $G_n - 1$.

Théorème 11.2 (Évolution markovienne)**.** *La suite* $(X_n)_{n \geqslant 0}$ *est une chaîne de Markov inhomogène sur* $\mathbb{Z}$ *issue de* 0 *dont les transitions sont décrites par les relations suivantes : pour* $-n \leqslant i \leqslant n$,

$$\mathbb{P}(X_{n+1} = i - 1 | X_n = i) = \frac{n + 2 + i}{2(n + 2)} \quad \textit{et} \quad \mathbb{P}(X_{n+1} = i + 1 | X_n = i) = \frac{n + 2 - i}{2(n + 2)}.$$

Démonstration. Pour tout $n \geqslant 1$,

$$\{X_n = i\} = \left\{ D_n = \frac{i + n}{2},\ G_n = \frac{i - n}{2} \right\}.$$

La probabilité $\mathbb{P}(X_{n+1} = i - 1 | X_n = i)$ est donc égale à la probabilité que la marche aléatoire simple S issue de 0 atteigne $(i-n)/2-1$ avant $(i+n)/2+1$. Or d'après le théorème 2.2, si a et b sont deux entiers distincts, le premier négatif, le second positif, et si $T_i = \inf\{n \geqslant 0,\ S_n = i\}$ pour $i = a, b$ et $T = T_a \wedge T_b$, alors, T_a et T_b sont finis presque sûrement, T est intégrable et

$$\mathbb{P}(T = T_a) = 1 - \mathbb{P}(T = T_b) = \frac{b}{b - a} \quad \text{et} \quad \mathbb{E}(T) = -ab.$$

□

Remarque 11.3 (Équilibrage)**.** *Lorsque* X_n *est strictement positif,* $X_{n+1} - X_n$ *a une probabilité plus forte de valoir -1 que de valoir +1. La tendance s'inverse lorsque* X_n *est strictement négatif. On peut donc penser que le processus aura davantage tendance à revenir en* 0 *qu'une marche aléatoire simple symétrique (dont l'accroissement vaut plus ou moins 1 avec probabilité* $1/2$*).*

11.3 Convergence presque sûre en dimension 1

On établit ici le premier résultat (relativement intuitif) de convergence en temps long pour le processus X. L'idée de la preuve est de rendre rigoureuse l'intuition de la remarque 11.3 en comparant (de manière déterministe) la valeur absolue de X à celle de la marche aléatoire simple.

Recueil de modèles stochastiques – Page 116/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Théorème 11.4 (Convergence presque sûre)**.** *Le processus* $(|X_n|)_{n\geqslant 0}$ *est une chaîne de Markov inhomogène et la suite* $(|X_n|/n)_{n\geqslant 0}$ *converge vers* 0 *presque sûrement. En particulier,*

$$\frac{X_n}{n} \xrightarrow[n\to\infty]{p.s.} 0 \quad d'où \quad \frac{G_n}{n} \xrightarrow[n\to\infty]{p.s.} -\frac{1}{2} \quad et \quad \frac{D_n}{n} \xrightarrow[n\to\infty]{p.s.} \frac{1}{2}.$$

Démonstration. En discutant les valeurs de X_n, on a que $|X_{n+1}| = 1$ si $|X_n| = 0$ et, que, si $|X_n| > 0$,

$$|X_{n+1}| = \begin{cases} |X_n| - 1 & \text{avec probabilité } \dfrac{1}{2} + \dfrac{|X_n|}{2(n+2)}, \\ |X_n| + 1 & \text{avec probabilité } \dfrac{1}{2} - \dfrac{|X_n|}{2(n+2)}. \end{cases}$$

Ainsi $(|X_n|)_{n\geqslant 0}$ est bien une chaîne de Markov (suite récurrente aléatoire) inhomogène. Pour établir que $|X_n|/n \to 0$ presque sûrement, on procède par couplage. Plus précisément, on construit deux processus $(Y_n)_{n\geqslant 0}$ et $(Z_n)_{n\geqslant 0}$ comme suit : $Y_0 = Z_0 = 0$ et pour $n \geqslant 0$,

$$Y_{n+1} = \begin{cases} Y_n + 2\mathbf{1}_{\{U_n<(n+2+Y_n)/2(n+2)\}} - 1 & \text{si } Y_n > 0, \\ 1 & \text{si } Y_n = 0, \end{cases}$$

et

$$Z_{n+1} = \begin{cases} Z_n + 2\mathbf{1}_{\{U_n<1/2\}} - 1 & \text{si } Z_n > 0, \\ 1 & \text{si } Z_n = 0, \end{cases}$$

où $(U_n)_{n\geqslant 0}$ est une suite de v.a.r. i.i.d. de loi uniforme sur $[0,1]$. Le processus $(Y_n)_{n\geqslant 0}$ a la loi de $(|X_n|)_{n\geqslant 0}$ tandis que le processus $(Z_n)_{n\geqslant 0}$ a la loi de la valeur absolue d'une marche aléatoire simple. Ce couplage est tel que presque sûrement, pour tout $n \in \mathbb{N}$, $0 \leqslant Y_n \leqslant Z_n$. De plus, si $0 < Y_n \leqslant Z_n$ alors, par construction, $Y_{n+1} \leqslant Z_{n+1}$. Enfin, si $0 = Y_n \leqslant Z_n$ alors, Z_n peut être nul et dans ce cas, $Y_{n+1} \leqslant Z_{n+1}$ ou $Z_n \geqslant 2$ car Z_n et Y_n ont même parité. Enfin, $(Z_n/n)_{n\geqslant 0}$ converge presque sûrement vers 0 en vertu de la loi forte des grands nombres. □

11.4 Convergence en loi en dimension 1

On s'intéresse à présent aux fluctuations de X_n/n autour de sa limite. L'objet de cette section est de donner des éléments de preuve du résultat suivant.

Théorème 11.5 (Convergence en loi)**.** *La convergence en loi suivante a lieu :*

$$\frac{X_n}{\sqrt{n}} \xrightarrow[n\to\infty]{\text{loi}} \mathcal{N}\left(0, \frac{1}{3}\right).$$

La vitesse de convergence en loi est la même que pour la marche aléatoire simple et la loi limite est encore gaussienne. Pourtant sa variance est plus petite que dans le cas identiquement distribué. Cela provient de la tendance auto-stabilisatrice de la chaîne X.

Avant d'aborder la démonstration du théorème 11.5, étudions le comportement des premiers moments de X_n/n quand n tend vers l'infini. La loi de X_n est symétrique donc tous les moments impairs de X_n sont nuls. Pour le moment d'ordre 2, on a, pour $n \geqslant 0$, on a

$$\begin{aligned} \mathbb{E}\big[X_{n+1}^2|\mathcal{F}_n\big] &= \frac{n+2-X_n}{2(n+2)}(X_n+1)^2 + \frac{n+2+X_n}{2(n+2)}(X_n-1)^2 \\ &= 1 + \frac{n}{n+s}X_n^2. \end{aligned}$$

En posant $x_n(2) = \mathbb{E}(X_n^2)$ et en prenant l'espérance dans la relation ci-dessus, on obtient :

$$x_{n+1}(2) = 1 + \frac{n}{n+2}x_n(2).$$

Recueil de modèles stochastiques – Page 117/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Une récurrence donne

$$x_n(2) = 1 + \frac{1}{n(n+1)} \sum_{k=1}^{n-1} k(k+1) + \frac{1}{n(n+1)} x_0(2),$$

et donc $(x_n(2)/n)_{n\geqslant 1}$ converge vers $1/3$. On peut établir de même que pour tout $k \geqslant 0$, la suite $(x_n(2k)/n^k)_{n\geqslant 1}$ admet une limite $\mu(2k)$ et que la suite $(\mu(2k))_{k\in\mathbb{N}}$ est solution de

$$\mu(0) = 1 \quad \text{et} \quad \mu(2k+2) = \frac{2k+1}{3}\mu(2k).$$

Or, la seule mesure dont les moments pairs sont comme ci-dessus et les moments impairs sont nuls est la loi $\mathcal{N}(0, 1/3)$. On peut ainsi déduire de cette convergence des moments de $X_n/\sqrt{n}$ la convergence en loi de $X_n/\sqrt{n}$ vers une mesure gaussienne centrée de variance $1/3$, en s'inspirant par exemple de l'approche utilisée pour les distributions spectrales dans le chapitre 16.

Un autre moyen d'établir la convergence annoncée est d'utiliser un raisonnement basé sur le *théorème limite central pour les martingales de carré intégrable*. Il existe plusieurs jeux d'hypothèses plus ou moins faciles à vérifier. Voici le théorème que nous utiliserons dans la suite.

Théorème 11.6 (TLC pour martingales)**.** *Soit $(M_n)_{n\geqslant 0}$ une martingale de carré intégrable et soit $(a_n)_{n\geqslant 0}$ une suite réelle, strictement positive, déterministe croissante vers l'infini. On pose*

$$\Delta M_n := M_n - M_{n-1}, \quad \langle M\rangle_0 = 0 \quad \textit{et} \quad \langle M\rangle_{n+1} = \langle M\rangle_n + \mathbb{E}\Big[(\Delta M_{n+1})^2 | \mathcal{F}_n\Big].$$

Supposons que

1. *il existe $\lambda \geqslant 0$ déterministe tel que $a_n^{-1}\langle M\rangle_n \xrightarrow[n\to\infty]{\mathbb{P}} \lambda$,*
2. *(condition de Lyapunov) il existe $\delta > 0$ tel que*

$$\frac{1}{(a_n)^{1+\delta/2}} \sum_{k=1}^{n} \mathbb{E}\Big(\Delta M_k^{2+\delta} | \mathcal{F}_{k-1}\Big) \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

Alors, on a

$$\frac{1}{\sqrt{a_n}} M_n \xrightarrow[n\to\infty]{\text{loi}} \mathcal{N}(0,\lambda) \quad \textit{et si } \lambda > 0 \quad \sqrt{a_n}\,\frac{M_n}{\langle M\rangle_n} \xrightarrow[n\to\infty]{\text{loi}} \mathcal{N}(0,\lambda^{-1}).$$

Notre martingale est fournie par lemme suivant.

Lemme 11.7 (Martingale dans le modèle)**.** *Le processus $(M_n)_{n\geqslant 0}$ défini par $M_n = (n+1)X_n$ est une martingale de carré intégrable qui vérifie*

$$\mathbb{E}\Big[(\Delta M_{n+1})^2 | \mathcal{F}_n\Big] = (n+2)^2 - X_n^2$$

$$\mathbb{E}\Big[(\Delta M_{n+1})^4 | \mathcal{F}_n\Big] = \big((n+2)^2 - X_n^2\big)\big((n+2)^2 + 3X_n^2\big),$$

$$\frac{\langle M\rangle_n}{n^3} \xrightarrow[n\to\infty]{p.s.} \frac{1}{3}.$$

Recueil de modèles stochastiques – Page 118/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Démonstration. Soit $\alpha \in \{1, 2, 4\}$. On a

$$\begin{aligned}\mathbb{E}((\Delta M_{n+1})^\alpha|\mathcal{F}_n) &= \mathbb{E}(((n+2)X_{n+1} - (n+1)X_n)^\alpha|X_n) \\ &= [X_n - (n+2)]^\alpha \frac{n+2+X_n}{2(n+2)} + [X_n + n + 2]^\alpha \frac{n+2-X_n}{2(n+2)},\end{aligned}$$

qui est nul si $\alpha = 1$, et donc $(M_n)_{n\geqslant 0}$ est une martingale. De plus, pour $\alpha \in \{2, 4\}$, on obtient

$$\mathbb{E}((\Delta M_{n+1})^\alpha|\mathcal{F}_n) = ((n+2)^2 - X_n^2)\frac{(n+2-X_n)^{\alpha-1} + (n+2+X_n)^{\alpha-1}}{2(n+2)},$$

ce qui donne deux premières relations attendues. En particulier, on a

$$\langle M\rangle_n = \sum_{k=0}^{n-1}((k+1)^2 - X_k^2) = \frac{n^3}{3} - \sum_{k=0}^{n-1} X_k^2 + o(n^3).$$

De plus

$$\frac{1}{n^3}\sum_{k=0}^{n-1} X_k^2 \leqslant \frac{1}{n}\sum_{k=1}^{n-1}\frac{X_k^2}{k^2}.$$

Puisque $(X_n/n)_{n\geqslant 1}$ converge vers 0 presque sûrement, le lemme de Cesàro permet de conclure. □

Pour pouvoir obtenir la convergence en loi de $(M_n)_{n\geqslant 0}$ correctement renormarlisée, il reste à vérifier la condition de Lyapunov du théorème 11.6 avec $a_n = n^3$. Avec $\delta = 2$, on a

$$\sum_{k=1}^{n}\mathbb{E}(\Delta M_k^4|\mathcal{F}_{k-1}) \leqslant \sum_{k=0}^{n} k^4 + 2\sum_{k=0}^{n}(k+1)^2 X_{k-1}^2 = \mathcal{O}(n^5).$$

On conclut en utilisant la convergence de $(X_n/n)_{n\geqslant 1}$ vers 0. La suite $(M_n/n^{3/2})_{n\geqslant 0}$ converge donc en loi vers $\mathcal{N}(0, 1/3)$. Ceci fournit immédiatement la convergence dans le théorème 11.5.

11.5 Dimensions supérieures

On se place à présent sur $\mathbb{Z}^d$ avec $d > 1$. Comme en dimension 1, on s'attend à ce que la forme limite de l'ensemble aléatoire devienne relativement régulière puisque la marche aléatoire issue de l'origine aura tendance probablement à atteindre l'ensemble aléatoire A_n^c en des points qui ont beaucoup de voisins dans A_n. La figure 11.1 fournit une réalisation de $A(1000)$ en dimension 2.

Pour tout $r \geqslant 0$ notons $\|x\|$ la norme euclidienne d'un point $x \in \mathbb{R}^d$ et

$$B(0, r) = \Big\{y \in \mathbb{R}^d \,:\, \|y\| < r\Big\} \quad \text{et} \quad \mathbb{B}(0, r) = B(0, r) \cap \mathbb{Z}^d.$$

Le cardinal d'un ensemble Λ de $\mathbb{Z}^d$ est noté $|\Lambda|$. Fixons un entier n et considérons l'ensemble aléatoire $A(|\mathbb{B}(0, n)|)$. Le rayon (aléatoire) de la plus grande boule incluse dans cet ensemble est

$$\sup\{r \geqslant 0 \,:\, \mathbb{B}(0, r) \subset A(|\mathbb{B}(0, n)|)\}$$

tandis que celui de la plus petite boule qui le contient est

$$\inf\{r \geqslant 0 \,:\, A(|\mathbb{B}(0, n)|) \subset \mathbb{B}(0, r)\}$$

Recueil de modèles stochastiques – Page 119/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

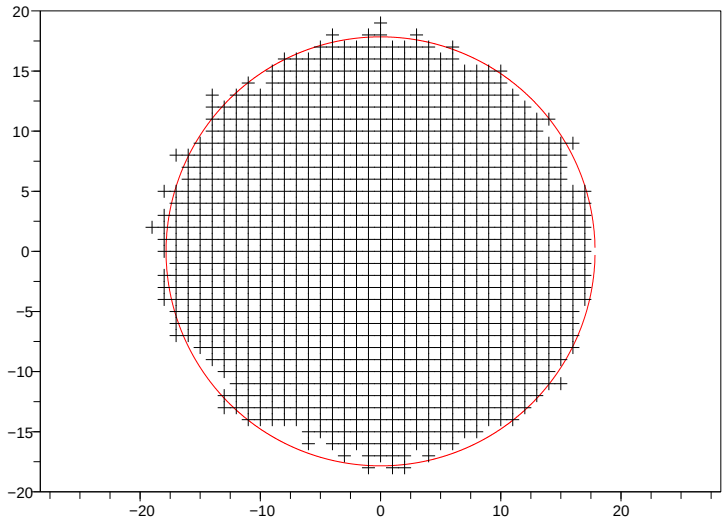


Figure 11.1 – Une réalisation de $A(1000)$ en dimension 2.

On définit alors l'erreur interne $\delta_I(n)$ comme

$$n - \delta_I(n) = \sup\{r \geqslant 0 \,:\, \mathbb{B}(0,r) \subset A(|\mathbb{B}(0,n)|)\},$$

et l'erreur externe $\delta_E(n)$ comme

$$n + \delta_E(n) = \inf\{r \geqslant 0 \,:\, A(|\mathbb{B}(0,n)|) \subset \mathbb{B}(0,r)\}.$$

Le théorème suivant montre non seulement que la forme asymptotique de $A(n)$ est sphérique mais fournit également une borne supérieure pour les fluctuations extrêmes de $A(n)$.

Théorème 11.8 (Forme et fluctuation)**.** *Il existe une constante β_d telle que presque sûrement*

— *pour $d \geqslant 3$:*

$$\limsup_{n\to\infty} \frac{\delta_I(n)}{\sqrt{\log n}} \leqslant \beta_d \quad \textit{et} \quad \limsup_{n\to\infty} \frac{\delta_E(n)}{\sqrt{\log n}} \leqslant \beta_d,$$

— *pour $d = 2$:*

$$\limsup_{n\to\infty} \frac{\delta_I(n)}{\log n} \leqslant \beta_2 \quad \textit{et} \quad \limsup_{n\to\infty} \frac{\delta_E(n)}{\log n} \leqslant \beta_2.$$

11.6 Notes et commentaires

L'une des première études de $(A(n))_{n\geqslant 0}$ est due à Maury Bramson, David Griffeath et Gregory Lawler dans [LBG92]. Il y est notamment établi que la forme asymptotique de $A(n)$ est sphérique. Il faudra attendre longtemps pour avoir une preuve complète pour les fluctuations autour de cette forme limite. On pourra retrouver le théorème 11.8 dans les articles de David Jerison, Lionel Levin, et Scott Scheffied [JLS13, JLS12] et de Amine Asselah et Alexandre Gaudillière [AG13b, AG13a].

Le modèle en dimension 1 peut être relié à un modèle de boules et d'urnes dit de Bernard Friedman décrit dans l'article [Fre65] de David Freedman [2]. Une boule tirée est remise dans

2. Prière de ne pas confondre !

Recueil de modèles stochastiques – Page 120/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

l'urne accompagnée de α boules de la même couleur et β boules de la couleur opposée. Si $\beta = 0$, on retrouve le modèle de l'*urne de Pólya* étudié dans le chapitre 7. Lorsque $\beta > 0$, on peut montrer que la proportion asymptotique de chaque couleur tend vers $1/2$ pour tout α.

Le *théorème limite central pour les martingales* 11.6 est tiré de [BC07, Thm 3.35].

Recueil de modèles stochastiques – Page 121/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Chapitre 12

Chaînes de Markov cachées

Recueil de modèles stochastiques – Page 123/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Mots-clés. Chaîne de Markov ; vecteur gaussien ; loi conditionnelle ; espérance conditionnelle ; processus auto-régressif ; estimateur du maximum de vraisemblance ; convergence de martingale.

Lorsque les modèles simples ne suffisent plus, on peut tenter de modéliser en déformant un modèle simple par une transformation. Ainsi par exemple une *chaîne de Markov cachée* est une projection d'une chaîne de Markov. Ce chapitre présente deux modèles de chaînes de Markov cachées. Le défi mathématique est alors d'obtenir à partir des observations partielles le plus d'informations possible sur la partie cachée. Il se ramène à un problème de calcul de lois conditionnelles. Il n'est en général pas possible de calculer explicitement ces lois. C'est cependant le cas dans les deux exemples présentés dans ce chapitre : l'un s'appuie sur les chaînes de Markov à espace d'états fini, l'autre sur un modèle gaussien.

12.1 Algorithme «forward-backward»

On modélise la structure d'un brin d'ADN [1] par une chaîne de caractères écrite dans un alphabet fini $\mathcal{A}$. Certaines parties de la chaîne correspondent à la suite des codes des acides aminés nécessaires à la fabrication d'une protéine, et on dit qu'elles sont codantes, tandis que d'autre parties de la chaîne ne sont pas codantes. Pour tenir compte de cette structure segmentée, on modélise la chaîne par une chaîne de Markov possédant deux régimes.

À un brin d'ADN de longueur l, noté $(X_1, \dots, X_l)$, et à valeurs dans $\mathcal{A}$, on adjoint le l-uplet $(U_1, \dots, U_l)$, à valeurs dans $\mathcal{U} = \{0, 1\}$. Si la position n est codante (respectivement. non codante) alors $U_n = 1$ (resp. $U_n = 0$). On modélise la loi de $(X, U) = ((X_n, U_n))_{n\in\mathbb{N}}$ par une chaîne de Markov homogène sur $\mathcal{A} \times \mathcal{U}$ de matrice de transition donnée par

$$P((x, u), (x', u')) := \mathbb{P}(X_{n+1} = x', U_{n+1} = u' \mid X_n = x, U_n = u) = \rho(u, u')\pi_{u'}(x, x'),$$

où ρ est une matrice de transition sur $\{0, 1\}$ et où π_0 et π_1 sont des matrice de transitions sur $\mathcal{A}$. On suppose que tous les coefficients des matrices ρ, π_0 et π_1 sont > 0. La composante U agit comme une sorte d'interrupteur ou de commutateur (markovien) à deux positions.

La chaîne (X, U) est irréductible, récurrente et apériodique car tous les coefficients de sa matrice de transition sont > 0. En général la projection $X = (X_n)_{n\in\mathbb{N}}$, qui modélise notre brin d'ADN, n'est pas une chaîne de Markov. Le modèle est rigide : la loi de la longueur d'un segment de la catégorie $u \in \mathcal{U}$ suit une loi géométrique de paramètre $1 - \rho(u, u)$.

On se place à présent dans la situation où la matrice de transition de (X, U) , c'est-à-dire le triplet ρ, π_0, π_1, est connue, par exemple grâce à une estimation menée au préalable (nous y

1. Acide désoxyribonucléique, base du code génétique de la plupart des organismes vivants.

reviendrons). La question est à présent la suivante : peut-on déterminer la loi de $(U_j)_{1\leqslant j\leqslant l}$ conditionnellement aux observations $(X_i)_{1\leqslant i\leqslant l}$? On dit que la chaîne de Markov (X, U) est cachée car la composante U n'est pas observée. Seule la composante X est observée.

Exemple 12.1. *Pour les applications numériques, les matrices suivantes sont utilisées :*

$$\rho = \begin{pmatrix} 0.95 & 0.05 \\ 0.1 & 0.9 \end{pmatrix}, \quad \pi_0 = \begin{pmatrix} 0.3 & 0.3 & 0.3 & 0.1 \\ 0.3 & 0.3 & 0.1 & 0.3 \\ 0.3 & 0.1 & 0.3 & 0.3 \\ 0.1 & 0.3 & 0.3 & 0.3 \end{pmatrix}, \quad \textit{et} \quad \pi_1 = \begin{pmatrix} 0.5 & 0.3 & 0.1 & 0.1 \\ 0.4 & 0.4 & 0.1 & 0.1 \\ 0.4 & 0.1 & 0.4 & 0.1 \\ 0.5 & 0.3 & 0.1 & 0.1 \end{pmatrix}.$$

Description de l'algorithme

On adopte la notation vectorielle $X_{1:i}$ et $x_{1:i}$ pour désigner les i-uplets $(X_1, \ldots, X_i)$ et $(x_1, \ldots, x_i)$. On définit les probabilités suivantes, pour tous $1 \leqslant i \leqslant l$, $x \in \mathcal{A}^l$, et $v \in \mathcal{U}$:

— *Probabilité de prédiction* : $P^i(v) := \mathbb{P}(U_i = v \,|\, X_{1:i-1} = x_{1:i-1})$. C'est la probabilité que l'état caché U_i soit en position v connaissant les observations $X_{1:i-1}$;

— *Probabilité de filtrage* : $F^i(v) := \mathbb{P}(U_i = v \,|\, X_{1:i} = x_{1:i})$. C'est la probabilité que l'état caché U_i soit en position v connaissant les observations $X_{1:i}$;

— *Probabilité de lissage* : $L^i(v) := \mathbb{P}(U_i = v \,|\, X_{1:l} = x_{1:l})$. C'est la probabilité que l'état caché U soit en position v connaissant les observations $X_{1:l}$.

Pour calculer les probabilités de lissage, on utilise un algorithme de type *forward-backward* (signifie progressif-rétrograde en anglais) : on calcule par récurrence, dans le sens croissant des indices, les probabilités de prédiction et de filtrage, puis on en déduit, toujours par récurrence mais descendante, les probabilités de lissage. Les relations de récurrence nécessaires à cet algorithme sont rassemblées dans le théorème suivant 12.2.

Théorème 12.2 (Relations de prédiction, filtrage, lissage)**.** *Pour tous* $1 \leqslant i \leqslant l$ *et* $v \in \mathcal{U}$,

$$\begin{aligned} P^i(v) &= \sum_{u\in\mathcal{U}} \rho(u,v) F^{i-1}(u), \\ F^i(v) &= \frac{\pi_v(x_{i-1}, x_i) P^i(v)}{\sum_{u\in\mathcal{U}} \pi_u(x_{i-1}, x_i) P^i(u)}, \\ L^{i-1}(u) &= F^{i-1}(u) \sum_{v\in\mathcal{U}} \rho(u,v) \frac{L^i(v)}{P^i(v)}. \end{aligned}$$

Démonstration. Commençons par l'équation de prédiction. L'idée est de faire intervenir U_{i-1} puis d'utiliser le fait que les lois $\mathcal{L}(U_i \,|\, U_{i-1}, X_{1:i-1})$ et $\mathcal{L}(U_i \,|\, U_{i-1})$ coïncident :

$$\begin{aligned} P^i(v) &= \sum_{u\in\mathcal{U}} \mathbb{P}(U_{i-1} = u, U_i = v \,|\, X_{1:i-1} = x_{1:i-1}) \\ &= \sum_{u\in\mathcal{U}} \mathbb{P}(U_i = v \,|\, U_{i-1} = u, X_{1:i-1} = x_{1:i-1}) \mathbb{P}(U_{i-1} = u \,|\, X_{1:i-1} = x_{1:i-1}) \\ &= \sum_{u\in\mathcal{U}} \rho(u,v) F^{i-1}. \end{aligned}$$

Pour l'équation de filtrage, on utilise de plus la relation $\mathbb{P}(A \,|\, B \cap C)\mathbb{P}(B \,|\, C) = \mathbb{P}(A \cap B \,|\, C)$:

$$\begin{aligned} F^i(v) &= \frac{\mathbb{P}(U_i = v, X_i = x_i \,|\, X_{1:i-1} = x_{1:i-1})}{\mathbb{P}(X_i = x_i \,|\, X_{1:i-1} = x_{1:i-1})} \\ &= \frac{\mathbb{P}(U_i = v, X_i = x_i \,|\, X_{1:i-1} = x_{1:i-1})}{\sum_{u\in\mathcal{U}} \mathbb{P}(U_i = u, X_i = x_i \,|\, X_{1:i-1} = x_{1:i-1})} \\ &= \frac{\mathbb{P}(X_i = x_i \,|\, U_i = v, X_{1:i-1} = x_{1:i-1}) \mathbb{P}(U_i = v, \,|\, X_{1:i-1} = x_{1:i-1})}{\sum_{u\in\mathcal{U}} \mathbb{P}(X_i = x_i \,|\, U_i = u, X_{1:i-1} = x_{1:i-1}) \mathbb{P}(U_i = u \,|\, X_{1:i-1} = x_{1:i-1})}. \end{aligned}$$

Recueil de modèles stochastiques – Page 124/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Pour conclure, il suffit de remarquer que par définition de la matrice de transition de (X, U),

$$\mathbb{P}(X_i = x_i \,|\, U_i = u, X_{1:i-1} = x_{1:i-1}) = \mathbb{P}(X_i = x_i \,|\, U_i = u, X_{i-1} = x_{i-1}) = \pi_{u_{i-1}}(x_i - 1, x_i).$$

Pour l'équation de lissage, on écrit

$$\mathbb{P}(U_{i-1} = u, U_i = v \,|\, X_{1:l} = x_{1:l}) = \mathbb{P}(U_{i-1} = u \,|\, U_i = v, X_{1:l} = x_{1:l})\mathbb{P}(U_i = v \,|\, X_{1:l} = x_{1:l}),$$

puis

$$\mathbb{P}(U_{i-1} = u \,|\, U_i = v, X_{1:l} = x_{1:l}) = \frac{\mathbb{P}(U_{i-1} = u, X_{1:i-1} = x_{1:i-1}, X_{i+1:l} = x_{i+1:l} \,|\, U_i = v, X_i = x_i)}{\mathbb{P}(X_{1:i-1} = x_{1:i-1}, X_{i+1:l} = x_{i+1:l} \,|\, U_i = v, X_i = x_i)}.$$

La propriété de Markov assure que, conditionnellement au présent, passé et futur sont indépendant. On obtient donc après simplification que

$$\mathbb{P}(U_{i-1} = u \,|\, U_i = v, X_{1:l} = x_{1:l}) = \mathbb{P}(U_{i-1} = u \,|\, U_i = v, X_{1:i} = x_{1:i}).$$

On écrit alors à nouveau, grâce à l'expression de la matrice de transition de (X, U),

$$\begin{aligned}\mathbb{P}(U_{i-1} = u \,|\, U_i = v, X_{1:i} = x_{1:i}) &= \frac{\mathbb{P}(U_{i-1} = u, U_i = v, X_i = x_i, \,|\, X_{1:i-1} = x_{1:i-1})}{\mathbb{P}(U_i = v, X_i = x_i \,|\, X_{1:i-1} = x_{1:i-1})} \\ &= \frac{F^{i-1}(u)\rho(u, v)\pi_v(x_{i-1}, x_i)}{P^i(v)\pi_v(x_{i-1}, x_i)}.\end{aligned}$$

On a donc obtenu

$$\mathbb{P}(U_{i-1} = u, U_i = v \,|\, X_{1:l} = x_{1:l}) = F^{i-1}(u)\rho(u, v)\frac{L^i(v)}{P^i(v)},$$

ce qui fournit la relation de lissage en sommant sur v. □

On initialise l'algorithme en choisissant pour $P^1(u)$ la loi initiale (par exemple la loi stationnaire). Les relations de prédiction et filtrage du théorème 12.2 permettent de calculer toutes les probabilités P^i et F^i par récurrence progressive. On en déduit ensuite les probabilités L^i par récurrence rétrograde grâce à la relation de lissage du théorème 12.2 avec l'initialisation $L^l(v) = F^l(v)$. La figure 12.1 illustre l'efficacité de l'algorithme : la plupart du temps la probabilité $L^i(v)$ est supérieure à 1/2 quand $U_i = v$ (dans 95,2% des cas exactement dans cet exemple). On voit toutefois dans cet exemple que l'algorithme rate une zone où U vaut 1 (aux alentours de 120) et qu'il a toujours un peu de retard aux changements de régimes. L'algorithme fonctionne d'autant mieux que les matrices π_0 et π_1 sont différentes (les deux régimes ont des comportements très différents) et que $\rho(0, 0)$ et $\rho(1, 1)$ sont grands (les plages entre deux changements de régime sont longues).

Estimation d'une matrice de transition

L'étude précédente suppose connues les matrices de transition. Il est toutefois important de savoir estimer la matrice de transition d'une chaîne de Markov à partir d'une de ses trajectoires. On considère une chaîne de Markov $(Z_n)_{n\in\mathbb{N}}$ sur l'espace d'états fini E de matrice de transition $\theta_* = (\theta_*(i, j))_{(i,j)\in E^2}$. On suppose que la chaîne est irréductible et récurrente, et on note μ_* sa loi invariante. On suppose la mesure initiale ν connue mais la matrice de transition θ_* inconnue. On note Θ l'ensemble des matrices de transitions sur E. Soit λ la mesure de comptage sur E. La loi de $(Z_0, \ldots, Z_n)$ a pour densité par rapport $\lambda^{\otimes n+1}$ la fonction $(z_0, \ldots, z_n) \in E^{n+1} \mapsto f_{\theta_*}(z_0, \ldots, z_n)$ donnée par

$$f_{\theta_*}(z_0, \ldots, z_n) = \mathbb{P}(Z_0 = z_0, \ldots, Z_n = z_n) = \nu(z_0)\theta_*(z_0, z_1) \cdots \theta_*(z_{n-1}, z_n).$$

Recueil de modèles stochastiques – Page 125/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

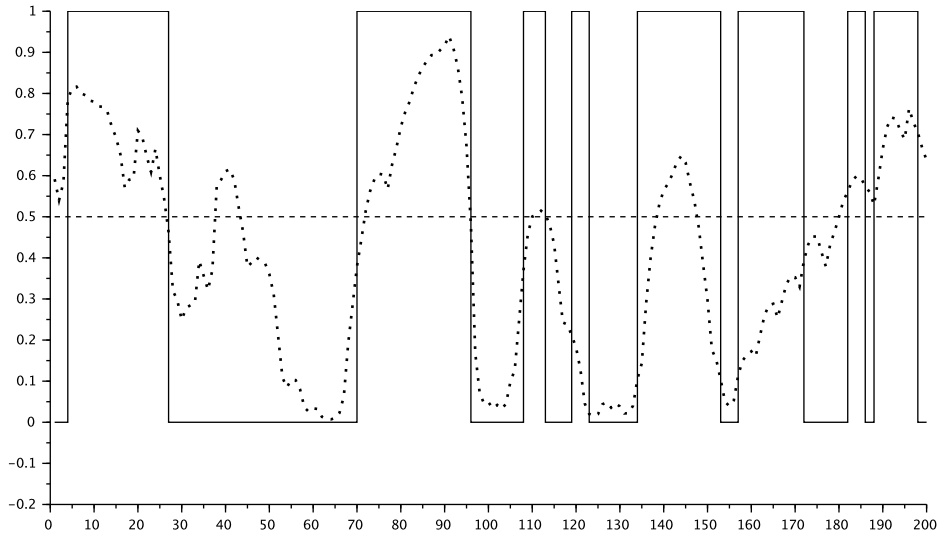


Figure 12.1 – Trajectoire de U (courbe continue) et probabilité de lissage pour $v = 0$ (courbe en pointillés) avec $l = 200$ et les transitions de l'exemple 12.1.

Pour tous $\theta \in \Theta$ et $z \in E^{n+1}$ on note $f_\theta(z_0, \ldots, z_n) := \nu(z_0)\theta(z_0, z_1) \cdots \theta(z_{n-1}, z_n)$. La quantité aléatoire $L_{Z_1,\ldots,Z_n}(\theta) := f_\theta(Z_0, \ldots, Z_n)$ est appelée *vraisemblance*[2] de θ pour l'échantillon $(Z_0, \ldots, Z_n)$. L'*estimateur de maximum de vraisemblance* $\widehat{\theta}_n$ de θ_* est par définition l'élément aléatoire de Θ qui maximise la fonction aléatoire $\theta \in \Theta \mapsto L_{Z_0,\ldots,Z_n}(\theta)$.

Lemme 12.3 (Maximum de vraisemblance)**.** *L'estimateur de maximum de vraisemblance $\widehat{\theta}_n$ de θ_* est donné pour tous $i, j \in E$ par*

$$\widehat{\theta}_n(i,j) = \begin{cases} \dfrac{N_n^{ij}}{N_n^i} & \text{si } N_n^i > 0, \\ 0 & \text{sinon,} \end{cases} \quad \text{où} \quad \begin{cases} N_n^i & := \sum_{p=0}^{n-1} \mathbf{1}_{\{Z_p = i\}}, \\ N_n^{ij} & := \sum_{p=0}^{n-1} \mathbf{1}_{\{Z_p = i, Z_{p+1} = j\}}. \end{cases}$$

Notons que si $\theta(i,j) = 0$ alors $\widehat{\theta}_n(i,j) = 0$, mais que la réciproque est fausse.

Démonstration. On numérote E de sorte que $E = \{1, \ldots, s\}$. Notons $L := L_{Z_1,\ldots,Z_n}$. Il est plus commode de maximiser $\log(L)$ plutôt que L. On doit maximiser $\theta \in \Theta \mapsto \log(L(\theta))$ sous les contraintes $\theta(i,j) \geqslant 0$ et $\sum_{k=1}^{s} \theta(i,k) = 1$ pour tous $1 \leqslant i, j \leqslant s$. Or on a

$$\partial_{\theta(i,j)}(\theta \mapsto \log(L(\theta))) = \frac{N_n^{ij}}{\theta(i,j)},$$

et les s conditions d'extrémalité reviennent à dire que pour tout $1 \leqslant i \leqslant s$ la fonction $j \mapsto N_n^{ij}/\theta(i,j)$ doit être constante. Ce qui donne le résultat en tenant compte de $\theta \in \Theta$. $\square$

L'estimateur $\hat{\theta}$ est intuitif puisqu'il estime $\theta(i,j)$ par la proportion de transitions de l'état i vers l'état j parmi les n sauts observés. Les propriétés asymptotiques de cet estimateur se déduisent du comportement des suites $(N_n^{ij})_n$ et $(N_n^i)_n$ qui sont établies ci-dessous.

2. En anglais, on dit *likelihood*, d'où la notation L.

Recueil de modèles stochastiques – Page 126/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Théorème 12.4 (Convergence et normalité asymptotique)**.** *Pour tout $x \in E$, sachant $\{X_0 = x\}$, l'estimateur $\widehat{\theta}_n$ est convergent et asymptotiquement normal : pour tous $i, j \in E$,*

$$\widehat{\theta}_n(i,j) \xrightarrow[n\to\infty]{p.s.} \theta(i,j) \quad \textit{et} \quad \sqrt{n\mu(i)}(\widehat{\theta}_n(i,j) - \theta(i,j)) \xrightarrow[n\to\infty]{\text{loi}} \mathcal{N}(0, \theta(i,j)(1-\theta(i,j))).$$

Démonstration. Les résultats découlent du fait que pour tous $x, i, j \in E$, sachant $\{X_0 = x\}$,

$$\frac{1}{n} N_n^i \xrightarrow[n\to\infty]{\text{p.s.}} \mu(i) \quad \text{et} \quad \frac{1}{n} N_n^{ij} \xrightarrow[n\to\infty]{\text{p.s.}} \mu(i)\theta(i,j)$$

tandis que

$$\frac{1}{\sqrt{n}}\big(N_n^{ij} - N_n^i \theta(i,j)\big) \xrightarrow[n\to\infty]{\text{loi}} \mathcal{N}(0, \mu(i)\theta(i,j)(1-\theta(i,j))).$$

Ces convergences p.s. découlent du *théorème ergodique* (*loi des grands nombres*) pour les chaînes Z et Y, où Y est la chaîne de Markov irréductible et récurrente d'espace d'états E^2 définie par $Y_n = (Z_n, Z_{n+1})$ et dont la loi invariante est $(i,j) \mapsto \mu(i)\theta(i,j)$. La convergence en loi découle du théorème limite central pour ces chaînes de Markov ergodiques. □

12.2 Filtre de Kalman

Un avion se déplace entre Paris et Londres, en tentant de suivre une trajectoire théorique définie par le plan de vol. L'avion est surveillé au sol par des contrôleurs aériens grâce à un radar qui reçoit un écho de l'avion à intervalles réguliers. La trajectoire effective de l'avion s'écarte de la trajectoire théorique pour de multiples raisons (météorologie, imprécision du pilote automatique, turbulences,...). On cherche donc à localiser l'avion au cours de son vol à partir des observations radars successives.

On note X_n l'écart entre la trajectoire théorique et la position de l'avion au temps n. Cet écart n'est pas connu. De plus, on note Y_n la mesure donnée par le radar au temps n. Cette mesure est entachée d'erreurs à cause de l'imprécision du radar. Le problème qui se pose à l'aiguilleur est d'estimer au mieux la position de l'avion au temps n au vu des observations $Y_0, \ldots, Y_n$. Pour simplifier l'étude, on supposera que l'objet observé évolue dans un espace de dimension 1. On modélise les suites aléatoires $(X_n)_{n\in\mathbb{N}}$ et $(Y_n)_{n\in\mathbb{N}}$ en posant

$$\begin{cases} X_0 = W_0, & \\ X_n = aX_{n-1} + W_n & n \geqslant 1, \\ Y_n = X_n + V_n & n \geqslant 0, \end{cases}$$

où a est un nombre réel déterministe, et où $(V_n)_{n\in\mathbb{N}}$ et $(W_n)_{n\in\mathbb{N}}$ sont des suites aléatoires indépendantes, avec $(V_n)_{n\in\mathbb{N}}$ i.i.d. de loi gaussienne $\mathcal{N}(0, \tau^2)$ et $(W_n)_{n\in\mathbb{N}}$ i.i.d. de loi gaussienne $\mathcal{N}(0, \sigma^2)$. Les variables aléatoires $(W_n)_{n\in\mathbb{N}}$ représentent les fluctuations instantanées de l'écart entre position théorique et position réelle. Les variables aléatoires $(V_n)_{n\in\mathbb{N}}$ modélisent les erreurs de mesure du radar. Le paramètre a modélise l'action du pilote.

En théorie du signal, on dit que $(X_n)_{n\in\mathbb{N}}$ est un *processus auto-régressif* d'ordre 1, noté AR(1), à bruit gaussien, et on a $X_n = \sum_{k=0}^n a^k W_{n-k} = \sum_{k=0}^n a^{n-k} W_k$ pour tout $n \in \mathbb{N}$.

À présent, le problème est d'estimer X_n sachant les observations $Y_0, \ldots, Y_n$. Le candidat naturel est l'espérance conditionnelle $\mathbb{E}(X_n \,|\, Y_0, \ldots, Y_n)$ puisque c'est la fonction des observations qui est la plus proche de X_n dans L^2 (moindre carrés). Or le caractère gaussien du modèle permet, grâce à la mécanique des vecteurs gaussiens, de calculer explicitement la loi conditionnelle $\mathcal{L}(X_n \,|\, Y_0, \ldots, Y_n)$, en particulier sa moyenne qui est $\mathbb{E}(X_n \,|\, Y_0, \ldots, Y_n)$.

La première remarque importante est que pour tout $n \geqslant 1$, le vecteur aléatoire

$$(X_0, \ldots, X_n, Y_0, \ldots, Y_n)$$

Recueil de modèles stochastiques – Page 127/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

est un vecteur aléatoire gaussien puisque toute combinaison linéaire de ses coordonnées est une combinaison linéaire des variables aléatoires gaussiennes indépendantes $(W_i)_{0\leqslant i\leqslant n}$ et $(V_i)_{0\leqslant i\leqslant n}$. En particulier, pour tout $n \geqslant 0$, $(X_n, Y_0, \dots, Y_n)$ est aussi un vecteur gaussien. De plus, la loi de X_n sachant $(Y_0, \dots, Y_n)$ est une loi gaussienne, dont on note la moyenne $\hat{X}_n$ et la variance P_n. On se propose de calculer ces quantités par récurrence. Comme pour les chaînes de Markov à espace d'états fini de la première partie du chapitre, la récurrence se fait en deux étapes. L'étape de prédiction consiste à exprimer la loi $\mathcal{L}(X_n \,|\, Y_0, \dots, Y_{n-1})$ en fonction de $\mathcal{L}(X_{n-1} \,|\, Y_0, \dots, Y_{n-1})$. Puis, dans l'étape de filtrage, on prend en compte l'observation Y_n pour exprimer $\mathcal{L}(X_n \,|\, Y_0, \dots, Y_n)$ en fonction de $\mathcal{L}(X_n \,|\, Y_0, \dots, Y_{n-1})$.

Lemme 12.5 (Lois conditionnelles)**.** *On a* $\mathcal{L}(X_n \,|\, Y_0, \dots, Y_n) = \mathcal{N}(X_n, P_n)$ *où*

$$\hat{X}_n = a\hat{X}_{n-1} + \frac{P_n}{\tau^2}(Y_n - a\hat{X}_{n-1}) \quad \textit{et} \quad P_n = \frac{a^2\tau^2 P_{n-1} + \sigma^2\tau^2}{a^2 P_{n-1} + \sigma^2 + \tau^2}.$$

Démonstration. Tout d'abord, on rappelle le résultat suivant sur les vecteurs gaussiens, que nous appelons *formule de Bayes conditionnelle* : si $(X, Y_0, \dots, Y_{n-1})$ est un vecteur gaussien dans $\mathbb{R}^{n+1}$ avec $\mathcal{L}(X \,|\, Y_0, \dots, Y_{n-1}) = \mathcal{N}(\mu, \gamma^2)$ et $\mathcal{L}(Y \,|\, Y_0, \dots, Y_{n-1}, X) = \mathcal{N}(X, \delta^2)$ alors

$$\mathcal{L}(X \,|\, Y_0, \dots, Y_{n-1}, Y) = \mathcal{N}\left(\rho^2\left(\frac{\mu}{\gamma^2} + \frac{Y}{\delta^2}\right), \rho^2\right) \quad \text{où} \quad \frac{1}{\rho^2} := \frac{1}{\gamma^2} + \frac{1}{\delta^2}.$$

À présent, et comme annoncé, on procède par récurrence, en trois étapes.

— **Initialisation.** Puisque $Y_0 = X_0 + V_0$, on a $\mathcal{L}(Y_0 \,|\, X_0) = \mathcal{N}(X_0, \tau^2)$. La formule de Bayes conditionnelle assure que $\mathcal{L}(X_0 \,|\, Y_0) = \mathcal{N}(\hat{X}_0, P_0)$ où

$$\hat{X}_0 = \frac{\sigma^2}{\sigma^2 + \tau^2} Y_0 \quad \text{et} \quad P_0 = \frac{\sigma^2\tau^2}{\sigma^2 + \tau^2}.$$

— **Prédiction.** On a $\mathcal{L}(X_{n-1} \,|\, Y_0, \dots, Y_{n-1}) = \mathcal{N}(\hat{X}_{n-1}, P_{n-1})$, et grâce au modèle,

$$\mathcal{L}(X_n \,|\, Y_0, \dots, Y_{n-1}) = \mathcal{N}(a\hat{X}_{n-1}, a^2 P_{n-1} + \sigma^2).$$

— **Filtrage.** D'après le modèle à nouveau, il vient

$$\mathcal{L}(Y_n \,|\, Y_0, \dots, Y_{n-1}, X_n) = \mathcal{N}(X_n, \tau^2).$$

On applique alors la formule de Bayes conditionnelle pour inverser le conditionnement entre Y_n et X_n : la loi $\mathcal{L}(X_n \,|\, Y_0, \dots, Y_n) = \mathcal{N}(\hat{X}_n, P_n)$ avec

$$\frac{1}{P_n} = \frac{1}{a^2 P_{n-1} + \sigma^2} + \frac{1}{\tau^2} \quad \text{et} \quad \hat{X}_n = P_n\left(\frac{a\hat{X}_{n-1}}{a^2 P_{n-1} + \sigma^2} + \frac{Y_n}{\tau^2}\right).$$

□

Remarque 12.6 (Gain suite aux observations)**.** *Un petit calcul permet d'obtenir la majoration* $P_n = \mathbb{E}\Big[(X_n - \hat{X}_n)^2\Big] \leqslant \mathbb{E}\big[(X_n - Y_n)^2\big] = \tau^2$. *Ceci n'est pas surprenant connaissant les propriétés de l'espérance conditionnelle mais souligne bien que l'on gagne effectivement à utiliser toutes les observations* $Y_0, \dots, Y_n$ *plutôt que de se contenter de la dernière* Y_n.

Estimation de certains paramètres du modèle

On suppose dans cette section qu'on observe les positions de l'avion sans erreurs, c'est-à-dire qu'on a accès à la suite $(X_i)_{1\leqslant i\leqslant n}$. On souhaite estimer les coefficients a et σ^2. La question n'est pas complètement évidente car les observations ne sont pas indépendantes.

Recueil de modèles stochastiques – Page 128/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Lemme 12.7 (Estimateurs de maximum de vraisemblance)**.** *L'estimateur de maximum de vraisemblance* $(\hat{a}, \hat{\sigma})$ *de* (a, σ) *est donné par*

$$\hat{a}_n = \frac{\sum_{k=1}^n X_{k-1}X_k}{\sum_{k=1}^n X_{k-1}^2} \quad \textit{et} \quad \hat{\sigma}_n^2 = \frac{1}{n}\sum_{k=1}^n (X_k - \hat{a}_n X_{k-1})^2.$$

Démonstration. Il suffit de maximiser le logarithme de la vraisemblance

$$L_{(X_1,\ldots,X_n)}(a,\sigma) = \frac{1}{(2\pi\sigma^2)^{n/2}} \prod_{k=1}^n \exp\left(-\frac{(X_k - aX_{k-1})^2}{2\sigma^2}\right).$$

□

Théorème 12.8 (Convergence et normalité asymptotique)**.** *Si* $-1 < a < 1$ *alors*

$$(\hat{a}_n, \hat{\sigma}_n) \xrightarrow[n\to\infty]{p.s.} (a,\sigma) \quad \textit{et} \quad \sqrt{n}(\hat{a} - a) \xrightarrow[n\to\infty]{\text{loi}} \mathcal{N}(0, 1-a^2) \quad \textit{et} \quad \sqrt{n}\big(\hat{\sigma}^2 - \sigma^2\big) \xrightarrow[n\to\infty]{\text{loi}} \mathcal{N}(0, 2\sigma^4).$$

Démonstration. On se contente d'établir les résultats sur $\hat{a}$, en supposant σ connu et fixé. On réécrit $\hat{a}_n$ de la manière suivante :

$$\hat{a}_n = a + \frac{\sum_{k=1}^n X_{k-1}W_k}{\sum_{k=1}^n X_{k-1}^2} = a + \frac{M_n}{\sum_{k=1}^n X_{k-1}}$$

où $M_n = \sum_{k=1}^n X_{k-1}W_k$, avec $M_0 = 0$. La suite $(M_n)_{n\in\mathbb{N}}$ est une martingale par rapport à la filtration $(\mathcal{F}_n)_n$ naturelle de W. Rappelons que la variation quadratique $(\langle M\rangle_n)_n$ de la martingale $(M_n)_{n\in\mathbb{N}}$ est la seule suite $(A_n)_{n\in\mathbb{N}}$ de variables aléatoires telle que $(M_n^2 - A_n)_{n\in\mathbb{N}}$ soit une martingale par rapport à $(\mathcal{F}_n)$. Or on a

$$\langle M\rangle_0 = 0 \quad \text{et, pour } n \geqslant 1, \quad \langle M\rangle_n = \sigma^2 \sum_{k=1}^n X_{k-1}^2.$$

À présent, on vérifie tout d'abord, en utilisant le fait que X est un AR(1) avec $|a| < 1$, que

$$\frac{\langle M\rangle_n}{n} \xrightarrow[n\to\infty]{\text{p.s.}} \frac{\sigma^4}{(1-a^2)},$$

(on a en particulier $\langle M\rangle_n \to +\infty$ p.s. quand $n \to \infty$), puis on utilise la loi forte des grands nombres et le *théorème limite central pour les martingales de carré intégrable*, qui donnent

$$\frac{M_n}{\langle M\rangle_n} \xrightarrow[n\to\infty]{\text{p.s.}} 0 \quad \text{et} \quad \frac{M_n}{\sqrt{\langle M\rangle_n}} \xrightarrow[n\to\infty]{\text{loi}} \mathcal{N}(0,1).$$

□

12.3 Notes et commentaires

L'algorithme de segmentation «forward-backward» remonte au moins aux travaux des années 1960 de Leonard Baum et Lloyd Welch, et peut être vu comme une instance du concept général de programmation dynamique. Les chaînes de Markov cachées ont été utilisées notamment pour la reconnaissance de la parole dans les années 1970, et pour la génomique à partir des années 1980. Le livre de Stéphane Robin, François Rodolphe, et Sophie Schbath [RRS05] propose une introduction accessible à l'utilisation des chaînes de Markov cachées en génomique. Bien que les modèles utilisés en pratique soient plus sophistiqués que celui présenté dans ce chapitre, notamment en ce qui concerne les espaces d'état $\mathcal{A}$ et $\mathcal{U}$, ils font

Recueil de modèles stochastiques – Page 129/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

appel aux mêmes concepts et outils. Cependant en pratique, on ne connaît pas en général les matrices de transition ni même le nombre d'états cachés pertinent pour rendre compte de la loi de la séquence, et il faut donc construire des algorithmes qui permettent en plus d'estimer ces paramètres de complexité. On pourra également consulter le livre de Jean-François Delmas et Benjamin Jourdain [DJ06] à ce sujet. D'autre part, il est possible de tester si une suite aléatoire est markovienne ou pas en utilisant par exemple N^i et N^{ij} et le test du χ^2, comme expliqué par exemple dans le livre de Didier Dacunha-Castelle et Marie Duflo [DCD83].

La partie sur le filtre de Kalman est inspirée du livre de David Williams [Wil91]. Le filtre de Kalman, présenté ici dans une version simple, est un grand classique de la théorie du signal, développé dès les années 1960 par Thorvald Nicolai Thiele et Peter Swerling, et par Rudolf Kalman et Richard Bucy, notamment pour les besoins du programme Apollo de la National Aeronautics and Space Administration. On trouvera dans le livre de Étienne Pardoux [Par07] l'expression du filtre de Kalman en dimension supérieure à 1. Enfin, la *loi forte des grands nombres* et le *théorème limite central pour les martingales* utilisé dans la preuve du théorème 12.8 sont tirés du livre [BC07].

Recueil de modèles stochastiques – Page 130/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Chapitre 13

Algorithme EM et mélanges

Recueil de modèles stochastiques – Page 131/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Mots-clés. Maximum de vraisemblance ; entropie relative ; formule de Bayes.

13.1 Nids de mouettes

En Bretagne cohabitent quatre espèces de mouettes différentes. Les ornithologues souhaitent estimer la proportion de mouettes de chaque espèce à partir de l'observation de la taille des nids. Contrairement aux oiseaux, les nids ne bougent pas, ce qui facilite le comptage. Malheureusement, les différentes espèces font des nids assez ressemblants : on ne sait pas par quelles espèces ils ont été construits :

> «rien ne ressemble plus à un nid de mouette, qu'un autre nid de mouette» [1].

On suppose en revanche que la distribution des nids est caractéristique d'une espèce. La distribution globale de la taille des nids apparaît comme le mélange de quatre lois de probabilité, chacune rendant compte de la répartition de la taille des nids pour chaque espèce d'oiseaux. Modélisons la distribution de la taille des nids construits par l'espèce j par une loi $\mu_j = \mathcal{N}(m(j), v(j))$ de densité $\gamma_{m(j),v(j)}$. La loi μ de la taille des nids de mouettes admet donc pour densité la fonction

$$x \mapsto \sum_{j=1}^{J} \alpha(j)\gamma_{m(j),v(j)}(x) = \sum_{j=1}^{J} \frac{\alpha(j)}{\sqrt{2\pi v(j)}} \exp\left(-\frac{(x-m(j))^2}{2v(j)}\right),$$

avec $\alpha_1 > 0, \ldots, \alpha_J > 0$ et $\alpha_1 + \cdots + \alpha_J = 1$. Cette combinaison convexe finie de densités de probabilité est appelée *mélange fini* [2] de populations. Si X désigne la taille du nid et Z l'espèce de l'oiseau qui l'a construit, la loi du couple (X, Z) est donnée de la manière suivante :

— la variable Z suit la loi discrète sur $\{1, \ldots, J\}$ de poids $\alpha(1), \ldots, \alpha(J)$:

$$\mathcal{L}(Z) = \sum_{j=1}^{J} \alpha(j)\delta_j$$

— pour tout $1 \leqslant j \leqslant J$, conditionnellement à $\{Z = j\}$, la variable X suit la loi $\mathcal{N}(m(j), v(j))$:

$$\mathcal{L}(X|Z = j) = \mathcal{N}(m(j), v(j)).$$

1. Poète breton anonyme.
2. «*Finite mixture*» en anglais.

Les paramètres $(\alpha_j, m(j), v(j))_{1\leqslant j\leqslant J}$ sont inconnus. Le vecteur α représente les coefficients du mélange tandis que les vecteurs m et v représentent respectivement les vecteurs des moyennes et des variances des lois gaussiennes. On souhaite estimer toutes ces quantités à la seule vue de la taille de n nids. Pour cela, on suppose qu'il existe θ dans l'ensemble

$$\Theta = \left\{\theta = (\alpha_j, m_j, \sigma_j^2)_{1\leqslant j\leqslant J} : \alpha_1 > 0, \ldots, \alpha_J > 0, \alpha_1 + \cdots + \alpha_J = 1\right\}$$

tel que les mesures $x_1, \ldots, x_n$ de la taille de n nids soient la réalisation d'un n-échantillon $X_1, \ldots, X_n$ de loi de densité

$$f_\theta(x) = \sum_{j=1}^{J} \alpha(j)\gamma_{m(j),v(j)}(x).$$

On note $\overline{X}$ le vecteur $(X_1, \ldots, X_n)$ et $\overline{Z}$ le vecteur $(Z_1, \ldots, Z_n)$. On introduit également les notations suivantes :

— densité au point (x, z) de la loi du couple (X, Z) par rapport à la mesure $d\lambda \otimes dN$:

$$h_\theta(x, z) = \alpha(z)\gamma_{m(z),v(z)}(x), \quad x \in \mathbb{R},\ z \in \{1, \ldots, J\},$$

— densité au point x de la loi de X par rapport à $d\lambda$:

$$f_\theta(x) = \sum_{j=1}^{J} \alpha(j)\gamma_{m(j),v(j)}(x), \quad x \in \mathbb{R},$$

— densité au point z de la loi de Z par rapport à la mesure de comptage dN sur $\mathbb{N}$:

$$g_\theta(z) = \alpha(z), \quad z \in \{1, \ldots, J\},$$

— densité au point x de la loi de X sachant Z par rapport à $d\lambda$:

$$f_\theta(x|Z = j) = \gamma_{m(j),v(j)}(x), \quad x \in \mathbb{R},$$

— densité au point z de la loi de Z sachant X par rapport à dN :

$$g_\theta(z|X = x) = \frac{\alpha(z)\gamma_{m(z),v(z)}(x)}{\sum_{j=1}^{J} \alpha(j)\gamma_{m(j),v(j)}(x)}, \quad z \in \{1, \ldots, J\}.$$

La dernière expression s'obtient grâce à la définition d'une probabilité conditionnelle : en effet, pour tout A borélien et tout $1 \leqslant z \leqslant J$,

$$\mathbb{P}(X \in A,\ Z = z) = \mathbb{P}(Z = z)\mathbb{P}(X \in A|Z = z) = \alpha(z) \int_A \gamma_{m(z),v(z)}(x)\, dx$$

et

$$\begin{aligned}\mathbb{P}(X \in A,\ Z = z) &= \mathbb{P}(Z = z|X \in A)\mathbb{P}(X \in A)\\ &= \mathbb{P}(Z = z|X \in A) \sum_{j=1}^{J} \alpha(j) \int_A \gamma_{m(j),v(j)}(x)\, dx,\end{aligned}$$

de sorte que si A est de la forme $[x - \varepsilon, x + \varepsilon]$ avec ε tendant vers 0, alors on obtient bien

$$\alpha(z)\gamma_{m(z),v(z)}(x) = \mathbb{P}(Z = z|X = x) \sum_{j=1}^{J} \alpha(j)\gamma_{m(j),v(j)}(x).$$

Recueil de modèles stochastiques – Page 132/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

13.2 Situation favorable mais irréaliste

On suppose dans cette section que l'on observe à la fois X et Z, c'est-à-dire à la fois la taille du nid et l'espèce qui a fait le nid. Estimer les paramètres inconnus est alors aisé grâce à la méthode du maximum de vraisemblance. La log-vraisemblance du modèle complet (logarithme de la densité de la loi de l'échantillon $(\overline{X}, \overline{Z})$ par rapport à la mesure $(d\lambda \otimes dN)^{\otimes n}$) s'écrit

$$\begin{aligned} L(\theta, \overline{X}, \overline{Z}) &= \log \prod_{i=1}^{n} h_\theta(X_i, Z_i) \\ &= \sum_{i=1}^{n} \left[\log \alpha(Z_i) + \log \gamma_{m(Z_i), v(Z_i)}(X_i)\right]. \end{aligned}$$

Pour un échantillon de taille n, notons, pour tout $1 \leqslant j \leqslant J$,

$$A_j = \{i = 1, \ldots, n,\ Z_i = j\} \quad \text{et} \quad C_j = \operatorname{card}(A_j).$$

En regroupant les termes en fonction des valeurs possibles de Z, on obtient

$$L(\theta, \overline{X}, \overline{Z}) = \sum_{j=1}^{J} C_j \log \alpha(j) + \sum_{j=1}^{J} \sum_{i \in A_j} \log \gamma_{m(j), v(j)}(X_i).$$

Théorème 13.1 (Estimation pour modèle complet)**.** *La vraisemblance complète atteint son maximum en un point unique $\hat{\theta} \in \Theta$ donné par :*

$$\hat{\alpha}(j) = \frac{C_j}{n}, \quad \hat{m}(j) = \frac{1}{C_j} \sum_{i \in A_j} X_i \quad \textit{et} \quad \hat{v}(j) = \frac{1}{C_j} \sum_{i \in A_j} (X_i - \hat{m}(j))^2.$$

Démonstration. Points critiques de $\theta \mapsto L(\theta, \overline{X}, \overline{Z})$ sous la contrainte $\alpha(1) + \cdots + \alpha(J) = 1$. □

Ce résultat ne répond pas à la question initiale puisque, en pratique, ne sont observées que les tailles des nids $x_1, \ldots, x_n$. Il faudrait plutôt étudier la log-vraisemblance de l'échantillon $X_1, \ldots, X_n$, notée L_{obs} pour *log-vraisemblance des observations*, qui s'écrit :

$$L_{obs}(\theta, \overline{X}) = \log \prod_{i=1}^{n} f_\theta(X_i) = \sum_{i=1}^{n} \log \left[\sum_{j=1}^{J} \alpha(j) \gamma_{m(j), v(j)}(X_i)\right].$$

Trouver un jeu de paramètres θ rendant cette quantité maximale n'est pas facile.

13.3 Résolution du vrai problème

Ne disposant pas de la log-vraisemblance complète L, on la remplace par son espérance conditionnelle sachant les observations : on définit la *log-vraisemblance conditionnelle des observations* sous la loi de paramètre $\tilde{\theta}$, que nous noterons $L_c(\theta, \tilde{\theta}, \overline{X})$, par :

$$L_c(\theta, \tilde{\theta}, \overline{X}) = \mathbb{E}_{\tilde{\theta}}(L(\theta, \overline{X}, \overline{Z}) | \overline{X}) = \sum_{i=1}^{n} \int \log h_\theta(X_i, z) g_{\tilde{\theta}}(z | X = X_i)\, dz.$$

Voici l'expression de la vraisemblance conditionnelle L_c dans notre cadre de mélange gaussien.

Recueil de modèles stochastiques – Page 133/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Théorème 13.2 (Log-vraisemblance conditionnelle)**.** *La fonction L_c est de la forme suivante :*

$$L_c(\theta, \tilde{\theta}, \overline{X}) = -\frac{n}{2}\log(2\pi) + \sum_{j=1}^{J}\left(\sum_{i=1}^{n} g_{\tilde{\theta}}(j|X = X_i)\right)\log\alpha(j)$$
$$-\frac{1}{2}\sum_{j=1}^{J}\left(\sum_{i=1}^{n} g_{\tilde{\theta}}(j|X = X_i)\right)\left(\log v(j) + \frac{(X_i - m(j))^2}{v(j)}\right).$$

Il est aisé de déterminer le maximum de la vraisemblance conditionnelle.

Théorème 13.3 (Maximum de la log-vraisemblance conditionnelle)**.** *La fonction $\theta \mapsto L_c(\theta, \tilde{\theta}, \overline{X})$ admet un unique maximum θ_M donné par*

$$\alpha_M(j) = \frac{1}{n}\sum_{i=1}^{n} g_{\tilde{\theta}}(j|X = X_i)$$
$$m_M(j) = \frac{\sum_{i=1}^{n} X_i g_{\tilde{\theta}}(j|X = X_i)}{\sum_{i=1}^{n} g_{\tilde{\theta}}(j|X = X_i)}$$
$$v_M(j) = \frac{\sum_{i=1}^{n}(X_i - m_{k+1}(j))^2 g_{\tilde{\theta}}(j|X = X_i)}{\sum_{i=1}^{n} g_{\tilde{\theta}}(j|X = X_i)}.$$

Démonstration. Sous la contrainte $\alpha(1) + \cdots + \alpha(J) = 1$ on trouve

$$\alpha_M(j) = \frac{\sum_{i=1}^{n} g_{\tilde{\theta}}(j|X = X_i)}{\sum_{l=1}^{J}\sum_{i=1}^{n} g_{\tilde{\theta}}(l|X = X_i)} = \frac{1}{n}\sum_{i=1}^{n} g_{\tilde{\theta}}(j|X = X_i),$$

tandis que (m_M, v_M) est point critique du dernier terme de la log-vraisemblance conditionnelle. □

L'algorithme EM consiste à répéter successivement deux étapes consécutives comme suit :
- étape E(xpectation) : étant donnée une valeur θ_k du paramètre, on calcule la log-vraisemblance conditionnelle des observations $L_c(\theta, \theta_k, \overline{X})$ avec le théorème 13.2,
- étape M(aximization) : grâce à le théorème 13.3, on choisit θ_{k+1} pour que la fonction $\theta \mapsto L_c(\theta, \theta_k, \overline{X})$ soit maximale au point θ_{k+1}.

En pratique, étant données les observations $x_1, \ldots, x_n$ et des valeurs initiales des paramètres, l'algorithme consiste à répéter les calculs suivants :
- à partir du paramètre θ_k, on calcule la matrice $H^{(k)}$ de taille $n \times J$ par

$$H_{ij}^{(k)} = \frac{\alpha_k(j)\gamma_{m_k(j),v_k(j)}(X_i)}{\sum_{l=1}^{J}\alpha_k(l)\gamma_{m_k(l),v_k(l)}(X_i)},$$

- on en déduit θ_{k+1} par les relations suivantes :

$$\alpha_{k+1}(j) = \frac{1}{n}\sum_{i=1}^{n} H_{ij}^{(k)}, \quad m_{k+1}(j) = \frac{\sum_{i=1}^{n} X_i H_{ij}^{(k)}}{\sum_{i=1}^{n} H_{ij}^{(k)}}, \quad v_{k+1}(j) = \frac{\sum_{i=1}^{n}(X_i - m_j)^2 H_{ij}^{(k)}}{\sum_{i=1}^{n} H_{ij}^{(k)}}.$$

Le résultat suivant montre que la log-vraisemblance L_{obs} est croissante le long de l'algorithme.

Théorème 13.4 (Croissance de la vraisemblance)**.** *La suite $(\theta_k)_{k\geqslant 0}$ construite par l'algorithme EM vérifie la propriété de stabilité numérique suivante :*

$$L_{obs}(\theta_{k+1}, \overline{X}) \geqslant L_{obs}(\theta_k, \overline{X}),$$

où $L_{obs}(\theta, \overline{X})$ est la log-vraisemblance des observations.

Recueil de modèles stochastiques – Page 134/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Démonstration. Posons

$$H(\theta, \theta_k) = \mathbb{E}_{\theta_k}(\log g_\theta(\overline{Z})|\overline{X}) = \sum_{i=1}^{n} \mathbb{E}_{\theta_k}(\log g_\theta(Z_i)|X_i).$$

Puisque $h_\theta(x, z) = f_\theta(x)g_\theta(z|X = x)$, la vraisemblance conditionnelle s'écrit encore

$$L_c(\theta, \theta_k, \overline{X}) = L_{obs}(\theta, \overline{X}) + H(\theta, \theta_k).$$

Lors de l'étape M, la fonction $\theta \mapsto L_c(\theta, \theta_k, \overline{X})$ est maximisée par rapport à θ donc, en particulier,

$$L_c(\theta_{k+1}, \theta_k, \overline{X}) \geqslant L_c(\theta_k, \theta_k, \overline{X}).$$

On en déduit que

$$L_{obs}(\theta_{k+1}, \overline{X}) - L_{obs}(\theta_k, \overline{X}) \geqslant H(\theta_k, \theta_k) - H(\theta_{k+1}, \theta_k).$$

Par définition de H,

$$H(\theta_k, \theta_k) - H(\theta_{k+1}, \theta_k) = \mathbb{E}_{\theta_k} \log \frac{g_{\theta_k}(\overline{Z}|\overline{X})}{g_{\theta_{k+1}}(\overline{Z}|\overline{X})}.$$

En notant μ_θ la loi de $\overline{Z}$ sachant $\overline{X}$ sous $\mathbb{P}_\theta$ on remarque que la quantité ci dessus n'est rien d'autre que l'entropie relative de μ_{θ_k} par rapport à $\mu_{\theta_{k+1}}$:

$$H(\theta_k, \theta_k) - H(\theta_{k+1}, \theta_k) = \int \log \frac{d\mu_{\theta_k}}{d\mu_{\theta_{k+1}}} d\mu_{\theta_k} = \int \frac{d\mu_{\theta_k}}{d\mu_{\theta_{k+1}}} \log \frac{d\mu_{\theta_k}}{d\mu_{\theta_{k+1}}} d\mu_{\theta_{k+1}}.$$

L'inégalité de Jensen assure que cette quantité est positive. □

Le théorème 13.4 montre que la suite $(L_{obs}(\theta_k, \overline{X}))_{k\geqslant 0}$ est croissante. Elle converge donc vers un maximum local ou un point selle de la vraisemblance $L_{obs}(\cdot, \overline{X})$.

Contrairement au cas où l'on observe en même temps X et Z, il peut exister des maxima locaux qui vont piéger l'algorithme, ce qui rend le résultat sensible au point de départ. On peut améliorer le comportement de l'algorithme en incorporant un peu d'aléa, comme par exemple dans l'algorithme du recuit simulé du chapitre 8, mais cela est un peu illusoire.

On ne sait pas combien d'itérations sont nécessaires pour approcher de manière satisfaisante un maximum local. Il n'y a pas de critère d'arrêt performant.

La figure 13.1 donne l'estimation du mélange de trois gaussiennes par l'algorithme EM avec les paramètres $\alpha = (0.4\ 0.4\ 0.2)$, $m = (-2\ 0\ 2)$, $v = (0.3\ 0.2\ 0.2)$, $n = 1000$, et 1000 itérations.

13.4 Notes et commentaires

Le théorème 13.3 se généralise aux familles exponentielles, de vraisemblance totale

$$L(\theta, \overline{x}, \overline{z}) = \sum_{i=1}^{n} \log h_\theta(x_i, z_i) = \sum_{i=1}^{n} \log g_\theta(z_i) + \sum_{i=1}^{n} \log f_\theta(x_i|Z = z_i).$$

Après intégration, il vient en rappelant que $g_\theta(z) = \alpha(z)$,

$$L_c(\theta, \tilde{\theta}, \overline{x}) = \sum_{j=1}^{J} \left(\sum_{i=1}^{n} g_{\tilde{\theta}}(j|X = x_i) \right) \log \alpha(j) + \sum_{j=1}^{J} \sum_{i=1}^{n} g_{\tilde{\theta}}(j|X = x_i) \log f_\theta(x_i|Z = j).$$

Recueil de modèles stochastiques – Page 135/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

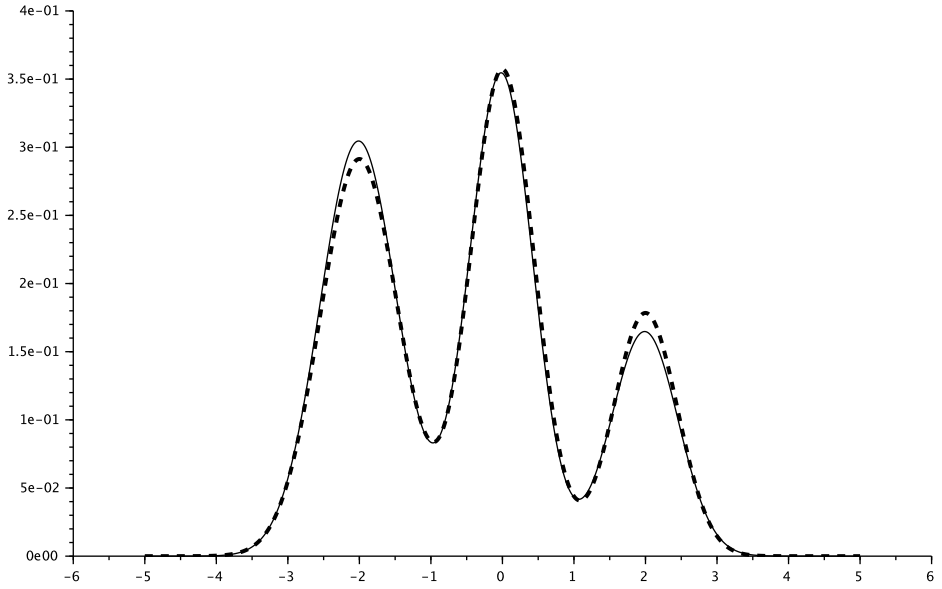


Figure 13.1 – Estimation du mélange de trois gaussiennes par l'algorithme EM.

Si $\theta = (\alpha, \beta)$ où α est encore la loi mélange et $\beta(j)$ représente les paramètres de la loi $\mathcal{L}(X|Z = j)$ on détermine séparément α_M (qui a toujours la même expression) et β_M qui s'obtient quasiment comme pour le maximum de la vraisemblance totale : on a seulement pondéré la contribution de l'observation X_i par le coefficient $g_{\tilde{\theta}}(j|X = X_i)$. Par exemple, si les composantes du mélange $(\mathcal{L}(X|Z = j))_{1\leqslant j\leqslant J}$ sont des lois exponentielles de paramètres $(\lambda(i))_{1\leqslant i\leqslant J}$, la log-vraisemblance conditionnelle s'écrit

$$L_c(\theta, \tilde{\theta}, \overline{X}) = \sum_{j=1}^{J} \left(\sum_{i=1}^{n} g_{\tilde{\theta}}(j|X = X_i) \right) \log \alpha(j) + \sum_{j=1}^{J} \sum_{i=1}^{n} g_{\tilde{\theta}}(j|X = X_i)(\log \lambda(j) - \lambda(j)x_i).$$

Un pas de l'algorithme EM a la forme suivante :

$$H_{ij}^{(k)} = \frac{\alpha_k(j)\lambda_k(j)e^{-\lambda_k(j)x_i}}{\sum_{l=1}^{J} \alpha_k(l)\lambda_k(l)e^{-\lambda_k(l)x_i}}, \quad \alpha_{k+1}(j) = \frac{1}{n}\sum_{i=1}^{n} H_{ij}^{(k)}, \quad \text{et} \quad \lambda_{k+1}(j) = \frac{\sum_{i=1}^{n} H_{ij}^{(k)}}{\sum_{i=1}^{n} X_i H_{ij}^{(k)}}.$$

Si de nombreux travaux existaient déjà sur cette question, c'est l'article [DLR77] de Arthur Dempster, Nan Laird, et Donald Rubin qui définit pour la première fois l'algorithme EM dans un cadre général. On trouvera dans la bibliographie de ce travail les références aux résultats antérieurs. Depuis, cet algorithme est très souvent utilisé dans des cadres et sous des formes variés. Lorsqu'il n'est pas possible de calculer la log-vraisemblance conditionnelle, on peut utiliser une méthode de Monte-Carlo (comme l'algorithme de Metropolis-Hastings du chapitre 8) pour obtenir une approximation de cette fonction dont on cherche ensuite un maximum numériquement.

Recueil de modèles stochastiques – Page 136/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Chapitre **14**

Records, extrêmes, et recrutements !

Mots-clés. Extrême ; martingale ; série centrée ; théorème de Lindeberg-Lévy.

Un cabinet de chasseurs de têtes souhaite constituer une équipe de choc grâce à un recrutement au fil de l'eau : chaque candidat obtient une note après son entretien et les recruteurs décident dans l'instant de l'engager ou pas. Le cahier des charges est de recruter une équipe à la fois excellente et fournie. Dans ce chapitre, on compare deux politiques de recrutement différentes : une personne est recrutée si sa note est supérieure aux notes (respectivement à la moyenne des notes) de tous les recrutés précédents. On modélise les notes des candidats successifs par une suite $(X_n)_{n\geqslant 1}$ de variables aléatoires réelles indépendantes et de même loi μ de fonction de répartition F. Afin d'éviter les cas particuliers, nous supposerons que F est continue sur $\mathbb{R}$. On note x_F l'extrémité droite du support de μ :

$$x_F = \sup\{x \in \mathbb{R} \,:\, F(x) < 1\}.$$

14.1 Élitisme

Pour tout $n \geqslant 1$, on note R_n le rang relatif de X_n dans l'échantillon $X_1, \ldots, X_n$. Il est donné par $R_n = 1 + \sum_{k=1}^{n} \mathbf{1}_{\{X_k > X_n\}}$. Le candidat n est recruté si son rang vaut 1,c'est-à-dire si sa note bat le précédent record. Le nombre de records, ou de personnes recrutées, jusqu'au temps n est donné par $Z_n = \sum_{i=1}^{n} \mathbf{1}_{\{R_i=1\}}$. On définit également les instants de record par récurrence :

$$T_1 = 1 \quad \text{et, pour } j \geqslant 1 \quad T_{j+1} = \inf\{n > T_j,\ X_n > X_i,\ \text{pour } i < n\}.$$

Lemme 14.1 (Loi des rangs successifs)**.** *Les variables aléatoires $(R_n)_{n\geqslant 1}$ sont indépendantes et, pour tout $n \in \mathbb{N}^*$, R_n suit la loi uniforme sur $\{1, \ldots, n\}$. De plus,*

$$\mathbb{E}(Z_n) = \sum_{k=1}^{n} \frac{1}{k} = \log(n) + \gamma + o(1) \quad \textit{et} \quad \mathbb{V}(Z_n) = \sum_{k=1}^{n} \frac{1}{k}\left(1 - \frac{1}{k}\right) = \log(n) + \gamma - \frac{\pi^2}{6} + o(1),$$

où γ désigne la constante d'Euler.

Démonstration. La fonction de répartition F de μ est continue donc pour tous entiers i et j distincts, $\mathbb{P}(X_i \neq X_j) = 1$. En particulier, pour $n \geqslant 1$, il existe une unique permutation (aléatoire) σ de $\{1, \ldots, n\}$ telle que

$$X_{\sigma(1)} < \cdots < X_{\sigma(n)}.$$

Recueil de modèles stochastiques – Page 137/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

De plus la loi de cette permutation est la loi uniforme sur $\mathcal{S}_n$. Enfin, à un vecteur $(R_1, R_2, \ldots, R_n)$ on associe la permutation σ de manière bijective. Ceci assure que

$$\mathbb{P}(R_1 = r_1, R_2 = r_2, \ldots, R_n = r_n) = \frac{1}{n!}$$

avec $1 \leqslant r_i \leqslant i$ pour $1 \leqslant i \leqslant n$. La v.a. Z_n est donc la somme de n variables aléatoires de Bernoulli indépendantes de paramètres 1, 1/2,..., $1/n$. Les expressions de son espérance et sa variance s'en déduisent. □

Théorème 14.2 (Asymptotique du nombre de records)**.**

$$\frac{Z_n}{\log(n)} \xrightarrow[n\to\infty]{p.s.} 1 \quad \textit{et} \quad \frac{Z_n - \log(n)}{\sqrt{\log(n)}} \xrightarrow[n\to\infty]{\text{loi}} \mathcal{N}(0,1).$$

Démonstration. La convergence presque sûre repose sur le théorème des séries centrées que nous rappelons ci-dessous.

Théorème 14.3 (Séries centrées)**.** *Soit $(Q_n)_{n\geqslant 1}$ une suite de variables aléatoires indépendantes et centrées et une suite (déterministe) $(b_n)_{n\geqslant 1}$ strictement croissante non bornée de réels strictement positifs.*

$$\textit{Si} \quad \sum_{n=1}^{\infty} \frac{\mathbb{E}(Q_n^2)}{b_n^2} < \infty, \quad \textit{alors} \quad \frac{1}{b_n}\sum_{k=1}^{n} Q_k \xrightarrow[n\to\infty]{p.s.} 0.$$

Reste à appliquer ce théorème à la suite Q définie par $Q_n = R_n - 1/n$ pour $n \geqslant 1$ avec $b_n = \log(n)$ en remarquant

$$\frac{\mathbb{V}(Q_n)}{b_n^2} = \frac{n-1}{n^2\log(n)^2}$$

est le terme général d'une série convergente.

La deuxième partie du théorème découle du *théorème limite central de Lindeberg-Lévy* rappelé ci-dessous.

Théorème 14.4 (Lindeberg-Lévy)**.** *Soit $(Y_i)_{i\geqslant 1}$ une suite de v.a.r. indépendantes de carré intégrable et centrées. Notons $S_n = Y_1 + \cdots + Y_n$ et s_n son écart-type. Si, pour tout $\varepsilon > 0$,*

$$\lim_{n\to\infty} \frac{1}{s_n^2}\sum_{i=1}^{n} \mathbb{E}\big[Y_i^2 \mathbf{1}_{|Y_i|>\varepsilon s_n}\big],$$

alors

$$\frac{S_n}{s_n} \xrightarrow[n\to\infty]{\text{loi}} \mathcal{N}(0,1).$$

On applique ce théorème aux variables $Y_n = \mathbf{1}_{\{R_n=1\}} - 1/n$ pour $n \geqslant 1$ en remarquant que $\mathbf{1}_{|Y_i|>\varepsilon s_n} = 0$ pour n assez grand. □

14.2 Au-dessus de la moyenne

On construit alors par récurrence les suites $(Y_n)_{n\geqslant 1}$, $(\overline{Y}_n)_{n\geqslant 1}$ et $(T_n)_{n\geqslant 1}$: $Y_1 = X_1$, $\overline{Y}_1 = Y_1$ et $T_1 = 1$, puis, pour tout $n \geqslant 1$, par

$$T_{n+1} = \inf\big\{i > T_n \,:\, X_i > \overline{Y}_n\big\}, \quad Y_{n+1} = X_{T_{n+1}} \quad \text{et} \quad \overline{Y}_{n+1} = \frac{1}{n+1}\sum_{i=1}^{n+1} Y_i.$$

Pour tout $n \geqslant 1$, Y_n est la performance du n-ième candidat retenu, T_n est le nombre de candidats auditionnés pour en retenir n et $\overline{Y}_n$ est la moyenne des performances des n premiers candidats retenus.

On peut également considérer

Recueil de modèles stochastiques – Page 138/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

— M_n est le nombre de candidats retenus après n auditions,
— A_n est la moyenne des performances des candidats retenus après n auditions.

Théorème 14.5 (Vers la qualité maximale)**.**

$$\overline{Y}_n \xrightarrow[n\to\infty]{p.s.} x_F.$$

Démonstration. Presque sûrement, les variables aléatoires $(X_n)_{n\geqslant 1}$ sont strictement inférieures à x_F. Ainsi, on montre par récurrence que Y_n et $\overline{Y}_n$ sont également strictement inférieures à x_F et que par suite T_{n+1} est fini. La suite $(Y_n)_{n\geqslant 1}$ est croissante et bornée par x_F, elle converge donc vers un réel x_∞. Celui-ci est nécessairement égal à x_F car le supremum d'une infinité de v.a. indépendantes de fonction de répartition F est égal à x_F. Ainsi la suite $(Y_n)_{n\geqslant 1}$ converge presque sûrement vers x_F. Le lemme de Cesàro assure qu'il en est de même pour $(\overline{Y}_n)_{n\geqslant 1}$. □

On s'intéresse alors au nombre de candidats qu'il faudra auditionner pour former une équipe de taille donnée. Le résultat suivant permet de conclure sous une hypothèse assez souvent vérifiée en pratique sur la convergence de la suite $(\overline{Y}_n)_n$.

Théorème 14.6. *Soit la suite de variables aléatoires $(P_n)_{n\geqslant 1}$ définie par $P_1 = 1$ et, pour $n \geqslant 2$, $P_n = 1 - F(\overline{Y}_{n-1})$. Supposons qu'il existe $\alpha \in (0,\infty)$ et W une variable aléatoire tels que $\mathbb{P}(W \in (0,\infty)) = 1$ et $(n^\alpha P_n)_{i\geqslant 1}$ converge presque sûrement vers W lorsque $i \to \infty$. Alors*

$$\frac{T_n}{n^{\alpha+1}} \xrightarrow[n\to\infty]{p.s.} \frac{1}{(\alpha+1)W}.$$

Démonstration. L'idée est d'appliquer le théorème 14.3 conditionnellement à la tribu $\overline{\mathcal{F}}$ engendrée par la suite $(\overline{Y}_n)_{n\geqslant 1}$ en posant, pour $n \geqslant 1$,

$$b_n = \sum_{i=1}^{n} \frac{1}{P_i} \quad \text{et} \quad Q_n = T_n - T_{n-1} - \frac{1}{P_n} \quad \text{avec} \quad T_0 = 0.$$

La suite $(b_n)_{n\geqslant 1}$ est strictement croissante. Elle tend vers $+\infty$ puisque $1/P_i$ est de l'ordre de i^α/W. Plus précisément, une comparaison série/intégrale assure que

$$b_n \sim \frac{n^{\alpha+1}}{(\alpha+1)W}.$$

D'autre part, $T_n - T_{n-1}$ suit la loi géométrique de paramètre P_n donc

$$\mathbb{E}\left(\frac{Q_n^2}{b_n^2}|\overline{\mathcal{F}}\right) = \frac{1-P_n}{b_n^2 P_n^2} \sim \frac{(\alpha+1)^2}{n^2}$$

qui est le terme général d'une série convergente. Le théorème des séries centrées assure donc que

$$\frac{1}{b_n}\sum_{k=1}^{n} Q_k = \frac{T_n}{b_n} - 1 \xrightarrow[n\to\infty]{\text{p.s.}} 0.$$

L'équivalent de b_n donne le résultat annoncé. □

Recueil de modèles stochastiques – Page 139/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

14.3 Cas de la loi exponentielle

Dans cette partie, on suppose que la fonction de répartition F est celle de la loi exponentielle de paramètre 1.

Théorème 14.7. *Supposons que les variables aléatoires $(X_n)_{n\geqslant 1}$ soient distribuées selon la loi exponentielle de paramètre 1. Alors il existe une variable aléatoire G de loi de Gumbel de fonction de répartition $x \mapsto \exp(-e^{-x})$ telle que*

1. *$(\overline{Y}_n - \log(n))_{n\geqslant 1}$ converge presque sûrement et dans L^2 vers G,*
2. *$(T_n/n^2)_{n\geqslant 1}$ converge presque sûrement vers $e^G/2$.*

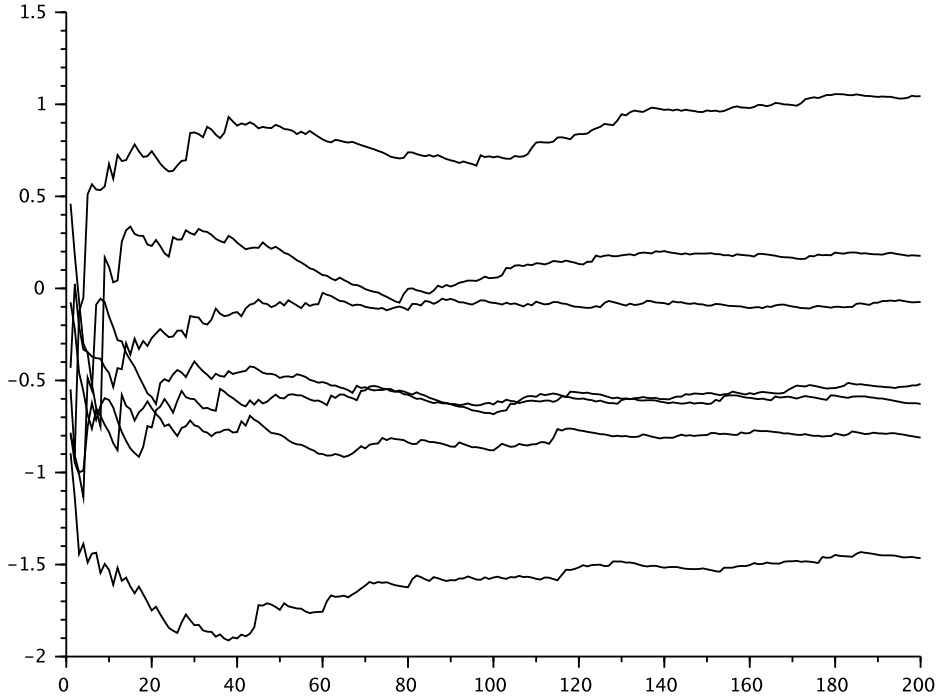


Figure 14.1 – Convergence p.s. de la suite $(\overline{Y}_n - \log n)_n$.

Démonstration. Avec la convention $\overline{Y}_0 = 0$, posons, pour tout $n \geqslant 0$,

$$\eta_n = Y_{n+1} - \overline{Y}_n = (n+1)(\overline{Y}_{n+1} - \overline{Y}_n).$$

On a donc $\overline{Y}_n = \sum_{j=1}^n \eta_{j-1}/j$. Notons $\mathcal{F}_n$ la tribu engendrée par $(Y_1, \ldots, Y_n)$. Pour tout $n \geqslant 1$,

$$\mathbb{P}(\eta_n > x|\mathcal{F}_n) = \mathbb{P}(Y_{n+1} > \overline{Y}_n + x|Y_{n+1} > \overline{Y}_n, \overline{Y}_n) = e^{-x}.$$

Les variables aléatoires $(\eta_n)_{n\geqslant 0}$ sont donc indépendantes et de même loi exponentielle de paramètre 1. Avec $h_n = \sum_{k=1}^n \frac{1}{k}$, on a

$$\overline{Y}_n - h_n = \sum_{k=1}^n \frac{\eta_{k-1} - 1}{k}.$$

Recueil de modèles stochastiques – Page 140/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

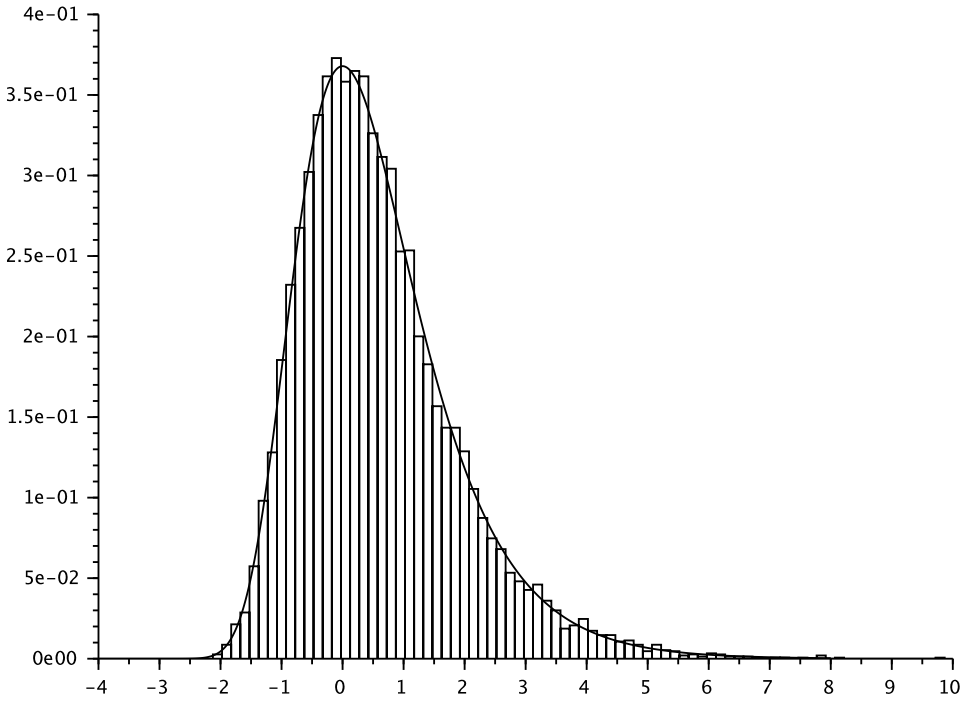


FIGURE 14.2 – Histogramme de la limite p.s. de la suite $(\overline{Y}_n - \log n)_n$ et densité de la loi de Gumbel.

La suite $(\overline{Y}_n - h_n)_{n\geqslant 1}$ est une martingale bornée dans L^2. Il existe donc une variable aléatoire Z telle que $\overline{Y}_n - h_n$ converge vers Z dans L^2 et presque sûrement. Reste à identifier la limite.

La transformée de Laplace de $\overline{Y}_n$ notée H_n vaut

$$H_n(\lambda) = \mathbb{E}\Big(e^{\lambda \overline{Y}_n}\Big) = \prod_{k=1}^{n} \frac{k}{k-\lambda} = n!\left(\prod_{k=1}^{n}(k-\lambda)\right)^{-1}.$$

Considérons à présent une suite $(Z_n)_{n\geqslant 1}$ de variables aléatoires indépendantes de même loi exponentielle de paramètre 1. Pour tout $n \geqslant 1$, notons $Z_{(n)} = \max(Z_1, Z_2, \ldots, Z_n)$. Une intégration par parties assure que

$$\begin{aligned} G_n(\lambda) = \mathbb{E}\Big(e^{\lambda Z_{(n)}}\Big) &= \int_0^\infty e^{\lambda z} n e^{-z}\big(1-e^{-z}\big)^{n-1}\,dz \\ &= \frac{n}{1-\lambda}\int_0^\infty e^{\lambda z}(n-1)e^{-2z}\big(1-e^{-z}\big)^{n-2}\,dz = \frac{n}{1-\lambda}G_{n-1}(\lambda-1). \end{aligned}$$

Une récurrence immédiate donne

$$G_n(\lambda) = \prod_{k=1}^{n} \frac{n+1-k}{k-\lambda} = n!\left(\prod_{k=1}^{n}(k-\lambda)\right)^{-1}.$$

Recueil de modèles stochastiques – Page 141/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

On en déduit que $\overline{Y}_n$ et $Z_{(n)}$ ont même loi[1]. Ainsi,

$$\begin{aligned}\mathbb{P}(\overline{Y}_n - \log(n) \leqslant x) &= \mathbb{P}(Z_{(n)} - \log(n) \leqslant x)\\ &= \left(1 - \frac{e^{-x}}{n}\right)^n \xrightarrow[n\to\infty]{} e^{-e^{-x}},\end{aligned}$$

ce qui conclut la preuve du point 1. De plus,

$$nP_n = n(1 - F(\overline{Y}_{n-1})) = e^{-(\overline{Y}_n - \log(n))} \xrightarrow[n\to\infty]{\text{p.s.}} e^{-G}.$$

Le théorème 14.6 fournit la convergence presque sûre de T_n/n^2 vers $e^G/2$. Pour tout $t > 0$,

$$\mathbb{P}(e^G/2 \leqslant t) = \mathbb{P}(G \leqslant \log(2t)) = \exp(-1/(2t)).$$

□

14.4 Cas de la loi de Pareto

On suppose ici que les variables aléatoires $(X_n)_{n\geqslant 1}$ sont distribuées selon la *loi de Pareto* de paramètre $\beta > 1$ c'est-à-dire qu'elles possèdent la fonction de répartition suivante :

$$F_\beta(x) = (1 - x^{-\beta})\mathbf{1}_{[1,+\infty[}(x).$$

Théorème 14.8 (Performance moyenne)**.** *Il existe une variable aléatoire Y telle que $\mathbb{P}(0 < Y < \infty) = 1$ et*

$$\frac{\overline{Y}_n}{n^{1/(\beta-1)}} \xrightarrow[n\to\infty]{p.s.} Y.$$

Démonstration. Remarquons tout d'abord que si X suit la loi de Pareto de paramètre β, alors la loi de X sachant que $X > c$ est la loi de cX. En notant $\mathcal{F}_n$ la tribu engendrée par $(Y_1, \ldots, Y_n)$, on a, pour tout $x > 1$,

$$\mathbb{P}(Y_n > x\overline{Y}_{n-1}|\mathcal{F}_{n-1}) = \mathbb{P}(X > x\overline{Y}_{n-1}|X > \overline{Y}_{n-1}, \overline{Y}_{n-1}) = x^{-\beta}.$$

où X est une v.a. de Pareto de paramètre β indépendante de $\mathcal{F}_{n-1}$. En d'autres termes, la loi de $Y_n/\overline{Y}_{n-1}$ sachant $\mathcal{F}_{n-1}$ est la loi de Pareto de paramètre β. En conséquence,

$$\mathbb{E}(\overline{Y}_n|\mathcal{F}_{n-1}) = \frac{(n-1)\overline{Y}_{n-1} + \mathbb{E}(Y_n|\overline{Y}_{n-1})}{n} = \frac{n - 1 + \mathbb{E}(X)}{n}\overline{Y}_{n-1}.$$

Posons, pour $n \geqslant 1$,

$$a_n = \frac{n - 1 + \mathbb{E}(X)}{n} = 1 + \frac{1}{n(\beta - 1)} \quad \text{et} \quad b_n = \left(\prod_{j=1}^n a_j\right)^{-1}.$$

La suite $(M_n)_{n\geqslant 1}$ définie par $M_n = b_n\overline{Y}_n$ est une martingale positive d'espérance 1. Elle converge donc presque sûrement vers une variable aléatoire M positive de L^1. Nous omettons ici la preuve assez technique du fait que $M \in (0, \infty)$ presque sûrement (voir la section Notes et commentaires). Enfin, on remarque que

$$b_n^{-1} = \exp\left(\sum_{j=1}^n \log\left(1 + \frac{1}{(\beta-1)j}\right)\right) = \exp\left(\frac{1}{\beta-1}\sum_{j=1}^n \frac{1}{j} + o(1)\right).$$

En d'autres termes, $n^{1/(\beta-1)}b_n$ converge vers un réel strictement positif. □

1. Ce résultat est démontré de deux autres façons dans la preuve du théorème 10.4 page 109.

Recueil de modèles stochastiques – Page 142/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Corollaire 14.9 (Temps de recrutement)**.** *Il existe une variable aléatoire T à valeurs dans $(0,\infty)$ telle que*

$$\frac{T_n}{n^{(2\beta-1)/(\beta-1)}} \xrightarrow[n\to\infty]{p.s.} T.$$

Démonstration. Pour tout $n \geqslant 1$, on a $P_n = (\overline{Y}_{n-1})^{-\beta}$. Le résultat précédent assure que $n^{\beta/(\beta-1)}P_n$ converge presque sûrement vers une variable aléatoire $W \in (0,\infty)$. On conclut alors avec le théorème 14.6. □

14.5 Notes et commentaires

Ce chapitre est essentiellement basé sur un article de John Preater [Pre00] pour le cas de la loi exponentielle, et sur un article de Abba Krieger, Moshe Pollak, et Ester Samuel-Cahn [KPSC08] pour le cas général. Les preuves du théorème des séries centrées (théorème 14.3) et du théorème de Lindeberg-Lévy (théorème 14.4) se trouvent par exemple dans l'incontournable livre de William Feller [Fel68].

Citons enfin le célèbre *problème des secrétaires*[2] dont on trouve une description et une mise en contexte savoureuse dans un article de Thomas Ferguson [Fer89]. Le contexte est le même que dans notre chapitre mais on ne recrute qu'une personne parmi n candidats. La question est de déterminer la stratégie fournissant avec la probabilité maximale le meilleur candidat. La première étape est de montrer que la stratégie optimale est nécessairement de la forme suivante : on fixe r entre 1 et n, on recale les $r-1$ premiers candidats puis on sélectionne le premier candidat qui obtient une meilleure note que ses prédécesseurs. Avec une telle stratégie, la probabilité de recruter le meilleur candidat est donnée par

$$p_n(r) = \sum_{j=r}^{n} \mathbb{P}(j \text{ est le meilleur et on le sélectionne}) = \sum_{j=r}^{n} \frac{1}{n}\frac{r-1}{j-1} = \frac{r-1}{n}\sum_{j=r}^{n}\frac{1}{j-1}.$$

Pour $x \in]0,1[$,

$$p_n(\lfloor xn \rfloor) = \frac{\lfloor xn \rfloor}{n}\sum_{j=\lfloor xn \rfloor}^{n}\frac{1}{j-1} \xrightarrow[n\to\infty]{} -x\log(x).$$

La probabilité maximale vaut donc $1/e$ pour un choix optimal de r de l'ordre de n/e.

2. Ce problème d'*arrêt optimal* est un grand classique de la théorie du contrôle stochastique et de la théorie de la décision. Il est également connu sous le nom du *problème du mariage* (entre autres).

Recueil de modèles stochastiques – Page 143/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Recueil de modèles stochastiques – Page 144/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

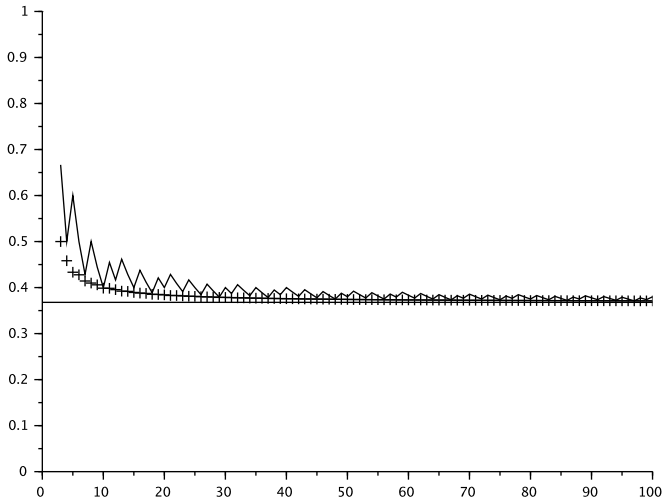


Figure 14.3 – Convergence de la probabilité optimale du meilleur choix (croix) et de la proportion idéale (trait continu) vers $1/e$.

Chapitre **15**

Polymères dirigés en environnement aléatoire

Mots-clés. Chaîne de Markov ; transformée de Laplace ; martingale ; loi du zéro-un ; phénomène de seuil ; espérance dépendant d'un paramètre.

Un *polymère* est une longue molécule constituée de briques élémentaires identiques appelées *monomères*. On s'intéresse à la modélisation de la structure d'un polymère constitué de monomères hydrophiles, placé dans une émulsion d'huile et d'eau. L'huile est présente sous la forme de gouttelettes microscopiques. On suppose que les monomères sont régulièrement espacés et que le polymère se déroule dans une direction privilégiée. On dit que le polymère est *dirigé*. Par analogie avec les processus, cette dimension est dite *temporelle*, et le polymère est vu comme un chemin dans $\mathbb{Z}^d$, c'est-à-dire une application de $\{0, 1, \ldots, n\}$ dans $\mathbb{Z}^d$ où n est la longueur du polymère. On positionne sur les points de $\mathbb{N} \times \mathbb{Z}^d$ des variables aléatoires i.i.d. qui quantifient la sensibilité de chaque monomère du polymère à la présence d'huile ou d'eau sur chaque site. Dans tout le chapitre, l'entier d est fixé supérieur ou égal à 1.

La sensibilité du polymère à son environnement est modélisée par un paramètre réel noté β introduit par la suite. Toute la question est de déterminer pour quelles valeurs de ce paramètre le comportement du polymère est sensiblement différent d'une marche aléatoire simple qui modélise la structure d'un polymère indifférent à son environnement.

Toutes les variables aléatoires de ce chapitre sont définies sur un même espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$. On note $\mathbb{E}$ l'espérance par rapport à la mesure $\mathbb{P}$.

15.1 Modèle et résultats principaux

On modélise un polymère insensible à son environnement par une marche aléatoire simple sur $\mathbb{Z}^d$, c'est-à-dire une suite $(S_n)_{n\geqslant 0}$ avec $S_0 = 0$ et pour $n \geqslant 1$,

$$S_n = \sum_{i=1}^{n} U_i$$

où $(U_n)_{n\geqslant 1}$ est une suite de variables aléatoires i.i.d. de loi uniforme sur $\{\pm e_i \,:\, 1 \leqslant i \leqslant d\}$, où $(e_1, \ldots, e_d)$ désigne la base canonique de $\mathbb{Z}^d$. On note $(\mathcal{F}_n)_{n\geqslant 0}$ la filtration naturelle de $(S_n)_{n\geqslant 0}$, et $\mathcal{F}$ la tribu engendrée par les variables aléatoires $(U_n)_{n\geqslant 1}$.

Il est commode de représenter les $n+1$ premiers termes de $(S_n)_{n\geqslant 0}$ par la variable aléatoire $S_{0:n} = (S_0, S_1, \ldots, S_n)$, qui prend ses valeurs dans Γ_n, l'ensemble des *chemins* issus de 0 et de longueur n dans $\mathbb{Z}^d$, c'est-à-dire l'ensemble des $\gamma = (\gamma_0, \gamma_1, \ldots, \gamma_n)$ avec $\gamma_0, \ldots, \gamma_n$ dans $\mathbb{Z}^d$, tels que $\gamma_0 = 0$ et $\gamma_{i+1} - \gamma_i \in \{\pm e_j \,:\, 1 \leqslant j \leqslant d\}$ pour tout $0 \leqslant i \leqslant n-1$.

Recueil de modèles stochastiques – Page 145/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

On modélise l'environnement par des variables aléatoires réelles $(\eta(n,x))_{n\geqslant 1,\, x\in\mathbb{Z}^d}$ i.i.d. non constantes. On suppose que $(U_n)_{n\geqslant 1}$ et $(\eta(n,x))_{n\geqslant 0,\, x\in\mathbb{Z}^d}$ sont indépendantes. On note $\mathcal{G}$ la tribu engendrée par $(\eta(i,x))_{i\geqslant 1,\, x\in\mathbb{Z}^d}$ et $\mathcal{G}_n$ la tribu engendrée par $(\eta(i,x))_{(i,x)\in E_n}$ où $E_n = \{(i,\gamma_i) \,:\, 1\leqslant i\leqslant n,\ \gamma\in\Gamma_n\}$ est l'environnement vu par les chemins de longueur n.

Pour tout $\beta\geqslant 0$, on définit la mesure de probabilité μ_n sur Γ_n en posant, pour tout $\gamma\in\Gamma_n$,

$$\mu_n(\gamma) = \frac{1}{Z_n(\beta)} e^{\beta H_n(\gamma)}\, \mathbb{P}(S_{0:n}=\gamma),$$

avec

$$H_n(\gamma) = \sum_{i=1}^{n} \eta(i,\gamma_i) \quad \text{et} \quad Z_n(\beta) = \mathbb{E}\Big(e^{\beta H_n(S_{0:n})}|\mathcal{G}\Big).$$

La quantité $H_n(S_{0:n})$ représente l'affinité de la configuration du polymère dans son environnement. Plus elle est élevée, plus la configuration du polymère est probable. Le paramètre β quantifie l'influence de l'environnement sur la structure du polymère. Il s'interprète comme l'inverse de la température. Si $\beta=0$ (la température est infinie) alors tous les chemins de longueur n ont le même poids : le polymère n'est pas sensible à l'environnement et se comporte comme la marche aléatoire simple. Si $\beta=+\infty$ (température nulle), le polymère ne charge que les chemins de longueur n qui maximisent H_n.

La quantité $Z_n(\beta)$ est une constante de normalisation appelée *fonction de partition* du système. On pose $\eta=\eta(0,0)$ et on suppose que la fonction suivante est bien définie :

$$\lambda : t\in\mathbb{R}\mapsto \lambda(t) = \log \mathbb{E}\big(e^{t\eta}\big)\in\mathbb{R}_+.$$

La *fonction de partition normalisée* est

$$W_n(\beta) = \mathbb{E}\Big(e^{\beta H_n(S_{0:n})-n\lambda(\beta)}|\mathcal{G}\Big).$$

L'*énergie libre* du système et son espérance sont définies par

$$q_n(\beta) = \frac{1}{n}\log Z_n(\beta) \quad \text{et} \quad u_n(\beta) = \mathbb{E}(q_n(\beta)).$$

Les quantités $Z_n(\beta)$, $q_n(\beta)$, $W_n(\beta)$ et μ_n sont aléatoires, et nous allons étudier leur comportement asymptotique quand n tend vers l'infini. Il y a plusieurs façons de mesurer l'influence de l'environnement sur la structure du polymère. Nous en présentons deux basées sur les comportements asymptotiques de $q_n(\beta)$ et $W_n(\beta)$. Les résultats sont présentés dans cette section, leurs démonstrations sont données dans les suivantes.

Nous commençons par le premier résultat significatif sur une dichotomie possible dans le comportement du polymère. Il est basé sur le comportement asymptotique de la fonction de partition normalisée $W_n(\beta)$.

Théorème 15.1 (Phénomène de seuil)**.** *Il existe une variable aléatoire $W_\infty(\beta)$ telle que*

$$W_n(\beta) \xrightarrow[n\to\infty]{} W_\infty(\beta) \quad \mathbb{P}-p.s.$$

De plus, il existe une constante $\bar\beta_c\in[0,+\infty]$ telle que, p.s.

$$\begin{cases} W_\infty(\beta)>0 & \text{si } \beta<\bar\beta_c, \text{ et on dit que le polymère est en «fort désordre»},\\ W_\infty(\beta)=0 & \text{si } \beta>\bar\beta_c, \text{ et on dit que le polymère est en «faible désordre»}.\end{cases}$$

Ce résultat sera démontré dans la section 15.2. Il donne la première notion de désordre. Présentons à présent l'autre notion de désordre. Elle s'appuie sur le résultat suivant.

Recueil de modèles stochastiques – Page 146/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Théorème 15.2 (Phénomène de seuil)**.** *Il existe une fonction u inférieure à λ telle que*

$$u_n(\beta) \xrightarrow[n\to\infty]{} u(\beta).$$

De plus, il existe une constante $\beta_c \in [0, +\infty]$ telle que

$$\begin{cases} u(\beta) = \lambda(\beta) & \text{si } \beta < \beta_c, \text{ et on dit que le polymère est en très faible désordre,} \\ u(\beta) < \lambda(\beta) & \text{si } \beta > \beta_c, \text{ et on dit que le polymère est en très fort désordre.} \end{cases}$$

Ce théorème est démontré dans la section 15.4. Les deux notions de désordre sont comparables.

Lemme 15.3 (Le très fort désordre implique le fort désordre)**.** *On a toujours*

$$0 \leqslant \bar{\beta}_c \leqslant \beta_c \leqslant +\infty.$$

Démonstration. Si $\beta < \bar{\beta}_c$ alors $\mathbb{P}(\{W_\infty(\beta) > 0\}) = 1$ et par suite

$$(u - \lambda)(\beta) = \lim_{n\to\infty} \frac{1}{n} \log W_n(\beta) = 0.$$

En d'autres termes, $\beta \leqslant \beta_c$. □

On a donc la situation suivante.

β	$(0, \bar{\beta}_c)$	$(\bar{\beta}_c, \beta_c)$	$(\beta_c, +\infty)$
Fort Désordre	*non*	*oui*	*oui*
Très Fort Désordre	*non*	*non*	*oui*

Les deux questions essentielles sont les suivantes :
- ces phénomènes de seuil (ont parle parfois en physique de transition de phase) ont-ils vraiment lieu, en d'autres termes, les paramètres critiques sont-ils finis et non nuls ?
- les deux paramètres critiques sont-ils différents ?

Les réponses à ces questions dépendent *a priori* de l'environnement et de la dimension d du polymère. Nous obtenons dans la section 15.3 une borne supérieure pour β_c qui est finie ou non selon la loi de l'environnement et la dimension d. Dans la section 15.5, nous montrons que si $d \geqslant 3$ alors quel que soit l'environnement, $\bar{\beta}_c$ est strictement positif. On pourra se reporter à la section Notes et commentaires pour le cas des petites dimensions qui s'avère beaucoup plus délicat.

Certaines démonstrations ci-dessous reposent sur la remarque simple suivante : si f et g sont croissantes, alors $f(X)$ et $g(X)$ sont positivement corrélées. Cette propriété, appelée inégalité FKG [1], est vraie pour des fonctions de plusieurs variables aléatoires indépendantes.

Lemme 15.4 (Inégalité FKG)**.** *Si $X = (X_i)_{1\leqslant i\leqslant k}$ sont des v.a.r. i.i.d. et si $f, g : \mathbb{R}^k \to \mathbb{R}$ sont deux fonctions de carré intégrable et croissantes au sens où $f(x) \leqslant f(y)$ si $x_i \leqslant y_i$ pour $1 \leqslant i \leqslant k$, alors les v.a.r. $f(X)$ et $g(X)$ sont positivement corrélées :*

$$\mathbb{E}[f(X)g(X)] \geqslant \mathbb{E}[f(X)]\,\mathbb{E}[g(X)].$$

Démonstration. Pour $k = 1$, si X' est une copie indépendante de X, on a

$$\mathbb{E}(f(X)g(X)) - \mathbb{E}f(X)\mathbb{E}g(X) = \frac{1}{2}\mathbb{E}\big[(f(X) - f(X'))(g(X) - g(X'))\big] \geqslant 0.$$

1. Provient de Fortuin, Kasteleyn et Ginibre, qui ont utilisé cette inégalité en physique mathématique.

Recueil de modèles stochastiques – Page 147/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Pour $k > 1$, on procède par récurrence en utilisant un conditionnement :

$$\begin{aligned}\mathbb{E}(f(X)g(X)) - \mathbb{E}f(X)\mathbb{E}g(X) &= \mathbb{E}(\mathbb{E}(f(X)g(X)|X_1) - \mathbb{E}(f(X)|X_1)\mathbb{E}(g(X)|X_1)) \\ &\quad + \mathbb{E}(\mathbb{E}(f(X)|X_1)\mathbb{E}(g(X)|X_1)) - \mathbb{E}f(X)\mathbb{E}g(X) \\ &= \mathbb{E}\mathrm{Cov}(f(X), g(X)|X_1) + \mathrm{Cov}(\mathbb{E}(f(X)|X_1), \mathbb{E}(g(X)|X_1)).\end{aligned}$$

Le premier terme est positif grâce à l'hypothèse de récurrence puisqu'à x_1 fixé f, g sont croissantes comme fonctions des $k-1$ variables $(x_2, \ldots, x_k)$. Le second l'est aussi puisque les fonctions $\mathbb{E}(f(X)|X_1 = \cdot)$ et $\mathbb{E}(g(X)|X_1 = \cdot)$ sont croissantes et que la propriété est vraie au rang 1. □

15.2 Fonction de partition normalisée

Cette section est consacrée à la preuve du théorème 15.1, en plusieurs étapes.

Lemme 15.5 (Convergence)**.** *La suite $(W_n(\beta))_{n\geqslant 1}$ converge p.s. vers une variable aléatoire $W_\infty(\beta)$ qui est positive et d'intégrale inférieure à 1.*

Démonstration. L'inégalité de Jensen assure que, pour $p \geqslant 1$,

$$\begin{aligned}\mathbb{E}(W_n(\beta)^p) &\leqslant \mathbb{E}\Big(e^{\beta p H_n(S_{0:n}) - np\lambda(\beta)}\Big) \\ &\leqslant \mathbb{E}\Big(e^{p\beta H_n(S_{0:n})}\Big)e^{-pn\lambda(\beta)}.\end{aligned}$$

Pour tout $\gamma \in \Gamma_n$, on a

$$\mathbb{E}\Big(e^{p\beta H_n(\gamma)}\Big) = \Big[\mathbb{E}\Big(e^{p\beta\eta}\Big)\Big]^n = e^{n\lambda(p\beta)}.$$

La variable aléatoire W_n^p est donc intégrable et $\mathbb{E}(W_n(\beta)^p) \leqslant e^{n(\lambda(p\beta) - p\lambda(\beta))}$. Pour $p = 1$, la majoration est une égalité : $\mathbb{E}(W_n(\beta)) = 1$. Pour tout $n \geqslant 1$,

$$\begin{aligned}\mathbb{E}(W_{n+1}(\beta)|\mathcal{G}_n) &= \frac{e^{-(n+1)\lambda(\beta)}}{(2d)^{n+1}} \sum_{\gamma\in\Gamma_{n+1}} \mathbb{E}\Big(e^{\beta\sum_{i=1}^{n+1}\eta(i,\gamma_i)}|\mathcal{G}_n\Big) \\ &= \frac{e^{-(n+1)\lambda(\beta)}}{(2d)^{n+1}} \sum_{\gamma\in\Gamma_{n+1}} e^{\beta\sum_{i=1}^{n}\eta(i,\gamma_i)}e^{\lambda(\beta)} \\ &= \frac{1}{(2d)^n} \sum_{\gamma\in\Gamma_n} e^{\beta H_n(\gamma) - n\lambda(\beta)} = W_n(\beta).\end{aligned}$$

La suite $(W_n(\beta))_{n\geqslant 1}$ est donc une $(\mathcal{G}_n)_n$-martingale positive. Elle converge donc p.s. vers une variable aléatoire positive $W_\infty(\beta)$. De plus, d'après le lemme de Fatou,

$$\mathbb{E}(W_\infty(\beta)) = \mathbb{E}\Big(\liminf_{n\to\infty} W_n(\beta)\Big) \leqslant \liminf_{n\to\infty} \mathbb{E}(W_n(\beta)) = 1,$$

ce qui est le résultat annoncé. □

Rien n'assure pour l'instant que $W_\infty(\beta)$ soit d'intégrale 1. Le résultat établit que $W_\infty(\beta)$ est soit nulle soit strictement positive p.s. .

Lemme 15.6 (Loi du zéro-un)**.** *Pour tout $\beta \geqslant 0$, $\mathbb{P}(W_\infty(\beta) = 0) \in \{0, 1\}$, et pour tout $\theta \in (0, 1)$,*

$$\lim_{n\to\infty} \mathbb{E}(W_n(\beta)^\theta) = \mathbb{E}(W_\infty(\beta)^\theta).$$

Recueil de modèles stochastiques – Page 148/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Démonstration. La propriété de Markov de la marche assure que

$$W_{n+m}(\beta) = \mathbb{E}\Big(e^{\beta H_n(S_{0:n})-n\lambda(\beta)} Z_{n,m}^{S_n} e^{-m\lambda(\beta)} | \mathcal{G}\Big)$$

où, pour $x \in \mathbb{Z}^d$ et n et m entiers strictement positifs,

$$Z_{n,m}^x(\beta) = \mathbb{E}\left[\exp\left(\beta \sum_{i=1}^m \eta(n+i, x+S_i)\right) | \mathcal{G}\right].$$

Ainsi,

$$W_\infty(\beta) = \lim_{m\to\infty} W_{n+m}(\beta) = \mathbb{E}\Big(e^{\beta H_n(S_{0:n})-n\lambda(\beta)} \lim_{m\to\infty}\Big(Z_{n,m}^{S_n} e^{-m\lambda(\beta)}\Big) | \mathcal{G}\Big).$$

Puisque $e^{\beta H_n(S_{0:n})-n\lambda(\beta)}$ est strictement positif p.s. , l'événement

$$\{W_\infty(\beta) = 0\} = \Big\{\lim_{m\to\infty}\Big(Z_{n,m}^x e^{-m\lambda(\beta)}\Big) = 0 \ : \ \forall x, \ \mathbb{P}(S_n = x) > 0\Big\}$$

appartient à la tribu $\sigma\big(\eta(j,x), \ j \geqslant n, \ x \in \mathbb{Z}^d\big)$. Ceci est vrai pour tout n. Il appartient donc à la tribu asymptotique.

$$\bigcap_{n\geqslant 1} \sigma\Big(\eta(j,x), \ j \geqslant n, \ x \in \mathbb{Z}^d\Big).$$

D'après la loi du zéro-un de Kolmogorov, tous les éléments de la tribu asymptotique sont de probabilité 0 ou 1. Puisque $W_n(\beta)$ est d'espérance 1, la suite $(W_n(\beta)^\theta)_{n\geqslant 0}$ est bornée dans $L^{1/\theta}$. De plus, elle converge vers $W_\infty(\beta)^\theta$ p.s. . On a donc

$$\lim_{n\to\infty} \mathbb{E}(W_n(\beta)^\theta) = \mathbb{E}(W_\infty(\beta)^\theta).$$

Cette limite est nulle si et seulement si $W_\infty(\beta)$ est nul presque sûrement. □

Il reste à établir une propriété de monotonie pour la dichotomie ci-dessus.

Lemme 15.7 (Monotonie)**.** *Pour tout $\theta \in (0,1)$, la fonction $\beta \mapsto \mathbb{E}(W_\infty(\beta)^\theta)$ décroît.*

Démonstration. On remarque tout d'abord que

$$\frac{\partial}{\partial\beta} W_n(\beta) = \mathbb{E}\Big((H_n(S_{0:n}) - \lambda'(\beta)) e^{H_n(S_{0:n})-n\lambda(\beta)} | \mathcal{G}\Big).$$

On a donc, après dérivation sous l'espérance,

$$\frac{\partial}{\partial\beta}\mathbb{E}(W_n(\beta)^\theta) = \theta\mathbb{E}\Big(W_n(\beta)^{\theta-1}(H_n(S_{0:n}) - n\lambda'(\beta)) e^{\beta H_n(S_{0:n})-n\lambda(\beta)}\Big).$$

L'inégalité FKG du lemme 15.4 appliquée à la mesure de probabilité de densité $e^{\beta H_n(\gamma)-n\lambda(\beta)}$ par rapport à $\mathbb{P}$ et aux fonctions $W_n(\beta)^{\theta-1}$ et $H_n(\gamma) - n\lambda'(\beta)$ qui sont des fonctions respectivement décroissante et croissante des variables $(\eta(k,x))_{(k,x)\in E_n}$ assure que

$$\mathbb{E}\Big(W_n(\beta)^{\theta-1}(H_n(\gamma) - n\lambda'(\beta)) e^{\beta H_n(\gamma)-n\lambda(\beta)}\Big)$$

est inférieur à

$$\mathbb{E}\Big(W_n(\beta)^{\theta-1} e^{\beta H_n(\gamma)-n\lambda(\beta)}\Big)\mathbb{E}\Big((H_n(\gamma) - n\lambda'(\beta)) e^{\beta H_n(\gamma)-n\lambda(\beta)}\Big).$$

On utilise alors que la quantité $\mathbb{E}\big((H_n(\gamma) - n\lambda'(\beta)) e^{\beta H_n(\gamma)-n\lambda(\beta)}\big)$ est nulle pour conclure. □

On peut à présent conclure. Par passage à la limite, on obtient que la fonction $\beta \mapsto \mathbb{E}(W_\infty(\beta)^\theta)$ est décroissante. Il existe donc $\bar\beta_c \in [0, +\infty]$ tel que, p.s. ,

$$\begin{cases} W_\infty(\beta) > 0 & \text{si } \beta < \bar\beta_c, \\ W_\infty(\beta) = 0 & \text{si } \beta > \bar\beta_c. \end{cases}$$

Ceci achève la démonstration du théorème 15.1.

Recueil de modèles stochastiques – Page 149/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

15.3 Borne inférieure en grandes dimensions

On se place ici en dimension $d \geqslant 3$. Dans ce cas, la marche aléatoire simple $(S_n)_{n\geqslant 0}$ est transiente. La probabilité qu'elle revienne en 0 est strictement inférieure à 1. Notons π_d cette probabilité.

Théorème 15.8 (Critère de non nullité)**.** *Si $d \geqslant 3$ et si β est tel que*

$$\tau(\beta) = \lambda(2\beta) - 2\lambda(\beta) < \log(1/\pi_d), \tag{15.1}$$

alors $W_\infty > 0$ p.s. .

Démonstration. Pour tout $\beta \geqslant 0$, on a

$$\tau'(\beta) = 2(\lambda'(2\beta) - \lambda'(\beta)) > 0$$

puisque λ est strictement convexe. Il en découle que τ est donc strictement croissante. Soit à présent une suite $(\tilde{U}_n)_{n\geqslant 1}$ de même loi que $(U_n)_{n\geqslant 1}$, de sorte que

$$(U_n)_{n\geqslant 1}, \quad (\tilde{U}_n)_{n\geqslant 1}, \quad (\eta(n,x))_{n\geqslant 0,\, x\in\mathbb{Z}^d}$$

soient indépendantes. Soit $(\tilde{S}_n)_{n\geqslant 1}$ la suite construite à partir de $(\tilde{U}_n)_{n\geqslant 1}$ à la manière de $(S_n)_{n\geqslant 1}$ à partir de $(U_n)_{n\geqslant 1}$. Par indépendance de (S_n) et $(\tilde{S}_n)$, on a

$$\begin{aligned} W_n(\beta)^2 &= \mathbb{E}\left(\prod_{k=1}^n e^{\beta\eta(k,S_k)-\lambda(\beta)}|\mathcal{G}\right)\mathbb{E}\left(\prod_{k=1}^n e^{\beta\eta(k,\tilde{S}_k)-\lambda(\beta)}|\mathcal{G}\right) \\ &= \mathbb{E}\left(\prod_{k=1}^n e^{\beta(\eta(k,S_k)+\eta(k,\tilde{S}_k))-2\lambda(\beta)}|\mathcal{G}\right). \end{aligned}$$

Prenons l'espérance :

$$\mathbb{E}\big(W_n^2\big) = \mathbb{E}\left[\mathbb{E}\left(\prod_{k=1}^n e^{\beta(\eta(k,S_k)+\eta(k,\tilde{S}_k))-2\lambda(\beta)}|\mathcal{F},\tilde{\mathcal{F}}\right)\right].$$

Or, pour tout $x, \tilde{x} \in \mathbb{Z}^d$,

$$\mathbb{E}\Big(e^{\beta(\eta(k,x)+\eta(k,\tilde{x}))-2\lambda(\beta)}\Big) = \begin{cases} e^{\tau(\beta)} & \text{si } x = \tilde{x}, \\ 1 & \text{si } x \neq \tilde{x}. \end{cases}$$

On a donc

$$\mathbb{E}\big(W_n(\beta)^2\big) = \mathbb{E}\Big(e^{\tau(\beta)I_n}\Big) \quad \text{où} \quad I_n = \sum_{k=0}^n \mathbf{1}_{\{S_k=\tilde{S}_k\}}.$$

On remarque alors que les suites $(S_k - \tilde{S}_k)_{k\geqslant 0}$ et $(S_{2k})_{k\geqslant 0}$ ont même loi puisque $-\tilde{U}_n$ et U_n ont même loi. La variable aléatoire I_n a donc même loi que le nombre N_{2n} de retours en 0 de S avant l'instant $2n$. Par convergence monotone,

$$\mathbb{E}\Big(e^{\tau(\beta)I_n}\Big) = \mathbb{E}\Big(e^{\tau(\beta)N_{2n}}\Big) \xrightarrow[n\to\infty]{} \mathbb{E}\Big(e^{\tau(\beta)N_\infty}\Big).$$

La propriété de Markov forte de S assure que la loi de N_∞ est la loi géométrique de paramètre $1-\pi_d$ et que, pour tout $t \in \mathbb{R}$,

$$\mathbb{E}\big(e^{tN_\infty}\big) = \sum_{k=1}^\infty (1-\pi_d)\pi_d^{k-1}e^{tk} = \begin{cases} \dfrac{(1-\pi_d)e^t}{1-\pi_d e^t} & \text{si } t < -\log\pi_d, \\ +\infty & \text{sinon.} \end{cases}$$

Recueil de modèles stochastiques – Page 150/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

On obtient ainsi une condition nécessaire et suffisante pour que la suite $(W_n(\beta))_{n\geqslant 1}$ soit bornée dans L^2 :

$$\sup_{n\geqslant 1} \mathbb{E}\big(W_n(\beta)^2\big) < \infty \Longleftrightarrow \tau(\beta) < \log(1/\pi_d).$$

Si $\tau(\beta) < \log(1/\pi_d)$, la martingale $(W_n(\beta))_{n\geqslant 1}$ converge dans L^2 et p.s. vers $W_\infty(\beta)$. En particulier, $\mathbb{E}(W_\infty(\beta)) = \lim_n \mathbb{E}(W_n(\beta)) = 1$ et par suite, $W_\infty(\beta) > 0$ p.s. . □

Corollaire 15.9 (Critère de non nullité)**.** *Si $d \geqslant 3$ alors pour tout environnement, $\bar{\beta}_c > 0$.*

Démonstration. La fonction τ est nulle en 0 et régulière sur $\mathbb{R}$. Elle est de plus strictement croissante car

$$\tau'(\beta) = 2(\lambda'(2\beta) - \lambda'(\beta)) > 0,$$

par stricte convexité de λ. La condition (15.1) est donc toujours vérifiée pour β assez petit et ce, quel que soit l'environnement. □

Exemple 15.10 (Environnement gaussien)**.** *Si l'environnement est gaussien, $\tau(\beta) = \beta^2$ et $\tau(\beta) < \log(1/\pi_d)$ si et seulement si $\beta < \sqrt{\log(1/\pi_d)}$. On a donc $\bar{\beta}_c > \sqrt{\log(1/\pi_d)}$.*

Exemple 15.11 (Environnement Bernoulli)**.** *Si η suit la loi de Bernoulli $\mathcal{B}(p)$, alors*

$$\lambda(\beta) = \log\Big(pe^\beta + 1 - p\Big) \quad \textit{et} \quad \lim_{\beta\to\infty} \tau(\beta) = -\log p.$$

Ainsi, dès que $p > \pi_d$, la condition (15.1) *est satisfaite pour tout $\beta \geqslant 0$, et $\bar{\beta}_c$ est infini.*

15.4 Énergie libre moyenne

L'objet principal de cette section est de démontrer le théorème 15.2. On démontre tout d'abord que l'énergie libre moyenne converge.

Lemme 15.12 (Convergence de l'énergie libre moyenne)**.** *Pour tout $\beta \geqslant 0$, il existe une constante $u(\beta)$ inférieure à $\lambda(\beta)$ telle que*

$$u_n(\beta) \xrightarrow[n\to+\infty]{} u(\beta).$$

Démonstration. L'inégalité de Jensen assure que

$$u_n(\beta) = \frac{1}{n}\mathbb{E}(\log Z_n(\beta)) \leqslant \frac{1}{n}\log \mathbb{E}(Z_n(\beta)).$$

Puisque $\mathbb{E}(W_n(\beta)) = 1$, on a $\mathbb{E}(Z_n(\beta)) = e^{n\lambda(\beta)}$ et $u_n(\beta) \leqslant \lambda(\beta)$.

Conditionnellement à $\mathcal{G}$, la propriété de Markov de la marche S assure que

$$\mathbb{E}\Big(e^{\beta H_{n+m}(S_{0:n+m})}|\mathcal{G}\Big) = \mathbb{E}\Big(e^{\beta H_n(S_{0:n})} Z_{n,m}^{S_n}|\mathcal{G}\Big).$$

Par définition de $Z_n(\beta)$ et μ_n, on a simplement

$$\begin{aligned}\mathbb{E}\Big(e^{\beta H_n(S_{0:n})} Z_{n,m}^{S_n}|\mathcal{G}\Big) &= \frac{\mathbb{E}\big(e^{\beta H_n(S_{0:n})} Z_{n,m}^{S_n}|\mathcal{G}\big)}{\mathbb{E}\big(e^{\beta H_n(S_{0:n})}|\mathcal{G}\big)} Z_n(\beta) \\ &= \mu_n\big(Z_{n,m}^{S_n}\big) Z_n(\beta).\end{aligned}$$

En d'autres termes,

$$\log Z_{n+m}(\beta) = \log Z_n(\beta) + \log \sum_{x\in\mathbb{Z}^d} \mu_n(S_n = x) Z_{n,m}^x.$$

Recueil de modèles stochastiques – Page 151/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

L'inégalité de Jensen pour la mesure μ_n assure que

$$\log Z_{n+m}(\beta) \geqslant \log Z_n(\beta) + \sum_{x\in\mathbb{Z}^d} \mu_n(S_n = x) \log Z^x_{n,m}.$$

La mesure aléatoire μ_n est $\mathcal{G}_n$-mesurable et $Z^x_{n,m}$ est indépendante de $\mathcal{G}_n$. On a donc

$$\mathbb{E}\left(\sum_{x\in\mathbb{Z}^d} \mu_n(S_n = x) \log Z^x_{n,m} | \mathcal{G}_n\right) = \sum_{x\in\mathbb{Z}^d} \mu_n(S_n = x)\mathbb{E}\big(\log Z^x_{n,m} | \mathcal{G}_n\big).$$

Les v.a. $Z^x_{n,m}$ et Z_m ont même loi car les v.a. $(\eta(n,x))_{n,x}$ sont i.i.d. En prenant l'espérance, on obtient que la suite $(\mathbb{E}(\log Z_m))_{n\geqslant 1}$ est sur-additive, et donc, d'après le lemme de Fekete, $(u_n(\beta))_{n\geqslant 1}$ converge dans $\mathbb{R} \cup \{+\infty\}$. Comme elle est bornée par $\lambda(\beta)$, sa limite $u(\beta)$ est finie. □

Le phénomène de seuil va découler d'une propriété de monotonie.

Lemme 15.13 (Monotonie)**.** *La fonction $\beta \mapsto u(\beta) - \lambda(\beta)$ est décroissante sur $\mathbb{R}^+$.*

Démonstration. On étudie tout d'abord les variations de $\beta \mapsto u_n(\beta) - \lambda(\beta)$. On a

$$\begin{aligned}\frac{\partial}{\partial\beta}\mathbb{E}(\log Z_n(\beta)) &= \mathbb{E}\left(\frac{\mathbb{E}\big(H_n(S_{0:n})e^{\beta H_n(S_{0:n})}|\mathcal{G}\big)}{Z_n(\beta)}\right)\\ &= \frac{1}{(2d)^n}\sum_{\gamma\in\Gamma_n}\mathbb{E}\left(\frac{H_n(\gamma)e^{\beta H_n(\gamma)}}{Z_n(\beta)}\right).\end{aligned}$$

Remarquons que $H_n(\gamma)$ (resp. $Z_n^{-1}(\beta)$) est une fonction croissante (resp. décroissante) des variables $(\eta(i,x))_{(i,x)\in E_n}$. D'autre part, la mesure de densité $e^{\beta H_n(\gamma) - n\lambda(\beta)}$ par rapport à la mesure $\mathbb{P}$ est une mesure de probabilité. L'inégalité FKG du lemme 15.4 assure donc

$$\mathbb{E}\left(\frac{H_n(\gamma)}{Z_n(\beta)}e^{\beta H_n(\gamma)-n\lambda(\beta)}\right) \leqslant \mathbb{E}\Big(H_n(\gamma)e^{\beta H_n(\gamma)-n\lambda(\beta)}\Big)\mathbb{E}\Big(Z_n(\beta)^{-1}e^{\beta H_n(\gamma)-n\lambda(\beta)}\Big).$$

On obtient ainsi

$$\begin{aligned}\frac{\partial}{\partial\beta}\mathbb{E}(\log Z_n(\beta)) &\leqslant \frac{1}{(2d)^n}\sum_{\gamma\in\Gamma_n} e^{-n\lambda(\beta)}\mathbb{E}\left(\frac{e^{\beta H_n(\gamma)}}{Z_n(\beta)}\right)\mathbb{E}\Big(H_n(\gamma)e^{\beta H_n(\gamma)}\Big)\\ &\leqslant n\lambda'(\beta)\mathbb{E}\left(\frac{1}{(2d)^n}\sum_{\gamma\in\Gamma_n}\frac{e^{\beta H_n(\gamma)}}{Z_n(\beta)}\right)\\ &\leqslant n\lambda'(\beta).\end{aligned}$$

En d'autres termes, $(u_n - \lambda)'(\beta) \leqslant 0$: pour tout $n \geqslant 1$ la fonction $u_n - \lambda$ est décroissante. C'est encore vrai en passant à la limite quand $n \to \infty$: la fonction $u - \lambda$ est décroissante sur $\mathbb{R}_+$. □

Les fonctions u et λ sont nulles en 0. Il existe donc $\beta_c \in [0, +\infty]$ tel que

$$\begin{cases} u(\beta) = \lambda(\beta) & \text{si } \beta < \beta_c,\\ u(\beta) < \lambda(\beta) & \text{si } \beta > \beta_c.\end{cases}$$

Recueil de modèles stochastiques – Page 152/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

15.5 Borne supérieure sur le paramètre critique

Le but de cette section est d'établir une borne supérieure sur β_c.

Lemme 15.14. *Les fonctions $(q_n)_{n\geqslant 1}$ et u vérifient les propriétés suivantes :*

1. *q_n est une fonction régulière, strictement convexe, $q_n(0) = 0$; u est convexe et nulle en 0.*
2. *$\beta \mapsto \beta^{-1} q_n(\beta)$ est croissante.*
3. *$\beta \mapsto \beta^{-1}(q_n(\beta) + \log(2d))$ est décroissante.*

Démonstration. La fonction $\beta \mapsto nq_n(\beta)$ est le logarithme de la transformée de Laplace de la variable aléatoire $H_n(S_{0:n})$ (les variables $\eta(\cdot,\cdot)$ étant fixées). La fonction q_n est donc de classe $\mathcal{C}^2$, convexe et nulle en 0. Ces propriétés sont stables par passage à l'espérance puis à la limite quand n tend vers l'infini : u est convexe sur $\mathbb{R}_+$ et nulle en 0. Puisque q_n est convexe, la fonction (pente de la corde entre $(0,0)$ et $(\beta, q_n(\beta))$) $\beta \mapsto q_n(\beta)/\beta = (q_n(\beta) - q_n(0))/\beta$ est croissante.

Pour le point 3. on dérive pour obtenir

$$\left(\frac{q_n(\beta) + \log(2d)}{\beta}\right)' = -\frac{1}{\beta^2}(q_n(\beta) + \log(2d)) + \frac{1}{n\beta}\mu_n(H_n).$$

Par ailleurs, l'entropie de μ_n définie par

$$\begin{aligned} h_n(\mu_n) &= \sum_{\gamma\in\Gamma_n} \frac{e^{\beta H(\gamma)}}{Z_n(2d)^n} \log\left(\frac{e^{\beta H(\gamma)}}{Z_n(2d)^n}\right) \\ &= \beta\,\mu_n(H_n) - (\log Z_n + n\log(2d)) \end{aligned}$$

est négative. Ainsi,

$$\left(\frac{q_n(\beta) + \log(2d)}{\beta}\right)' = \frac{1}{n\beta^2} h_n(\mu_n) \leqslant 0.$$

La fonction $\beta \mapsto \beta^{-1}(q_n(\beta) + \log(2d))$ est décroissante sur $\mathbb{R}_+^*$. □

Le lemme 15.14 fournit une majoration de u et de β_c.

Théorème 15.15 (Majorations)**.** *Si la condition suivante est vérifiée :*

$$\lim_{\beta\to+\infty} \beta\lambda'(\beta) - \lambda(\beta) > \log(2d),$$

alors l'équation $\beta\lambda'(\beta) = \lambda(\beta) + \log(2d)$ admet une unique racine β_1 et, pour tout $\beta > \beta_1$,

$$u(\beta) \leqslant \frac{\beta}{\beta_1}(\lambda(\beta_1) + \log(2d)) - \log(2d),$$

Enfin, on a $\beta_c \leqslant \beta_1$.

Démonstration. Puisque λ est strictement convexe, $\beta \mapsto \beta\lambda'(\beta) - \lambda(\beta)$ est strictement croissante sur $\mathbb{R}_+$ (sa dérivée vaut $\beta \mapsto \beta\lambda''(\beta)$). Elle est de plus nulle en 0. L'équation $\beta\lambda'(\beta) = \lambda(\beta) + \log(2d)$ a donc au plus une solution. Elle existe si et seulement si la condition du théorème est satisfaite.

Le lemme 15.14 assure que la fonction $\beta \mapsto (u_n(\beta) + \log(2d))/\beta$ est décroissante. Comme de plus $u_n \leqslant \lambda$, on obtient

$$\begin{aligned} \frac{u_n(\beta) + \log(2d)}{\beta} &= \inf_{\alpha\in]0,\beta]} \frac{u_n(\alpha) + \log(2d)}{\alpha} \\ &\leqslant \inf_{\alpha\in]0,\beta]} \frac{\lambda(\alpha) + \log(2d)}{\alpha}. \end{aligned}$$

Recueil de modèles stochastiques – Page 153/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

La fonction $\alpha \mapsto (\lambda(\alpha) + \log(2d))/\alpha$ atteint son minimum sur $[0, \beta]$ en β ou au point où sa dérivée s'annule à savoir en β_1. Donc, pour tout $\beta > \beta_1$,

$$u(\beta) \leqslant \frac{(\lambda(\beta_1) + \log(2d))}{\beta_1}\beta - \log(2d).$$

Il reste à remarquer que pour $\beta > \beta_1$,

$$\frac{(\lambda(\beta_1) + \log(2d))}{\beta_1}\beta - \log(2d) < \lambda(\beta)$$

puisque λ est strictement convexe et qu'il y a égalité en β_1. □

Exemple 15.16 (Environnement gaussien et environnement Bernoulli)**.** *Si $\eta \sim \mathcal{N}(0,1)$ alors $\beta\lambda'(\beta) - \lambda(\beta) = \beta^2/2$, et dans ce cas $\beta_1 = \sqrt{2\log(2d)}$. En revanche, si $\eta \sim \mathcal{B}(p)$ alors*

$$\beta\lambda'(\beta) - \lambda(\beta) = \frac{p\beta e^\beta}{pe^\beta + 1 - p} - \log\left(pe^\beta + 1 - p\right) \underset{\beta\to\infty}{\sim} -\log p,$$

et β_1 est fini si et seulement si $2dp < 1$.

15.6 Notes et commentaires

Ce modèle de polymère dirigé en environnement aléatoire a été introduit par John Imbrie et Thomas Spencer [IS88]. La notion de fort désordre a été étudiée en détail par Erwin Bolthausen dans [Bol89], qui démontre en particulier les théorèmes 15.1 et 15.8. Plus tard la notion de très fort désordre a été étudiée notamment par Francis Comets et Nobuo Yoshida [CY06], qui démontrent le théorème 15.2. En utilisant une inégalité de concentration pour q_n autour de sa moyenne, il est possible (voir [CH02, LW09]) d'améliorer les conclusions du théorème 15.2, en démontrant que pour tout $\beta \geqslant 0$,

$$q_n(\beta) \xrightarrow[n\to\infty]{} u(\beta) \quad \text{p.s. et dans } L^1.$$

En dimension 1 et 2, on sait démontrer que $\beta_c = 0$. Philippe Carmona et Yueyun Hu ont établi dans [CH02] que $\bar{\beta}_c = 0$ pour $d = 1$ et $d = 2$ lorsque l'environnement est gaussien. Le cas général est traité par Francis Comets, Tokuzo Shiga et Nobuo Yoshida dans [CSY03]. Par la suite, Francis Comets et Vincent Vargas ont établi dans [CV06] que β_c est nul pour $d = 1$. Enfin, Hubert Lacoin a démontré dans [Lac10] que c'est encore vrai pour $d = 2$. La question de l'égalité des deux paramètres critiques en dimension supérieure ou égale à 3 semble toujours ouverte à l'heure actuelle.

Pour conclure, signalons que par soucis de clarté nous avons modifié légèrement les notations classiques utilisées dans les articles de recherche cités plus haut. En effet, la mesure de référence est en général écrite comme le produit de deux mesures, l'une pour l'environnement et l'autre pour la loi de la marche. Il faut alors introduire deux signes d'espérance, qui correspondent aux espérances conditionnelles par rapport aux tribus $\mathcal{F}$ et $\mathcal{G}$ dans le présent chapitre.

Recueil de modèles stochastiques – Page 154/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Chapitre **16**

Matrices aléatoires

Recueil de modèles stochastiques – Page 155/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Mots-clés. Matrice aléatoire ; matrice de covariance ; spectre ; convergence étroite ; théorème des moments ; fonction caractéristique ; combinatoire ; nombres de Catalan ; théorème de Wigner ; loi du demi-cercle ; théorème de Marchenko-Pastur ; loi du quart de cercle ; vecteurs gaussiens.

Nous commençons par une motivation statistique. Soit $X_1, \dots, X_n$ des vecteurs colonnes 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 $1 \leqslant i, j \leqslant d$ et tout $1 \leqslant k \leqslant n$ ou encore

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

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

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

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

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

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 $(Y_{ij})_{i,j\geqslant 1}$ de v.a.r. i.i.d. de moyenne 0 et de variance 1, et pour tout entier $n \geqslant 1$, on considère la matrice aléatoire $d_n \times d_n$ symétrique

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

où $Y = (Y_{ij})_{1\leqslant i\leqslant d_n, 1\leqslant j\leqslant n}$. Si $n \mapsto d_n$ est constante et égale à d alors $\frac{1}{n}YY^\top$ converge p.s. vers I. Il est naturel de chercher à comprendre le comportement de la matrice $\frac{1}{n}YY^\top$ lorsque d_n dépend de n, par exemple en étudiant son spectre. L'analyse de la matrice symétrique $YY^\top$ 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.

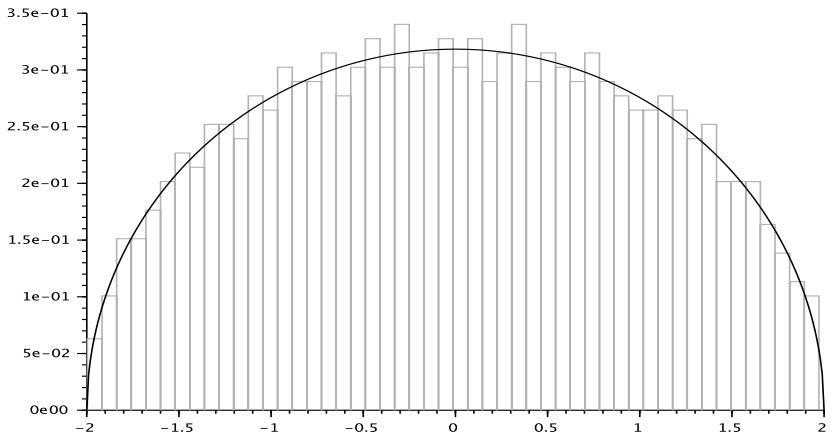


Figure 16.1 – Illustration du théorème 16.1 : densité de la loi du demi-cercle μ_1^{DC} et histogramme des valeurs propres d'une matrice symétrique de taille $n = 1000$ avec la loi uniforme sur $[-\sqrt{3/n}, \sqrt{3/n}]$ (dont la variance vaut $1/n$).

16.1 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 $\lambda_{n,1}, \ldots, \lambda_{n,n}$ les valeurs propres de la matrice symétrique $\frac{1}{\sqrt{n}}M$, ordonnées de sorte que $\lambda_{n,1} \geqslant \cdots \geqslant \lambda_{n,n}$. 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-Cantelli [1], mis à part que les atomes $\lambda_{n,1}, \ldots, \lambda_{n,n}$ sont ici des variables aléatoires dépendantes [2]. Pour tout borélien B de $\mathbb{R}$ on a

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

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

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

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

Théorème 16.1 (Théorème de Wigner)**.** *Presque sûrement, la suite de mesures de probabilités* $(\mu_n)_{n\geqslant 1}$ *converge étroitement* [3] *vers la loi du demi-cercle* μ_σ^{DC} *de densité*

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

1. Si $(Y_n)_{n\geqslant 1}$ 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 $n \to \infty$, 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.

Recueil de modèles stochastiques – Page 156/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Le nom *loi du demi-cercle* utilisé pour désigner μ_σ^{DC} vient du 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 $I \subset \mathbb{R}$,

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

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 $n \to \infty$. 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

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

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

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

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 $r > 0$,

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

Par conséquent, presque sûrement, pour tout $\varepsilon > 0$ il existe un compact $K_\varepsilon = [-r_\varepsilon, r_\varepsilon]$ tel que $\mu_n(K_\varepsilon) \leqslant \varepsilon$ pour tout $n \geqslant 1$. Ainsi, presque sûrement, la suite $(\mu_n)_{n\geqslant 1}$ 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[4] $(\mathbb{E}\mu_n)_{n\geqslant 1}$. La réduction du théorème de Wigner au théorème de Wigner faible se fait en utilisant des techniques regroupées dans la section 16.6 pour faciliter la première lecture.

Théorème 16.2 (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 $f : \mathbb{R} \to \mathbb{R}$,*

$$\lim_{n\to\infty} \mathbb{E}\int f\, d\mu_n = \int f\, d\mu_1^{\mathrm{DC}}.$$

La preuve du théorème 16.2 consiste à se ramener à des fonctions test f polynomiales (théorème 16.3 ci-dessous), puis à établir la convergence des moments (théorème 16.6 ci-dessous), en s'inspirant de la méthode utilisée ci-dessus pour les moments d'ordre 1 et 2.

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.

Recueil de modèles stochastiques – Page 157/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

16.2 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 16.3 (Convergence des moments et convergence étroite)**.** *Si $\mu, \mu_1, \mu_2, \dots$ sont des éléments de $\mathcal{P}$ vérifiant, pour tout $P \in \mathbb{R}[X]$,*

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

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

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

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

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

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

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

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

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

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

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

D'un autre côté, on sait que

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

et on obtient donc

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

En utilisant cela pour P^2, qui est $\geqslant 0$, 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 $\mu = \nu$. □

Recueil de modèles stochastiques – Page 158/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

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 16.4 (Analycité 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 $\overline{\lim}_{n\to\infty} \frac{1}{n}|\kappa_n|^{1/n} < \infty$ 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 $(\kappa_n)_{n\geqslant 1}$. Le rayon de convergence r de la série entière $\sum_n a_n z^n$ associée à la suite de nombres complexes $(a_n)_{n\geqslant 0}$ est donné par la formule de Hadamard $r^{-1} = \overline{\lim}_n |a_n|^{\frac{1}{n}}$. Ainsi, 1$\Leftrightarrow$ 3 (prendre $a_n = i^n\kappa_n/n!$). D'autre part, comme pour tout $n \in \mathbb{N}$, $s, t \in \mathbb{R}$,

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

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

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

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 $\geqslant r$. De proche en proche, on obtient que φ est caractérisée par ses dérivées en zéro. Si μ est à support compact, alors $\sup_n |\kappa_n|^{\frac{1}{n}} < \infty$ et donc 3 a lieu (via Stirling $n! = (n/e)^n\sqrt{2\pi n}(1 + \mathcal{O}(1/n))$). □

16.3 Moments de la loi du demi-cercle

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

Théorème 16.5 (Moments de la loi du demi-cercle)**.** *Pour tout entier $r \geqslant 0$,*

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

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.

Recueil de modèles stochastiques – Page 159/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

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

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

À présent, en utilisant le changement de variables $x = 2\cos(u)$ 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. □

16.4 Convergence des moments et théorème de Wigner faible

Théorème 16.6 (Convergence des moments)**.** *Sous les hypothèses du théorème de Wigner faible 16.2 on a, pour tout entier $r \geqslant 1$,*

$$\mathbb{E}\int x^r d\mu_n \underset{n\to\infty}{\longrightarrow} \int x^r\, d\mu_\sigma^{\mathrm{DC}}.$$

Démonstration. Les moments de μ_1^{DC} sont donnés par le théorème 16.5. Comme par hypothèse, les coefficients de la matrice aléatoire M sont centrés et réduits, on a

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

et

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

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

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

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 $n \to \infty$ vers 0, qui est le moment d'ordre 3 de μ_1^{DC}.

Plus généralement, pour le moment d'ordre $r \geqslant 1$, on écrit (avec $i_{r+1} := i_1$)

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

Nous allons associer à chaque r-uplet d'indices $i_1,\dots,i_r$ un graphe orienté, obtenu en positionnant leurs valeurs sur une droite horizontale figurant $\mathbb{N}^*$, cf. figure 16.2. 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 $1 \leqslant k \leqslant r+1$ 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

Recueil de modèles stochastiques – Page 160/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

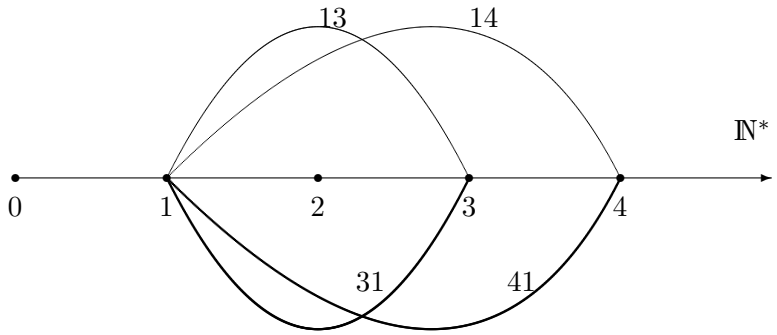


FIGURE 16.2 – Graphe pour $r = 4$ et $i_1 = i_3 = 1$, $i_2 = 4$, $i_4 = 3$, correspondant à $\mathbb{E}(M_{14}M_{41}M_{13}M_{31})$. Les arrêtes successives sont 14, 41, 13, 31. On a $t = 3$.

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). Afin de calculer les contributions, on distingue trois types $\mathcal{G}_1$, $\mathcal{G}_2$ et $\mathcal{G}_3$ 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 de type 3 alors il 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 $2t \leqslant r+1$, qui donne la majoration $n(n-1)\cdots(n-t+1) \leqslant n^t \leqslant n^{1/2+r/2}$. D'autre part, on a $\mathrm{card}(\mathcal{G}_3) \leqslant \mathrm{card}(G(r,t)) \leqslant C^r = \mathcal{O}(1)$, et comme les coefficients de M sont bornés, on a également $\mathbb{E}(M_G) \leqslant C^r = \mathcal{O}(1)$, d'où enfin

$$\sum_{G \in \mathcal{G}_3} \frac{n(n-1)\cdots(n-t+1)}{n^{1+r/2}} \mathbb{E}(M_G) = \mathcal{O}(n^{-1/2}) = o_{n\to\infty}(1);$$

— *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 $\mathrm{card}(\mathcal{G}_1)$. Si r est impair alors $\mathrm{card}(\mathcal{G}_1) = 0$, 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 2.5 page 16). iIl y en a donc $\frac{1}{s+1}\binom{2s}{s}$, et

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

□

Recueil de modèles stochastiques – Page 161/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

16.5 Retour aux matrices de covariance empiriques

Considérons le modèle de covariance du début du chapitre. Soit $\lambda_{n,1} \leqslant \cdots \leqslant \lambda_{n,d_n}$ le spectre de $\frac{1}{n}YY^\top$ 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.

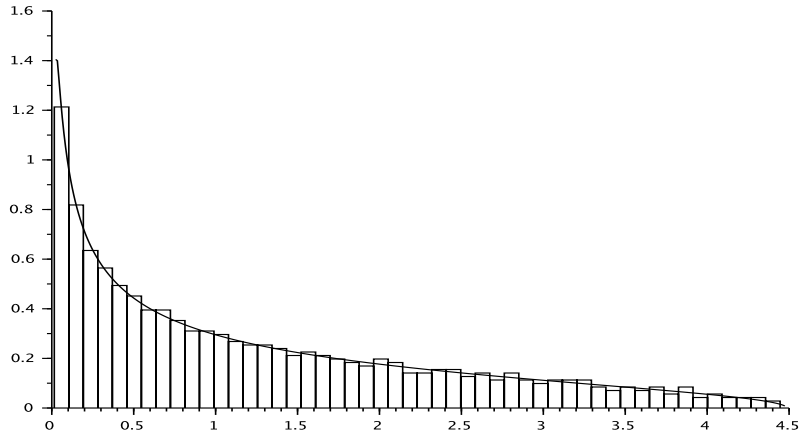


Figure 16.3 – Illustration du théorème 16.7 : histogramme des valeurs propres non nulles d'une matrice de covariance empirique avec $n = 800$ et $d = 1000$ et loi de Marchenko-Pastur.

Théorème 16.7 (Marchenko-Pastur)**.** *Si* $\Sigma = I_{d_n}$ *et si* $\lim_{n\to\infty} \frac{d_n}{n} = \rho \in]0, \infty[$ *alors, presque sûrement, pour toute fonction continue et bornée* $f : \mathbb{R} \to \mathbb{R}$*,*

$$\int f \, d\mu_n \underset{n\to\infty}{\longrightarrow} \int f \, d\mu_\rho^{\mathrm{MP}}$$

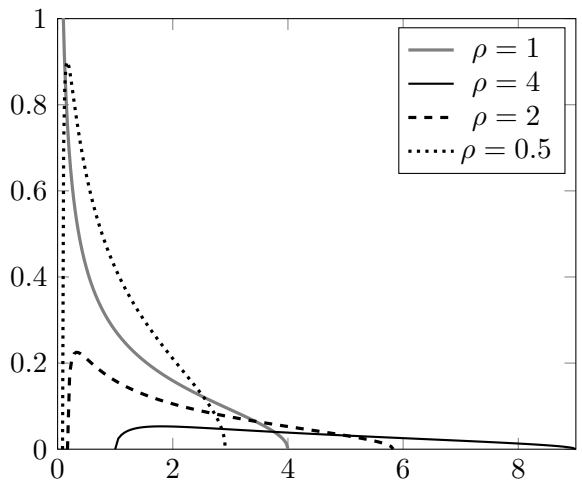


Figure 16.4 – Partie à densité de μ_ρ^{MP} pour différentes valeurs de ρ.

Recueil de modèles stochastiques – Page 162/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

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

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

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

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

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 $\rho \leqslant 1$. Le théorème de Marchenko-Pastur affirme que p.s., pour tout $x \in \mathbb{R}$, avec $x \neq 0$ si $\rho > 1$, en notant $I =]-\infty, x]$,

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

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 $n \to \infty$ 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).

16.6 Quelques outils

Cette section présente brièvement quelques outils de portée très générale qui permettent en particulier de réduire le théorème de Wigner 16.1 au théorème de Wigner faible 16.2.

Lemme 16.8 (Formules variationnelles min-max de Courant-Fischer)**.** *Soit* $A \in \mathcal{M}_n(\mathbb{C})$ *hermitienne de valeurs propres* $\lambda_1(A) \geqslant \cdots \geqslant \lambda_n(A)$. *Alors pour tout* $1 \leqslant k \leqslant n$, *en notant* $\mathcal{G}_k$ *l'ensemble des sous-espaces vectoriels de* $\mathbb{C}^n$ *de dimension* k, *on a*

$$\lambda_k(A) = \max_{V\in\mathcal{G}_k} \min_{\substack{x\in V\\ \|x\|_2=1}} \langle Ax, x\rangle = \min_{V\in\mathcal{G}_{n-k+1}} \max_{\substack{x\in V\\ \|x\|_2=1}} \langle Ax, x\rangle.$$

Les cas $k = 1$ et $k = n$ sont familiers. Ces formules fournissent un entrelacement pour le spectre des perturbations additives : si $A, B \in \mathcal{M}_n(\mathbb{C})$ sont hermitiennes de valeurs propres $\lambda_1(A) \geqslant \cdots \geqslant \lambda_n(A)$ et $\lambda_1(B) \geqslant \cdots \geqslant \lambda_n(B)$ et si $r := \mathrm{rang}(B-A)$ alors pour tout $1 \leqslant k \leqslant n$,

$$\lambda_{k+r}(A) \leqslant \lambda_k(B) \leqslant \lambda_{k-r}(A)$$

avec la convention $\lambda_k(A) = +\infty$ si $k < 1$ et $\lambda_k(A) = -\infty$ si $k > n$. Il en découle que si

$$F_A(t) := \frac{\mathrm{card}\{1 \leqslant k \leqslant n : \lambda_k(A) \leqslant t\}}{n} \quad \text{et} \quad F_B(t) := \frac{\mathrm{card}\{1 \leqslant k \leqslant n : \lambda_k(A) \leqslant t\}}{n}$$

sont les fonctions de répartition des mesures de comptage

$$\mu_A := \frac{1}{n}\sum_{k=1}^n \delta_{\lambda_k(A)} \quad \text{et} \quad \mu_B := \frac{1}{n}\sum_{k=1}^n \delta_{\lambda_k(B)}$$

Recueil de modèles stochastiques – Page 163/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

alors

$$\|F_A - F_B\|_\infty := \sup_{t\in\mathbb{R}} |F_A(t) - F_B(t)| \leqslant \frac{\operatorname{rang}(A-B)}{n}.$$

Le membre de gauche est la distance de Kolmogorov-Smirnov entre les lois de probabilités discrètes μ_A et μ_B. Voici un exemple d'application : si H est une matrice aléatoire hermitienne telle que $(H_{i,j})_{1\leqslant i,j\leqslant n}$ sont de même moyenne m alors on a $\operatorname{rang}(\mathbb{E}(H)) = \operatorname{rang}(m1_n) \leqslant 1$ et donc $\mu_{H-\mathbb{E}(H)}$ et μ_H ont les mêmes limites en loi !

Lemme 16.9 (Inégalité de Hoffman-Wielandt)**.** *Si $A, B \in \mathcal{M}_n(\mathbb{C})$ sont hermitiennes de spectres respectifs $\lambda_1(A) \geqslant \cdots \geqslant \lambda_n(A)$ et $\lambda_1(B) \geqslant \cdots \geqslant \lambda_n(B)$ alors*

$$\sum_{k=1}^n (\lambda_k(A) - \lambda_k(B))^2 \leqslant \sum_{i=1}^n \sum_{j=1}^n |A_{ij} - B_{ij}|^2.$$

Notons $\mathcal{P}_2$ l'ensemble des lois sur $\mathbb{R}$ possédant un moment d'ordre 2 fini. La distance de couplage (ou de Wasserstein) d sur $\mathcal{P}_2$ est définie pour tous $\nu_1, \nu_2 \in \mathcal{P}_2$ par

$$d(\nu_1,\nu_2)^2 := \inf_{\substack{(X_1,X_2)\\ X_1\sim\nu_1, X_2\sim\nu_2}} \mathbb{E}(|X_1 - X_2|^2)$$

où l'infimum porte sur l'ensemble des couples de variables aléatoires de lois marginales ν_1 et ν_2. La convergence pour d entraîne la convergence en loi (ainsi que la convergence des deux premiers moments). Dans le cas de loi discrètes de la forme $\nu_1 = \frac{1}{n}\sum_{k=1}^n \delta_{a_k}$ et $\nu_2 = \frac{1}{n}\sum_{k=1}^n \delta_{b_k}$ avec $a_1 \leqslant \cdots \leqslant a_n$ et $b_1 \leqslant \cdots \leqslant b_n$, on trouve

$$d(\nu_1,\nu_2)^2 = \frac{1}{n}\sum_{k=1}^n (a_i - b_i)^2.$$

Par ailleurs $\mathcal{M}_n(\mathbb{C})$ est un espace de Hilbert pour le produit scalaire $\langle A, B\rangle := \operatorname{Tr}(AB^*)$. La norme associée est appelée norme de Schur, de Frobenius, ou encore de Hilbert-Schmidt : $\|A\|_{\mathrm{HS}}^2 = \operatorname{Tr}(AA^*) = \sum_{i=1}^n\sum_{j=1}^n |A_{ij}|^2 = \sum_{i=1}^n \lambda_i(A)^2$. L'inégalité de Hoffman-Wielandt s'écrit

$$d(\mu_A,\mu_B)^2 \leqslant \frac{\|A-B\|_{\mathrm{HS}}^2}{n}.$$

Voici un exemple d'application : si H est une matrice hermitienne aléatoire à valeurs dans $\mathcal{M}_n(\mathbb{C})$ telle que $(H_{ii})_{1\leqslant i\leqslant j\leqslant n}$ sont i.i.d. de carré intégrable, alors pour tout réel $C > 0$, en désignant par H_C la matrice tronquée définie par $(H_C)_{ij} := H_{ij}\mathbb{1}_{\{|H_{ij}|\leqslant C\}}$, et en utilisant l'inégalité ci-dessus et la loi forte des grands nombres, on a

$$d(\mu_H,\mu_{H_C})^2 \leqslant \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n |H_{ij}|^2\mathbb{1}_{\{|H_{ij}|>C\}} \xrightarrow[n\to\infty]{\text{p.s.}} \mathbb{E}(|H_{12}|^2\mathbb{1}_{\{|H_{12}|>C\}}).$$

Par convergence dominée ou monotone, le membre de droite peut être rendu arbitrairement proche de 0 en choisissant C assez grand. Cela permet de se ramener, dans la preuve du théorème de Wigner, au cas où les coefficients de H sont bornés. La même méthode permet d'établir que la diagonale de H ne joue pas de rôle dans le théorème de Wigner. On peut en particulier la supposer nulle, ou de variance arbitraire.

Lemme 16.10 (Inégalité de concentration de Azuma-Hoeffding)**.** *Si $G(X_1,\dots,X_n)$ est une variable aléatoire intégrable fonction de vecteurs aléatoires $X_1,\dots,X_n$ indépendants (pas forcément de même dimension), alors pour tout $r \geqslant 0$,*

$$\mathbb{P}(|G(X_1,\dots,X_n) - \mathbb{E}(G(X_1,\dots,X_n))| \geqslant r) \leqslant 2\exp\left(-\frac{r^2}{2(c_1^2+\cdots+c_n^2)}\right)$$

où $c_k := \sup_{(x,x')\in\mathcal{D}_k} |G(x) - G(x')|$ et $\mathcal{D}_k := \{(x,x') : x_i = x'_i \text{ si } i \neq k\}$.

Recueil de modèles stochastiques – Page 164/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Ce résultat, de même nature que l'inégalité familière de Bienaymé-Tchebychev, est démontré dans le chapitre 17.1 (lemme 17.3). Donnons un exemple d'application aux matrices aléatoires. Soit $f : \mathbb{R} \to \mathbb{R}$ de classe $\mathcal{C}^1$ et à support compact, et soient $A, B \in \mathcal{M}_n(\mathbb{C})$ hermitiennes. Le lemme 16.8 donne $\|F_A - F_B\|_\infty \leqslant r/n$ où $r := \operatorname{rang}(A - B)$, et par intégration par parties,

$$\left|\int f\,d\mu_A - \int f\,d\mu_B\right| = \left|\int_{-\infty}^{+\infty} f'(t)(F_A(t) - F_B(t))\,dt\right| \leqslant \frac{r}{n}\int_{-\infty}^{+\infty} |f'(t)|\,dt := \frac{r}{n}\|f'\|_1.$$

On note $A(x_1, \ldots, x_n) \in \mathcal{M}_n(\mathbb{C})$ la matrice hermitienne dont les lignes du triangle supérieur sont $x_1, \ldots, x_n$ (ce sont des vecteurs de $\mathbb{C}^n, \mathbb{C}^{n-1}, \ldots, \mathbb{C}$), et on pose

$$G(x_1, \ldots, x_n) = \int f\,d\mu_{A(x_1,\ldots,x_n)}.$$

On observe que

$$\operatorname{rang}\big\{A(x_1, \ldots, x_{k-1}, x_k, x_{k+1}, \ldots, x_n) - A(x_1, \ldots, x_{k-1}, x'_k, x_{k+1}, \ldots, x_n)\big\} \leqslant 2$$

et donc

$$|G(x_1, \ldots, x_{k-1}, x_k, x_{k+1}, \ldots, x_n) - G(x_1, \ldots, x_{k-1}, x'_k, x_{k+1}, \ldots, x_n)| \leqslant \frac{2}{n}\|f'\|_1.$$

À présent, si H est une matrice hermitienne aléatoire à valeurs dans $\mathcal{M}_n(\mathbb{C})$ telle que les lignes de son triangle supérieur $(H_{ij})_{1\leqslant i\leqslant j\leqslant n}$ sont indépendantes (ce sont des vecteurs aléatoires de $\mathbb{C}^n, \mathbb{C}^{n-1}, \ldots, \mathbb{C}$), alors, grâce à l'inégalité de Azuma-Hoeffding, pour toute fonction f de classe $\mathcal{C}^1$ et à support compact, et pour tout réel $r \geqslant 0$,

$$\mathbb{P}\left(\left|\int f\,d\mu_H - \mathbb{E}\int f\,d\mu_H\right| \geqslant r\right) \leqslant 2\exp\left(-\frac{nr^2}{8\|f'\|_1^2}\right).$$

Cette inégalité exprime le fait que la distribution spectrale empirique est très concentrée autour de son espérance. C'est pour cette raison que les fluctuations sont si faibles dans les simulations, dès que la dimension dépasse quelques dizaines. La variable aléatoire $\int f\,d\mu_H - \mathbb{E}\int f\,d\mu_H$ converge en probabilité vers 0. Pour tirer parti de la rapidité de convergence, on pose par exemple $r = n^{-1/2+\varepsilon}$ avec $\varepsilon > 0$ petit fixé, de sorte que la borne exponentielle dans le membre de droite de l'inégalité soit sommable en n. Le *lemme de Borel-Cantelli*[5] entraîne alors que la convergence de la variable aléatoire vers 0 a lieu presque sûrement. Cela permet de ramener l'étude de la convergence en loi presque sûre de μ_H à celle de la convergence en loi de $\mathbb{E}\mu_H$.

16.7 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 John Wishart (années 1920) sur les matrices de covariance empiriques des échantillons gaussiens, à John von Neumann et Herman Goldstine (années 1940) sur l'analyse numérique matricielle, et à Eugene Wigner, Freeman Dyson, et Madan Lal Mehta (années 1950-1960) sur les niveaux d'énergie des noyaux atomiques en physique mathématique. Le théorème de Marchenko-Pastur a été obtenu par Vladimir Marchenko et Leonid 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 limite central*. 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

5. Si $\mathbb{E}(\sum_n \mathbb{1}_{A_n}) = \sum_n \mathbb{P}(A_n) < \infty$ alors $\mathbb{P}(\sum_n \mathbb{1}_{A_n} < \infty) = 1$. Note : $\{\sum_n \mathbb{1}_{A_n} = \infty\} = \cap_n \cup_{m\geqslant n} A_m$.

Recueil de modèles stochastiques – Page 165/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

la brillante *théorie des probabilités libres* développée à l'origine par Dan-Virgil 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 des hypothèses plus restrictives. 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 $n \to \infty$. 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. $\lim_{n\to\infty} \lambda_n = -2\sigma$ et $\lim_{n\to\infty} \lambda_1 = 2\sigma$.

La partie combinatoire de la preuve du théorème de Wigner est détaillée par exemple dans le livre de Zhidong Bai et Jack Silverstein [BS10], et dans le livre de Greg W. Anderson, Alice Guionnet et Ofer Zeitouni [AGZ10]. Le théorème de Wigner peut également être obtenu par une méthode analytique, en utilisant la transformée de Cauchy-Stieltjes, faisant apparaître la loi du demi-cercle comme solution d'une équation de point fixe. Le théorème de Wigner est également présenté dans les notes de cours de Mireille Capitaine [Cap12] et celles de Charles Bordenave [Bor14b]. Le contenu de notre section 16.6 est tiré de [Cha13]. Un panorama sur la décomposition en valeurs singulières des matrices se trouve dans [CGLP12]. Notons enfin qu'au delà du théorème 16.4, il existe une condition suffisante pour que $\mu \in \mathcal{P}$ soit caractérisée par ses moments $(\kappa_n)_{n\geqslant 1}$, connue sous le nom de condition de Carleman, et liée à la quasi-analycité[6] de la transformée de Fourier : $\sum_n \kappa_{2n}^{-1/(2n)} = \infty$.

6. Une fonction $f : \mathbb{R} \to \mathbb{R}$ 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.

Recueil de modèles stochastiques – Page 166/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Chapitre 17

Problème du voyageur de commerce

Recueil de modèles stochastiques – Page 167/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Mots-clés. Optimisation combinatoire randomisée ; concentration de la mesure ; sous-additivité ; poissonisation.

Voici trois problèmes phares de l'*optimisation combinatoire randomisée* :
— 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 œuvre certaines techniques. Le *problème du voyageur de commerce* consiste à trouver une tournée, c'est-à-dire un chemin circulaire [1], de longueur minimale passant par des points prescrits $X_1, \ldots, X_n \in \mathbb{R}^d$, $n, d \geqslant 2$. Cela revient à résoudre sur le groupe symétrique

$$\min_{\sigma \in \Sigma_n} \sum_{k=1}^{n} \left| X_{\sigma(k)} - X_{\sigma(k+1)} \right|$$

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. L'explosion combinatoire du groupe symétrique fait que dès que n dépasse quelques dizaines, il n'est plus possible de tester toutes les permutations pour trouver la meilleure. Il est cependant possible d'utiliser un algorithme d'optimisation globale comme par exemple le *recuit simulé*, abordé dans la section 8.5, pour produire en un temps raisonnable une solution approchée : une permutation pour laquelle le minimum est (presque) atteint.

Plutôt que de rechercher une permutation optimale, nous nous intéressons dans ce chapitre à la valeur du minimum, et à son comportement lorsque n est grand et les points $X_1, \ldots, X_n$ sont des variables aléatoires indépendantes et de même loi μ sur $\mathbb{R}^d$. On note $L_n = L_n(X_1, \ldots, X_n)$ la longueur minimale de la tournée, qui est une fonction de $X_1, \ldots, X_n$.

Théorème 17.1 (Bearwood-Halton-Hammersley)**.** *Il existe une constante* $0 < \gamma_d < \infty$ *qui dépend de* $d \geqslant 2$ *telle que si* μ *est à support compact et de densité* f *alors*

$$\frac{L_n(X_1, \ldots, X_n)}{n^{(d-1)/d}} \xrightarrow[n \to \infty]{p.s.} \gamma_d \int_{\mathbb{R}^d} f(x)^{(d-1)/d} \, dx.$$

En particulier $L_n(X_1, \ldots, X_n)$ 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 montrer que la variable aléatoire $L_n(X_1, \ldots, X_n)$ est d'autant plus concentrée autour de son espérance $\mathbb{E}(L_n(X_1, \ldots, X_n))$ que n est grand, puis que cette espérance est d'ordre $n^{(d-1)/d}$.

1. On parle parfois de *circuit hamiltonien*.

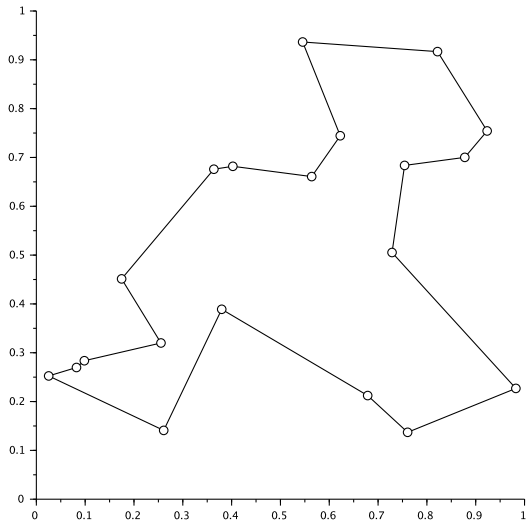


Figure 17.1 – Trajet le plus court pour $n = 20$ points uniformément répartis dans le carré unité obtenu par l'algorithme stochastique du recuit simulé (chapitre 8).

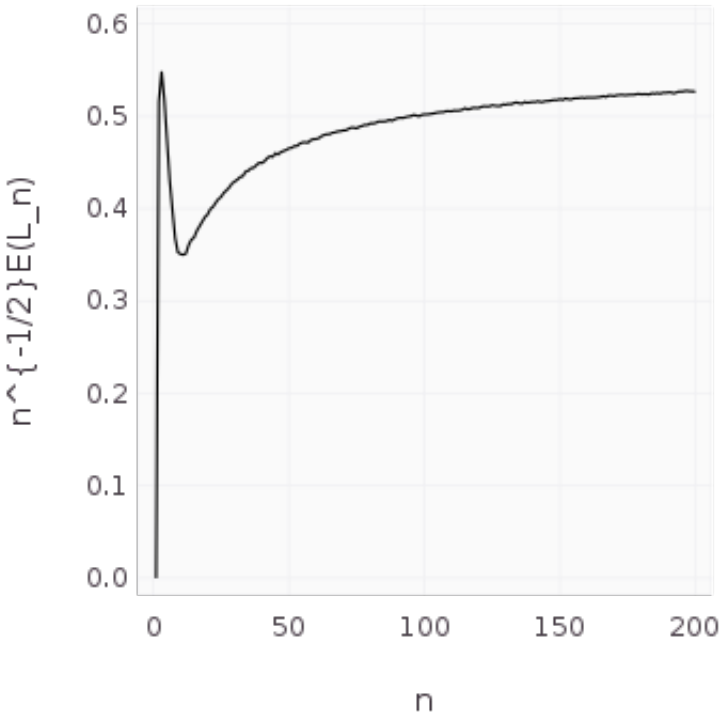


Figure 17.2 – Tracé d'une approximation de la courbe $n \mapsto n^{-1/2}\mathbb{E}(L_n)$ dans le cas uniforme, avec une méthode de Monte-Carlo et l'algorithme stochastique du recuit simulé (chapitre 8).

Recueil de modèles stochastiques – Page 168/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

17.1 Concentration pour le cas uniforme

Théorème 17.2 (Concentration autour de la moyenne)**.** *Si μ est la loi uniforme sur le cube $[0,1]^d$ alors pour tout $n \geqslant 1$ et tout réel $t \geqslant 0$ on a*

$$\mathbb{P}(|L_n - \mathbb{E}(L_n)| \geqslant t) \leqslant 2\exp\begin{cases} -\dfrac{ct^2}{\log(n)} & \text{si } d=2, \\ -\dfrac{ct^2}{n^{(d-2)/d}} & \text{si } d>2, \end{cases}$$

où $c > 0$ est une constante qui ne dépend que de la dimension d. Ainsi,

$$\mathbb{P}\left(\left|\frac{L_n}{n^{(d-1)/d}} - \mathbb{E}\left(\frac{L_n}{n^{(d-1)/d}}\right)\right| \geqslant t\right) \leqslant 2\exp\begin{cases} -\dfrac{cnt^2}{\log(n)} & \text{si } d=2, \\ -cnt^2 & \text{si } d>2, \end{cases}$$

La preuve du théorème 17.2 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 concentration 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 $\operatorname{osc}(f) = \sup f - \inf f = \text{diamètre}(\text{support}(f))$, et on a alors $\operatorname{osc}(f) \leqslant 2\|f\|_\infty$.

Lemme 17.3 (Inégalité de concentration de Azuma-Hoeffding)**.** *Si $Y : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ est une variable aléatoire intégrable, alors pour tout $r \geqslant 0$,*

$$\mathbb{P}(|Y - \mathbb{E}(Y)| \geqslant r) \leqslant 2\exp\left(-\frac{2r^2}{\operatorname{osc}(d_1)^2 + \cdots + \operatorname{osc}(d_n)^2}\right)$$

pour la décomposition télescopique de Doob en somme de différences de martingales

$$Y - \mathbb{E}(Y) = \sum_{k=1}^{n} \mathbb{E}(Y \mid \mathcal{F}_k) - \mathbb{E}(Y \mid \mathcal{F}_{k-1}) = \sum_{k=1}^{n} d_k$$

associée à une filtration arbitraire d'interpolation $\{\varnothing, \Omega\} = \mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_n = \mathcal{F}$.

Démonstration. Soit U une v.a.r. telle que $\mathbb{E}(U) = 0$ et $a \leqslant U \leqslant b$. La convexité de $u \mapsto e^u$ donne, pour tout $t \geqslant 0$ et tout $x \in [a,b]$,

$$e^{tx} \leqslant \frac{x-a}{b-a}e^{tb} + \frac{b-x}{b-a}e^{ta}.$$

En posant $p = -a/(b-a)$ et $f(u) = -pu + \log(1-p+pe^u)$ il vient donc

$$\mathbb{E}(e^{tU}) \leqslant \frac{b}{b-a}e^{ta} - \frac{a}{b-a}e^{tb} = e^{ta}\Big((1-p) + pe^{t(b-a)}\Big) = e^{f(t(b-a))}.$$

À présent on a

$$f'(u) = -p + \frac{pe^u}{1-p+pe^u} \quad \text{et} \quad f''(u) = p(1-p)\frac{e^u}{(1-p+e^u)^2} \leqslant \frac{1}{4}.$$

Recueil de modèles stochastiques – Page 169/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Comme $f(0) = f'(0) = 0$, on en déduit que $f(u) \leqslant u^2/8$ et donc

$$\mathbb{E}(e^{tU}) \leqslant e^{\frac{t^2}{8}(b-a)^2}.$$

Appliquée à $U = d_k = \mathbb{E}(Y \mid \mathcal{F}_k) - \mathbb{E}(Y \mid \mathcal{F}_{k-1})$ sachant $\mathcal{F}_{k-1}$, cela donne

$$\mathbb{E}(e^{td_k} \mid \mathcal{F}_{k-1}) \leqslant e^{\frac{t^2}{8}\operatorname{osc}(d_k)^2}.$$

Ensuite, en écrivant la somme télescopique $Y - \mathbb{E}(Y) = d_n + \cdots + d_1$ on obtient

$$\mathbb{E}(e^{t(Y-\mathbb{E}(Y))}) = \mathbb{E}(e^{t(d_{n-1}+\cdots+d_1)}\mathbb{E}(e^{td_n} \mid \mathcal{F}_{n-1})) \leqslant \cdots \leqslant e^{\frac{t^2}{8}(\operatorname{osc}(d_1)^2+\cdots+\operatorname{osc}(d_n)^2)}.$$

À présent, pour tout $t, r > 0$, avec $c := \operatorname{osc}(d_1)^2 + \cdots + \operatorname{osc}(d_n)^2$, par l'inégalité de Markov,

$$\begin{aligned}\mathbb{P}(Y - \mathbb{E}(Y) \geqslant r) &= \mathbb{P}(e^{t(Y-\mathbb{E}(Y))} \geqslant e^{tr}) \\ &\leqslant e^{-tr}\mathbb{E}(e^{t(Y-\mathbb{E}(Y))}) \\ &\leqslant e^{-tr+\frac{ct^2}{8}}.\end{aligned}$$

Ainsi, pour tout $r \geqslant 0$,

$$\mathbb{P}(Y - \mathbb{E}(Y) \geqslant r) \leqslant e^{\inf_{t>0}\left(-tr+\frac{ct^2}{8}\right)} = e^{-\frac{2r^2}{c}}.$$

En utilisant cela pour Y et $-Y$, on obtient le résultat souhaité pour $\mathbb{P}(|Y - \mathbb{E}(Y)| \geqslant r)$. □

Lemme 17.4 (Lemme géométrique)**.** *Il existe une constante $c_d > 0$ telle que si $X_1, \ldots, X_k$ sont i.i.d. de loi uniforme sur $[0,1]^d$ alors pour tout $x \in [0,1]^d$,*

$$g_k(x) := \mathbb{E}\left(\min_{1\leqslant i\leqslant k} |X_i - x|\right) \leqslant c_d k^{-1/d}.$$

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 $r \leqslant 1$, la valeur du minimum est $2^{-d}|B(0,r)| = 2^{-d}|B(0,1)|r^d$ (dessin). Si $1 < r \leqslant \sqrt{d}$, la valeur du minimum est difficile à calculer. Elle reste néanmoins supérieure ou égale à celle du cas $r = 1$. Ainsi, pour tout $0 < r \leqslant \sqrt{d}$, ce volume minimal est $\geqslant 2^{-d}|B(0,1)|(r/\sqrt{d})^d = a_d r^d$. Donc pour tout $x \in [0,1]^d$ et tout $0 < r \leqslant \sqrt{d}$, en utilisant à la fin $(1-u)^k \leqslant e^{-ku}$,

$$\begin{aligned}\mathbb{P}\left(\min_{1\leqslant i\leqslant k} |X_i - x| \geqslant r\right) &= \prod_{i=1}^{k} \mathbb{P}(X_i \in B(x,r)^c) \\ &= \Big(1 - |B(x,r) \cap [0,1]^d|\Big)^k \\ &\leqslant \Big(1 - a_d r^d\Big)^k \\ &\leqslant \exp\Big(-a_d k r^d\Big).\end{aligned}$$

Ceci reste valable si $r > \sqrt{d}$ car dans ce cas $\mathbb{P}(\min_{1\leqslant i\leqslant k} |X_i - x| \geqslant r) = 0$. À présent,

$$\mathbb{E}\left(\min_{1\leqslant i\leqslant k} |X_i - x|\right) = \int_0^\infty \mathbb{P}\left(\min_{1\leqslant i\leqslant k} |X_i - x| \geqslant r\right) dr \quad \text{et} \quad \int_0^\infty e^{-br^d}\, dr = \frac{\Gamma(1/d)}{db^{1/d}}.$$

□

Recueil de modèles stochastiques – Page 170/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Preuve du théorème 17.2. Appliquons l'inégalité d'Azuma-Hoeffding du lemme 17.3 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 $\mathcal{F}_0 = \{\varnothing, \Omega\}$. 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 $x \in \mathbb{R}^d$,

$$L(S) \leqslant L(S \cup \{x\}) \leqslant L(S) + 2 \min_{y \in S} |x - y|.$$

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

$$\begin{aligned} \big|L_n(x_1, \dots, x_k, \dots, x_n) - L_n(x_1, \dots, x'_k, \dots, x_n)\big| &\leqslant 2 \min_{i \neq k} \big|x'_k - x_i\big| + 2 \min_{i \neq k} |x_k - x_i| \\ &\leqslant 2 \min_{i > k} \big|x'_k - x_i\big| + 2 \min_{i > k} |x_k - x_i|. \end{aligned}$$

En posant $g_k(x) := \mathbb{E}\left(\min_{1 \leqslant i \leqslant k} |X_i - x|\right)$ on obtient pour $1 \leqslant k \leqslant n-1$,

$$|d_k| \leqslant 2\mathbb{E}\left(\min_{i > k} |X_k - X_i| + \min_{i > k} \big|X'_k - X_i\big| \,\middle|\, \mathcal{F}_k\right) = 2g_{n-k}(X_k) + 2\mathbb{E}(g_{n-k}(X'_k)).$$

Le lemme géométrique 17.4 donne, pour $d \geqslant 2$ et $1 \leqslant k \leqslant n-1$,

$$\|d_k\|_\infty \leqslant c_d (n-k)^{-1/d}.$$

Comme $\|d_n\|_\infty \leqslant c_d$ pour une autre constante car les X_i sont bornées, on obtient

$$\sum_{k=1}^{n} \|d_k\|_\infty^2 \leqslant \begin{cases} c_d \log(n) & \text{pour } d = 2, \\ c_d n^{(d-2)/d} & \text{pour } d > 2. \end{cases}$$

□

17.2 Évaluation de la moyenne du cas uniforme

Lemme 17.5 (Un bon début)**.** *Si μ est la loi uniforme sur $[0,1]^d$ alors pour tout $n \geqslant 1$,*

$$c_d^- n^{(d-1)/d} \leqslant \mathbb{E}(L_n(X_1, \dots, X_n)) \leqslant c_d^+ n^{(d-1)/d}$$

où $0 < c_d^- \leqslant c_d^+ < \infty$ 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 tiroirs*[2] entraîne alors que pour tous $x_1, \dots, x_n \in [0,1]^d$ on a

$$\min_{1 \leqslant i \neq j \leqslant n} |x_i - x_j| \leqslant c_d n^{-1/d}$$

2. *Pigeonhole principle* en anglais : si on dispose n objets dans m boites avec $n > m$ alors au moins l'une des boites contient deux objets ou plus.

Recueil de modèles stochastiques – Page 171/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

où c_d est une constante qui peut dépendre de la dimension d. Par conséquent, si ℓ_n est le maximum sur $x_1, \ldots, x_n$ de la longueur de tournée minimale pour $x_1, \ldots, x_n$, alors

$$\ell_n \leqslant \ell_{n-1} + 2c_d n^{-1/d}$$

qui donne, pour une nouvelle constante c_d qui peut dépendre de la dimension d,

$$\|L_n(X_1, \ldots, X_n)\|_\infty \leqslant c_d n^{-1/d+1} = c_d n^{(d-1)/d}. \tag{17.1}$$

Par ailleurs, en reprenant la preuve du lemme 17.4, 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 $0 < r \leqslant 1/2$,

$$\mathbb{P}\left(\min_{1 \leqslant i \leqslant n-1} |X_i - x| \geqslant r\right) \geqslant (1 - \omega_d r^d)^{n-1}$$

avec $\omega_d := |B(0,1)|$. En utilisant l'inégalité $(1-u)^\alpha \geqslant 1 - \alpha u$ pour $0 \leqslant u \leqslant 1/\alpha$, on en déduit que

$$\mathbb{E}\left(\min_{1 \leqslant i \leqslant n-1} |X_i - x|\right) \geqslant \int_0^{1/2} (1 - \omega_d r^d)^{n-1}\, dr \geqslant c_d n^{-1/d}.$$

Ainsi,

$$\min_{1 \leqslant j \leqslant n} \mathbb{E}\left(\min_{1 \leqslant i \neq j \leqslant n} |X_i - X_j|\right) = \min_{1 \leqslant j \leqslant n} \mathbb{E}\left[\mathbb{E}\left(\min_{1 \leqslant i \neq j \leqslant n} |X_i - X_j| \,|\, X_j\right)\right] \geqslant c_d n^{-1/d}$$

où c_d est une constante qui dépend de la dimension d. Or le circuit optimal à travers $X_1, \ldots, X_n$ possède n arêtes et passe par chacun des n sommets. On dispose donc d'une écriture $L_n(X_1, \ldots, X_n) = \sum_{i=1}^n |X_i - X_{j_i}|$ pour des entiers aléatoires $j_1, \ldots, j_n$, ce qui donne

$$\mathbb{E}(L_n(X_1, \ldots, X_n)) \geqslant \sum_{j=1}^n \mathbb{E}\left(\min_{1 \leqslant i \neq j \leqslant n} |X_i - X_j|\right) \geqslant n c_d n^{-1/d} = c_d n^{(d-1)/d}.$$

□

Théorème 17.6 (Le bon résultat)**.** *Si μ est la loi uniforme sur $[0,1]^d$ alors*

$$\lim_{n \to \infty} \frac{\mathbb{E}(L_n(X_1, \ldots, X_n))}{n^{(d-1)/d}} = \gamma_d$$

où $0 < \gamma_d < \infty$ est un réel qui dépend 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 $N \sim \mathrm{Poi}(t^{d/(d-1)})$. Par ailleurs si $N \sim \mathrm{Poi}(n)$ 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 $S = \{x_1, \ldots, x_n\} \subset \mathbb{R}^d$, avec la convention $L(S) = 0$ si $\mathrm{card}(S) \leqslant 2$. Soit P un processus ponctuel de Poisson sur $\mathbb{R}^d$ de mesure d'intensité Lebesgue. Soit $(Z_t)_{t \geqslant 0}$ 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 $n \geqslant 0$,

$$\mathrm{Loi}(P \,|\, \mathrm{card}(P \cap [0,t]^d) = n) = \mathrm{Unif}([0,t]^d)^{\otimes n}.$$

Recueil de modèles stochastiques – Page 172/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

D'autre part, $L(tS) = tL(S)$ et $\mathrm{card}(P \cap [0,t]^d) \sim \mathrm{Poi}(t^d)$, ce qui donne

$$\mathbb{E}(Z_t) = \mathbb{E}(\mathbb{E}(Z_t \,|\, \mathrm{card}(P \cap [0,t]^d))) = e^{-t^d} \sum_{n=0}^{\infty} t a_n \frac{t^{dn}}{n!} \quad \text{où} \quad a_n := \mathbb{E}(L_n(X_1, \ldots, X_n)).$$

Le facteur t devant a_n vient du fait que a_n concerne $[0,1]^d$ et non pas $[0,t]^d$. Le lemme 17.5 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 $(a_n)_{n \geqslant 1}$ 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 $\lim_{t\to\infty} t^{-1}\mathbb{E}(Z_t)$ existe, ce qui suffira.

Sous-additivité. Soit $C_1, \ldots, C_{k^d}$ une partition (aux bords près) de $[0,t]^d$ en k^d cubes similaires à $[0, t/k]^d$. Pour tous $x_1, \ldots, x_n \in \mathbb{R}^d$, les k^d tournées des ensembles $S_i = \{x_1, \ldots, x_n\} \cap C_i$ peuvent être concaténées pour former une tournée pour $\{x_1, \ldots, x_n\} \cap [0,t]^d$, 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 (17.1) ce qui donne une grande tournée en chapelet. Ainsi, il existe une constante $c > 0$ telle que pour tout $k \geqslant 1$ et tout $t \geqslant 0$,

$$L(\{x_1, \ldots, x_n\} \cap [0,t]^d) \leqslant \sum_{i=1}^{k^d} L(\{x_1, \ldots, x_n\} \cap C_i) + ctk^{d-1}.$$

Cela donne $\mathbb{E}(Z_t) \leqslant k^d \mathbb{E}(Z_{t/k}) + ctk^{d-1}$, d'où l'inégalité

$$\frac{\mathbb{E}(Z_{tk})}{(tk)^d} \leqslant \frac{\mathbb{E}(Z_t)}{t^d} + ct^{1-d} \tag{17.2}$$

En prenant $t = 1$ on obtient

$$0 \leqslant \gamma := \varliminf_{k\to\infty} \frac{\mathbb{E}(Z_k)}{k^d} \leqslant \mathbb{E}(Z_1) + c < \infty.$$

Par définition de γ, pour tout $\varepsilon > 0$ on peut choisir k_0 assez grand pour que

$$\frac{\mathbb{E}(Z_{k_0})}{k_0^d} + ck_0^{1-d} \leqslant \gamma + \varepsilon.$$

Comme $t \mapsto \mathbb{E}(Z_t)$ est continue, on peut choisir $\delta > 0$ tel que pour tout $k_0 < t < k_0 + \delta$,

$$\frac{\mathbb{E}(Z_t)}{t^d} + ct^{1-d} \leqslant \gamma + 2\varepsilon. \tag{17.3}$$

Grâce à (17.2) on voit que (17.3) a lieu pour tout $kk_0 < t < k(k_0 + \delta)$. Or pour $k > k_0/\delta$ les intervalles $I_k :=]kk_0, k(k_0 + \delta)[$ et I_{k+1} se recouvrent. On en déduit que (17.3) a lieu pour tout $t \geqslant \lfloor k_0^2/\delta \rfloor$, ce qui implique en particulier

$$\varlimsup_{t\to\infty} \frac{\mathbb{E}(Z_t)}{t^d} \leqslant \varliminf_{t\to\infty} \frac{\mathbb{E}(Z_t)}{t^d} + 2\varepsilon = \gamma + 2\varepsilon.$$

Comme $\varepsilon > 0$ est arbitraire, on obtient donc

$$\lim_{t\to\infty} \frac{\mathbb{E}(Z_t)}{t^d} = \gamma.$$

Ainsi, le développement en série de $\mathbb{E}(Z_t)$ donne, après le changement de variable $u = t^d$,

$$\sum_{k=0}^{\infty} e^{-u} a_k \frac{u^k}{k!} \sim_{u\to\infty} \gamma u^{(d-1)/d}. \tag{17.4}$$

Recueil de modèles stochastiques – Page 173/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Dépoissonisation. On commence par montrer que $n \mapsto a_n$ est assez régulière. Comme il est possible de concaténer un circuit pour $X_1, \ldots, X_m$ et un circuit pour $X_{m+1}, \ldots, X_{m+n}$ avec un coût inférieur à $2\sqrt{d}$ (diamètre du cube) on en déduit que

$$a_{n+m} \leqslant a_n + a_m + 2\sqrt{d}$$

ce qui donne $0 \leqslant a_n - a_k \leqslant a_{n-k} + 2\sqrt{d}$ pour tout $0 \leqslant k \leqslant n$. Or le lemme 17.5 donne $a_n = \mathcal{O}(n^{(d-1)/d})$ et on en déduit

$$|a_n - a_k| \leqslant |a_{n-k} + 2\sqrt{d}| \leqslant c|n-k|^{(d-1)/d}.$$

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

$$\begin{aligned}
|a_n - \mathbb{E}(a_N)| &\leqslant \sum_{k=0}^{\infty} |a_n - a_k| e^{-n} \frac{n^k}{k!} \\
&\leqslant c \sum_{k=0}^{\infty} |n-k|^{(d-1)/d} e^{-n} \frac{n^k}{k!} \\
&= c\mathbb{E}(|N - \mathbb{E}(N)|^{(d-1)/d}) \\
&\leqslant c\mathbb{E}(|N - \mathbb{E}(N)|^2)^{(d-1)/(2d)} \\
&= c\mathbb{V}(N)^{(d-1)/(2d)} \\
&= \mathcal{O}(n^{(d-1)/(2d)}) = o(n^{(d-1)/d}).
\end{aligned}$$

Or (17.4) avec $t = n$ donne $\mathbb{E}(a_N) = \sum_{k=0}^{\infty} a_k e^{-n} n^k / k! \sim_{n\to\infty} \gamma n^{(d-1)/d}$ et donc

$$\lim_{n\to\infty} \frac{\mathbb{E}(L_n(X_1, \ldots, X_n))}{n^{(d-1)/d}} = \lim_{n\to\infty} \frac{a_n}{n^{(d-1)/d}} = \gamma.$$

Le fait que $\gamma > 0$ découle de la borne inférieure du lemme 17.5. □

17.3 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 17.2 et le *lemme de Borel-Cantelli*, presque sûrement,

$$L_n - \mathbb{E}L_n = \begin{cases} \mathcal{O}_{n\to\infty}(\log(n)) & \text{si } d = 2, \\ \mathcal{O}_{n\to\infty}(n^{(d-2)/(2d)} \sqrt{\log(n)}) & \text{si } d > 2 \end{cases}$$

tandis que par le théorème 17.6 on a $\mathbb{E}(L_n(X_1, \ldots, X_n)) \sim_{n\to\infty} \gamma n^{(d-1)/d}$ pour un $0 < \gamma < \infty$. Ceci achève la preuve du théorème 17.1 dans le cas uniforme.

17.4 Notes et commentaires

Lorsque $d = 2$ et $X_1, \ldots, X_n$ sont i.i.d. uniforme sur $[0,1]^2$, on peut se convaincre intuitivement que probablement $L_n(X_1, \ldots, X_n)$ est de l'ordre de $\sqrt{n}$ lorsque n est grand en parcourant le carré par zigzag parallèles régulièrement espacés, par exemple horizontaux : il y a approximativement $\sqrt{n}$ lignes de longueur unité contenant chacune approximativement $\sqrt{n}$ points. Plus rigoureusement, cela conduit à aborder le problème du voyageur de commerce en dimension $d = 2$ en utilisant une courbe qui «remplit l'espace».

Le contenu de ce chapitre est directement inspiré du joli livre de Michael Steele [Ste97]. On pourra également consulter sur ce thème le livre de Joseph Yukich [Yuk98] ainsi que le livre de

Recueil de modèles stochastiques – Page 174/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Marc Mézard et Andrea Montanari [MM09]. L'inégalité de concentration de Azuma-Hoeffding est étudiée par Colin McDiarmid dans [McD89, McD98]. Le théorème 17.1 a été obtenu par John Hammersley et ses élèves Jillian Beardwood et John Halton vers 1959 et peut être démontré par réduction au cas uniforme sur le cube $[0,1]^d$. On sait par ailleurs que $\gamma_2 \approx 0,7$ et que $\gamma_d \sim_{d\to\infty} \sqrt{d2\pi e}$.

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 inexploitables. 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, comme expliqué dans le chapitre 8. Contrairement à TSP, les deux autres problèmes phares de l'optimisation combinatoire (MST et EMM) sont résolubles en temps polynomial. Pour en savoir plus, on renvoie par exemple aux livres [AHU75, ACG$^+$99] ainsi qu'au cours [Bor14a].

Recueil de modèles stochastiques – Page 175/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Recueil de modèles stochastiques – Page 176/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Chapitre 18

File d'attente M/M/Infini

Recueil de modèles stochastiques – Page 177/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Mots-clés. Loi exponentielle ; loi de Poisson ; file d'attente ; chaîne de Markov à temps continu ; mesure invariante ; fonction génératrice ; convergence à l'équilibre.

On considère un fragment de kryptonite, constitué d'atomes radioactifs susceptibles de se décomposer en atomes instables qui se décomposent eux-mêmes pour donner une forme stable et inerte. La première décomposition est très lente et, durant l'expérience, peu d'atomes de kryptonite se seront décomposés. Nous notons 0, 1 et 2 les états successifs dans lesquels les atomes sont susceptibles de passer. Un atome donné se décompose indépendamment des autres (pas de réactions en chaîne) et sans que son âge n'ait d'influence sur son taux de décomposition (pas de vieillissement). Le séjour d'un atome dans la forme instable est beaucoup plus court mais obéit au même principe d'absence de vieillissement et d'indépendance par rapport aux autres atomes. Les instants d'apparition d'un atome sous forme instable 1 sont modélisés par les temps de saut d'un *processus de Poisson* homogène d'intensité λ. Un atome instable se décomposera à nouveau (pour rejoindre la forme inerte 2) au bout d'un temps exponentiel de paramètre μ. Le processus de Poisson et les temps de séjour dans l'état instable de tous les atomes sont supposés indépendants. On s'intéresse à X_t le nombre d'atomes au temps t présents sous la forme instable.

18.1 Lois exponentielles

Le résultat suivant justifie le fait que l'on modélise le temps de vie d'un atome dans un état donné par une loi exponentielle.

Lemme 18.1 (Absence de mémoire)**.** *Si S et T sont deux variables aléatoires exponentielles indépendantes de paramètres respectifs λ et μ, alors*

$$\mathcal{L}(T - S \,|\, T > S) = \mathrm{Exp}(\mu) = \mathcal{L}(T)$$

c'est-à-dire que $\mathbb{P}(T > S + t \,|\, T > S) = e^{-\mu t}$ pour tout $t \geqslant 0$.

Démonstration. $\mathbb{P}(T > S) = \frac{\lambda}{\lambda+\mu}$ et $\mathbb{P}(T > S + t, T > S) = \mathbb{P}(T > S + t) = \frac{\lambda}{\lambda+\mu} e^{-\mu t}$. □

Lemme 18.2 (Horloges exponentielles en compétition)**.** *Si $(T_k)_{k\in\mathcal{K}}$ une collection finie ou dénombrable de variables aléatoires indépendantes de lois exponentielles de paramètres respectifs $(\lambda_k)_{k\in\mathcal{K}}$ tels que $\lambda = \sum_{k\in\mathcal{K}} \lambda_k < \infty$, alors $T = \inf_{k\in\mathcal{K}} T_k$ suit la loi exponentielle $\mathrm{Exp}(\lambda)$, et de plus, presque sûrement l'infimum est atteint en un unique entier aléatoire K, indépendant de T, dont la loi est donnée par $\mathbb{P}(K = k) = \lambda_k/\lambda$ pour tout $k \in \mathcal{K}$.*

Démonstration. On détermine la loi du couple (K,T) en écrivant, pour tous $k \in \mathcal{K}$ et $t \geqslant 0$,

$$\begin{aligned}\mathbb{P}(K=k, T \geqslant t) &= \mathbb{P}(T_k \geqslant t, T_j > T_k, \forall j \neq k) = \int_t^\infty \lambda_k e^{-\lambda_k s} \mathbb{P}(T_j > s, \forall j \neq k)\, ds \\ &= \int_t^\infty \lambda_k e^{-\lambda_k s} \prod_{j \neq k} e^{-\lambda_j s}\, ds = \int_t^\infty \lambda_k e^{-\lambda s}\, ds = \frac{\lambda_k}{\lambda} e^{-\lambda t}.\end{aligned}$$

Les lois de K et de T s'obtiennent en prenant $t=0$ et en sommant sur k respectivement. □

18.2 Temps de demi-vie d'un fragment de métal radioactif

Considérons un fragment de métal contenant un grand nombre N d'atomes radioactifs dans l'état 0 et supposons que chaque atome passe dans l'état 1 après un temps exponentiel de paramètre α (petit). Notons $(T_n^{(N)})_{1 \leqslant n \leqslant N}$ les instants où un atome passe de l'état 0 à l'état 1, avec $T_0^{(N)} = 0$ et, pour $1 \leqslant n \leqslant N$, $\tau_n^{(N)} = T_n^{(N)} - T_{n-1}^{(N)}$. Les variables aléatoires $(\tau_n^{(N)})_{1 \leqslant n \leqslant N}$ sont indépendantes de lois exponentielles de paramètres respectifs $((N-n+1)\alpha)_{1 \leqslant n \leqslant N}$. On s'intéresse au temps de demi-vie $T_{N/2}^{(N)}$ du fragment, c'est-à-dire au temps au bout du quel la moitié moitié des atomes se sont décomposés.

Théorème 18.3 (Convergence du temps de demi-vie)**.**

$$T_{N/2}^{(N)} \underset{N \to \infty}{\overset{p.s.}{\longrightarrow}} \frac{\log 2}{\alpha}.$$

Démonstration. On omet la notation (N) et on abrège par exemple T_n par $T_n^{(N)}$. On a

$$\begin{aligned}\mathbb{E}\big(T_{N/2}\big) &= \sum_{n=1}^{N/2} \mathbb{E}(\tau_n) = \frac{1}{\alpha} \sum_{n=1}^{N/2} \frac{1}{N-n+1} \\ &= \frac{1}{\alpha}(H(N) - H(N/2)) = \frac{\log 2}{\alpha} + o(1),\end{aligned}$$

où $H(n)$ est la série harmonique. De plus,

$$T_{N/2} - \mathbb{E}(T_{N/2}) = \sum_{n=1}^{N/2} (\tau_n - \mathbb{E}(\tau_n)),$$

où les variables $(\tau_n - \mathbb{E}(\tau_n))_n$ sont indépendantes et centrées et $\tau_n - \mathbb{E}(\tau_n)$ a la même loi que $(E-1)/\alpha(N+1-n)$ où E suit la loi exponentielle de paramètre 1. On a donc

$$\begin{aligned}\mathbb{E}\Big[\big(T_{N/2} - \mathbb{E}(T_{N/2})\big)^4\Big] &= \sum_{n=1}^{N/2} \mathbb{E}\big[(\tau_n - \mathbb{E}(\tau_n))^4\big] + 6 \sum_{1 \leqslant m \neq n \leqslant N/2} \mathbb{E}\big[(\tau_n - \mathbb{E}(\tau_n))^2\big] \mathbb{E}\big[(\tau_m - \mathbb{E}(\tau_m))^2\big] \\ &= \frac{\mathbb{E}\big[(E-1)^4\big]}{\alpha^4} \sum_{n=N/2+1}^{N} \frac{1}{n^4} + 6 \frac{\big(\mathbb{E}\big[(E-1)^2\big]\big)^2}{\alpha^4} \sum_{1+N/2 \leqslant m \neq n \leqslant N} \frac{1}{n^2} \frac{1}{m^2} \\ &= \frac{9}{\alpha^4} \sum_{n=N/2+1}^{N} \frac{1}{n^4} + \frac{6}{\alpha^4} \left(\sum_{n=N/2+1}^{N} \frac{1}{n^2} \right)^2,\end{aligned}$$

puisque $\mathbb{E}((E-1)^2) = 1$ et $\mathbb{E}((E-1)^4) = 9$. Ainsi

$$\mathbb{E}\Big[\big(T_{N/2} - \mathbb{E}(T_{N/2})\big)^4\Big] = \mathcal{O}\big(N^{-2}\big).$$

Recueil de modèles stochastiques – Page 178/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

L'inégalité de Markov assure alors que, pour tout $r \geqslant 0$,

$$\mathbb{P}\big(\big|T_{N/2} - \mathbb{E}(T_{N/2})\big| \geqslant r\big) \leqslant \frac{C}{r^4 N^2}.$$

On en déduit que la suite $(T_{N/2})_{N\geqslant 2}$ converge presque sûrement vers $\log(2)/\alpha$. □

Voici quelques exemples concrets de temps de demi-vie moyens, en années :

Carbone 14	Plutonium 239	Iode 129	Uranium 235	Uranium 238	Thorium 232
5734	25×10^3	17×10^6	710×10^6	4.5×10^9	14×10^9

18.3 Loi du nombre d'atomes

Cette section est dédiée au calcul de la loi du nombre X_t d'atomes dans l'état 1 au temps t. Pour tous $t \in \mathbb{R}_+$ et $n \in \mathbb{N}$, on note $p_n(t) = \mathbb{P}(X_t = n)$. Le résultat classique suivant affirme que ces fonctions sont solutions d'un système (infini) d'équations différentielles linéaires.

Théorème 18.4 (Équations de Chapman-Kolmogorov)**.** *Les fonctions $(p_n)_{n\in\mathbb{N}}$ sont de classe $\mathcal{C}^1$ et satisfont le système d'équations différentielles linéaires suivant :*

$$\forall n \in \mathbb{N}, \quad p_n'(t) = -(\lambda + n\mu)p_n(t) + \lambda p_{n-1}(t) + (n+1)\mu p_{n+1}(t),$$

avec la convention $p_{-1}(t) = 0$ pour tout $t \geqslant 0$, la loi initiale $(p_n(0))_{n\in\mathbb{N}}$ étant donnée.

Démonstration. Pour calculer $p_n(t+h)$, on remarque que si $X_{t+h} = n$ alors l'une des conditions incompatibles suivantes est réalisée :

1. $X_u = n$ pour tout $u \in [t, t+h]$;
2. $X_t = n-1$ et une seule transition a lieu $(n-1 \to n)$ dans l'intervalle de temps $[t, t+h]$;
3. $X_t = n+1$ et une seule transition a lieu $(n+1 \to n)$ dans l'intervalle de temps $[t, t+h]$;
4. dans l'intervalle de temps $[t, t+h]$, au moins deux transitions ont lieu.

D'après les lemmes 18.1 et 18.2, le processus reste dans l'état n un temps aléatoire de loi exponentielle de paramètre $\lambda+n\mu$ puis saute de n à $n-1$ ou $n+1$ avec probabilités respectives $n\mu/(\lambda+n\mu)$ et $\lambda/(\lambda+n\mu)$. En conditionnant par l'évènement $\{X_t = n\}$, on en déduit l'égalité

$$p_n(t+h) = p_n(t)\{1 - \lambda h - n\mu h\} + \lambda h p_{n-1}(t) + (n+1)\mu h p_{n+1}(t) + o(h).$$

Il reste à faire tendre h vers 0. □

Plutôt que de chercher à résoudre le système d'équation du théorème 18.4, on peut déterminer la fonction génératrice de X_t en montrant qu'elle est solution d'une équation aux dérivées partielles. Pour tout $t \in \mathbb{R}_+$ et tout $s \in [-1, 1]$, on pose

$$G(s,t) := \mathbb{E}\big(s^{X_t}\big) = \sum_{n=0}^{\infty} p_n(t) s^n.$$

Théorème 18.5 (Expression de la fonction génératrice de X_t)**.** *Si G_0 est la fonction génératrice de X_0 alors, pour tout $t \geqslant 0$ et tout $s \in [-1, 1]$,*

$$G(s,t) = \exp\left(\frac{\lambda}{\mu}(1 - e^{-\mu t})(s-1)\right) G_0(1 - e^{-\mu t} + s e^{-\mu t}).$$

Recueil de modèles stochastiques – Page 179/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Démonstration. On a, en vertu des équations du théorème 18.4,

$$\begin{aligned}\partial_t G(s,t) &= \sum_{n=0}^{\infty} p_n'(t)s^n \\ &= \sum_{n=0}^{\infty} (-(\lambda+n\mu)p_n(t) + \lambda p_{n-1}(t) + (n+1)\mu p_{n+1}(t))s^n \\ &= -\lambda(1-s)\sum_{n=0}^{\infty} p_n(t)s^n + \mu(1-s)\sum_{n=0}^{\infty} np_n(t)s^{n-1}.\end{aligned}$$

La fonction G est donc solution de l'équation aux dérivées partielles suivante :

$$\partial_t G(s,t) = (1-s)[\mu\partial_s G(s,t) - \lambda G(s,t)].$$

On pose $H = \log(G)$ puis on effectue le changement de variables $(s,\tau) = (s,(1-s)e^{-\mu t})$. En notant $\tilde{H}$ la fonction telle que $H(s,t) = \tilde{H}(s,\tau)$, la fonction $\tilde{H}$ s'écrit sous la forme

$$\tilde{H}(s,\tau) = \frac{\lambda}{\mu}s + h(\tau),$$

où h reste à déterminer. Ceci revient à dire que H est de la forme

$$H(s,t) = \frac{\lambda}{\mu}s + h((1-s)e^{-\mu t}).$$

On conclut en remarquant que la fonction h est caractérisée par le fait que $H(s,0)$ est égal à $\log G_0(s)$ pour tout $s \in [-1,1]$. □

La loi de X_t n'est en général pas une loi classique, sauf dans certains cas remarquables.

Corollaire 18.6 (Loi de X_t)**.** *Pour tout $k \in \mathbb{N}$ et tout $t \geqslant 0$,*

$$\mathcal{L}(X_t \,|\, X_0 = k) = \mathrm{Poi}\left(\frac{\lambda}{\mu}(1-e^{-\mu t})\right) * \mathrm{Bin}(k, e^{-\mu t}).$$

avec la convention $\mathrm{Bin}(0,p) = \delta_0$*. De plus, si* $X_0 \sim \mathrm{Poi}(\alpha)$ *alors*

$$X_t \sim \mathrm{Poi}\left(\frac{\lambda}{\mu}(1-e^{-\mu t}) + \alpha e^{-\mu t}\right).$$

Voici une preuve du corollaire 18.6 plus probabiliste que le théorème 18.5.

Démonstration alternative du corollaire 18.6. Supposons tout d'abord que $X_0 = 0$. Notons $(N_t)_{t\geqslant 0}$ le processus de Poisson de paramètre λ régissant l'apparition des atomes instables. Pour $t > 0$, la variable aléatoire N_t suit la loi de Poisson de paramètre λt. De plus, sachant que $N_t = n$, la loi des n instants de saut est celle d'un n-échantillon réordonné de variables aléatoires indépendantes de même loi uniforme sur $[0,t]$. D'autre part, un atome apparu à un instant aléatoire de loi uniforme sur $[0,t]$ est encore présent à l'instant t avec probabilité

$$q(t) := \frac{1}{t}\int_0^t (1-F_\mu(s))\,ds = \frac{1}{\mu t}(1-e^{-\mu t}),$$

où F_μ est la fonction de répartition de la loi $\mathrm{Exp}(\mu)$. Si $X_0 = 0$, on obtient donc, par indépendance des atomes, que pour tout $k \in \mathbb{N}$,

$$\begin{aligned}\mathbb{P}(X_t = k) &= \sum_{n=k}^{\infty} \mathbb{P}(X_t = k | N_t = n)\mathbb{P}(N_t = n) \\ &= \sum_{n=k}^{\infty} \binom{n}{k} q(t)^k (1-q(t))^{n-k} e^{-\lambda t}\frac{(\lambda t)^n}{n!} \\ &= e^{-\lambda t q(t)}\frac{(\lambda t q(t))^k}{k!}.\end{aligned}$$

Recueil de modèles stochastiques – Page 180/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Donc la loi de X_t sachant que $X_0 = 0$ est la loi de Poisson de paramètre $\lambda(1 - e^{-\mu t})/\mu$.

Enfin, le nombre d'atomes présents au temps t est la somme de ceux présents à l'instant initial et non encore désintégrés au temps t et de ceux qui sont arrivés ensuite et sont encore présents. Par indépendance des désintégrations, la loi de X_t sachant que $X_0 = k$ est la convolution de la loi du nombre d'atomes non encore désintégrés à l'instant t parmi les k atomes initiaux et de la loi de X_t sachant que $X_0 = 0$, d'où le résultat. □

Remarque 18.7 (Loi de désintégration quelconque)**.** *Si les temps de désintégration des atomes instables ne sont plus de loi exponentielle mais sont indépendants et de même loi de fonction de répartition F, on montre de même que la loi de X_t sachant que $X_0 = 0$ est une loi de Poisson de paramètre $\lambda \int_0^t (1 - F(s))\, ds$.*

Le processus $(X_t)_{t \geqslant 0}$ est un processus de Markov, c'est-à-dire une chaîne de Markov à temps continu. On lui associe de manière naturelle une famille d'opérateurs de la façon suivante. Définissons le semi-groupe $(P_t)_{t \geqslant 0}$ associé à $(X_t)_{t \geqslant 0}$.

Pour tout $t \geqslant 0$ et toute fonction $f : \mathbb{N} \to \mathbb{R}$, on définit la fonction $P_t f$ par

$$P_t f(n) = \mathbb{E}(f(X_t) | X_0 = n).$$

Si ν est une mesure de probabilité sur $\mathbb{N}$ on notera νP_t la loi de X_t sachant que X_0 suit la loi ν. On a, pour toute fonction f bornée,

$$\nu P_t f = \sum_{n \in \mathbb{N}} \nu(n) P_t f(n).$$

D'après le corollaire 18.6, la mesure de probabilité $P_t(\cdot)(n)$ est la loi

$$\mathrm{Bin}(n, e^{-\mu t}) * \mathrm{Poi}((\lambda/\mu)(1 - e^{-\mu t})).$$

Lemme 18.8 (Propriétés essentielles du semi-groupe)**.** *Le semi-groupe $(P_t)_{t \geqslant 0}$ satisfait les propriétés suivantes :*

1. *Élément neutre : $P_0 = I$ c'est-à-dire que $P_0 f = f$ pour toute fonction f ;*
2. *Positivité et normalisation : si $f \geqslant 0$ alors $P_t f \geqslant 0$, et $P_t 1 = 1$;*
3. *Propriété de semi-groupe : pour toute fonction f et tous $s, t \geqslant 0$,*
$$P_{t+s} f = (P_t \circ P_s) f;$$
4. *Générateur infinitésimal : pour toute fonction f,*
$$\partial_t P_t f = L P_t f = P_t L f;$$
où L est appelé générateur infinitésimal *de X et est défini par*
$$L f(n) = \lambda(f(n+1) - f(n)) + n\mu(f(n-1) - f(n)).$$

Formellement, tout se passe comme si $P_t = e^{tL}$.

Démonstration. Les deux premiers points sont évidents. Le troisième découle de la formule explicite de P_t. C'est une reformulation de la propriété de Markov. La dernière assertion s'obtient grâce à 3. et un raisonnement analogue à celui de la preuve de le théorème 18.4. □

Remarque 18.9 (Lire la dynamique du processus dans le générateur)**.** *Si on définit le noyau de transition $Q(n, \cdot) = \frac{\lambda}{\lambda + n\mu} \delta_{n+1} + \frac{n\mu}{\lambda + n\mu} \delta_{n-1}$ et la «temporisation» $D(n) = \lambda + n\mu$ alors*

$$L f(n) = D(n) \int (f(m) - f(n))\, Q(n, dm)$$

On lit dans le générateur infinitésimal un algorithme qui permet de simuler les trajectoires du processus : le processus X saute de n vers $n - 1$ au taux μn et vers $n + 1$ au taux λ.

On peut associer à tout processus de Markov raisonnable son semi-groupe et son générateur infinitésimal. Si ce dernier est souvent explicite, il est rare que ce soit le cas pour les mesures $P_t(\cdot)(x)$. En ce sens, le processus étudié dans ce chapitre est assez remarquable.

Recueil de modèles stochastiques – Page 181/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

18.4 Comportement en temps long

Notons $\rho = \lambda/\mu$ et π la loi de Poisson Poi(ρ). Si X_0 suit la loi π, le corollaire 18.6 assure que c'est encore le cas pour X_t pour tout $t \geqslant 0$. La mesure π est donc invariante sous l'action de P_t. De manière équivalente, $\pi L = 0$. Comme pour tous les processus de naissance et mort, la probabilité invariante π, si elle existe, est de plus réversible pour L (ou P_t), c'est-à-dire que L et P_t sont auto-adjoints dans $L^2(\pi)$. Enfin, pour toute loi initiale de X_0, la loi de X_t converge vers π lorsque $t \to \infty$. On souhaite ici quantifier cette convergence.

Soit $\mathcal{P}_1(\mathbb{N})$ l'ensemble des mesures de probabilité sur $\mathbb{N}$ possédant un moment d'ordre 1. La *distance de Wasserstein* sur $\mathcal{P}_1(\mathbb{N})$ est définie pour tous $\nu, \tilde{\nu} \in \mathcal{P}_1(\mathbb{N})$ par

$$W_1(\nu, \tilde{\nu}) = \inf\Big\{\mathbb{E}|X - \tilde{X}|,\ X \sim \nu,\ \tilde{X} \sim \tilde{\nu}\Big\}.$$

Cette distance fait penser à la *distance en variation totale* introduite dans le chapitre 1, et utilisée par exemple dans le théorème 1.8 page 5. Pour des variables aléatoires entières, $X \neq \tilde{X}$ implique $|X - \tilde{X}| \geqslant 1$. On a donc

$$d_{\mathrm{VT}}(\nu, \tilde{\nu}) \leqslant W_1(\nu, \tilde{\nu}).$$

Théorème 18.10 (Contraction en distance de Wasserstein)**.** *Si $\nu, \tilde{\nu} \in \mathcal{P}_1(\mathbb{N})$ alors*

$$W_1(\nu P_t, \tilde{\nu} P_t) \leqslant e^{-\mu t} W_1(\nu, \tilde{\nu}),$$

où, pour tout $\eta \in \mathcal{P}(\mathbb{N})$, ηP_t est la loi de X_t si X_0 suit la loi η.

Démonstration. Supposons dans un premier temps que $\nu = \delta_m$ et $\tilde{\nu} = \delta_n$ avec $m < n$. Soit X_t de loi $\delta_m P_t$ et B_t de loi $\mathrm{Bin}(n - m, e^{-\mu t})$ indépendante de X_t. D'après le corollaire 18.6, la variable aléatoire $\tilde{X}_t = X_t + B_t$ suit la loi $\delta_n P_t$. On a donc

$$\mathbb{E}|X_t - \tilde{X}_t| = \mathbb{E}(B_t) = (n - m)e^{-\mu t}.$$

Comme $(X_t, \tilde{X}_t)$ est un couplage de $(\delta_m P_t, \delta_n P_t)$, on a

$$W_1(\delta_m P_t, \delta_n P_t) \leqslant e^{-\mu t}(n - m) = e^{-\mu t} W_1(\delta_m, \delta_n).$$

Concluons à présent dans le cas général. Soit ν et $\tilde{\nu}$ dans $\mathcal{P}_1(\mathbb{N})$. Soit $(X_0, \tilde{X}_0)$ un couplage de ν et $\tilde{\nu}$. En procédant comme ci-dessous, on peut construire (X_t) et $(\tilde{X}_t)$ tels que

$$\mathcal{L}(|X_t - \tilde{X}_t| | X_0, \tilde{X}_0) = \mathrm{Bin}(|X_0 - \tilde{X}_0|, e^{-\mu t}).$$

On en déduit donc

$$W_1(\nu P_t, \tilde{\nu} P_t) \leqslant e^{-\mu t}\mathbb{E}|X_0 - \tilde{X}_0|.$$

Reste à prendre l'infimum sur tous les couplages pour obtenir la propriété de contraction. □

Ce résultat de contraction fournit une estimation pour la convergence de la loi de X_t vers la mesure invariante $\pi = \mathcal{P}(\rho)$ de X.

Corollaire 18.11 (Convergence à l'équilibre)**.** *Si $\nu \in \mathcal{P}_1(\mathbb{N})$ alors*

$$W_1(\nu P_t, \pi) \leqslant e^{-\mu t} W_1(\nu, \pi).$$

Remarque 18.12 (Optimalité du taux de convergence)**.** *Le taux de convergence est optimal puisque, si $\nu = \delta_n$ avec n supérieur à la moyenne ρ de la mesure invariante,*

$$(n - \rho)e^{-\mu t} = \mathbb{E}_n(X_t) - \rho \leqslant W_1(\delta_n P_t, \pi) \leqslant \mathbb{E}|n - X|e^{-\mu t},$$

où $X \sim \pi$. Ce taux de convergence ne dépend pas du taux de naissance λ.

Recueil de modèles stochastiques – Page 182/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Dans la même veine, on peut montrer que la fonction $P_t f$ converge vers la fonction constante égale à l'intégrale de f pour π en montrant que son gradient (discret) converge vers 0.

Théorème 18.13 (Commutation)**.** *Soit f de $\mathbb{N}$ dans $\mathbb{R}$. Alors, pour tout $t \geqslant 0$,*

$$DP_t f(n) = e^{-\mu t} P_t(Df)(n),$$

où Dg est la fonction définie sur $\mathbb{N}$ par $Dg(n) = g(n+1) - g(n)$.

Démonstration. D'après le corollaire 18.6,

$$P_t f(n+1) = \mathbb{E} f(Z_1 + Z_2 + \cdots + Z_{n+1} + Y),$$

où $Z_1, X_2, \ldots, Z_{n+1}$ sont des variables aléatoires indépendantes de loi $\mathrm{Ber}(e^{-\mu t})$ et indépendantes de Y de loi de Poisson de paramètre $\rho(1 - e^{-\mu t})$. On a donc

$$\begin{aligned} P_t f(n+1) &= \mathbb{E}(f(Z_1 + Z_2 + \cdots + Z_{n+1} + Y)) \\ &= e^{-\mu t}\mathbb{E}(f(Z_1 + \cdots + Z_n + 1 + Y)) + (1 - e^{-\mu t})\mathbb{E}(f(Z_1 + \cdots + Z_n + Y)) \\ &= e^{-\mu t} P_t(Df)(n) + P_t f(n), \end{aligned}$$

ce qui fournit le résultat. □

Si f est une fonction 1-lipschitzienne, au sens où $|Df(n)| \leqslant 1$ pour tout $n \in \mathbb{N}$ alors le théorème précédent assure que $P_t f$ est une fonction $e^{-\mu t}$-lipschitzienne. On retrouve ainsi la convergence pour la distance de Wasserstein par la formulation duale de cette distance[1] :

$$W_1(\nu, \tilde{\nu}) = \sup\left\{\int f\, d\nu - \int f\, d\tilde{\nu} \ : \ f \text{ 1-lipschitzienne}\right\}.$$

Remarque 18.14 (Files d'attentes, vie et mort)**.** *Le processus X modélise également le nombre de clients en attente de service lorsque les arrivées des clients sont séparés par des temps exponentiels de paramètre λ, lorsque les temps de service sont exponentiels de paramètre μ, toutes ces variables étant indépendantes, et lorsque chaque client a un serveur dédié immédiatement à son arrivée, ce qui revient à dire qu'il y a une infinité de serveurs. On parle alors de* file d'attente $\mathrm{M/M/\infty}$*, où le premier M indique l'absence de mémoire des durées entre les arrivées, et le second M l'absence de mémoire des durées de service. On peut également limiter le nombre de serveurs à un entier $s \geqslant 1$ ou introduire une salle d'attente qui limite la capacité maximale de la file à un entier K, ce qui donne lieu à un processus appelé M/M/s/K. Le processus M/M/K/K modélise un parking à K places. Le processus M/M/1 modélise le nombre de clients en attente de service devant un unique serveur. La figure 18.1 compare les trajectoires des files M/M/1 et M/M/∞. Ces files d'attentes font partie des* processus de vie et de mort[2]*, qui sont des processus de Markov à temps continu et d'espace d'états $\mathbb{N}$. Leur générateur infinitésimal a la forme suivante :*

$$Lf(n) = \lambda_n(f(n+1) - f(n)) + \mu_n \mathbb{1}_{n>0}(f(n-1) - f(n)).$$

Le cas où λ_n et μ_n sont linéaires en n correspond au processus de Yule.

1. Due à Kantorovitch et Rubinstein.
2. «*Birth and death processes*» en anglais.

Recueil de modèles stochastiques – Page 183/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

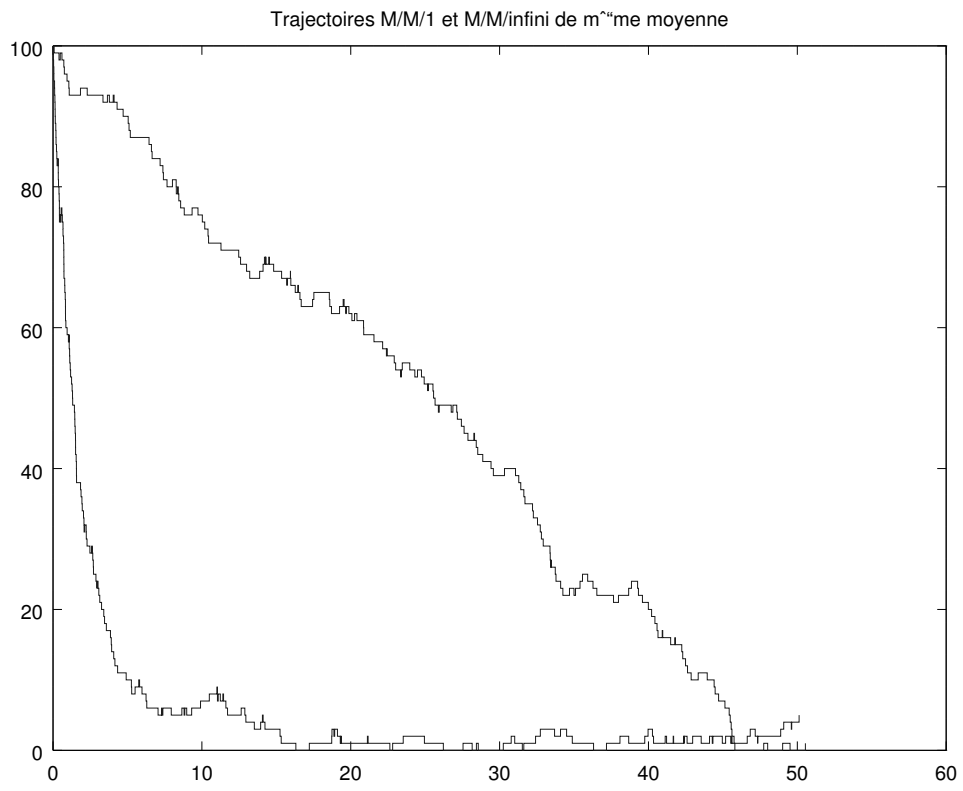


Figure 18.1 – Un début de trajectoire de file d'attente M/M/1 de paramètres 1 et 3, et de file d'attente M/M/∞ de paramètres 1 et .5. Les deux trajectoires sont issues de 100. Le premier processus a pour loi invariante la loi géométrique sur $\mathbb{N}$ de paramètre 1/3, et le second la loi de Poisson de paramètre $1/.5 = 2$. Dans les deux cas, la loi invariante a pour moyenne 2. La première trajectoire à un comportement linéaire et la seconde un comportement exponentiel.

18.5 Notes et commentaires

Les résultats classiques concernant les lois exponentielles sont disponibles par exemple dans le livre de James Norris [Nor98]. Dans le livre de Geoffrey Grimmett et David Stirzaker [GS01], une version à temps discret du processus X est étudiée. Pour l'anecdote, elle peut modéliser par exemple le nombre de coquilles présentes dans un manuscrit après n relectures : à chaque relecture, l'auteur corrige chaque coquille avec probabilité p (indépendamment des autres) et en ajoute par mégarde un nombre aléatoire de loi de Poisson (*loi des événements rares* ou *loi des petits nombres*).

Dans le livre de Philippe Robert [Rob00], un résultat emblématique de résultats plus généraux, est établi. Sous une renormalisation convenable, on peut montrer que ce processus à espace d'états discret ressemble au *processus d'Ornstein-Uhlenbeck* étudié dans le chapitre 24.

On peut établir de nombreux autres résultats pour le processus X. On trouvera dans [Cha06] des inégalités fonctionnelles permettant de fournir d'autres propriétés de comportement en temps long. La relation de commutation du théorème 18.13 est une propriété remarquable du semi-groupe. Le processus d'Ornstein-Uhlenbeck satisfait la même relation en remplaçant le gradient discret D par un gradient usuel sur $\mathbb{R}^d$, et cette liaison est une instance d'un phénomène plus général qui fait l'objet du chapitre 26. On peut obtenir des versions un peu plus faibles pour des processus plus généraux qui fournissent des estimations pour la convergence vers la mesure invariante, voir par exemple l'article de Pietro Caputo, Paolo Dai

Recueil de modèles stochastiques – Page 184/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Pra, et Gustavo Posta [CDPP09].

Recueil de modèles stochastiques – Page 185/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Recueil de modèles stochastiques – Page 186/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Chapitre **19**

Ruine d'une compagnie d'assurance

Mots-clés. Processus de Poisson ; processus de Poisson composé ; transformée de Laplace ; martingale.

Une nouvelle compagnie d'assurance veut entrer sur le marché. Elle souhaite évaluer sa probabilité de faillite en fonction du capital initial investi et du prix de ses polices d'assurance. On suppose que la compagnie a un capital initial c et on note R_t sa réserve à l'instant t. La compagnie d'assurance

— perçoit des cotisations de ses clients que l'on supposera mensualisées, uniformément réparties sur l'année et constantes au cours du temps : les recettes de la compagnie pendant un temps t sont donc égales à pt où p est le taux de cotisations par unité de temps ;

— verse des primes à ses assurés sinistrés en fonction des dommages qu'ils subissent.

On modélise l'apparition et les coûts des sinistres de la manière suivante :

— les coûts des sinistres $(X_k)_{k\geqslant 1}$ sont des variables aléatoires indépendantes, de même loi ν, d'espérance commune λ et de variance finie ;

— les temps d'apparition $(T_n)_{n\geqslant 1}$ des sinistres sont distribués selon un processus de Poisson $(N_t)_{t\geqslant 0}$ d'intensité μ. Les intervalles de temps entre deux sinistres, notés $(\Delta_n)_{n\geqslant 1}$ avec $\Delta_n = T_n - T_{n-1}$, sont i.i.d. de loi $\mathrm{Exp}(\mu)$.

La réserve de la compagnie d'assurance à l'instant t est par conséquent :

$$R_t = c + pt - \sum_{k=1}^{N_t} X_k.$$

Pour simplifier légèrement les calculs, on considérera plutôt la quantité

$$\tilde{R}_t = c - R_t = \sum_{k=1}^{N_t} X_k - pt.$$

La ruine survient lorsque la réserve R descend sous 0 ou que $\tilde{R}$ excède c. L'instant de ruine, ou de défaut, de la compagnie est défini par $\tau(c) = \inf\{t \geqslant 0 : R_t < 0\} = \inf\{t \geqslant 0 : \tilde{R}_t > c\}$. La probabilité de ruine de la compagnie est donnée par

$$\psi(c) = \mathbb{P}\left(\sup_{t\geqslant 0} \tilde{R}_t > c | \tilde{R}_0 = 0\right) = \mathbb{P}(\tau(c) < \infty).$$

On notera enfin $Y(c)$ le découvert en cas de ruine : $Y(c) = -R_{\tau(c)} = \tilde{R}_{\tau(c)} - c$.

Recueil de modèles stochastiques – Page 187/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

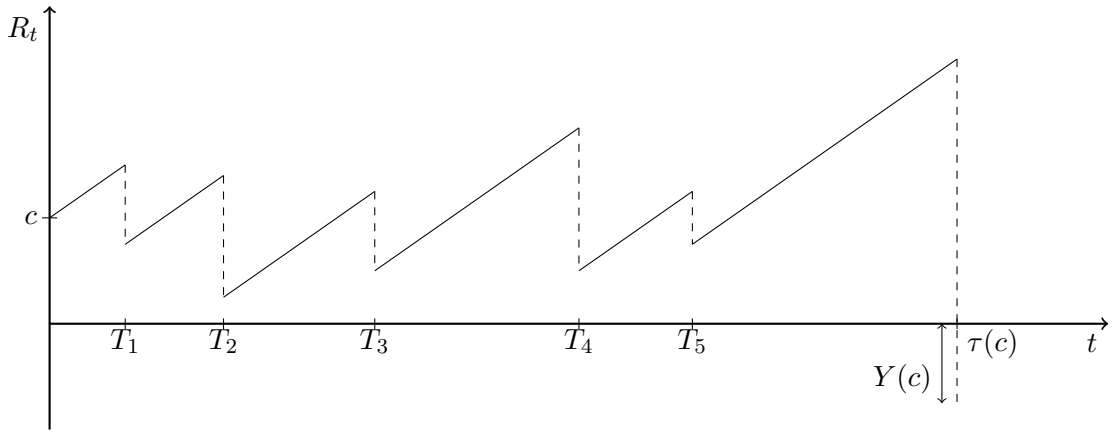


Figure 19.1 – Une trajectoire menant à la ruine au temps $\tau(c)$ avec un découvert $Y(c)$.

Remarque 19.1 (Processus de Markov et générateur infinitésimal)**.** *On peut montrer que $\tilde{R}$ est un processus de Markov sur $\mathbb{R}$ de générateur infinitésimal*

$$Lf(x) = -pf'(x) + \mu \int_0^\infty (f(x+y) - f(x))\, \nu(dy).$$

Ce processus évolue de manière déterministe entre ses temps de saut. On parle de processus de Markov déterministe par morceaux.

19.1 Processus de Poisson composé

Rappelons qu'il est facile d'estimer l'intensité d'un processus de Poisson à partir d'une de ses trajectoires comme le montre le résultat suivant.

Lemme 19.2 (Temps long)**.** *Soit N un processus de Poisson d'intensité μ alors*

$$\frac{N_t}{t} \xrightarrow[t\to\infty]{p.s.} \mu \quad \text{et} \quad \sqrt{t}\left(\frac{N_t}{t} - \mu\right) \xrightarrow[t\to\infty]{\text{loi}} \mathcal{N}(0,\mu),$$

Démonstration. Pour tout $i \in \mathbb{N}^*$, posons $Y_i = N_i - N_{i-1}$. La variable Y_i représente le nombre de sauts du processus de Poisson dans l'intervalle de temps $[i-1, i[$. Le processus de Poisson est à accroissements stationnaires et indépendants. En particulier, les variables aléatoires $(Y_i)_{i\geqslant 1}$ sont indépendantes de même loi de Poisson de paramètre μ. Donc, en vertu de la loi des grands nombres, $N_{\lfloor t\rfloor}/\lfloor t\rfloor \xrightarrow[t\to\infty]{\text{p.s.}} \mu$. On remarque ensuite l'encadrement suivant

$$\frac{N_{\lfloor t\rfloor}}{\lfloor t\rfloor}\frac{\lfloor t\rfloor}{t} \leqslant \frac{N_t}{t} \leqslant \frac{N_{\lfloor t\rfloor+1}}{\lfloor t\rfloor+1}\frac{\lfloor t\rfloor+1}{t}$$

qui permet de conclure que N_t/t tend vers μ presque sûrement. La convergence en loi se déduit du calcul explicite de la fonction caractéristique de $\sqrt{t}(N_t/t - \mu)$ puisque N_t suit la loi $\mathrm{Poi}(\mu t)$. □

Soit $(N_t)_{t\geqslant 0}$ un *processus de Poisson* d'intensité μ et $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires i.i.d. de loi ν indépendante de N. Le processus C défini par

$$\forall t \geqslant 0, \quad C_t = \sum_{k=1}^{N_t} X_k,$$

est appelé *processus de Poisson composé* de paramètre μ et de loi de saut ν.

Recueil de modèles stochastiques – Page 188/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Lemme 19.3 (Fonction caractéristique)**.** *La fonction caractéristique de C_t est donnée par*

$$\varphi_{C_t}(u) = \exp(\mu t(\varphi(u) - 1)),$$

où φ est la fonction caractéristique de ν.

Si ν admet un moment d'ordre 2 et que l'on note λ et a ses deux premiers moments alors C_t admet un moment d'ordre 2 et

$$\mathbb{E}(C_t) = \mu\lambda t, \quad \mathbb{V}(C_t) = \mu a t, \quad \frac{C_t}{t} \xrightarrow[t\to\infty]{p.s.} \mu\lambda \quad et \quad \sqrt{t}\left(\frac{C_t}{t} - \mu\lambda\right) \xrightarrow[t\to\infty]{\text{loi}} \mathcal{N}(0, \mu a).$$

Démonstration. L'expression des moments de C_t et le résultat de convergence en loi découlent de l'expression de la fonction caractéristique de C_t. La convergence presque sûre s'obtient en utilisant le lemme 19.2 et la *loi des grands nombres.* □

19.2 La ruine est-elle presque sûre ?

Comme les trajectoires de $\tilde{R}$ sont décroissantes entre les sauts, le franchissement du niveau c ne peut intervenir qu'à un instant de saut. Définissons la suite $(S_n)_{n\geqslant 0}$ par $S_n = \tilde{R}_{T_n}$. On dit que $(S_n)_{n\in\mathbb{N}}$ est la chaîne incluse de $(\tilde{R}_t)_{t\in\mathbb{R}_+}$. Par construction,

$$S_0 = 0 \quad \text{et} \quad \forall n \geqslant 1, \quad S_n = \sum_{k=1}^{n} (X_k - p\Delta_k).$$

En particulier, S_n s'écrit comme la somme de n variables aléatoires i.i.d. de moyenne $\lambda - p/\mu$. La quantité $\lambda\mu$ représente le montant moyen des sinistres par unité de temps et sa position par rapport à p dicte, comme le montre le résultat suivant, la valeur de la probabilité de ruine qui s'écrit encore

$$\psi(c) = \mathbb{P}\left(\sup_{n\geqslant 0} S_n > c\right).$$

Théorème 19.4 (CNS de ruine presque sure)**.**

$$p \leqslant \lambda\mu \quad \Longleftrightarrow \psi(c) = 1.$$

Démonstration. Si $p < \lambda\mu$, le résultat découle de la *loi forte des grands nombres* appliquée à la suite $(X_k - p\Delta_k)_{k\geqslant 1}$. Si $p = \lambda\mu$, la loi du logarithme itéré appliqué aux variables aléatoires i.i.d. $(X_k - p\Delta_k)_{1\leqslant k\leqslant n}$ qui sont centrées et de variance σ^2 finie assure que

$$\liminf_{n\to\infty} \frac{S_n}{\sigma\sqrt{2n\log(\log(n))}} = -1 \quad \text{p.s.} \quad \text{et} \quad \limsup_{n\to\infty} \frac{S_n}{\sigma\sqrt{2n\log(\log(n))}} = +1 \quad \text{p.s.}$$

Enfin, si $p > \lambda\mu$, la *loi forte des grands nombres* assure ici que $S_n \xrightarrow[n\to\infty]{\text{p.s.}} -\infty$. Supposons que le temps d'arrêt $\tau_1 = \inf\{n > 0 : S_n > c\}$ soit fini presque sûrement. La marche aléatoire $(S_{n+\tau_1} - S_{\tau_1})_{n\geqslant 0}$ est de même loi que $(S_n)_{n\geqslant 0}$. Par conséquent, le temps d'arrêt $\tau_2 = \inf\{n > 0 : S_{n+\tau_1} - S_{\tau_1} > c\}$ est lui aussi fini p.s. En itérant le procédé, on montre que $\limsup_{n\to+\infty} S_n = +\infty$, d'où la contradiction. □

Remarque 19.5. *Comme le suggère l'intuition, la compagnie doit s'assurer que $p > \lambda\mu$ pour éviter une ruine presque sûre. Le lemme 19.3 permet, à partir d'un historique des sinistres, d'estimer les paramètres λ et μ. Ceci fournira des informations sur la façon de choisir p.*

Recueil de modèles stochastiques – Page 189/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

19.3 Expression de la probabilité de ruine

On suppose dans la suite que la condition $p > \lambda\mu$ est satisfaite et on cherche à exprimer la probabilité de ruine $\psi(c)$. On introduit la transformée de Laplace H de la loi des sinistres ν :

$$\forall x \in \mathbb{R}, \quad H(x) = \mathbb{E}\big(e^{xX_1}\big) \in [0, +\infty].$$

Excluant le cas où X_1 est constante presque sûrement, la fonction H est strictement convexe. Elle est donc finie sur un intervalle I. Celui-ci contient $]-\infty, 0]$ car X_1 est une variable aléatoire positive.

Lemme 19.6 (Allure de la transformé de Laplace des sinistres)**.** *Supposons que l'intervalle I sur lequel la transformée de Laplace H est finie soit un ouvert. Alors la fonction $x \mapsto H(x) - px/\mu - 1$ s'annule en 0 et en un unique réel strictement positif noté u.*

Démonstration. La fonction H est de classe $\mathcal{C}^\infty$ sur I et tend vers $+\infty$ au bord droit de I. Soit $x_0 > 0$ tel que $\mathbb{P}(X \geqslant x_0) > 0$. Ainsi, pour tout $x \in I$,

$$H(x) \geqslant e^{xx_0}\mathbb{P}(X \geqslant x_0).$$

De plus, $H'(0) = \mathbb{E}(X_1) = \lambda < p/\mu$. Ainsi, la fonction $x \mapsto H(x) - px/\mu - 1$ est strictement convexe sur I, s'annule en 0, admet une dérivée strictement négative en 0 et tend vers l'infini au bord droit de I. □

Théorème 19.7 (Probabilité de ruine)**.** *Sous les hypothèses du lemme 19.6, la probabilité de ruine est donnée par*

$$\psi(c) = \frac{e^{-uc}}{\mathbb{E}\big(e^{uY(c)}|\tau(c) < +\infty\big)}$$

où $Y(c) = S_{\tau(c)} - c$ est le découvert de la compagnie à l'instant de ruine.

Démonstration. La suite $(e^{uS_n})_{n\geqslant 0}$ est une martingale positive pour sa filtration naturelle $(\mathcal{F}_n)_{n\geqslant 0}$. En effet,

$$\mathbb{E}\big(e^{uS_{n+1}}|\mathcal{F}_n\big) = e^{uS_n}\mathbb{E}\Big(e^{u(X_{n+1}-p\Delta_{n+1})}\Big).$$

Puisque les v.a. X_{n+1} et Δ_{n+1} sont indépendantes et que Δ_{n+1} suit la loi $\mathrm{Exp}(\mu)$, on a

$$\mathbb{E}\Big(e^{u(X_{n+1}-p\Delta_{n+1})}\Big) = H(u)\frac{\mu}{\mu - pu} = 1.$$

En particulier, pour tout $n \geqslant 0$, $E[e^{uS_n}] = 1$. Soit $n > 0$, alors $\tau(c) \wedge n$ est un temps d'arrêt borné et

$$1 = \mathbb{E}\big(e^{uS_{\tau(c)\wedge n}}\big) = \mathbb{E}\big(e^{uS_{\tau(c)}}\mathbf{1}_{\{\tau(c)<n\}}\big) + \mathbb{E}\big(e^{uS_n}\mathbf{1}_{\{\tau(c)\geqslant n\}}\big).$$

Sous l'hypothèse $p > \lambda\mu$, $S_n \xrightarrow[n\to\infty]{\text{p.s.}} -\infty$. De plus, sur $\{\tau(c) \geqslant n\}$, on a $S_n \leqslant c$. Ainsi, par convergence dominée, la seconde espérance ci-dessus converge vers 0 lorsque $n \to +\infty$. La première espérance converge par convergence monotone. Par conséquent,

$$1 = \mathbb{E}\big(e^{uS_{\tau(c)}}\mathbf{1}_{\{\tau(c)<+\infty\}}\big) = e^{uc}\mathbb{E}\Big(e^{uY(c)}|\tau(c) < +\infty\Big)\mathbb{P}(\tau(c) < +\infty),$$

ce qui fournit le résultat annoncé. □

Remarque 19.8 (Majoration de la probabilité de ruine)**.** *D'après le théorème précédent, on a toujours $\psi(c) \leqslant e^{-uc}$.*

Recueil de modèles stochastiques – Page 190/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

La transformée de Laplace de la mesure de densité $x \mapsto e^{-x}/(1+x^2)\mathbb{1}_{[0,\infty[}(x)$ est finie au point x ssi $x \leqslant 1$. Si $H(1) < 1 + p/\mu$, la fonction $x \mapsto H(x) - px/\mu - 1$ ne s'annule qu'en 0 : cette fonction, infinie sur $]1, +\infty[$, est strictement négative sur $]0, 1]$. Dans ce cas, le théorème 19.7 ne s'applique pas. Cependant, la suite $(e^{S_n})_{n\geqslant 0}$ est encore une sur-martingale :

$$\mathbb{E}\big(e^{S_{n+1}}|\mathcal{F}_n\big) = e^{S_n}\frac{H(1)}{1+px/\mu} < e^{S_n}.$$

On en déduit, en procédant comme dans la preuve du théorème 19.7 que

$$\psi(c) \leqslant \frac{e^{-c}}{\mathbb{E}\big(e^{Y(c)}|\tau(c) < +\infty\big)}.$$

19.4 Sinistres exponentiels

La probabilité de ruine est explicite pour des coûts de sinistres exponentiellement distribués.

Théorème 19.9 (Probabilité de ruine)**.** *Si la distribution des coûts des sinistres est la loi exponentielle de paramètre* $1/\lambda$*, ou de moyenne* λ *avec* $p > \lambda\mu$*, alors, pour tout* $c \geqslant 0$*, on a*

$$\psi(c) = \frac{\lambda\mu}{p}\exp\left(-\frac{p-\lambda\mu}{p\lambda}c\right).$$

Démonstration. Pour $x < 1/\lambda$, on a

$$H(x) = \frac{1}{1-\lambda x} \quad \text{et} \quad u = \frac{p-\mu\lambda}{p\lambda} < \frac{1}{\lambda}.$$

Il suffit de montrer que, sachant que $\tau(c) < +\infty$, la variable $Y(c) = S_{\tau(c)} - c$ suit la loi exponentielle d'espérance λ. Si $\tau(c) = n$ alors $Y(c) = S_n - c = X_n - p\Delta_n + S_{n-1} - c$. On a donc

$$\begin{aligned}
\mathbb{P}(Y(c) > y, \tau(c) = n) &= \mathbb{P}\Big(Y(c) > y, \max_{1\leqslant i\leqslant n-1} S_i \leqslant c, S_n > c\Big)\\
&= \mathbb{P}\Big(S_n > c + y, \max_{1\leqslant i\leqslant n-1} S_i \leqslant c\Big)\\
&= \mathbb{P}\Big(X_n > y + p\Delta_n + c - S_{n-1}, \max_{1\leqslant i\leqslant n-1} S_i \leqslant c\Big).
\end{aligned}$$

Puisque X_n suit la loi $\mathrm{Exp}(1/\lambda)$ et est indépendante de $(\Delta_n, S_0, \ldots, S_{n-1})$, on obtient

$$\begin{aligned}
\mathbb{P}(Y(c) > y, \tau(c) = n) &= \mathbb{P}(X_n > y)\mathbb{P}\Big(X_n > p\Delta_n + c - S_{n-1}, \max_{1\leqslant i\leqslant n-1} S_i \leqslant c\Big)\\
&= e^{-y/\lambda}\mathbb{P}\Big(S_n > c, \max_{1\leqslant i\leqslant n-1} S_i \leqslant c\Big)\\
&= e^{-y/\lambda}\mathbb{P}(\tau(c) = n).
\end{aligned}$$

Ainsi, $\mathbb{P}(Y(c) > y|\tau(c) < +\infty) = e^{-y/\lambda}$ pour tout $y > 0$. Ceci assure que

$$\mathbb{E}\Big(e^{uY(c)}|\tau(c) < +\infty\Big) = H(u).$$

On conclut en appliquant le théorème 19.7. □

Recueil de modèles stochastiques – Page 191/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Considérons deux compagnies qui assurent le même type de risques et pratiquent les mêmes politiques de tarif et d'indemnisation. Supposons que la clientèle de la seconde est plus grosse d'un facteur ρ que la première. Si (λ, μ, p) désigne les paramètres de la première compagnie, ceux de la seconde sont donnés par $(\lambda, \rho\mu, \rho p)$. On suppose que les sinistres sont toujours exponentiellement distribués. Par homogénéité, la probabilité de ruine de la compagnie $i = 1, 2$ est donnée par le théorème 19.9 :

$$\psi(c_i) = \frac{\lambda\mu}{p} \exp\left(-\frac{p - \lambda\mu}{p\lambda} c_i\right).$$

Supposons à présent qu'elles s'unissent pour ne former qu'une seule compagnie. On supposera à nouveau que les sinistres assurés par chacune des compagnies sont indépendants. Les paramètres de la nouvelle entité sont $(\lambda, (1+\rho)\mu, (1+\rho)p)$ et le capital initial vaut $c_1 + c_2$. Une fois de plus, la probabilité de ruine est donc donnée par $\psi(c_1 + c_2)$ qui est inférieure aux probabilités de ruine de chacune des deux compagnies. Les assureurs ont donc intérêt à se regrouper pour mutualiser les risques.

19.5 Notes et commentaires

Pour la définition et les propriétés essentielles des processus de Poisson, on pourra consulter par exemple le livre de Dominique Foata et Aimé Fuchs [FF04]. Les probabilités sont un des outils essentiels de l'actuariat, nom donné à la science de l'assurance. Pour aller plus loin, on pourra notamment consulter les ouvrages de Søren Asmussen [Asm00] ou de Arthur Charpentier et Michel Denuit [CD04].

Recueil de modèles stochastiques – Page 192/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Chapitre **20**

Naissances et assassinats

Recueil de modèles stochastiques – Page 193/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Mots-clés. Processus de Poisson ; processus de Galton-Watson ; théorème de Chernov.

On s'intéresse à l'extinction ou à la survie d'une population organisée en clans. Chaque individu vivant donne naissance à des enfants selon un *processus de Poisson* et ne peut pas mourir tant que son père est en vie. Dès que celui-ci disparaît tous ses enfants sont exposés et meurent après un temps aléatoire propre de loi F. Une configuration à un instant donné est une collection finie d'arbres généalogiques munis d'un patriarche qui seul peut être assassiné.

Le patriarche de la population entière sera noté $(\varnothing)$. Chacun de ses descendants est identifié à un k-uplet d'entiers strictement positifs pour $k \geqslant 1$. Par exemple, (1) est le premier enfant de l'ancêtre, tandis que $(3,2)$ est le deuxième enfant du troisième enfant de l'ancêtre. Le nombre de coordonnées indique la génération de l'individu. Soit

$$\mathcal{N} = \cup_{n \geqslant 0} (\mathbb{N}^*)^n$$

l'ensemble des n-uplets d'entiers strictement positifs, avec la convention $(\mathbb{N}^*)^0 = \varnothing$. Pour $\mathbf{n} \in \mathcal{N}$, notons $k(\mathbf{n})$ le nombre de coordonées de $\mathbf{n}$ et $\sigma(\mathbf{n}) = n_1 + \cdots + n_{k(\mathbf{n})}$ le nombre d'individus qui ont dû naître pour que $\mathbf{n}$ naisse. On adopte les conventions naturelles suivantes : $k(\varnothing) = 0 = \sigma(\varnothing)$.

Soit $(X^{\mathbf{n}})_{\mathbf{n} \in \mathcal{N}}$ une famille de processus de Poisson indépendants de même intensité λ. On notera $(T_k^{\mathbf{n}})_{k \geqslant 1}$ les temps de saut du processus $X^{\mathbf{n}}$. Soit $(K_{\mathbf{n}})_{\mathbf{n} \in \mathcal{N}}$ une famille de variables aléatoires indépendantes strictement positives de même loi F. On suppose que les familles $(X^{\mathbf{n}})_{\mathbf{n} \in \mathcal{N}}$ et $(K_{\mathbf{n}})_{\mathbf{n} \in \mathcal{N}}$ sont indépendantes.

Le système de particules est initialisé à $t = 0$ avec un unique ancêtre étiqueté par $\varnothing$. Cet individu donne naissance à des enfants aux instants de saut du processus de Poisson $X^{\varnothing}$ qui entrent dans le système à leur tour avec les indices $(1), (2), \ldots$ dans l'ordre de leurs naissances. Chaque nouvel individu $\mathbf{n}$ entrant dans le système commence à donner des descendants aux instants de saut du processus $X^{\mathbf{n}}$; les descendants de $\mathbf{n}$ sont indicés par $(\mathbf{n}, 1), (\mathbf{n}, 2), \ldots$ toujours dans l'ordre chronologique.

Le patriarche est exposé dès le temps 0. Il donne naissance à des enfants jusqu'au temps $Y_{\varnothing} = K_{\varnothing}$ où il disparaît. Chacun de ses enfants devient le patriarche de son clan et est à son tour exposé. Soit $\mathbf{n} = (n_1, \ldots, n_{k-1}, n_k) \in \mathcal{N}$ et $\mathbf{n}' = (n_1, \ldots, n_{k-1})$. Quand $\mathbf{n}'$ est supprimé, au temps $Y_{\mathbf{n}'}$, $\mathbf{n}$ devient exposé ; il continue à avoir des descendants jusqu'au temps $Y_{\mathbf{n}} = Y_{\mathbf{n}'} + K_{\mathbf{n}}$ puis est à son tour supprimé.

Exemple 20.1. *Dire que* $(3,2)$ *est né, c'est dire que*

— $\varnothing$ *a eu au moins trois enfants avant de mourir :* $T_3^{\varnothing}$ *intervient avant la mort de* $\varnothing$ *à l'instant* $K^{\varnothing}$*,*

— (3) *a eu au moins deux enfants avant de mourir* i.e. $T_2^{(3)}$ *intervient avant la mort de* (3) *à l'instant* $K^{\varnothing} + K^{(3)}$.

On a donc

$$\{(3,2) \text{ est né}\} = \left\{T_3^{\varnothing} < K^{\varnothing},\ T_3^{\varnothing} + T_2^{(3)} < K^{\varnothing} + K^{(3)}\right\}.$$

On dit que le système de naissances et meurtres est stable si, avec probabilité 1, il existe un temps fini à partir duquel le système est vide. Sinon on dit que le système est instable.

Théorème 20.2 (Étude de la stabilité)**.** *Soit $\mathcal{B}$ le processus de naissances et meurtres d'intensité λ dont la transformée de Laplace ϕ est finie sur un voisinage de l'origine, et soit*

$$\alpha = \min_{u>0} \frac{\lambda\phi(u)}{u}.$$

Alors on dispose des conditions suffisantes de stabilité et d'instabilité suivantes :

— *le processus $\mathcal{B}$ est stable si $\alpha < 1$,*

— *le processus $\mathcal{B}$ est instable si $\alpha > 1$.*

Les sections 20.1 et 20.2 sont consacrées à la preuve des deux conclusions du théorème 20.2.

20.1 Condition suffisante de stabilité

Considérons le nombre moyen de descendants de $\varnothing$:

$$m = \sum_{\mathbf{n}} \mathbb{P}(\mathbf{n} \text{ est né}) \in [0, +\infty].$$

Si $m < \infty$ alors le processus $\mathcal{B}$ est stable. Montons que c'est bien le cas dès que $\alpha < 1$.

À l'individu $\mathbf{n} = (n_1, \ldots, n_k)$, on associe sa lignée ancestrale formée des individus $\mathbf{n}(i) = (n_1, \ldots, n_i)$ pour $0 \leqslant i \leqslant k$ (avec $\mathbf{n}(0) = \varnothing$). L'individu $\mathbf{n}$ voit le jour si, pour tout $0 \leqslant i \leqslant k$, son ancêtre $\mathbf{n}(i-1)$ à la génération i a eu au moins n_i descendants :

$$\{\mathbf{n} \text{ est né}\} = \left\{\sum_{i=1}^{j} T_{n_i}^{\mathbf{n}(i-1)} < \sum_{i=1}^{j} K^{\mathbf{n}(i-1)},\ j = 1, 2, \ldots, k(\mathbf{n})\right\}.$$

En particulier,

$$\mathbb{P}(\mathbf{n} \text{ est né}) \leqslant \mathbb{P}\left(\sum_{i=1}^{k(\mathbf{n})} T_{n_i}^{\mathbf{n}(i-1)} < \sum_{i=1}^{k(\mathbf{n})} K^{\mathbf{n}(i-1)}\right).$$

L'inégalité de Markov assure que, pour tout $u > 0$,

$$\mathbb{P}\left(\sum_{i=1}^{k(\mathbf{n})} T_{n_i}^{\mathbf{n}(i-1)} < \sum_{i=1}^{k(\mathbf{n})} K^{\mathbf{n}(i-1)}\right) \leqslant \mathbb{E}\left(\exp\left(u\sum_{i=1}^{k(\mathbf{n})} K^{\mathbf{n}(i-1)} - u\sum_{i=1}^{k(\mathbf{n})} T_{n_i}^{\mathbf{n}(i-1)}\right)\right).$$

L'indépendance des processus de naissances et de meurtres permet d'écrire

$$\mathbb{E}\left(\exp\left(u\sum_{i=1}^{k(\mathbf{n})} K^{\mathbf{n}(i-1)} - u\sum_{i=1}^{k(\mathbf{n})} T_{n_i}^{\mathbf{n}(i-1)}\right)\right) = \phi(u)^{k(\mathbf{n})} \prod_{i=1}^{k(\mathbf{n})} \left(\frac{\lambda}{\lambda+u}\right)^{n_i}.$$

En sommant sur $\mathbf{n}$, on obtient

$$\sum_{\mathbf{n}} \mathbb{P}(\mathbf{n} \text{ est né}) \leqslant \sum_{k=1}^{\infty} \phi(u)^k \left[\sum_{\mathbf{n},\, k(\mathbf{n})=k} \left(\frac{\lambda}{\lambda+u}\right)^{\sigma(\mathbf{n})}\right].$$

Recueil de modèles stochastiques – Page 194/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

D'autre part,

$$\begin{aligned}\sum_{\mathbf{n},\, k(\mathbf{n})=k} \left(\frac{\lambda}{\lambda+u}\right)^{\sigma(\mathbf{n})} &= \sum_{m\geqslant k} \mathrm{card}\{\mathbf{n} \,:\, k(\mathbf{n})=k,\ \sigma(\mathbf{n})=m\}\left(\frac{\lambda}{\lambda+u}\right)^{m} \\ &= \left(\frac{\lambda}{u}\right)^{k} \sum_{m\geqslant k} \mathrm{card}\{\mathbf{n} \,:\, k(\mathbf{n})=k,\ \sigma(\mathbf{n})=m\}\left(\frac{u}{u+\lambda}\right)^{k}\left(\frac{\lambda}{u+\lambda}\right)^{m-k}.\end{aligned}$$

L'entier $\mathrm{card}\{\mathbf{n} \,:\, k(\mathbf{n})=k,\ \sigma(\mathbf{n})=m\}$ représente le nombre de m-uplets à coordonées 0 ou 1 contenant exactement k 1. Le terme général de la dernière somme ci-dessus est donc la probabilité que dans un schéma de Bernoulli de probabilité de succès $u/(u+\lambda)$ le k^{e} succès survienne au temps m (c'est la loi binomiale négative). On retrouve au passage

$$\mathrm{card}\{\mathbf{n} \,:\, k(\mathbf{n})=k,\ \sigma(\mathbf{n})=m\} = \binom{m-1}{k-1}.$$

On a donc

$$\sum_{\mathbf{n},\, k(\mathbf{n})=k} \left(\frac{\lambda}{\lambda+u}\right)^{\sigma(\mathbf{n})} = \left(\frac{\lambda}{u}\right)^{k}.$$

En d'autres termes, on a donc montré que

$$m \leqslant \sum_{k\geqslant 1} \left(\frac{\lambda\phi(u)}{u}\right)^{k}.$$

Si $\alpha < 1$, alors le nombre moyen de particules créées est fini et le processus $\mathcal{B}$ est stable.

20.2 Condition suffisante d'instabilité

Démontrons à présent la seconde partie du théorème 20.2. L'idée principale est de montrer que, si $\alpha > 1$, il existe $k \geqslant 1$ tel que

$$\sum_{\mathbf{n},\, k(\mathbf{n})=k} \mathbb{P}(\mathbf{n} \text{ est né}) > 1$$

puis de conclure grâce à un argument de type branchement.

Soit $(N_i)_{i\geqslant 1}$ une suite de variables aléatoires indépendantes de loi géométrique de paramètre p indépendante du processus $\mathcal{B}$. À la génération k, on considère l'individu $\mathbf{N} = (N_1, \ldots, N_k)$ ainsi que sa ligné ancestrale formée des individus $\mathbf{N}(i) = (N_1, \ldots, N_i)$ pour $0 \leqslant i \leqslant k$. Exprimons de deux manières différentes la probabilité que $\mathbf{N}$ naisse. D'une part,

$$\begin{aligned}\mathbb{P}(\mathbf{N} \text{ est né}) &= \sum_{\mathbf{n},\, k(\mathbf{n})=k} \mathbb{P}\Big(\mathbf{N} \text{ est né}\Big|\mathbf{N}=\mathbf{n}\Big)\mathbb{P}(\mathbf{N}=\mathbf{n}) \\ &= \sum_{\mathbf{n},\, k(\mathbf{n})=k} \mathbb{P}(\mathbf{n} \text{ est né}) p^{k}(1-p)^{n_1+\cdots+n_k-k} \\ &= \left(\frac{p}{1-p}\right)^{k} \sum_{\mathbf{n},\, k(\mathbf{n})=k} \mathbb{P}(\mathbf{n} \text{ est né})(1-p)^{\sigma(\mathbf{n})}.\end{aligned}$$

D'autre part, posons

$$Z_i = K^{\mathbf{N}(i-1)} - T_{N_i}^{\mathbf{N}(i-1)},$$

Recueil de modèles stochastiques – Page 195/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

pour tout $i \geqslant 1$. Comme dans la première partie de la preuve,

$$\begin{aligned}\mathbb{P}(\mathbf{N} \text{ est né}) &= \mathbb{P}\left(\sum_{i=1}^{j} T_{N_i}^{\mathbf{N}(i-1)} < \sum_{i=1}^{j} K^{\mathbf{N}(i-1)},\ j = 1, 2, \ldots, k\right) \\ &= \mathbb{P}\left(\sum_{i=1}^{j} Z_i > 0,\ j = 1, 2, \ldots, k\right).\end{aligned}$$

Pour estimer cette probabilité, on utilise le lemme suivant, démontré en fin de section.

Lemme 20.3 (Déviation asymptotique)**.** *Si $(X_i)_{i\geqslant 1}$ est une suite de v.a.r. i.i.d. avec $\mathbb{E}(X_1) < 0$, $\mathbb{P}(X_1 > 0) > 0$, et si la transformée de Laplace ψ de X_1 est finie sur un voisinage de 0, alors*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\min_{1\leqslant k\leqslant n} \sum_{i=1}^{k} X_i > 0\right) = \log(\rho), \quad \text{où} \quad \rho = \min_{u>0} \psi(u).$$

La variable $T_{N_i}^{\mathbf{N}(i-1)}$ suit la loi $\mathrm{Exp}(\lambda p)$ car somme géométrique de v.a. i.i.d. de loi exponentielle. Les variables aléatoires $(Z_i)_{i\geqslant 1}$ sont i.i.d. avec pour transformée de Laplace et moyenne

$$\mathbb{E}\big(e^{uZ_1}\big) = \phi(u)\frac{\lambda p}{\lambda p + u} \quad \text{et} \quad \mathbb{E}(Z_1) = \phi'(0) - \frac{1}{p\lambda}.$$

Ainsi, l'espérance de Z_1 est strictement négative pour p assez petit. De plus, si l'on note $\alpha = 1 + \varepsilon$, on peut choisir p suffisamment petit pour que

$$\min_{u>0} \mathbb{E}\big(e^{uZ_1}\big) = p \min_{u>0} \frac{\lambda\phi(u)}{\lambda p + u} > p(1 + \varepsilon/2).$$

Le lemme 20.3 appliqué à la suite $(Z_n)_{n\geqslant 1}$ assure que, pour k assez grand,

$$\mathbb{P}\left(\min_{1\leqslant j\leqslant k} \sum_{i=1}^{j} Z_i > 0\right) > p^k(1 + \varepsilon/2)^k.$$

On obtient donc

$$\left(\frac{p}{1-p}\right)^k \sum_{\mathbf{n},\, k(\mathbf{n})=k} \mathbb{P}(\mathbf{n} \text{ est né}) q^{\sigma(\mathbf{n})} > p^k(1 + \varepsilon/2)^k.$$

On divise par p^k et on fait tendre p vers 0. Le théorème de convergence monotone assure que

$$\sum_{\mathbf{n},\, k(\mathbf{n})=k} \mathbb{P}(\mathbf{n} \text{ est né}) \geqslant (1 + \varepsilon/2)^k > 1,$$

ce qui fournit la relation

$$\sum_{\mathbf{n},\, k(\mathbf{n})=k} \mathbb{P}(\mathbf{n} \text{ est né}) > 1.$$

Achevons la démonstration du théorème 20.2. On dira qu'une particule est *spéciale* si c'est l'ancêtre ou si c'est une particule de la génération nk qui descend d'une particule spéciale à la génération $(n-1)k$ dont le père est mort. La relation obtenue ci-dessus nous apprend que le nombre moyen de particules spéciales à la génération k est strictement supérieur à $(1+\varepsilon/2)^k$. De même, le nombre moyen de particules spéciales à la génération nk est supérieur à $(1 + \varepsilon/2)^{nk}$. Ainsi, le processus de comptage des particules spéciales est un processus de branchement de Galton-Watson sur-critique. Avec une probabilité strictement positive, ce processus ne s'éteint pas (voir le théorème 5.6 page 5.6). C'est donc également le cas pour le processus $\mathcal{B}$.

Terminons cette section par la démonstration du lemme 20.3.

Recueil de modèles stochastiques – Page 196/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Démonstration du lemme 20.3. Sous les hypothèses du lemme 20.3, le théorème de Chernov

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\sum_{i=1}^{n} X_i > 0\right) = \log(\rho)$$

assure que le résultat du lemme 20.3 est vrai pour l'instant terminal. Montrons à présent que [1]

$$\mathbb{P}\left(\min_{1\leqslant k\leqslant n} \sum_{i=1}^{k} X_i > 0\right) \geqslant \frac{1}{n}\mathbb{P}\left(\sum_{i=1}^{n} X_i > 0\right).$$

Pour tout $l \in \{1, \dots, n\}$ notons σ_l la permutation circulaire $\sigma_l = (l, \dots, n, 1, \dots, l-1)$ et posons

$$S_k^{\sigma_l} = \sum_{i=\sigma_1(1)}^{\sigma_l(k)} X_i \quad \text{et} \quad M_n^{\sigma_l} = \min_{1\leqslant k\leqslant n} S_k^{\sigma_l}.$$

La somme $S_n^{\sigma_l} = X_1 + \cdots + X_n$ ne dépend pas de l. C'est également le cas pour les probabilités

$$\mathbb{P}(M_n^{\sigma_l} > 0 \,|\, S_n > 0)$$

puisque $(X_1, \dots, X_n)$ et $(X_{\sigma_l(1)}, \dots, X_{\sigma_l(n)})$ ont même loi pour tout $l \in \{1, \dots, n\}$. On a donc

$$\mathbb{P}\left(\min_{1\leqslant k\leqslant n} \sum_{i=1}^{k} X_i > 0 \,\middle|\, S_n > 0\right) = \frac{1}{n}\mathbb{E}\left(\sum_{l=1}^{n} \mathbf{1}_{\left\{M_n^{\sigma_l} > 0\right\}} \,|\, S_n > 0\right).$$

Il reste à remarquer que sur l'évènement $\{S_n > 0\}$, l'une au moins des variables aléatoires $M_n^{\sigma_l}$ est strictement positive. Cette propriété n'a rien d'aléatoire. Notons

$$K = \begin{cases} \max\left\{1 \leqslant k \leqslant n-1 : S_k^{\sigma_1} = M_n^{\sigma_1}\right\} & \text{si } M_n^{\sigma_1} \leqslant 0, \\ 0 & \text{sinon.} \end{cases}$$

Notons que $K \leqslant n-1$ puisque $S_n > 0$. De plus, K peut être constant égal 0. C'est par exemple le cas si X_1 est à valeurs dans $\{-10, 1\}$ et que $n = 5$: sur l'évènement $\{S_5 > 0\}$, toutes les variables aléatoires $(X_i)_{1\leqslant i\leqslant 5}$ doivent être positives (égales à 1) et $K = 0$. Par construction, $M_n^{\sigma_{K+1}} > 0$ (faire un dessin et comparer avec la figure 5.3 page 5.3). Ceci fournit la minoration annoncée. □

20.3 Notes et commentaires

Le modèle présenté a été introduit dans un article [AK90] de David Aldous et William Krebs. Le lien avec un modèle de file d'attente est abordé dans [TPH86]. Le théorème de Chernov est démontré par exemple dans un article [DZ10] de John Tsitsiklis, Christos Papadimitriou, et Pierre Humblet.

Le théorème 20.2 laisse entière la question de la stabilité éventuelle du processus $\mathcal{B}$ au point critique. Une réponse précise peut être apportée dans un cas particulier important. Supposons que la loi de meurtre soit exponentielle. Par changement d'échelle de temps, on peut supposer sans perte de généralité que son paramètre vaut 1. Le théorème 20.2 assure que le processus $\mathcal{B}$ est stable si $0 < \lambda < 1/4$ et instable si $\lambda > 1/4$. On note $\mathbb{E}_\lambda$ l'espérance par rapport à la loi de $\mathcal{B}_\lambda$ et N le nombre total de particules qui sont nées, ancêtre compris. L'article [Bor08] de Charles Bordenave montre notamment que les moments de N satisfont les propriétés d'intégrabilité suivantes :

1. Cette inégalité évoque l'identité du principe de rotation du lemme 5.16 concernant une marche aléatoire et intervenant dans l'étude de la loi de la taille totale des arbres de Galton-Watson sous-critiques.

Recueil de modèles stochastiques – Page 197/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

— si $\lambda \in (0, 1/4]$,

$$\mathbb{E}_\lambda(N) = \frac{2}{1+\sqrt{1-4\lambda}},$$

— si $\lambda \in (0, 2/9]$,

$$\mathbb{E}_\lambda(N^2) = \frac{2}{3\sqrt{1-4\lambda}-1},$$

— pour tout $p \geqslant 2$, $\mathbb{E}_\lambda(N^p) < \infty$ si et seulement si $\lambda \in (0, p(p+1)^{-2})$.

En particulier, si $\lambda = 1/4$, le processus $\mathcal{B}$ est stable puisque $\mathbb{E}_{1/4}N = 2$. Il est aussi montré dans [Bor08] que la loi de N a une queue polynomiale pour tout $\lambda \in (0, 1/4)$. En effet, pour tout $\varepsilon > 0$, il existe C tel que pour tout $t > 0$,

$$\mathbb{P}_\lambda(N > t) \leqslant Ct^{-\gamma(\lambda)+\varepsilon} \quad \text{avec} \quad \gamma(\lambda) = \frac{1+\sqrt{1-4\lambda}}{1-\sqrt{1-4\lambda}}.$$

Les processus de branchement de type Galton-Watson (étudiés au chapitre 5) en régime sous-critique ne présentent pas ce comportement à queues lourdes : il existe une constante $c > 0$ telle que $\mathbb{E}_\lambda \exp(cN) < \infty$ pour $\lambda < 1$. D'autre part, le nombre moyen d'individus est donné par $\mathbb{E}_\lambda(N) = (1-\lambda)^{-1}$: il explose donc quand λ tend vers la valeur critique 1.

Recueil de modèles stochastiques – Page 198/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Chapitre 21

Croissance et fragmentation

Recueil de modèles stochastiques – Page 199/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Mots-clés. Processus de Poisson ; processus de Markov déterministe par morceaux ; générateur infinitésimal ; couplage en Wasserstein ; processus auto-régressif ; loi exponentielle ; loi géométrique ; processus de branchement à temps continu de Yule.

21.1 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 $N = (N_t)_{t\geqslant 0}$ un *processus de Poisson* issu de 0 et d'intensité λ comptant les signaux de congestion. On note $(T_n)_{n\geqslant 0}$ ses temps de saut successifs, avec la convention $T_0 = 0$. Si $E_n = T_n - T_{n-1}$ pour tout $n \geqslant 1$ alors $T_n = E_1 + \cdots + E_n$ et la suite $E = (E_n)_{n\geqslant 1}$ est constituée de v.a.r. i.i.d. de loi exponentielle de moyenne $1/\lambda$. On a $N_t = \sum_{n\geqslant 1} \mathbb{1}_{[0,t]}(T_n)$ pour tout $t \in \mathbb{R}_+$. Soit $Q = (Q_n)_{n\geqslant 1}$ une suite de variables aléatoires indépendantes à valeurs dans $[0,1[$ de même loi $\mathcal{Q}$. Soit X_0 une variable aléatoire positive. On suppose que X_0, E, et Q sont indépendants. On définit le processus $X = (X_t)_{t\geqslant 0}$ à temps continu et à espace d'états $\mathbb{R}_+$ de la manière suivante :

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

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 $[0,1]$. On dit qu'il s'agit d'un *processus de Markov déterministe par morceaux* [1] et plus particulièrement à celles des processus à croissance linéaire et décroissance multiplicative [2]. Le processus X est connu sous le nom de *processus de taille de fenêtre TCP* [3] car en informatique, le débit maximal sortant est réglé par une taille de « fenêtre TCP ».

1. «*Piecewise Deterministic Markov Processes (PDMP)*» en anglais.
2. «*Additive Increase Multiplicative Decrease (AIMD)*» en anglais, ou encore «*Growth Collapse*» lorsque $\mathcal{Q} = \delta_0$.
3. «*TCP window-size process*» en anglais.

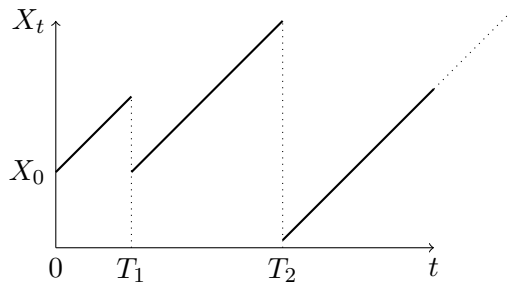


Figure 21.1 – Allure d'une trajectoire typique du processus X.

Recueil de modèles stochastiques – Page 200/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Théorème 21.1 (Générateur infinitésimal)**.** *Pour tout $f \in \mathcal{C}^1(\mathbb{R}_+, \mathbb{R})$ et $x \geqslant 0$ on a*

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

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 $X_t \leqslant x + t$, 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 $t \leqslant 1$. On a, par définition de X et N,

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

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 $t \to 0$) provient de la continuité de f, et donc

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

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

$$\big|\mathbb{E}\big(f(X_t)\mathbb{1}_{\{N_t \geqslant 2\}} \,|\, X_0 = x\big)\big| \leqslant \left(\sup_{u \in [0, x+1]} |f(u)|\right) \mathbb{P}(N_t \geqslant 2) = \mathcal{O}(t^2).$$

Finalement, la collecte des trois termes donne bien le résultat voulu. □

Pour tout $t \geqslant 0$ on considère l'opérateur P_t qui à toute fonction $f : \mathbb{R}_+ \to \mathbb{R}$ 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 $[0, x+t]$. On a $P_0 = I$ et la propriété de Markov entraîne que $(P_t)_{t \geqslant 0}$ 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 : $P_t(f) \geqslant 0$ si $f \geqslant 0$ et $P_t(1) = 1$. On dit qu'une fonction mesurable et bornée $f : \mathbb{R}_+ \to \mathbb{R}$ appartient au domaine du générateur L du processus lorsque la limite suivante existe pour tout $x \geqslant 0$:

$$L(f)(x) = \lim_{t \to 0^+} \frac{P_t(f)(x) - f(x)}{t} = \partial_{t=0} P_t(f)(x).$$

Tout se passe comme si «$P_t = e^{tL}$». On peut établir que la propriété de semi-groupe donne, pour tout $t \geqslant 0$ et toute fonction f de classe $\mathcal{C}^1$,

$$\partial_t P_t(f) = \lim_{\varepsilon \to 0} P_t\left(\frac{P_\varepsilon f - f}{t}\right) = P_t(Lf) \quad \text{et} \quad \partial_t P_t(f) = \lim_{\varepsilon \to 0} \frac{P_\varepsilon(P_t f) - P_t f}{t} = L(P_t f).$$

On dit qu'il s'agit des équations de Chapman-Kolmogorov *progressive* et *rétrograde* respectivement (*forward* et *backward* en anglais).

Le théorème 21.1 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 21.2 (Moments et loi invariante)**.** *Pour tout $n \in \mathbb{N}$, tout $x \geqslant 0$ et tout $t \geqslant 0$,*

$$\mathbb{E}(X_t^n \,|\, X_0 = x) = \frac{n!}{\prod_{k=1}^n \theta_k} + n! \sum_{m=1}^n \left(\sum_{k=0}^m \frac{x^k}{k!} \prod_{\substack{j=k \\ j \neq m}}^n \frac{1}{\theta_j - \theta_m} \right) e^{-\theta_m t}$$

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 $t \to \infty$ 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 $n \geqslant 0$. Enfin, 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 $n \geqslant 1$, le théorème 21.1 donne

$$Lf_n(x) = nx^{n-1} + \lambda x^n \int_0^1 (q^n - 1) \, \mathcal{Q}(dq) = nf_{n-1}(x) - \theta_n f_n(x)$$

et donc, en utilisant l'équation de Chapman-Kolmogorov progressive (forward),

$$\alpha_n'(t) = \partial_t P_t(f_n)(x) = P_t(Lf_n)(x) = P_t(nf_{n-1} - \theta_n f_n)(x) = n\alpha_{n-1}(t) - \theta_n \alpha_n(t).$$

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 $(\alpha_n)_{n \geqslant 0}$. Ainsi, conditionnellement à $\{X_0 = x\}$, chaque moment de X_t converge quand $t \to \infty$ 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 16. On prendra garde au fait qu'on se trouve ici sur $\mathbb{R}_+$ et non pas sur $\mathbb{R}$ tout entier. □

Recueil de modèles stochastiques – Page 201/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Remarque 21.3 (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 paramère λ.*

Remarque 21.4 (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 $x \geqslant 0$. 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 $\mathrm{dist}(x + t, \mathrm{supp}(\mu_t)) = (1 - q)x$. *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 $0 = U_0 < U_1 < \cdots < U_n < U_{n+1} = t$ 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 $[q^n(x + t), q^n x + t]$. 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 21.5 (Irréductibilité)**.** *Pour tout $x \geqslant 0$ et conditionnellement à l'événement $\{X_0 = x\}$, le processus X entre dans tout intervalle en un temps fini avec probabilité positive. En d'autre 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.*

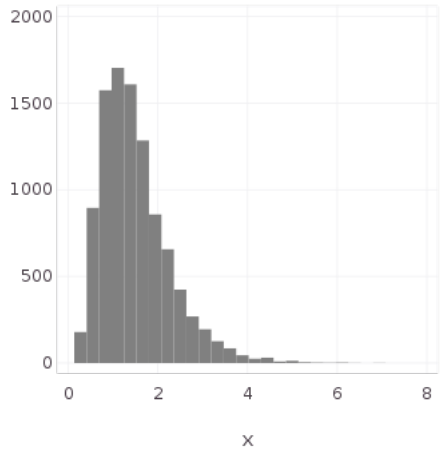


Figure 21.2 – Histogramme d'un échantillon de X_t pour $t = 100$, $\lambda = 1$, $\mathcal{Q} = \delta_{1/2}$.

Soit $p \geqslant 1$ 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 Wasserstein[4] W_p sur l'ensemble des lois sur E possédant un moment d'ordre p fini est définie par

$$W_p(\mu_1, \mu_2) = \inf_{\substack{(X,Y) \\ X_1 \sim \mu_1 \\ X_2 \sim \mu_2}} \mathbb{E}(d(X, Y)^p)^{1/p} = \left(\inf_{\Pi \in \mathrm{M}(\mu_1, \mu_2)} \int_{E \times E} d(x, y)^p \, \Pi(dx, dy) \right)^{1/p}$$

4. On parle également de distance de couplage, de transport, de Kantorovich, de Mallows, de Fréchet, etc.

Recueil de modèles stochastiques – Page 202/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

où $\mathrm{M}(\mu_1, \mu_2)$ 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 $\leqslant p$ de μ_n convergent vers ceux de μ.

Théorème 21.6 (Comportement en temps long par couplage)**.** *Soient $X = (X_t)_{t \geqslant 0}$ et $Y = (Y_t)_{t \geqslant 0}$ deux processus TCP de même paramètres λ et $\mathcal{Q}$, 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 $p \geqslant 1$ fini alors pour tout $t \geqslant 0$ et en posant $\theta_p := \lambda(1 - \mathbb{E}(Q_1^p))$ on a*

$$W_p(\mathcal{L}(X_t), \mathcal{L}(Y_t)) \leqslant W_p(\mathcal{L}(X_0), \mathcal{L}(Y_0)) \exp\left(-\frac{\theta_p t}{p}\right).$$

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 $t \geqslant 0$, on a

$$\begin{aligned}
\mathbb{E}(|X_t - Y_t|^p) &= \sum_{k=0}^{\infty} \mathbb{E}\big(|X_t - Y_t|^p \mathbb{1}_{\{N_t = k\}}\big) \\
&= \sum_{k=0}^{\infty} \mathbb{E}\big(|x - y|^p Q_1^p \dots Q_k^p \mathbb{1}_{\{N_t = k\}}\big) \\
&= |x - y|^p \sum_{k=0}^{\infty} \mathbb{E}(Q_1^p)^k \mathbb{P}(N_t = k) \\
&= |x - y|^p e^{-\lambda t(1 - \mathbb{E}(Q_1^p))}.
\end{aligned}$$

Par conséquent, si à présent X et Y sont deux processus TCP de paramètres λ et $\mathcal{Q}$ alors pour tout couplage Π de leurs lois initiales $\mathcal{L}(X_0)$ et $\mathcal{L}(Y_0)$, pour tout $t \geqslant 0$ et tout $p \geqslant 1$,

$$W_p(\mathcal{L}(X_t), \mathcal{L}(Y_t))^p \leqslant e^{-\theta_p t} \int_{\mathbb{R}_+ \times \mathbb{R}_+} |x - y|^p \, \Pi(dx, dy).$$

Il ne reste plus qu'à prendre l'infimum sur tous les couplages Π pour obtenir le résultat. □

La chaîne incluse $\widehat{X} := (\widehat{X}_n)_{n \geqslant 0} = (X_{T_n})_{n \geqslant 0}$ 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 $x \geqslant 0$ 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 21.7 (Convergence en loi de la chaîne incluse)**.** *Quelque soit X_0 la chaîne incluse $(\widehat{X}_n)_{n \geqslant 0}$ 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

$$\widehat{X}_n = Q_n \cdots Q_1 X_0 + \sum_{k=1}^{n} Q_n \cdots Q_{n-k+1} E_{n-k+1} \stackrel{d}{=} Q_1 \cdots Q_n X_0 + \sum_{k=1}^{n} Q_1 \cdots Q_k E_k.$$

Recueil de modèles stochastiques – Page 203/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

On a $\mathbb{E}(Q_1 \cdots Q_n) = \mathbb{E}(Q)^n$. Or $\mathbb{E}(Q) < 1$ car $\mathbb{P}(0 \leqslant Q < 1) = 1$ 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 $\mathbb{P}(Q_1 \cdots Q_n \to 0) = 1$. Par conséquent, $Q_1 \cdots Q_n X_0 \to 0$ p.s. D'autre part, par convergence monotone, $\sum_{k=1}^n Q_1 \ldots Q_k E_k$ 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. □

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.

21.2 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 $(E_n)_{n \geqslant 1}$ de v.a.r. i.i.d. de loi exponentielle de paramètre 1. On construit la trajectoire du processus $(\widetilde{X}_t)_{t \geqslant 0}$ partant de $\widetilde{X}_0 = x$ en posant $\widetilde{X}_s = x_0 + s$ pour tout $0 \leqslant s < \widetilde{T}_1$ où le premier temps de saut $\widetilde{T}_1$ est défini par l'équation (instant ou l'$\int$ 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 $\widetilde{X}_{\widetilde{T}_1} = Q_1 \widetilde{X}_{\widetilde{T}_1^-} = Q_1(x_0 + \widetilde{T}_1)$, et plus généralement, on pose

$$\widetilde{X}_t = \begin{cases} \widetilde{X}_{\widetilde{T}_n} + t - \widetilde{T}_n & \text{si } \widetilde{T}_n \leqslant t < \widetilde{T}_{n+1}, \\ Q_{n+1}(\widetilde{X}_{\widetilde{T}_n} + \widetilde{T}_{n+1} - \widetilde{T}_n) & \text{si } t = \widetilde{T}_{n+1}, \end{cases}$$

où la suite croissante des temps de saut $(\widetilde{T}_n)_{n \geqslant 0}$ est définie par $\widetilde{T}_0 = 0$ et

$$\int_{\widetilde{T}_n}^{\widetilde{T}_{n+1}} \lambda(\widetilde{X}_s)\, ds = \int_0^{\widetilde{T}_{n+1} - \widetilde{T}_n} \lambda(\widetilde{X}_{\widetilde{T}_n} + s)\, ds = E_{n+1}.$$

Par exemple, le cas spécial $\lambda(x) = x$ donne

$$\widetilde{T}_{n+1} - \widetilde{T}_n = \sqrt{2E_{n+1} + (\widetilde{X}_{\widetilde{T}_n})^2} - \widetilde{X}_{\widetilde{T}_n}.$$

Cette construction fournit la trajectoire $t \mapsto \widetilde{X}_t$ sur l'intervalle $[0, \widetilde{T}_\infty[$ où $\widetilde{T}_\infty := \lim_{n \to \infty} \widetilde{T}_n$ est le *temps d'explosion* du processus. Si λ est une fonction bornée, alors par comparaison au cas constant, on obtient $\mathbb{P}(\widetilde{T}_\infty = \infty) = 1$ 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 $(\widetilde{X}_t)_{t \geqslant 0}$ est

$$\widetilde{L}(f)(x) = f'(x) + \lambda(x) \int_0^1 (f(qx) - f(x))\, \mathcal{Q}(dq).$$

Plus généralement, on peut construire les trajectoires d'un processus de générateur

$$\widetilde{L}(f)(x) = G(f)(x) + \lambda(x) \int (f(y) - f(x))\, K(x, dy)$$

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.

Recueil de modèles stochastiques – Page 204/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

21.3 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 polymère [5] 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 $t \geqslant 0$. 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 $t \geqslant 0$, l'arbre possède N_t branches, et $N_0 = 1$. Le processus $(N_t)_{t\geqslant 0}$ 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.

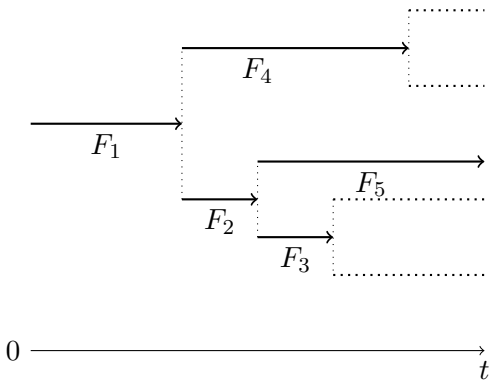


FIGURE 21.3 – Début de l'arbre binaire définissant le processus N.

Théorème 21.8 (Taille de la population)**.** *Pour tout $t \geqslant 0$ la taille de la population N_t suit la loi géométrique sur $\mathbb{N}^*$ de paramètre $e^{-\lambda t}$. En particulier,*

$$\mathbb{E}(N_t) = e^{\lambda t} \quad \textit{et} \quad \mathbb{V}(N_t) = (1 - e^{-\lambda t})e^{2\lambda t}.$$

Démonstration. La propriété d'absence de mémoire des lois exponentielles permet d'établir que $(N_t)_{t\geqslant 0}$ est le processus de comptage issu de $N_0 = 1$ de la suite de v.a.r. $(S_n)_{n\geqslant 1}$ où $S_n = F_1 + \cdots + F_n$ avec $(F_n)_{n\geqslant 1}$ suite de v.a.r. indépendantes et F_n de loi exponentielle de paramètre $n\lambda$ pour tout $n \geqslant 1$. On a $N_t - 1 = \sum_{n=1}^{\infty} \mathbb{1}_{[0,t]}(S_n)$. Or la v.a.r. S_n a la loi de $\max(E_1, \ldots, E_n)$ où $E_1, \ldots, E_n$ sont i.i.d. de loi exponentielle de paramètre λ (propriété également utilisée dans la preuve du théorème 10.4). Donc, pour tout entier $n \geqslant 0$,

$$\mathbb{P}(N_t - 1 \geqslant n) = \mathbb{P}(S_n \leqslant t) = \mathbb{P}(E_1 \leqslant t) \cdots \mathbb{P}(E_n \leqslant t) = (1 - e^{-\lambda t})^n.$$

□

Considérons le cas où chaque fragmentation donne lieu à $k + 1$ fragments où $k \geqslant 1$ est fixé. On note $(N_t^{(k)})_{t\geqslant 0}$ le processus de comptage associé. Pour tout $n \geqslant 2$, 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 $(N_t^{(k)})_{t\geqslant 0}$ est un processus de Yule de loi de reproduction

5. Longue molécule constituée d'une chaîne de molécules plus petites identiques appelées monomères.

Recueil de modèles stochastiques – Page 205/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

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 $f : \mathbb{N} \to \mathbb{R}$ et $n \in \mathbb{N}$ par

$$L^{(k)}(f)(n) = \lim_{t \to 0^+} \frac{\mathbb{E}(f(N_t^{(k)}) \,|\, N_0^{(k)} = n) - f(n)}{t} = n\lambda(f(n+k) - f(n)).$$

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 21.9 (Processus de Yule)**.** *Supposons que $N_0^{(k)} \sim \delta_1$. Pour tout réel $t \geqslant 0$ et tout entier $k \geqslant 1$ si $Z_1, \ldots, Z_k$ sont des v.a.r. i.i.d. de même loi que $N_t^{(k)}$ alors $(Z_1 + \cdots + Z_k)/k$ suit la loi géométrique sur $\mathbb{N}^*$ de paramètre $e^{-k\lambda t}$. En particulier,*

$$\mathbb{E}(N_t^{(k)}) = e^{k\lambda t} \quad \textit{et} \quad \mathbb{V}(N_t^{(k)}) = k(1 - e^{-k\lambda t})e^{2k\lambda t}.$$

De plus $e^{-k\lambda t} N_t^{(k)}$ converge en loi quand $t \to \infty$ vers la loi Gamma$(1/k, 1/k)$ *de densité*

$$x \mapsto \frac{(1/k)^{1/k}}{\Gamma(1/k)} x^{1/k-1} e^{-x/k} \mathbb{1}_{\mathbb{R}_+}(x).$$

Il est possible d'établir que le processus $(e^{-k\lambda t} N_t^{(k)})_{t \geqslant 0}$ 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

$$\begin{aligned}\mathbb{E}\Big(\exp\Big(\theta e^{-kt} N_t^{(k)}\Big)\Big) &= \exp\Big(\theta e^{-kt}\Big) e^{-t} \Big(1 - \Big(1 - e^{-kt}\Big) \exp\Big(\theta k e^{-kt}\Big)\Big)^{-1/k} \\ &= \Big(e^{kt} \exp\Big(-\theta k e^{-kt}\Big) - e^{kt} + 1\Big)^{-1/k}\end{aligned}$$

converge, lorsque $t \to \infty$, vers la transformée de Laplace de la loi Gamma$(1/k, 1/k)$:

$$(1 - k\theta)^{-1/k} = \left(\frac{1/k}{1/k - \theta}\right)^{1/k}.$$

□

Recueil de modèles stochastiques – Page 206/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Remarque 21.10 (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 $t \geqslant 0$, 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 $(\mathbb{N}^*)^0 = \varnothing$, 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 $f : \mathbb{R}_+ \to \mathbb{R}$, on a $\mathbb{E}\Big((1/N_t^{(k)}) \sum_{u \in G_t} f(X_{u,t})\Big) = \mathbb{E}(f(X_{U_t,t}))$ 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 apparaître[6].

21.4 Notes et commentaires

Le véritable modèle de taille de fenêtre TCP est plus complexe et incorpore notamment une taille maximale de fenêtre. Il est étudié par Teunis Ott et Johannes Kamperman dans [OK08], voire également les écrits de Philippe Robert [Rob05, Rob10]. Le comportement en temps long est étudié dans l'article [CMP10]. Des aspects graphiques sont explorés dans l'article [GR10] de Carl Graham et Philippe Robert. La version branchante est étudiée par Vincent Bansaye, Jean-François Delmas, Laurence Marsalle, et Viet Chi Tran dans [BDMT11, Clo13]. Le processus de Yule doit son nom à Udny Yule et est étudié dans les livres [Har02, AN04]. Un lien avec des équations d'évolution de transport-fragmentation se trouve dans le livre de Benoît Perthame [Per07]. Une théorie de la fragmentation se trouve dans le livre de Jean Bertoin [Ber06].

Les distances de Wasserstein sont au cœur de la théorie du transport optimal de la mesure, objet du livre de Villani [Vil03b]. 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 :

$$W_p(\mu_1,\mu_2) = \left(\int_0^1 |F_{\mu_1}^{-1}(x) - F_{\mu_2}^{-1}(x)|^p \, dx\right)^{1/p} = \|F_{\mu_1} - F_{\mu_2}\|_p.$$

6. 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.

Recueil de modèles stochastiques – Page 207/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Recueil de modèles stochastiques – Page 208/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Chapitre 22

Modèle du télégraphe

Recueil de modèles stochastiques – Page 209/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Mots-clés. Processus de Poisson ; chaîne de Markov à temps continu ; générateur infinitésimal.

On s'intéresse au mouvement d'une bactérie qui possède de quatre à six flagelles répartis sur sa surface et se déplace en les faisant tourner à la manière d'un tire-bouchon. Quand ils tournent dans le sens anti-horaire, les flagelles forment un faisceau propulsif qui produit un déplacement en quasi-ligne droite. Quand ils tournent dans le sens des aiguilles d'une montre, les flagelles sont décorellés et produisent une rotation de la cellule sans déplacement significatif. La cellule alterne donc les déplacements en ligne droite ou *runs* et les rotations sur place. La durée moyenne d'un run est d'environ une seconde tandis que le temps moyen d'une rotation est de l'ordre du dixième de seconde. Dans la suite, on néglige le temps de rotation et on suppose que le mouvement s'effectue en dimension un. Dans un milieu isotrope, le taux d'apparition d'une rotation au cours d'un run est essentiellement indépendant du temps passé depuis la dernière rotation : la longueur d'un run peut donc être modélisée par une loi exponentielle de paramètre α.

Notons X_t et V_t la position et la vitesse de la bactérie à l'instant t. La vitesse de la bactérie prend ses valeurs dans l'ensemble $\{-1,+1\}$ après un changement d'échelle adéquat. Soit $(N_t)_{t\geqslant 0}$ un processus de Poisson d'intensité α et (V_0, X_0) une variable aléatoire à valeurs dans $E = \{-1,+1\} \times \mathbb{R}$ indépendante de $(N_t)_{t\geqslant 0}$. On définit le processus $((V_t, X_t))_{t\geqslant 0}$ à valeurs dans E en posant, pour tout $t \geqslant 0$

$$V_t = (-1)^{N_t} V_0 \quad \text{et} \quad X_t = X_0 + \int_0^t V_s \, ds.$$

On souhaite savoir comment évolue la répartition des bactéries au cours du temps connaissant la répartition à l'instant initial.

Tout d'abord, le processus des vitesses $(V_t)_{t\geqslant}$ est un processus de Markov sur $\{-1,+1\}$ dont on peut déterminer entièrement la loi. Pour tout $t \geqslant 0$, notons

$$p_+(t) = \mathbb{P}(V_t = +1) \quad \text{et} \quad p_-(t) = \mathbb{P}(V_t = -1).$$

Ces fonctions sont solutions des équations différentielles suivantes :

$$p'_+(t) = \alpha(p_-(t) - p_+(t)) \quad \text{et} \quad p'_-(t) = \alpha(p_+(t) - p_-(t)).$$

Théorème 22.1 (Loi de la vitesse)**.** *Ce processus admet une unique mesure invariante :* $\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_{+1}$. *De plus, pour tout* $t \geqslant 0$, *on connaît la loi de* V_t *en fonction de celle de* V_0 *:*

$$p_+(t) = \frac{1}{2} + \frac{p_+(0) - p_-(0)}{2} e^{-2\alpha t} \quad \textit{et} \quad p_-(t) = \frac{1}{2} - \frac{p_+(0) - p_-(0)}{2} e^{-2\alpha t}.$$

Démonstration. Il suffit de remarquer que

$$(p_+ - p_-)'(t) = -2\alpha(p_+ - p_-)(t) \quad \text{et} \quad (p_+ + p_-)'(t) = 0$$

et de résoudre ces équations différentielles linéaires. □

Nous verrons notamment dans la suite que V et (X, V) héritent de la propriété de Markov du *processus de Poisson* N, ce qui n'est pas le cas du processus X. Dans la section 22.1, on étudie les propriétés markoviennes du processus complet (V, X) héritées de celles du processus de Poisson. Nous faisons enstuite le lien entre ce processus et une équation aux dérivées partielles dite du télégraphe dans la section 22.2.

22.1 Aspects markoviens du processus complet

Nous étudions dans cette section le semi-groupe de Markov et le générateur infinitésimal associé au processus (V, X). L'approche est similaire à celle adoptée pour le processus de croissance-fragmentation dans le chapitre 21 et pour la file d'attente M/M/∞ dans le chapitre 18. Pour tout $t \geqslant 0$ et toute fonction f mesurable bornée de E dans $\mathbb{R}$, on note

$$P_t f(v, x) = \mathbb{E}(f(V_t, X_t)|V_0 = v, X_0 = x) = \mathbb{E}\left[f\left((-1)^{N_t} v, x + v \int_0^t (-1)^{N_s}\, ds\right)\right].$$

Pour $k \in \mathbb{N}$, on note $\mathcal{C}_E^k$ l'ensemble des fonctions $f : E \to \mathbb{R}$ mesurables bornées qui sont de classe $\mathcal{C}^k$ en la deuxième variable à dérivées successives bornées.

Théorème 22.2 (Semi-groupe de Markov)**.** *La famille $(P_t)_{t\geqslant 0}$ vérifie les propriété suivantes :*
- *pour tout $t \geqslant 0$, si $\mathbf{1}$ est la fonction constante égale à 1, alors $P_t\mathbf{1} = \mathbf{1}$;*
- *si f est positive, alors $P_t f$ l'est aussi ;*
- *pour toute fonction $f \in \mathcal{C}_E^k$ et tout $t \geqslant 0$, $P_t f \in \mathcal{C}_E^k$;*
- *pour toute fonction $f \in \mathcal{C}_E^1$, $P_t f$ converge uniformément vers f quand t tend vers 0 ;*
- *pour tous $t, s \geqslant 0$ et toute fonction f mesurable bornée de E dans $\mathbb{R}$,*

$$P_{t+s} f(v, x) = P_t \circ P_s f(v, x).$$

On dit que $(P_t)_{t\geqslant 0}$ est un semi-groupe de Markov.

Démonstration. Les deux premiers points sont évidents. Pour établir le troisième, on écrit

$$\begin{aligned}|P_t f(v, x) - f(v, x)| &\leqslant \mathbb{E}(|f(V_t, X_t) - f(v, x)||V_0 = v, X_0 = x)\\ &\leqslant \mathbb{E}\big(|f(V_t, X_t) - f(v, x)|\mathbf{1}_{\{N_t=0\}}|V_0 = v, X_0 = x\big)\\ &\quad + \mathbb{E}\big(|f(V_t, X_t) - f(v, x)|\mathbf{1}_{\{N_t\geqslant 1\}}|V_0 = v, X_0 = x\big)\end{aligned}$$

Si $N_t = 0$, alors, pour tout $s \in [0, t]$, $V_s = v$ et $X_t = x + tv$. On a donc

$$\mathbb{E}\big(|f(V_t, X_t) - f(v, x)|\mathbf{1}_{\{N_t=0\}}|V_0 = v, X_0 = x\big) = |f(v, x + vt) - f(v, x)| \leqslant \|\partial_x f\|_\infty t.$$

D'autre part,

$$\mathbb{E}\big(|f(V_t, X_t) - f(v, x)|\mathbf{1}_{\{N_t\geqslant 1\}}|V_0 = v, X_0 = x\big) \leqslant 2\|f\|_\infty \mathbb{P}(N_t \geqslant 1) = 2\|f\|_\infty \big(1 - e^{-\alpha t}\big).$$

Démontrons la dernière propriété. Pour tous $s, t \geqslant 0$, on a

$$P_{t+s} f(v, x) = \mathbb{E}\{\mathbb{E}\{f(V_{t+s}, X_{t+s})|X_0, V_0, X_t, V_t\}|X_0 = x, V_0 = v\}.$$

Recueil de modèles stochastiques – Page 210/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Par définition de (V, X) on écrit

$$V_{t+s} = (-1)^{N_{t+s}-N_t} V_t \quad \text{et} \quad X_{t+s} = X_t + V_t \int_0^s (-1)^{N_{t+u}-N_t}\, du.$$

On utilise ensuite le fait que le processus de Poisson est à accroissements indépendants et stationnaires, et, en particulier, que $(N_{t+s} - N_t)_{s \geqslant 0}$ est un processus de Poisson de paramètre α indépendant de $(N_u)_{0 \leqslant u \leqslant t}$ pour obtenir que

$$\mathbb{E}\Big(f\Big((-1)^{N_{t+s}-N_t} V_t, X_t + V_t \int_t^{t+s} (-1)^{N_u - N_t}\, du \Big) | X_0, V_0, X_t, V_t \Big) = P_s f(V_t, X_t).$$

On retrouve alors la relation de semi-groupe $P_{t+s} f(v, x) = P_t \circ P_s f(v, x)$. □

Théorème 22.3 (Générateur infinitésimal)**.** *Pour tout $t \geqslant 0$ et toute fonction $f \in \mathcal{C}_E^1$,*

$$\left\| \frac{P_{t+h} f - P_t f}{h} - A P_t f \right\|_\infty \xrightarrow[h \to 0]{} 0,$$

où l'opérateur A est défini par

$$Af(v, x) = \alpha(f(-v, x) - f(v, x)) + v \partial_x f(v, x),$$

pour toute fonction $f \in \mathcal{C}_E^1$. En particulier, pour tout $t \geqslant 0$ et pour toute fonction $f \in \mathcal{C}_E^1$,

$$\partial_t P_t f(v, x) = A P_t f(v, x) = P_t(Af)(v, x). \tag{22.1}$$

Démonstration. Démontrons dans un premier temps le résultat pour $t = 0$. Soit $f \in \mathcal{C}_E^1$. Alors

$$\begin{aligned} P_h f(v, x) = &\mathbb{E}\big(f(V_h, X_h) \mathbf{1}_{\{N_h = 0\}} | X_0 = x, V_0 = v \big) \\ &+ \mathbb{E}\big(f(V_h, X_h) \mathbf{1}_{\{N_h = 1\}} | X_0 = x, V_0 = v \big) \\ &+ \mathbb{E}\big(f(V_h, X_h) \mathbf{1}_{\{N_h \geqslant 2\}} | X_0 = x, V_0 = v \big). \end{aligned}$$

On étudie ces trois espérances séparément. La dernière est bornée en valeur absolue par

$$\|f\|_\infty \mathbb{P}(N_h \geqslant 2) = \|f\|_\infty (1 - e^{-\alpha h}(1 + \alpha h)) = O(h^2).$$

La première vaut

$$\begin{aligned} \mathbb{E}\big(f(v, x + vh) \mathbf{1}_{\{N_h = 0\}} \big) &= f(v, x + vh) e^{-\alpha h} \\ &= f(v, x) + h(v \partial_x f(v, x) - \alpha f(v, x)) + o(h). \end{aligned}$$

Enfin, la deuxième espérance s'écrit

$$\mathbb{E}\Big[f\Big(-v, x + v \int_0^h (-1)^{N_u}\, du \Big) \mathbf{1}_{\{N_h = 1\}} \Big] = \alpha h e^{-\alpha h} (f(-v, x) + o(h))$$

puisque le terme intégral est borné (en valeur absolue) par h. On a ainsi obtenu

$$P_h f(v, x) = f(v, x) + h A f(v, x) + o(h),$$

ce qui achève le cas $t = 0$. Pour étendre le résultat à tout temps $t \geqslant 0$, on utilise la propriété de semi-groupe et la première partie de la preuve :

$$\frac{P_{t+h} f - P_t f}{h} - A P_t f = \Big(\frac{P_h - P_0}{h} - A \Big) P_t f \xrightarrow[h \to 0]{\|\cdot\|_\infty} 0,$$

puisque que $P_t f \in \mathcal{C}_E^1$. De même, l'écriture alternative

$$\frac{P_{t+h} f - P_t f}{h} = P_t \Big(\frac{P_h - P_0}{h} \Big) f;$$

fournit la relation $\partial_t P_t f = P_t A f$. □

Recueil de modèles stochastiques – Page 211/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

22.2 Lien entre équation et processus

Soit ϕ une fonction de classe $\mathcal{C}^1$ sur $\mathbb{R}$. Définissons u^+ et u^- sur $\mathbb{R}^+ \times \mathbb{R}$ par

$$\begin{cases} u^+(t,x) := P_t\phi(1,x) = \mathbb{E}(\phi(X_t)|V_0 = 1, X_0 = x), \\ u^-(t,x) := P_t\phi(-1,x) = \mathbb{E}(\phi(X_t)|V_0 = -1, X_0 = x), \end{cases}$$

où l'on a commis l'abus de notation suivant : on note encore ϕ l'application $(v,x) \mapsto \phi(x)$. Par définition de u^+ et u^-, la propriété (22.1) assure que

$$\begin{aligned} \partial_t u^+(t,x) &= \partial_t P_t\phi(1,x) = AP_t\phi(1,x) \\ &= \alpha(P_t\phi(-1,x) - P_t\phi(1,x)) + \partial_x P_t\phi(1,x) \\ &= \alpha(u^-(t,x) - u^+(t,x)) + \partial_x u^+(t,x). \end{aligned}$$

On peut faire de même pour u^- et en déduire que le couple (u^+, u^-) est solution du système d'équations :

$$\begin{cases} \partial_t u^+(t,x) = \alpha(u^-(t,x) - u^+(t,x)) + \partial_x u^+(t,x) \\ \partial_t u^-(t,x) = \alpha(u^+(t,x) - u^-(t,x)) - \partial_x u^-(t,x). \end{cases} \tag{22.2}$$

En posant $u = (u^+ + u^-)/2$ et $\tilde{u} = (u^+ - u^-)/2$, le couple $(u, \tilde{u})$ est solution du système :

$$\begin{cases} \partial_t u(t,x) = w\partial_x \tilde{u}(t,x), \\ \partial_t \tilde{u}(t,x) = -2\alpha\tilde{u}(t,x) + \partial_x u(t,x). \end{cases}$$

Dériver la première équation par rapport à t et la seconde par rapport à x permet de supprimer le terme $\partial_t\partial_x\tilde{u}(t,x)$ pour obtenir

$$\partial^2_{tt} u(t,x) = -2\alpha w\partial_x\tilde{u}(t,x) + w^2\partial^2_{xx} u(t,x).$$

Il reste alors à remplacer $\partial_x\tilde{u}(t,x)$ par $\partial_t u(t,x)$ pour obtenir que

$$\partial^2_{tt} u(t,x) - w^2\partial^2_{xx} u(t,x) + 2\alpha\partial_t u(t,x) = 0.$$

On a montré le résultat suivant.

Théorème 22.4 (Représentation probabiliste de la solution d'une ÉDP)**.** *La fonction u définie par $u(t,x) = \mathbb{E}(\phi(x + X_t))$ pour tout $t \geqslant 0$ et $x \in \mathbb{R}$ où $X_0 = 0$ et V_0 suit la loi $\frac{1}{2}\delta_{-w} + \frac{1}{2}\delta_w$ est solution de l'*équation du télégraphe *:*

$$\begin{cases} \partial^2_{tt} u(t,x) - \partial^2_{xx} u(t,x) + 2\alpha\partial_t u(t,x) = 0 & \textit{pour tous } t \geqslant 0,\ x \in \mathbb{R}, \\ u(0,x) = \phi(x) & \textit{pour tout } x \in \mathbb{R}, \\ \partial_t u(0,x) = 0 & \textit{pour tout } x \in \mathbb{R}. \end{cases}$$

L'équation aux dérivées partielles du théorème 22.4 est une superposition de deux équations classiques : l'équation des ondes ou des cordes vibrantes, et l'équation de la chaleur. La première décrit la propagation d'un signal sans amortissement tandis que la seconde traduit un phénomène de diffusion.

22.3 Modification ergodique

Dans cette section, on modifie légèrement la dynamique du processus (V, X). La vitesse est toujours à valeurs dans $\{-1, +1\}$ et X l'intégrale de V. Toutefois, on complique légèrement le mécanisme de saut de V. Pour simplifier, on construit par récurrence les positions aux instants de saut de V et on complète ensuite par interpolation.

Recueil de modèles stochastiques – Page 212/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

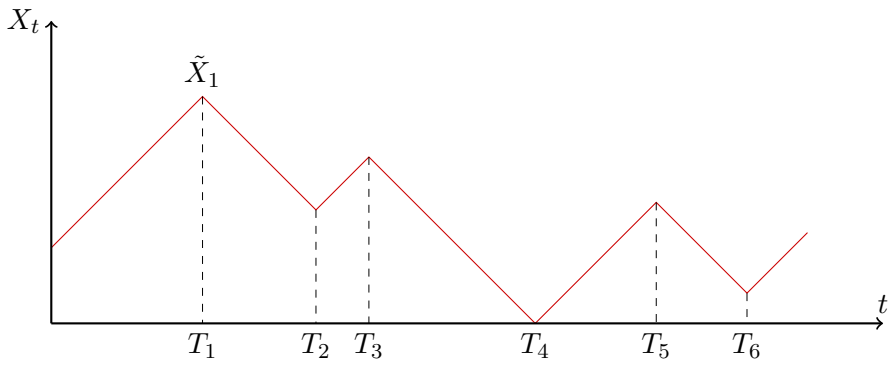


FIGURE 22.1 – Une trajectoire.

Soit $(E_n)_{n\geqslant 1}$ une suite de variables aléatoires indépendantes de loi exponentielle de paramètre 1 indépendante de $(\tilde{V}_0, \tilde{X}_0) \in \{-1, +1\} \times \mathbb{R}^+$. On note $T_0 = 0$. On définit par récurrence les quantités suivantes :

$$T_n = T_{n-1} + \begin{cases} E_n/b & \text{si } \tilde{V}_{n-1} = 1, \\ (E_n/a) \wedge \tilde{X}_{n-1} & \text{si } \tilde{V}_{n-1} = -1, \end{cases}$$

puis

$$\tilde{V}_n = -\tilde{V}_{n-1} \quad \text{et} \quad \tilde{X}_n = \tilde{X}_{n-1} + \tilde{V}_{n-1}(T_n - T_{n-1}).$$

On construit alors le processus (X, V) en posant $V = \tilde{V}$ et X comme interpolation linéaire de $\tilde{X}$. Pour résumer, le taux de saut de V est donné par

$$b\mathbf{1}_{\{v=1\}} + \left(a + \frac{1}{\mathbf{1}_{\{x=0\}}}\right)\mathbf{1}_{\{v=-1\}}.$$

Une *excursion* de (V, X) est une trajectoire issue de $(+1, 0)$ arrêtée à l'instant

$$S = \inf\{t > 0 \,:\, X_t = 0\}.$$

On appelle ψ la transformée de Laplace de S :

$$\psi \;:\; \lambda \in \mathbb{R} \mapsto \psi(\lambda) = \mathbb{E}_{(0,1)}\left(e^{\lambda S}\right).$$

Lemme 22.5 (Longueur d'une excursion)**.** *Le domaine de définition de ψ est $]-\infty, \lambda_c]$ où*

$$\lambda_c = \frac{a+b}{2} - \sqrt{ab} = \frac{(\sqrt{b} - \sqrt{a})^2}{2}.$$

De plus, si $\lambda \leqslant \lambda_c$, on a

$$\psi(\lambda) = \frac{a + b - 2\lambda - \sqrt{(a+b-2\lambda)^2 - 4ab}}{2a}.$$

En particulier, $\psi(\lambda_c) = \sqrt{b/a}$ et $\mathbb{E}_{(1,0)}(S) = 2/(b-a)$.

Démonstration. Le premier temps de saut $T_1 = E$ pour V lorsque $(V_0, X_0) = (1, 0)$ suit la loi exponentielle de paramètre b. Si V ne saute pas entre E et $2E$ alors $S = 2E$. Dans le cas contraire, (V, X) fait une excursion au dessus de X_{T_2} qui a même loi que X. Après cette excursion, (V, X) se trouve en $(-1, X_{T_2})$. Encore une fois, de deux choses l'une : X atteint 0 avant que V ne saute ou non (dans ce cas, une autre excursion se greffe). La figure 22.2

Recueil de modèles stochastiques – Page 213/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

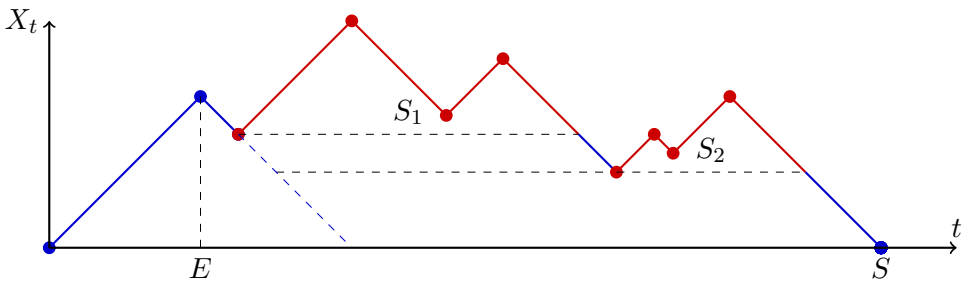


Figure 22.2 – Décomposition d'une excursion. Les points signalent les sauts de la vitesse.

Recueil de modèles stochastiques – Page 214/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

illustre ce découpage. Le nombre N d'excursions incluses dans la descente depuis E suit la loi $\mathcal{P}(aN)$. On peut donc écrire

$$S \overset{\text{loi}}{=} 2E + \sum_{k=1}^{N} S_k,$$

où E, N et $(S_k)_{k\geqslant 1}$ sont indépendantes. Ainsi donc, on a

$$\begin{aligned}\psi(\lambda) &= \mathbb{E}\Big(\mathbb{E}\Big(e^{2\lambda E+\lambda\sum_{k=1}^{N} S_k}|E,N\Big)\Big) = \Big(e^{2\lambda E}\mathbb{E}\big(\psi(\lambda)^N|E\big)\Big)\\ &= \mathbb{E}\Big(e^{2\lambda E}e^{aE(\psi(\lambda)-1)}\Big) = \frac{b}{b+a-2\lambda-a\psi(\lambda)}\end{aligned}$$

pour tout λ dans le domaine de définition de ψ (qui contient $]-\infty,0]$), d'où la formule. □

Plus de précision est possible sur la structure d'une excursion issue de $(+1,0)$. Ceci va permettre d'exprimer la quantité suivante en termes de variables aléatoires classiques :

$$I = \int_0^S e^{\lambda X_s} g(V_s)\,ds$$

où g est une fonction sur $\{-1,+1\}$. Le premier temps de saut E suit la loi exponentielle de paramètre b. Conditionnellement, à l'événement $\{E=x\}$, le nombre N d'excursions « greffées » sur le retour à 0 est distribué selon la loi de Poisson de paramètre a. Conditionnellement à $\{E=x, N=k\}$ les k hauteurs auxquelles elles ont lieu sont distribuées comme $(xU_{(k,k)}, xU_{(k,k-1)},\dots,xU_{(k,1)})$ où les variables aléatoires $(U_i)_{i\geqslant 1}$ sont indépendantes et de même loi uniforme sur $[0,1]$ et $(U_{(1,k)},\dots,U_{(k,k)})$ représente l'échantillon réordonné (dans l'ordre croissant). Soient $\Big((V_t^{(i)}, X_t^{(i)})_{0\leqslant t\leqslant S^{(i)}}\Big)_{i\geqslant 1}$ des trajectoires indépendantes sur une excursion issue de $(+1,0)$. On a alors

$$I = (g(1)+g(-1))\int_0^E e^{\lambda y}\,dy + \sum_{i=1}^{N}\int_0^{S_i} e^{\lambda(EU_{(i,N)}+X_s^{(i)})} g(V_s^{(i)})\,ds.$$

On a alors

$$\mathbb{E}\Big(\int_0^{S_i} e^{\lambda(EU_{(i,N)}+X_s^{(i)})} g(V_s^{(i)})\,ds|E,N,(U_i)_{i\geqslant 1}\Big) = \mathbb{E}(I)e^{\lambda EU_{(i,N)}}.$$

On en déduit la relation suivante

$$\begin{aligned}\mathbb{E}(I) &= \mathbb{E}\Big(\frac{g(1)+g(-1)}{\lambda}(e^{\lambda E}-1) + \mathbb{E}(I)\sum_{i=1}^{N} e^{\lambda EU_{(i,N)}}\Big)\\ &= \frac{g(1)+g(-1)}{b-\lambda} + \mathbb{E}(I)\mathbb{E}\Big(\sum_{i=1}^{N} e^{\lambda EU_i}\Big).\end{aligned}$$

Enfin, par indépendance des variables aléatoires considérées,

$$\begin{aligned}\mathbb{E}\left(\sum_{i=1}^{N} e^{\lambda E U_i}|E\right) &= \mathbb{E}\Big(N e^{\lambda E U_1}|E\Big)\\ &= \mathbb{E}\left(\frac{N}{\lambda E}(e^{\lambda E}-1)|E\right)\\ &= \frac{a}{\lambda}(e^{\lambda E}-1).\end{aligned}$$

Ainsi, pour tout $\lambda < b-a$, on obtient

$$\mathbb{E}(I) = \frac{g(1)+g(-1)}{b-a-\lambda}.$$

Le théorème ergodique montre que si ν est la mesure invariante de (V,X) alors

$$\int_{\{-1,+1\}\times\mathbb{R}^+} e^{\lambda x} g(v)\nu(dv,dx) = \frac{\mathbb{E}(I)}{\mathbb{E}(S)} = \frac{g(1)+g(-1)}{2}\frac{b-a}{b-a-\lambda}.$$

Ainsi, la mesure invariante ν est la loi $(1/2)(\delta_{-1}+\delta_{+1})\otimes \mathrm{Exp}(b-a)$.

22.4 Notes et commentaires

De nombreux modèles mathématiques pour ces problèmes de déplacement de bactéries se trouvent par exemple dans les articles de Hans Otmer, Steven Dunbar, et Wolfgang Alt [ODA88], et de Radek Erban et Hang Othmer [EO05]. Le théorème 22.4 est dû à Marc Kac [Kac74]. L'équation aux dérivées partielles du théorème 22.4 est appelée équation du télégraphe car elle modélise également l'évolution de la tension dans un cable coaxial. On trouvera dans l'article de Samuel Herman et Pierre Vallois [HV10] le lien avec les marches aléatoires persistantes.

Recueil de modèles stochastiques – Page 215/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Chapitre 23

Problème de Dirichlet

Recueil de modèles stochastiques – Page 217/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Mots-clés. Chaîne de Markov ; martingale ; méthode de Monte-Carlo ; mouvement brownien.

Ce chapitre est consacré au problème de Dirichlet continu, en liaison avec le mouvement brownien. La version discrète est abordée dans le chapitre 2 consacré aux marches aléatoires. La motivation est l'étude de la répartition de la température dans un solide. Considérons donc un solide homogène identifié à l'adhérence $\overline{D}$ d'un ouvert connexe borné D de $\mathbb{R}^d$ dont on fixe la température de surface non nécessairement constante égale à une fonction b. La température à l'intérieur du solide converge vers un équilibre thermique que l'on souhaite exprimer en fonction de D et b. Intuitivement, la température en un point intérieur du solide est égale à la moyenne des températures prises sur toute boule centrée en ce point et incluse dans le solide : la température ϕ vérifie la formule de la moyenne. En analyse, on dit que ϕ est *harmonique* sur D si elle est localement intégrable et vérifie la *formule de la moyenne* :

$$\forall x \in D,\ r < \mathrm{dist}(x, \partial D), \quad \phi(x) = \frac{1}{|B(0,r)|} \int_{B(0,r)} \phi(x+y)\, dy.$$

La température devrait être continue sur l'ensemble du solide si b l'est au bord ∂D de D. On peut donc formuler ainsi le *problème de Dirichlet* : une fonction ϕ définie sur $\overline{D}$ à valeurs dans $\mathbb{R}$ est solution du problème de Dirichlet :

1. pour tout point x de ∂D, $\phi(x) = b(x)$,
2. ϕ est continue sur $\overline{D}$,
3. ϕ harmonique sur D.

Il existe de nombreux théorèmes étudiant l'existence et/ou l'unicité de la solution du problème de Dirichlet mais bien sûr la solution n'est en général pas explicite. Dans certains cas très particuliers, il est cependant possible de donner une représentation intégrale relativement explicite de la solution. Dans les autres cas, des méthodes numériques, déterministes ou probabilistes, permettent de résoudre le problème de manière approchée.

Théorème 23.1 (Fonctions harmoniques)**.** *Pour toute fonction $\phi : D \to \mathbb{R}$ les trois propriétés suivantes sont équivalente :*

1. *Moyenne sur les boules : ϕ est harmonique ;*
2. *Moyenne sur les sphères : ϕ est localement intégrable sur D et vérifie*

$$\forall x \in D,\ r < \mathrm{dist}(x, \partial D), \quad \phi(x) = \int_{S(0,r)} \phi(x+y)\, \sigma_r(dy),$$

où σ_r est la mesure de probabilité uniforme sur la sphère centrée de rayon r ;

3. Laplacien nul : ϕ est de classe $\mathcal{C}^\infty$ sur D et son laplacien est nul sur D.

De plus, si ϕ est harmonique sur D alors toutes ses dérivées partielles sont harmoniques sur D. Enfin, si $|\phi|$ est bornée sur $\overline{D}$, alors il en va de même pour toutes les dérivées partielles ϕ.

Le théorème 23.1 permet de reformuler le problème de Dirichlet en termes d'équations aux dérivées partielles (ÉDP).

Corollaire 23.2 (Reformulation du problème de Dirichlet en terme d'ÉDP)**.** *Une fonction ϕ est solution du problème de Dirichlet sur D avec conditions aux bords b, si et seulement si, elle est de classe $\mathcal{C}^\infty$ sur D et continue sur $\overline{D}$ et vérifie*

$$\begin{cases} \Delta\phi(x) = 0 & \textit{pour tout } x \in D, \\ \phi(x) = b(x) & \textit{pour tout } x \in \partial D. \end{cases}$$

Preuve du théorème 23.1. Montrons que la seconde propriété implique la première. Supposons sans perte de généralité que $x = 0$. Grâce à un changement de variables polaires, on a

$$\begin{aligned} \frac{1}{|B(0,r)|}\int_{B(0,r)} h(y)\,dy &= \frac{1}{r^d|B(0,1)|}\int_0^r \int_{\partial B(0,s)} h(y)\,\sigma_s(dy)|\partial B(0,s)|\,ds \\ &= \frac{d}{r^d}\int_0^r \int_{\partial B(0,s)} h(y)\,\sigma_s(dy)s^{d-1}\,ds. \end{aligned}$$

Ainsi, si h vérifie la formule de la moyenne sur les sphères, alors h vérifie la formule de la moyenne sur les boules. Réciproquement, si h est harmonique, l'expression ci-dessus donne

$$r^d h(0) = d\int_0^r \int_{\partial B(0,s)} h(y)\,\sigma_s(dy)s^{d-1}\,ds.$$

Une dérivation par rapport à r donne la formule de la moyenne sur les sphères.

Montrons que les deux dernières propriétés sont équivalentes. Soit $x \in D$, $\mathcal{V}$ un voisinage de x inclus dans D et $\varepsilon < \mathrm{dist}(\mathcal{V}, \partial D)$. La fonction Ψ_ε donnée par $x \mapsto c(\varepsilon)\exp((|x|^2 - \varepsilon^2)^{-1})\mathbf{1}_{\{|x|<\varepsilon\}}$ est de classe $\mathcal{C}^\infty$, positive, d'intégrale 1 pour $c(\varepsilon)$ bien choisi, à support dans la boule $\overline{B(0,\varepsilon)}$ et radiale. La fonction $\phi * \Psi_\varepsilon$ est donc de classe $\mathcal{C}^\infty$ sur $\mathcal{V}$. De plus, pour tout $y \in \mathcal{V}$, la formule de la moyenne sur les sphères assure que

$$\begin{aligned} \phi * \Psi_\varepsilon(y) &= \int_{B(0,\varepsilon)} \phi(y+u)\Psi_\varepsilon(u)du \\ &= \int_0^\varepsilon \left(\int_{S(0,s)} \phi(y+sz)\,\sigma_s(dz)\right)|S(0,s)|\Psi_\varepsilon(s)ds \\ &= \phi(y)\int_{B(0,\varepsilon)} \Phi_\varepsilon(z)\,dz, \end{aligned}$$

puisque Ψ_ε est radiale. Elle est de plus d'intégrale 1 : ϕ et $\phi * \Psi_\varepsilon$ coïncident sur $\mathcal{V}$ et ϕ est ainsi de classe $\mathcal{C}^\infty$ sur $\mathcal{V}$. Enfin, pour $\|y\|$ assez petit, on a

$$\phi(x+y) = \phi(x) + \nabla\phi(x)\cdot y + \frac{1}{2}(\mathrm{Hess}(\phi)(x)y)\cdot y + o(\|y\|^2).$$

En intégrant cette relation sur la sphère centrée de rayon r contre la mesure uniforme et en faisant tendre r vers 0, on obtient que $\Delta\phi(x) = 0$, tandis que la réciproque s'obtient en intégrant la relation sur une petite boule.

Il reste à établir les propriétés de régularité automatiques des dérivées. La propagation de l'harmonicité s'obtient par récurrence en dérivant la formule de la moyenne. De plus, pour tout multi-indice $\alpha \in \mathbb{N}^d$, il existe existe $K(\varepsilon,\alpha)$ tel que, pour tout $y \in \mathcal{V}$,

$$|\partial_\alpha\phi(y)| = |\phi * \partial_\alpha\Psi_\varepsilon| \leqslant K(\varepsilon,\alpha)\sup_{z\in\overline{D}}|\phi(z)|.$$

Un argument de compacité assure le résultat. □

Recueil de modèles stochastiques – Page 218/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

23.1 Interprétation probabiliste

La théorie des probabilités fournit les outils nécessaires à l'étude du problème de Dirichlet sous sa forme ÉDP. Elle fournit notamment une fonction qui pourrait être solution. Soit $(B_t)_{t\geqslant 0}$ est un *mouvement brownien* sur $\mathbb{R}^d$ et $\tau = \inf\{t > 0 : B_t \notin D\}$ est le temps de sortie de D pour $(B_t)_{t\geqslant 0}$. On notera $\mathbb{P}_x$ la loi de $(B_t)_{t\geqslant 0}$ sachant que $B_0 = x$ et $\mathbb{E}_x$ l'espérance par rapport à cette loi. Le mouvement brownien est un processus de Markov fort et la formule d'Itô assure que si f est une fonction de classe $\mathcal{C}^2$ de $\mathbb{R}^d$ dans $\mathbb{R}$ à dérivées bornées, alors

$$f(B_t) = f(B_0) + \int_0^t \frac{1}{2}\Delta f(B_s)\,ds + \int_0^t \nabla f(B_s)\,dB_s.$$

L'intégrale stochastique ci-dessus est une martingale. La clé du lien entre probabilités et analyse est contenue dans la remarque suivante : $(f(B_t))_{t\geqslant 0}$ est une martingale si et seulement si (ou presque) f est une fonction harmonique.

Théorème 23.3 (Harmonicité de la représentation probabiliste)**.** *La fonction v définie sur D par $v(x) = \mathbb{E}_x[b(B_\tau)]$ est harmonique sur D.*

Démonstration. Soit $x \in D$ et $r < \operatorname{dist}(x, \partial D)$. Notons $\tau_r = \inf\{t \geqslant 0 : B_t \notin B(x,r)\}$ le temps de sortie de la boule centrée en x de rayon r pour le mouvement brownien issu de x. Voir la figure 23.1. La propriété de Markov forte de $(B_t)_{t\geqslant 0}$ assure alors que

$$v(x) = \mathbb{E}_x[\mathbb{E}_x(b(B_\tau)|\mathcal{F}_{\tau_r})] = \mathbb{E}_x\left[\mathbb{E}_{X_{\tau_r}}(b(B_\tau))\right] = \mathbb{E}_x(v(B_{\tau_r})).$$

Enfin, comme la loi du mouvement brownien est invariante par rotation, la v.a. B_{τ_r} est distribuée selon la loi uniforme sur $S(0,r)$. Cette propriété assure que la fonction v vérifie la formule de la moyenne sur les sphères. □

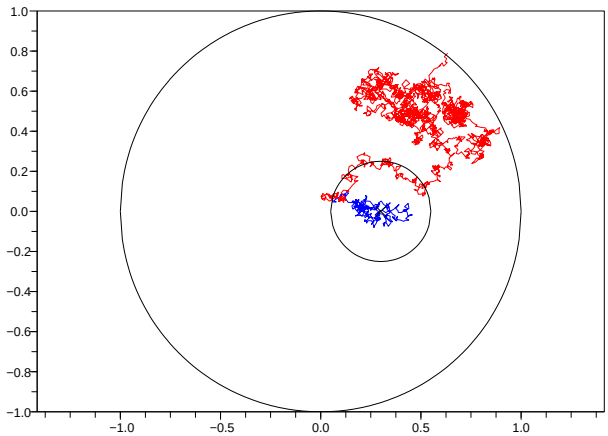


FIGURE 23.1 – Découpage d'une trajectoire sortant de D après être sortie de $B(x,r)$.

Remarque 23.4 (Noyau de Poisson)**.** *La représentation probabiliste ci-dessus montre que la solution du problème de Dirichlet en un point x de D s'écrit comme la moyenne des valeurs aux bords pondérées par la loi du lieu de sortie d'un mouvement brownien issu de x. La densité de cette mesure est appelée* noyau de Poisson *de D.*

L'unicité de la solution du problème de Dirichlet se déduit de l'interprétation probabiliste.

Recueil de modèles stochastiques – Page 219/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Théorème 23.5 (Unicité de la solution)**.** *Soit ϕ une solution du problème de Dirichlet sur D de condition aux bords b. Alors, cette solution est unique et s'écrit*

$$\forall x \in D, \quad \phi(x) = \mathbb{E}_x[b(B_\tau)].$$

Démonstration. La formule d'Itô fournit la décomposition suivante :

$$\phi(B_{t\wedge\tau}) = \phi(x) + \int_0^{t\wedge\tau} \frac{1}{2}\Delta\phi(B_s)\,ds + M_{t\wedge\tau}$$

où $(M_{t\wedge\tau})_{t\geqslant 0}$ est une martingale bornée d'après les propriétés de régularités automatiques des fonctions harmoniques (théorème 23.1). En effet, ϕ est une fonction harmonique sur D et bornée sur $\overline{D}$. Il en est alors de même pour ses dérivées. En prenant l'espérance et en faisant tendre t vers l'infini, on obtient le résultat annoncé. □

Remarque 23.6 (Preuve de l'unicité par principe du maximum)**.** *Une fonction h harmonique sur un domaine D atteint nécessairement maximum sur un compact K en un point du bord de K. Ce* principe du maximum *fournit l'unicité de la solution du problème de Dirichlet : la différence de deux solutions est une fonction harmonique nulle sur ∂D et donc nulle sur D.*

Il est délicat d'établir la continuité sur $\overline{D}$ de la fonction v donnée par le théorème 23.3. Cela fait intervenir les propriétés géométriques du domaine D.

23.2 Cas de la boule unité

Il existe très peu de domaines D pour lesquels une expression explicite de la solution du problème de Dirichlet est disponible. Cette section est consacrée au cas des boules euclidiennes.

Théorème 23.7 (Représentation intégrale de la solution dans une boule)**.** *La solution du problème de Dirichlet sur la boule centrée de rayon r dans $\mathbb{R}^d$ de condition aux bords b s'écrit*

$$\phi(x) = \int_{S(0,r)} b(y)P_r(x,y)\,\sigma_r(dy), \quad \text{où} \quad P_r(x,y) = \frac{r^{d-2}(r^2 - |x|^2)}{|y-x|^d}.$$

Ici encore $\phi(x)$ s'écrit comme la moyenne de b sur ∂D contre la densité $P(x,y)$ par rapport à la mesure uniforme sur $S(0,r)$. Cette mesure donne plus de poids aux points de la sphère qui sont proches de x, ce qui correspond bien à l'intuition en termes de température du solide.

Le problème de Dirichlet sur une boule de $\mathbb{R}^d$ se ramène ainsi au calcul d'une intégrale sur la sphère de dimension $d-1$. Lorsque $d \gg 1$ les méthodes de calcul approché déterministes deviennent très difficiles à mettre en place car elles demandent de discrétiser la sphère. La méthode de Monte-Carlo reste elle très simple à utiliser grâce au résultat suivant.

Théorème 23.8 (Simulation de la loi uniforme sur la sphère)**.** *Si X est un vecteur aléatoire dans $\mathbb{R}^d$ de loi $\mathcal{N}_d(0, I_d)$, le vecteur aléatoire $Y = X/|X|$ suit la loi uniforme sur $S(0,1)$.*

Démonstration. La v.a. Y est définie p.s. puisque l'ensemble $\{X = 0\}$ est de mesure nulle. De plus, la loi de X est invariante par toute rotation : si O est une matrice orthogonale alors OX a même loi que X. Il en est donc de même pour Y. Or la seule loi de probabilité sur la sphère unité munie de sa tribu borélienne qui soit invariante par rotation est la loi uniforme. □

Remarque 23.9 (Algorithme de Box-Muller)**.** *Si U est une v.a.r. de loi uniforme sur $[0,1]$ et V une v.a.r. de loi exponentielle de paramètre 1/2 sont indépendantes alors*

$$X = \sqrt{V}\cos(2\pi U) \quad \text{et} \quad Y = \sqrt{V}\sin(2\pi U)$$

sont indépendantes et de même loi gaussienne centrée réduite. Cette propriété a longtemps été utilisée par les logiciels de calcul scientifique avant d'être délaissé pour des procédures plus efficaces comme l'algorithme du Ziggurat de Marsaglia basé sur une méthode de polygonalisation.

Recueil de modèles stochastiques – Page 220/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

23.3 Approches numériques

Dans le cas d'un domaine général, on ne dispose pas d'une expression explicite du noyau de Poisson. Il faut donc envisager des méthodes de résolutions approchées. Pour cela, on discrétise le domaine D. Soit $h > 0$ le pas de discrétisation. On définit la grille D_h comme l'intersection de D et du réseau $h\mathbb{Z}^d$:

$$D_h := D \cap h\mathbb{Z}^d = \Big\{x \in D,\ \exists (i_1, \dots, i_d) \in \mathbb{Z}^d,\ x = (hi_1, \dots, hi_d)\Big\}.$$

On dit que $x, y \in h\mathbb{Z}^d$ sont voisins, et on note $x \sim y$, si $|x - y| = h$. La frontière de D_h s'écrit

$$\partial_h D_h = \{y \in D_h^c \,:\, \exists x \in D_h,\ x \sim y\}.$$

L'opérateur laplacien est lui remplacé par l'opérateur aux différences :

$$\Delta_h f(x) = \frac{1}{2d} \sum_{y \sim x} (f(y) - f(x)).$$

On dit que ϕ_h est h-harmonique sur D_h si $\Delta_h \phi_h = 0$ sur D_h. Il convient également de définir une fonction b_h sur $\partial_h D_h$ qui soit une approximation raisonnable de b définie sur ∂D. Le problème de Dirichlet discret est alors le suivant : trouver ϕ_h sur $D_h \cup \partial_h D_h$ telle que

$$\begin{cases} \Delta_h \phi_h(x) = 0 & \text{pour tout } x \in D_h, \\ \phi_h(x) = b_h(x) & \text{pour tout } x \in \partial_h D_h. \end{cases}$$

C'est exactement le problème de Dirichlet discret étudié dans le chapitre 2, qui possède une solution probabiliste dans le même esprit que celle du problème de Dirichlet continu, avec la marche aléatoire simple en lieu et place du mouvement brownien. Cette substitution est justifiée par la remarque suivante. Si f est une fonction de classe $\mathcal{C}^2$ de $\mathbb{R}^d$ dans $\mathbb{R}$, alors

$$\Delta_h f(x) = \frac{h^2}{2d} \Delta f(x) + o(h^2).$$

Ainsi, quand h est petit, Δ_h apparaît comme une approximation du laplacien. D'autre part, dire que $\Delta_h \phi_h(x) = 0$ revient à dire que $\phi_h(x)$ est la moyenne (avec la mesure uniforme) des valeurs de ϕ_h en les $2d$ voisins de x, ce qui est un analogue de la propriété de la moyenne.

Remarque 23.10 (Approche déterministe matricielle versus approche probabiliste)**.** *Puisque D_h est un ensemble fini, résoudre le problème de Dirichlet discret revient à résoudre un système linéaire de* $\mathrm{card}(D_h)$ *équations à* $\mathrm{card}(D_h)$ *inconnues. On peut montrer que ce système possède une unique solution. Sa résolution est simplifiée par le fait la matrice associée au système est très creuse : au plus $2d + 1$ inconnues sont impliquées dans chaque équation. Cette méthode directe permet, au terme des calculs, de déterminer simultanément la valeur de la solution ϕ_h en tout point de la grille D_h. Cette méthode est payante lorsque l'on veut déterminer la température du corps en tous les points du solide.*

23.4 Demi-plan supérieur : du discret au continu

Le but de cette section est d'illustrer dans un cas particulier la convergence de la solution du problème de Dirichlet discret vers la solution du problème de Dirichlet continu formulé dans le corollaire 23.2. Même s'il n'est pas borné, le demi-supérieur dans $\mathbb{R}^2$ possède un noyau de Poisson explicite. Il est également possible de déterminer la loi du lieu sortie de la marche aléatoire sur le réseau de $h\mathbb{Z}^2$. Soit N un entier. Le pas de discrétisation est choisi de la forme $h = 1/N$. On note $(X_n^N, Y_n^N)_n$ la marche aléatoire sur $h\mathbb{Z}^2$ issue de $(X_0^N, Y_0^N) = (0, 1)$ telle

Recueil de modèles stochastiques – Page 221/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

que les suites de variables aléatoires $(X_{n+1}^N - X_n^N)$ et $(Y_{n+1}^N - Y_n^N)$ soient indépendantes entre elles, formées de variables indépendantes de même loi :

$$\mathbb{P}((X_{n+1}^N - X_n^N) = -1/N) = \mathbb{P}((X_{n+1}^N - X_n^N) = 1/N) = 1/2,$$
$$\mathbb{P}((Y_{n+1}^N - Y_n^N) = -1/N) = \mathbb{P}((Y_{n+1}^N - X_n^Y) = 1/N) = 1/2.$$

À chaque pas, la marche choisit au hasard entre les 4 points à distance $\sqrt{2}/N$ (nord-ouest, nord-est, sud-ouest et sud-est). Cette modification légère de dynamique de la marche aléatoire permet d'assurer que $(X_n^N)_n$ et $(Y_n^N)_n$ sont deux marches aléatoires simples indépendantes, ce qui simplifiera le raisonnement par la suite.

On s'intéresse à l'instant et à l'abscisse de sortie du demi-plan supérieur pour la marche aléatoire ainsi définie. L'instant de sortie est la variable aléatoire T^N définie par :

$$T^N = \inf\{k \geqslant 0 \,:\, Y_k^N = 0\}.$$

L'abscisse de sortie U^N est l'abscisse de la marche aléatoire à l'instant de sortie T^N :

$$U^N = X_{T^N}^N.$$

Théorème 23.11 (Convergence en loi du lieu de sortie)**.** *La suite $(U^N)_{N\geqslant 1}$ converge en loi, quand N tend vers l'infini, vers la loi de Cauchy de densité :*

$$\frac{1}{\pi}\frac{1}{1+x^2}.$$

Démonstration. Notons S_k le temps d'atteinte de 0 pour la marche aléatoire simple $(Z_n)_{n\geqslant 0}$ sur $\mathbb{Z}$ issue de 1 et G_k sa fonction génératrice du temps d'atteinte. La propriété de Markov forte assure que

$$S_k \overset{\text{loi}}{=} S_1^1 + \cdots + S_1^k$$

où $(S_1^i)_{1\leqslant i\leqslant k}$ sont indépendants de même loi que S_1. On a donc $G_k(s) = G_1(s)^k$. De même,

$$S_1 \overset{\text{loi}}{=} 1 + ZS_2$$

où Z est une v.a. de loi de Bernoulli de paramètre 1/2 indépendante de S_2, d'où

$$G_1(s) = \frac{s}{2} + \frac{s}{2}G_2(s), \qquad s \in [0,1[.$$

Pour tout s, $G_1(s)$ est solution de $sx^2 - 2x + s = 0$ et vaut donc

$$G_1(s) = \frac{1-\sqrt{1-s^2}}{s}$$

puisque $G_1(s) < 1$. En conclusion, T^N a même loi que S_N et sa fonction génératrice G_N de T^N est égale à $G_N(s) = G_1(s)^N$.

On peut à présent déterminer la fonction caractéristique de U_N. Pour tout $t \in \mathbb{R}$,

$$\varphi_{U^N}(t) = \mathbb{E}\Big(e^{itU^N}\Big) = \frac{1-\sqrt{1-\cos(t/N)^2}}{\cos(t/N)}^N,$$

puisque la fonction caractéristique de X_n^N vaut $\cos(t/N)^n$. Enfin,

$$\varphi_{U^N}(t) = \frac{1-|\sin(t/N)|}{\cos(t/N)}^N \underset{N\to\infty}{\sim} \left(1-\frac{|t|}{N}\right)^N \xrightarrow[N\to\infty]{} e^{-|t|}.$$

Le membre de droite est la fonction caractéristique de la loi de Cauchy standard. □

Remarque 23.12 (Noyau de Poisson du demi-plan)**.** *Le problème de Dirichlet sur le demi-plan supérieur admet la représentation suivante :*

$$\forall (x,y) \in \mathbb{R}\times\mathbb{R}_+^*, \quad u(x,y) = \int_{\mathbb{R}} b(z)P_H(x,y,z)\,dz, \quad \textit{avec} \quad P((x,y),z) = \frac{1}{\pi}\frac{y}{y^2+(x-z)^2}.$$

Recueil de modèles stochastiques – Page 222/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

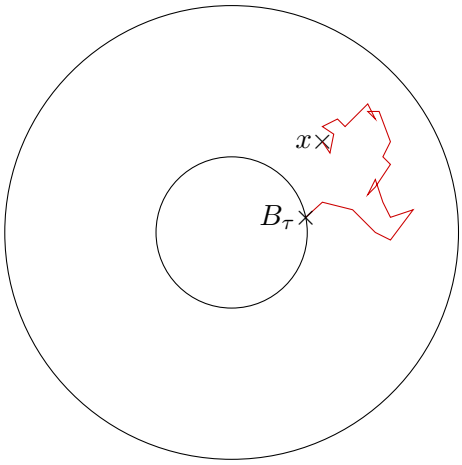


FIGURE 23.2 – Une trajectoire brownienne issue de x arrêtée hors d'une couronne.

23.5 Récurrence, transience, et fonctions harmoniques

La connaissance des fonctions harmoniques radiales sur $\mathbb{R}^d$ permet d'obtenir des propriétés de récurrence et transience pour le mouvement brownien. Pour $r < |x| < R$, notons

$$T_{B(0,r)} = \inf\{t \geqslant 0 \,:\, B_t \in B(0,r)\} \quad \text{et} \quad \tau_{B(0,R)} = \inf\{t \geqslant 0 \,:\, B_t \notin B(0,R)\}.$$

Le temps $S = T_{B(0,r)} \wedge \tau_{B(0,R)}$ est le temps de sortie de la couronne $C(0,r,R)$.

Théorème 23.13 (Récurrence ou transience du mouvement brownien)**.** *Soit $(B_t)_{t\geqslant 0}$ un mouvement brownien sur $\mathbb{R}^d$.*

1. *Si $d = 2$ et $r < |x| < R$, alors*
   a) $\mathbb{P}_x(T_{B(0,r)} < \tau_{B(0,R)}) = \dfrac{\log(R) - \log(|x|)}{\log(R) - \log(r)}$,
   b) $\mathbb{P}_x(T_{\{0\}} < \infty) = 0$,
   c) $\mathbb{P}_x(T_{B(0,r)} < \infty) = 1$,
   d) $\mathbb{P}_x(T_{B(0,r)} < \infty \ i.s.) = 1$.
2. *Si $d \geqslant 3$ et $r < |x| < R$, alors*
   a) $\mathbb{P}_x(T_{B(0,r)} < \tau_{B(0,R)}) = \dfrac{\|x\|^{2-d} - R^{2-d}}{r^{2-d} - R^{2-d}}$,
   b) $\mathbb{P}_x(T_{\{0\}} < \infty) = 0$,
   c) $\mathbb{P}_x(T_{B(0,r)} < \infty) = \dfrac{|x|^{2-d}}{r^{2-d}}$.

Le théorème 23.13 établit la récurrence du mouvement brownien dans $\mathbb{R}^2$: s'il n'atteint aucun point donné, il atteint presque sûrement tout voisinage de ce point une infinité de fois ! Ceci n'est plus vrai en dimensions supérieures.

Démonstration. Soit ϕ la fonction $x \mapsto \log(|x|)$, respectivement $x \mapsto |x|^{2-d}$, pour $d \geqslant 2$, respectivement $d \geqslant 3$. Dans les deux cas, ϕ est une fonction harmonique radiale sur $\mathbb{R}^d$ privé de l'origine. Notons $S = T_{B(0,r)} \wedge \tau_{B(0,R)}$. Le processus $(\phi(B_{t\wedge S}))_{t\geqslant 0}$ est une martingale bornée. De plus, S est fini presque sûrement puisque la première coordonnée de (B_t) prend des valeurs arbitrairement grandes. Le théorème d'arrêt assure donc que

$$\phi(x) = \mathbb{E}_x(\phi(B_S)) = \phi(r)\mathbb{P}_x\big(T_{B(0,r)} < \tau_{B(0,R)}\big) + \phi(R)\mathbb{P}_x\big(T_{B(0,r)} > \tau_{B(0,R)}\big)$$

Recueil de modèles stochastiques – Page 223/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

que l'on écrit encore

$$\mathbb{P}_x\big(T_{B(0,r)} < \tau_{B(0,R)}\big) = \frac{\phi(x) - \phi(R)}{\phi(r) - \phi(R)}.$$

Pour obtenir 1.(b) et 2.(b) on prend les limites successives $r \to 0$ puis $R \to \infty$. Le point 1.(d) se déduit de 1.(c) et de la propriété de Markov forte. □

23.6 Notes et commentaires

Le problème de Dirichlet continu date du dix-neuvième siècle et doit son nom au mathématicien Peter Gustav Lejeune Dirichlet. La solution probabiliste du problème de Dirichlet continu date du milieu du vingtième siècle et est due à Shizuo Kakutani. Le problème de Dirichlet continu se situe à la confluence des probabilités (mouvement brownien), de la théorie du potentiel, de l'analyse harmonique, et des équations aux dérivées partielles, confluence fortement développée par Paul Lévy et Joseph Léo Doob au cours du vingtième siècle. Des aspects variés sont développés notamment dans le livre de Richard Bass [Bas95]. La version discrétisée constitue une ouverture vers l'analyse numérique matricielle (matrices creuses) et l'analyse numérique des équations aux dérivées partielles, ainsi que vers l'analyse numérique stochastique comme les schémas d'Euler stochastiques (voir la section 26.5 page 253). La solution explicite du problème de Dirichlet dans le cas de la boule se trouve par exemple dans le livre de Walter Rudin [Rud87]. L'existence et la continuité sur $\overline{D}$ sont abordées dans le livre de Richard Bass [Bas95]. Notre section 23.4 est inspirée du livre de Bernard Ycart [Yca02]. La plupart des résultats sont généralisables à des opérateurs différentiels bien au delà du laplacien usuel.

Recueil de modèles stochastiques – Page 224/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Chapitre 24

Processus d'Ornstein-Uhlenbeck

Recueil de modèles stochastiques – Page 225/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Mots-clés. Diffusion ; mouvement brownien ; processus d'Ornstein-Uhlenbeck ; équation différentielle stochastique ; processus gaussien ; semi-groupe de Markov ; vecteur gaussien.

Le *processus d'Ornstein-Uhlenbeck* $(X_t)_{t\geqslant 0}$ sur $\mathbb{R}$ est la solution de l'équation différentielle stochastique linéaire suivante :

$$X_t = X_0 - \int_0^t X_s\,ds + \sqrt{2}B_t,$$

où $(B_t)_{t\geqslant 0}$ est un mouvement brownien standard sur $\mathbb{R}$ indépendant de X_0. Le facteur $\sqrt{2}$ rendra certaines formules plus belles. On utilise aussi la version abrégée suivante de l'ÉDS :

$$dX_t = -X_t\,dt + \sqrt{2}\,dB_t.$$

Théorème 24.1 (Changement d'échelle)**.** *Soit $\lambda > 0$ et $\sigma > 0$. Si X est un processus d'Ornstein-Uhlenbeck alors le processus $\tilde{X}$ défini par*

$$\tilde{X}_t = \frac{\sigma}{\sqrt{2\lambda}} X_{\lambda t}$$

est solution de l'équation différentielle stochastique

$$\tilde{X}_t = \tilde{X}_0 - \lambda \int_0^t \tilde{X}_s\,ds + \sigma W_t$$

où $(W_t)_{t\geqslant 0}$ est un mouvement brownien standard.

Démonstration. Pour tout $t \geqslant 0$, en notant $W_t = \sqrt{\lambda} B_{\lambda t}$,

$$\begin{aligned}
\frac{\sigma}{\sqrt{2\lambda}} X_{\lambda t} &= \frac{\sigma}{\sqrt{2\lambda}} X_0 - \int_0^{\lambda t} \frac{\sigma}{\sqrt{2\lambda}} X_s\,ds + \frac{\sigma}{\sqrt{\lambda}} B_{\lambda t} \\
&= \tilde{X}_0 - \lambda \int_0^{\lambda t} \tilde{X}_s\,ds + \frac{\sigma}{\sqrt{\lambda}} B_{\lambda t} \\
&\overset{\mathcal{L}}{=} \tilde{X}_0 - \lambda \int_0^{\lambda t} \tilde{X}_s\,ds + \sigma W_t.
\end{aligned}$$

Le processus $(W_t)_{t\geqslant 0}$ est bien un mouvement brownien standard (par invariance d'échelle). □

Théorème 24.2 (Vecteur gaussien)**.** *Si $X_0 = x$, alors $X_t \sim \mathcal{N}(xe^{-t}, 1 - e^{-2t})$. Plus généralement, le processus $(X_t)_{t\geqslant 0}$ est un processus gaussien de covariance donnée pour tous $0 \leqslant s \leqslant t$ par*

$$\mathbf{Cov}(X_s, X_t) = (1 - e^{-2s})e^{-(t-s)}.$$

Démonstration. En appliquant la formule d'Itô au processus $t \mapsto e^t X_t$ avec $X_0 = x$, on obtient

$$X_t = xe^{-t} + \int_0^t e^{s-t} dB_s.$$

On en déduit que la loi de X_t est la loi gaussienne de moyenne xe^{-t} et de variance $1 - e^{-2t}$. Pour calculer la covariance de X_s et X_t, on écrit de même que

$$X_t = X_s e^{-(t-s)} + Y_t$$

où Y_t est une variable aléatoire centrée indépendante de X_s. □

Soit $\mathcal{A}$ l'ensemble des fonctions de classe $\mathcal{C}^\infty$ sur $\mathbb{R}$ dont toutes les dérivées sont à croissance lente[1]. Le semi-groupe d'Ornstein-Uhlenbeck sur $\mathbb{R}$ est la famille $(N_t)_{t\in\mathbb{R}_+}$ d'opérateurs de $\mathcal{A}$ dans $\mathcal{A}$ définis pour tout $t \geqslant 0$ et tout $f \in \mathcal{A}$ par

$$N_t(f)(x) = \mathbb{E}(f(X_t)|X_0 = x).$$

D'après ce qui précède, on obtient la *formule de Mehler* : pour tout $t \geqslant 0$ et tout $f \in \mathcal{A}$,

$$N_t(f)(x) = \int_{\mathbb{R}} f\Big(e^{-t}x + \sqrt{1-e^{-2t}}y\Big)\,\gamma(dy),$$

où γ désigne la mesure gaussienne standard sur $\mathbb{R}$.

Théorème 24.3 (Propriété de semi-goupe)**.** *La famille $(N_t)_{t\geqslant 0}$ vérifie les propriétés suivantes :*

1. *pour toute fonction $f \in \mathcal{A}$, $N_0 f = f$,*
2. *pour tous $s, t \geqslant 0$ et $f \in \mathcal{A}$, $(N_s \circ N_t)(f) = N_{t+s} f$.*

Démonstration. Le premier point est évident. Pour établir le second, notons $\sigma_t = \sqrt{1-e^{-2t}}$. Soient X et Y deux variables aléatoires indépendantes de loi γ. Par définition de N_t,

$$\begin{aligned}(N_s \circ N_t)(f)(x) &= \mathbb{E}\big[N_t(f)(e^{-s}x + \sigma_s X)\big] \\ &= \mathbb{E}\big[f(e^{-t}(e^{-s}x + \sigma_s X) + \sigma_t Y)\big] \\ &= \mathbb{E}\big[f(e^{-s-t}x + e^{-t}\sigma_s X + \sigma_t Y)\big].\end{aligned}$$

Il ne reste plus qu'à observer que $e^{-t}\sigma_s X + \sigma_t Y \sim \mathcal{N}(0, \sigma_{s+t}^2)$ et à utiliser la formule de Mehler. □

Examinons à présent l'évolution infinitésimale, sur un temps très court, du processus.

Théorème 24.4 (Générateur infinitésimal)**.** *Pour tout $f \in \mathcal{A}$ et tout $x \in \mathbb{R}$,*

$$\lim_{t\to 0^+} \frac{N_t f(x) - N_0 f(x)}{t} = (Af)(x) \quad \text{où} \quad (Af)(x) = f''(x) - xf'(x).$$

Plus généralement, pour tout $f \in \mathcal{A}$ et tout $t \geqslant 0$, on a l'équation aux dérivées partielles

$$\partial_t N_t f = A(N_t f) = N_t(Af).$$

L'équation aux dérivées partielles est parfois appelée *équation de la chaleur avec dérive.* L'opérateur A est le *générateur infinitésimal* du semi-groupe d'Ornstein-Uhlenbeck. Il est important de penser à la formule formelle $N_t = e^{tA}$, qui fait apparaître que $N_0 = I$ et $\partial_t N_t = AN_t$.

1. Une fonction $f : \mathbb{R} \to \mathbb{R}$ est à croissance lente s'il existe un polynôme P tel que $|f| \leqslant P$.

Recueil de modèles stochastiques – Page 226/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Démonstration. Si $f \in \mathcal{A}$, par la formule de Taylor à l'ordre 2 avec reste borné au voisinage de x,

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2} f''(x) + \mathcal{O}(|h|^3).$$

Remarquons dans un premier temps que pour $h = (e^{-t}-1)x + \sigma_t y$, on a

$$\int h\,\gamma(dy) = (e^{-t}-1)x, \quad \int h^2\,\gamma(dy) = (e^{-t}-1)^2x^2 + \sigma_t^2 \quad \text{et} \quad \int |h|^3\,\gamma(dy) = O(t^{3/2}).$$

Le développement limité ci-dessus assure que

$$N_t f(x) = f(x) + (e^{-t}-1)xf'(x) + \frac{1}{2}\big((e^{-t}-1)^2x^2 + \sigma_t^2\big)f''(x) + O(t^{3/2}).$$

On a donc bien la limite annoncée :

$$\lim_{t\to 0} \frac{1}{t}(N_t f(x) - f(x)) = f''(x) - xf'(x).$$

On peut aussi obtenir ce résultat grâce à la formule d'Itô : si $X_0 = x$, alors pour tout $t \geqslant 0$,

$$f(X_t) = f(x) + \int_0^t Af(X_s)\,ds + \sqrt{2}\int_0^t f'(X_s)\,dB_s.$$

Puisque l'intégrale stochastique est une martingale, on obtient en prenant l'espérance

$$N_t f(x) = f(x) + \int_0^t N_s(Af)(x)\,ds.$$

Il reste à utiliser que $N_s(Af)(x) \to Af(x)$ quand $s \to 0$ pour obtenir la première propriété du théorème. La seconde se déduit de la première en utilisant la propriété de semi-groupe. □

D'après la formule de Mehler, $N_t f(x)$ converge vers $\int f d\gamma$ quand t tend vers l'infini pour tout $x \in \mathbb{R}$ et $f \in \mathcal{A}$. De plus, si X_0 suit la loi γ, c'est encore le cas pour X_t. La mesure γ est donc une mesure invariante du processus d'Ornstein-Uhlenbeck. Si $X_0 \sim \gamma$, alors pour tous $t \geqslant 0$ et $f \in \mathcal{A}$,

$$\int N_t f\,d\gamma = \mathbb{E}[(N_t f)(X_0)] = \mathbb{E}(f(X_t)) = \mathbb{E}(f(X_0)) = \int f\,d\gamma.$$

Le résultat suivant va plus loin que l'invariance.

Théorème 24.5 (Symétrie et réversibilité)**.** *Les opérateurs $(N_t)_{t\geqslant 0}$ et A sont auto-adjoints dans $L^2(\gamma)$. En particulier, pour toutes fonctions $f, g \in \mathcal{A}$,*

$$\int Afg\,d\gamma = \int fAg\,d\gamma = -\int f'g'd\,\gamma.$$

La mesure invariante γ est réversible : si X_0 suit la loi γ, alors pour tout $T \geqslant 0$, le processus renversé $(X_{T-t})_{t\in[0,T]}$ a la même loi que $(X_t)_{t\in[0,T]}$.

Démonstration. Par définition du semi-groupe d'Ornstein-Uhlenbeck,

$$\int fN_t g\,d\gamma = \mathbb{E}\big[f(X)g\big(e^{-t}X + \sigma_t Y\big)\big],$$

où X et Y sont deux variables aléatoires indépendantes de loi γ. Le couple $(X, e^{-t}X + \sigma_t Y)$ est un vecteur gaussien centré de matrice de covariance

$$\begin{pmatrix} 1 & e^{-t} \\ e^{-t} & 1 \end{pmatrix}.$$

On peut donc échanger les rôles de f et g dans l'identité précédente, ce qui assure que N_t est auto-adjoint dans $L^2(\gamma)$. Enfin, une intégration par parties fournit la formule du théorème. □

Recueil de modèles stochastiques – Page 227/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Grâce aux propriétés ci-dessus, on peut facilement montrer que la loi de X_t est absolument continue avec une densité régulière dès que $t > 0$, quelque soit X_0.

Théorème 24.6 (Équation d'évolution de la densité de la loi)**.** *Soit ν_t la loi de X_t. Quelque soit ν_0, pour tout $t > 0$, la loi ν_t admet une densité de classe $\mathcal{C}^\infty$ par rapport à la mesure de Lebesgue. De plus, si ν_0 est absolument continue de densité v_0 par rapport à la mesure γ alors ν_t l'est aussi et sa densité v_t par rapport à γ est solution de l'équation aux dérivées partielles suivante :*

$$\partial_t v_t(x) = A v_t(x)$$

Démonstration. Si g est une fonction régulière, la propriété de symétrie de N_t assure que

$$\mathbb{E}g(X_t) = \int \mathbb{E}(g(X_t)|X_0 = x)v_0(x)\gamma(dx) = \int N_t g(x)v_0(x)\gamma(dx) = \int g(x)N_t(v_0)(x)\gamma(dx).$$

Par définition de v_t on a $v_t = N_t(v_0)$. Par définition de N_t, la fonction $x \mapsto N_t(v_0)(x)$ est de classe $\mathcal{C}^\infty$ (même si v_0 ne l'est pas). On conclut grâce au théorème 24.4. □

Remarque 24.7 (Équation d'évolution et aux dérivées partielles de Fokker-Planck)**.** *On peut montrer par un changement de variables que la densité u_t de la loi de X_t par rapport à la mesure de Lebesgue est solution de l'équation aux dérivées partielles suivante*

$$\partial_t u_t(x) = \partial^2_{xx} u_t(x) + \partial_x(x u_t(x)).$$

Cette équation aux dérivées partielles est parfois appelée équation de Fokker-Planck*. Elle exprime l'évolution temporelle d'une densité spatiale, tout comme l'équation d'Euler en mécanique des fluides, et l'équation de Boltzmann en théorie cinétique des gaz.*

24.1 Décomposition spectrale

On se propose à présent d'étudier l'action du semi-groupe $(N_t)_{t\geqslant 0}$ sur les polynômes. Les polynômes forment un sous-espace dense de $L^2(\gamma)$. En effet, soit $f \in L^2(\gamma)$. La transformée de Laplace de la mesure signée $\nu(dx) = f(x)\gamma(dx)$ est finie sur $\mathbb{R}$ puisque, pour $\lambda \in \mathbb{R}$, on a

$$\left|\int e^{\lambda x}\,\nu(dx)\right| \leqslant \int f^2\,d\gamma \int e^{2\lambda x}\,\gamma(dx) < \infty.$$

La transformée de Laplace de ν est donc en particulier analytique au voisinage de 0. Si f est orthogonale à tous les polynômes, toutes les dérivées de cette transformée sont nulles en 0 et par suite elle est également nulle. Ceci implique que f est nulle γ-presque sûrement. On peut donc trouver une base hilbertienne de $L^2(\gamma)$ formée de polynômes. On lira avec profit le théorème 16.4.

Les polynômes de Hermite moniques $(H_n)_{n\geqslant 0}$ sont définis par leur série génératrice :

$$G(s,x) = e^{sx-\frac{1}{2}s^2} = \sum_{n=0}^{\infty} \frac{s^n}{n!} H_n(x),$$

c'est-à-dire $H_n(x) = \partial^n_{s=0} G(s,x)$. On a $H_0(x) = 1$, $H_1(x) = x$, $H_2(x) = x^2 - 1$, etc. Ils vérifient l'équation de récurrence à trois termes $H_{n+1}(x) = xH_n(x) - nH_{n-1}(x)$, l'équation de récurrence différentielle $H'_n(x) = nH_{n-1}(x)$, et l'équation différentielle $H''_n(x) - xH'_n(x) + nH_n(x) = 0$. Il se trouve que les polynômes de Hermite sont les polynômes orthogonaux pour la loi gaussienne γ, et diagonalisent le processus d'Ornstein-Uhlenbeck, c'est-à-dire les opérateurs N_t et A.

Théorème 24.8 (Polynômes de Hermite et processus d'Ornstein-Uhlenbeck)**.** *La suite $(H_n/\sqrt{n!})_{n\geqslant 0}$ est une base hilbertienne de $L^2(\gamma)$. De plus H_n est vecteur propre de N_t (respectivement de A) associé à la valeur propre e^{-nt} (respectivement $-n$).*

Recueil de modèles stochastiques – Page 228/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Démonstration. Soient $s, t \in \mathbb{R}$ fixés. Appliquons N_t à la fonction $x \mapsto G(s,x)$:

$$N_t(G(s,\cdot))(x) = \exp(se^{-t}x - s^2/2)\mathbb{E}[\exp(s\sigma_t Y)]$$

où Y suit la loi γ. La transformée de Laplace de Y vaut $\mathbb{E}(\exp(\lambda Y)) = \exp(\lambda^2/2)$ donc

$$N_t(G(s,\cdot))(x) = G(se^{-t}, x).$$

Par définition des polynômes de Hermite, on obtient

$$\begin{aligned} N_t(H_n)(x) &= N_t(\partial^n_{s=0} G(s,\cdot))(x) = \partial^n_{s=0} N_t(G(s,\cdot))(x) \\ &= \partial^n_{s=0} G(se^{-t}, x) = e^{-nt}\partial^n_{s=0} G(se^{-t}, x) = e^{-nt} H_n(x). \end{aligned}$$

Le polynôme de Hermite H_n est donc vecteur propre de N_t, de valeur propre e^{-nt}. En appliquant le théorème 24.4 à la fonction H_n, on obtient que H_n est vecteur propre de A associé à la valeur propre $-n$. La symétrie de N_t par rapport à γ, assure que, pour tous entiers m et n et tout $t > 0$,

$$e^{-mt}\int H_m H_n \, d\gamma = \int N_t(H_m) H_n \, d\gamma = \int H_m N_t(H_n) \, d\gamma = e^{-nt}\int H_m H_n \, d\gamma.$$

En particulier, H_n et H_m sont orthogonaux dans $L^2(\gamma)$ dès que $n \neq m$. Pour voir enfin que $(H_n/\sqrt{n!})_{n\geqslant 0}$ est bien une base orthonormée de $L^2(\gamma)$, on peut utiliser la formule de Plancherel

$$\sum_{n=0}^{\infty} \frac{s^{2n}}{n!^2}\|H_n\|_2^2 = \int G(s,x)^2 \, \gamma(dx) = \exp(-s^2)\int e^{-2sx} \, \gamma(dx) = \exp(s^2) = \sum_{n=0}^{\infty} \frac{s^{2n}}{n!},$$

qui donne la valeur de la norme $\|H_n\|_2^2 = n!$ en identifiant les deux séries. □

Le théorème 24.8 permet d'obtenir la décomposition de $N_t f$ sur la base hilbertienne des polynômes de Hermite à partir de celle de f, et fournit en particulier une estimation pour la convergence de $N_t(f)$ vers une fonction constante dans $L^2(\gamma)$.

Théorème 24.9 (Décomposition spectrale et convergence dans $L^2(\gamma)$)**.** *Pour tout $f \in L^2(\gamma)$, si*

$$f = \sum_{n\geqslant 0} a_n H_n \quad \text{où} \quad a_n = \int f H_n \, d\gamma \quad \text{avec} \quad \sum_n n! a_n^2 < \infty,$$

alors, pour tout $t \geqslant 0$,

$$N_t f = \sum_{n\geqslant 0} e^{-nt} a_n H_n.$$

De plus, pour toute fonction $f \in L^2(\gamma)$, la convergence exponentielle suivante a lieu :

$$\|N_t f - \gamma(f)\|_{L^2(\gamma)} \leqslant e^{-t}\|f - \gamma(f)\|_{L^2(\gamma)} \quad \text{où} \quad \gamma(f) = \int f \, d\gamma.$$

Démonstration. La première partie découle du théorème 24.8. Pour la seconde partie, on remarque tout d'abord que $H_0(x) = 1$. Ensuite, pour tout $t \geqslant 0$,

$$a_0 = \int f \, d\gamma = \int f N_t(H_0) \, d\gamma = \int N_t f H_0 \, d\gamma = \int N_t f \, d\gamma.$$

Ainsi, pour tout $t \geqslant 0$,

$$\begin{aligned} \|N_t f - \gamma(f)\|^2_{L^2(\gamma)} &= \int (N_t f - a_0)^2 \, d\gamma = \sum_{n\geqslant 1} a_n^2 e^{-2nt} n! \\ &\leqslant e^{-2t}\sum_{n\geqslant 1} a_n^2 n! = e^{-2t}\|f - \gamma(f)\|^2_{L^2(\gamma)}, \end{aligned}$$

ce qui est le résultat annoncé. □

Recueil de modèles stochastiques – Page 229/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Remarque 24.10 (Trou spectral : mind the gap !). *Le taux de convergence exponentiel (égal à 1) du théorème 24.9 correspond à l'écart entre les deux premières valeurs propres de A, à savoir* 0 *et* -1. *Cette propriété se généralise à tous les processus de Markov réversibles dont le générateur infinitésimal possède la propriété dite de trou spectral :* 0 *est valeur propre simple et isolée.*

Remarque 24.11 (Lois de conservation et vitesse de convergence). *Si* $f \perp \mathrm{Vect}\{H_1, \dots, H_{k-1}\}$ *dans* $L^2(\gamma)$ *alors il en va de même pour* N_t *(le semi-groupe préserve la décomposition spectrale) et*

$$\|N_t f - \gamma(f)\|_{L^2(\gamma)} \leqslant e^{-kt} \|f - \gamma(f)\|_{L^2(\gamma)} \quad \text{où} \quad \gamma(f) = \int f \, d\gamma.$$

24.2 Entropie, couplage, et temps long

On souhaite à présent étudier la convergence de la loi de X_t vers sa mesure invariante quand $t \to \infty$ en termes d'entropie relative et de distance de Wasserstein. Soit μ_1 et μ_2 deux mesures de probabilité sur $\mathbb{R}^d$. L'entropie relative de μ_1 par rapport à μ_2 est définie par [2]

$$\mathrm{Ent}(\mu_1|\mu_2) = \begin{cases} \displaystyle\int \frac{d\mu_1}{d\mu_2} \log \frac{d\mu_1}{d\mu_2} \, d\mu_2 & \text{si } \mu_1 \ll< \mu_2, \\ +\infty & \text{sinon.} \end{cases}$$

Ce n'est pas une distance mais c'est une fonction positive, nulle si et seulement si [3] $\mu = \nu$. La distance de Wasserstein d'ordre 2 entre μ_1 et μ_2 est donnée par

$$W_2(\mu_1, \mu_2) = \inf_{X_1 \sim \mu_1,\, X_2 \sim \mu_2} \sqrt{\mathbb{E}\Big(|X_1 - X_2|^2\Big)},$$

où l'infimum porte sur les couples (X_1, X_2) de variables aléatoires dont les lois marginales sont μ_1 et μ_2. Les distances de Wassertein de tous ordres sont définies dans le chapitre 21.

Ces deux quantités ont une expression explicite si μ et ν sont deux mesures gaussiennes.

Théorème 24.12 (Gaussiennes). *Si* $\mu_1 = \mathcal{N}(m_1, \Sigma_1)$ *et* $\mu_2 = \mathcal{N}(m_2, \Sigma_2)$ *sur* $\mathbb{R}^d$ *alors*

$$W_2(\mu_1, \mu_2)^2 = \|m_1 - m_2\|_2^2 + \mathrm{Tr}(\Sigma_1 \Sigma_2 - 2(\Sigma_1^{1/2} \Sigma_2 \Sigma_1^{1/2})^{1/2}).$$

D'autre part, si les matrices de covariance sont inversibles alors

$$\mathrm{Ent}(\mu_2|\mu_1) = \frac{1}{2} \log \left(\frac{\det \Sigma_1}{\det \Sigma_2} \right) + \mathrm{Tr}\,(\Sigma_1^{-1} \Sigma_2) + (m_1 - m_2)^T \Sigma_1^{-1} (m_1 - m_2).$$

Dans le cas contraire, l'entropie relative est infinie.

Peut-on quantifier la vitesse de convergence du processus d'Ornstein-Uhlenbeck à l'aide de l'entropie ou de la distance de Wasserstein, au delà du cas où la loi initiale est gaussienne ? Le cas de la distance de Wasserstein est simple. Celui de l'entropie fait l'objet de la section suivante.

Théorème 24.13 (Convergence en distance de Wasserstein). *Si* ν_0 *et* $\tilde{\nu}_0$ *sont deux mesures de probabilité de moment d'ordre* 2 *fini, et si* ν_t *(respectivement* $\tilde{\nu}_t$*) est la loi de* X_t *lorsque* $X_0 \sim \nu_0$ *(respectivement* $\tilde{\nu}_0$*), alors pour tout* $t \geqslant 0$,

$$W_2(\nu_t, \tilde{\nu}_t) \leqslant e^{-t} W_2(\nu_0, \tilde{\nu}_0)$$

En particulier, si $\tilde{\nu}_0 = \gamma$ *alors pour tout* $t \geqslant 0$,

$$W_2(\nu_t, \gamma) \leqslant e^{-t} W_2(\nu_0, \gamma).$$

2. En physique statistique, $\mathrm{Ent}(\mu|\gamma)$ est une énergie libre si μ, γ sont vues comme des mesures de Boltzmann.
3. Grâce à l'inégalité de Jensen pour la fonction strictement convexe $x \in \mathbb{R}_+ \mapsto u \log(u)$ avec $0 \log(0) = 0$.

Recueil de modèles stochastiques – Page 230/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Démonstration. Considérons tout d'abord le cas où $\nu_0 = \delta_x$ et $\tilde{\nu}_0 = \delta_{\tilde{x}}$. En prenant le même mouvement brownien, on construit deux processus X et $\tilde{X}$ couplés issus de x et $\tilde{x}$. On a alors

$$X_t - \tilde{X}_t = x - \tilde{x} - \int_0^t (X_s - \tilde{X}_s)\, ds.$$

On en déduit que

$$\left|X_t - \tilde{X}_t\right|^2 = e^{-2t}|x - \tilde{x}|^2.$$

Par définition de $W_2(\nu_t, \tilde{\nu}_t)$, on obtient

$$W_2(\nu_t, \tilde{\nu}_t) \leqslant e^{-t}|x - \tilde{x}|.$$

On passe au cas général en conditionnant, et en considérant l'infimum sur le couplage initial. □

24.3 Inégalité de Sobolev logarithmique

On dit qu'une mesure de probabilité μ sur $\mathbb{R}^d$ vérifie une inégalité de Sobolev logarithmique de constante $c > 0$ si, pour toute fonction $f : \mathbb{R}^n \to \mathbb{R}$ de classe $\mathcal{C}^1$ positive et à support compact,

$$\mathbf{Ent}_\mu(f) := \int f \log f \, d\mu - \int f \, d\mu \log \int f \, d\mu \leqslant c \int \frac{|\nabla f|^2}{f} \, d\mu.$$

Note : $\mathbf{Ent}_\mu f = \mathrm{Ent}(\nu|\mu)$ où $d\nu = f\, d\mu$. La constante c optimale est la plus petite possible.

Théorème 24.14 (Inégalité de Sobolev logarithmique gaussienne)**.** *Pour tout $t \geqslant 0$ et tout $x \in \mathbb{R}$, la mesure de probabilité $N_t(\cdot)(x) = \mathcal{N}(xe^{-t}, 1 - e^{-2t})$ vérifie une inégalité de Sobolev logarithmique de constante optimale $(1 - e^{-2t})/2$. En particulier, la mesure γ vérifie une inégalité de Sobolev logarithmique de constante optimale $1/2$.*

Démonstration. Soit $t \geqslant 0$ et f régulière. On définit la fonction α sur $[0, t]$ en posant

$$\alpha(s) = N_s(\Psi(N_{t-s}f)) \quad \text{où} \quad \Psi(x) = x \log x.$$

Dans la suite, on notera $g = N_{t-s}f$. Remarquons que

$$\alpha(t) - \alpha(0) = N_t(f \log f) - N_t f \log N_t f = \mathbf{Ent}_{N_t(\cdot)(x)}(f).$$

D'autre part, grâce au théorème 24.4, on a

$$\alpha'(s) = N_s\big(A(\Psi(g)) - Ag\Psi'(g)\big).$$

Un calcul direct montre que

$$A(\Psi(g)) - Ag\Psi'(g) = \frac{|\nabla g|^2}{g}.$$

On a donc

$$\alpha'(s) = N_s\left(\frac{|\nabla N_{t-s}f|^2}{N_{t-s}f}\right).$$

Par définition de N_t, le semi-groupe vérifie la relation de commutation suivante :

$$\nabla N_t f = e^{-t} N_t(\nabla f),$$

Recueil de modèles stochastiques – Page 231/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

qui fournit l'expression suivante pour $\alpha'(s)$:

$$\alpha'(s) = e^{-(t-s)} N_s\left(\frac{(N_{t-s}|\nabla f|)^2}{N_{t-s} f}\right).$$

L'inégalité de Cauchy-Schwarz assure que

$$(N_t|\nabla f|)^2 = \left[N_t\left(\frac{|\nabla f|}{\sqrt{f}}\sqrt{f}\right)\right]^2 \leqslant N_t\left(\frac{|\nabla f|^2}{f}\right) N_t(f).$$

Ainsi, on obtient, grâce à la relation de semi-groupe, la majoration suivante

$$\alpha'(s) \leqslant e^{-2(t-s)} N_t\left(\frac{|\nabla f|^2}{f}\right).$$

Une intégration entre 0 et t donne à présent l'inégalité de Sobolev logarithmique pour N_t. On en déduit celle pour γ par passage à la limite $t \to \infty$. On remarque enfin que les fonctions $x \mapsto e^{\lambda x}$ saturent l'inégalité de Sobolev logarithmique : la constante 1/2 est optimale. □

Remarque 24.15 (Forme L^2 plutôt que L^1 de l'inégalité de Sobolev logarithmique)**.** *On préfère parfois définir l'inégalité de Sobolev logarithmique sous forme L^2 : pour toute fonction f régulière,*

$$\int f^2 \log f^2 \, d\mu - \int f^2 \, d\mu \log \int f^2 \, d\mu \leqslant 4c \int |\nabla f|^2 \, d\mu.$$

Cette inégalité de Sobolev logarithmique permet d'obtenir une vitesse de convergence exponentielle de l'entropie relative de la loi de X_t par rapport à la mesure invariante γ.

Théorème 24.16 (Convergence exponentielle de l'entropie)**.** *Supposons que la loi μ_0 de X_0 soit d'entropie relative finie par rapport à γ. Alors, en notant μ_t la loi de X_t, la fonction $t \in \mathbb{R}_+ \mapsto \mathrm{Ent}(\mu_t|\gamma)$ décroît et pour tout $t \geqslant 0$,*

$$\mathrm{Ent}(\mu_t|\gamma) \leqslant e^{-t}\mathrm{Ent}(\mu_0|\gamma).$$

Démonstration. Notons v_t la densité de μ_t par rapport à γ. On a alors

$$\mathrm{Ent}(\mu_t|\gamma) = \int v_t \log v_t d\gamma.$$

D'après le théorème 24.6, on a

$$\begin{aligned}\frac{d}{dt}\int v_t \log v_t d\gamma &= \int \partial_t v_t (1 + \log v_t) d\gamma \\ &= \int A v_t (1 + \log v_t) d\gamma \\ &= \int v_t A (1 + \log v_t) d\gamma.\end{aligned}$$

Or, pour toute fonction régulière g, un calcul direct montre que

$$A(1 + \log g) = \frac{Ag}{g} - \frac{g'^2}{g^2}.$$

D'après la propriété d'invariance de A, on a $\int Ag d\gamma = 0$. On a donc

$$\frac{d}{dt}\mathrm{Ent}(\mu_t|\gamma) = -\int \frac{(\partial_x v_t)^2}{v_t} d\gamma.$$

Recueil de modèles stochastiques – Page 232/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

L'inégalité de Sobolev logarithmique appliquée à la fonction v_t assure que

$$\frac{d}{dt}\mathrm{Ent}(\mu_t|\gamma) \leqslant -2\mathrm{Ent}(\mu_t|\gamma).$$

Le lemme de Gronwall fournit enfin la relation souhaitée. □

Remarque 24.17. *Le théorème 24.6 assure que même si la loi de X_0 n'a pas de densité, la loi de X_t a toujours une densité dès que $t > 0$. On peut donc appliquer le théorème 24.16 dès que $t > 0$.*

24.4 Notes et commentaires

Tout ce qui concerne le processus d'Ornstein-Uhlenbeck se trouve dans les dix premières pages du cours de Dominique Bakry [Bak94]. L'expression de la distance de Wasserstein entre deux mesures gaussiennes dans le théorème 24.12 est établie notamment dans un article [GS84] de Clark Givens et Rae Shortt. Le processus d'Ornstein-Uhlenbeck est l'illustre représentant d'une importante classe de processus stochastiques : les diffusions de Kolmogorov. Soit U une fonction sur $\mathbb{R}^d$ de classe $\mathcal{C}^2$ et $(B_t)_{t\geqslant 0}$ un mouvement brownien standard à valeurs dans $\mathbb{R}^d$. On appelle équation de Langevin l'équation différentielle stochastique suivante :

$$dX_t = \sqrt{2}dB_t - \nabla U(X_t)dt.$$

La solution $(X_t)_{t\geqslant 0}$ de cette équation différentielle stochastique est appelée processus de diffusion de Kolmogorov associée au potentiel U. Comme le processus d'Ornstein-Uhlenbeck, c'est un processus de Markov. La version intégrale de l'équation différentielle stochastique s'écrit

$$X_t = X_0 + \sqrt{2}B_t - \int_0^t \nabla U(X_s)\,ds,$$

Son générateur infinitésimal agit sur les fonctions f de classe $\mathcal{C}^2$ de la manière suivante :

$$Lf = \Delta f - \nabla\langle U, \nabla f\rangle.$$

Si e^{-U} est intégrable alors la mesure de probabilité

$$\mu(dx) = \frac{1}{Z}e^{-U(x)}\,dx \quad \text{avec} \quad Z = \int e^{-U(x)}\,dx,$$

est la mesure de probabilité invariante associée à X. L'opérateur L est même auto-adjoint dans $L^2(\mu)$. La décomposition spectrale du semi-groupe associé n'est pas explicite au-delà du cas gaussien, mais il est toutefois possible d'obtenir des analogues des théorèmes 24.9 ou 24.16. On peut par exemple montrer que s'il existe $\lambda > 0$ tel que

$$\forall x \in \mathbb{R}^d, \quad \mathrm{Hess}\,U(x) \geqslant \lambda \mathrm{I}_d,$$

ce qui revient à dire que μ est uniformément log-concave, alors pour tout $t \geqslant 0$,

$$\mathrm{Ent}(\mu_t|\mu) \leqslant e^{-\lambda t}\mathrm{Ent}(\mu_0|\mu)$$

où μ_t est la loi de X_t. C'est un cas particulier du critère dit de Γ_2 établi par Dominique Bakry et Michel Émery [BÉ85]. Les processus de Kolmogorov sont étudiés en détail dans les livres [Roy99, ABC+00, BGL14].

Recueil de modèles stochastiques – Page 233/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Chapitre 25

Modèles de diffusion

Recueil de modèles stochastiques – Page 235/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Mots-clés. Diffusion ; mouvement brownien ; équation cinétique ; vecteur gaussien.

25.1 Modèle de Langevin

On considère une particule de poussière dans un fluide comme l'eau. La particule est énorme par rapport aux molécules du fluide. L'échelle de taille de la particule est dite *macroscopique* : ses mouvements peuvent être observés, au moins à l'aide d'une bonne loupe, et mesurés. Les molécules qui composent le fluide sont quant à elles beaucoup trop petites et surtout beaucoup trop nombreuses pour être observées une à une. Elles se situent à une échelle dite *microscopique*. À chaque instant, d'innombrables chocs ont lieu entre la particule et les molécules qui l'entourent. Ces chocs se répartissent en moyenne de manière uniforme sur la surface de la particule. Pourtant à un instant donné il peut y avoir des déviations par rapport à cette répartition uniforme. Dans le modèle de Langevin, ces fluctuations sont modélisées par un mouvement brownien. On recherche un modèle pour l'évolution du couple position/vitesse (X_t, V_t) de la particule au temps t qui découle du principe fondamental de la dynamique. Les forces qui s'exercent sur la particule sont les suivantes :

— le poids de la particule ;
— la poussée d'Archimède ;
— la force de frottement qui s'oppose au déplacement de la particule ;
— la résultante des chocs de molécules d'eau.

Le poids de la particule et de la poussée d'Archimède se compensent et nous négligerons leur résultante en considérant l'évolution de la particule dans le plan horizontal. On suppose que l'intensité de la force de frottement est proportionnelle à la vitesse de la particule. Elle s'écrit donc $-\lambda V_t$ puisqu'elle s'oppose au mouvement. La modélisation de l'effet des chocs des molécules d'eau demande un peu plus d'attention. Entre les instants t et $t+\alpha$ avec α grand dans l'échelle de temps microscopique, la particule subit un nombre de chocs $N\alpha$ où N, le nombre de chocs par unité de temps, est supposé très grand. On adopte une approche «statistique» ou «stochastique» et on modélise les chocs par des variables aléatoires $(\varepsilon_i)_{i\geqslant 1}$ indépendantes et identiquement distribuées centrées de variance τ^2. On pose $\sigma^2 = N\tau^2$. La force exercée pendant la durée α est donnée par

$$\sum_{i=1}^{N\alpha} \varepsilon_i = \sqrt{\sigma^2\alpha}\, \underbrace{\frac{1}{\sqrt{\tau^2 N\alpha}} \sum_{i=1}^{N\alpha} \varepsilon_i}_{Y}.$$

Si $N\alpha$ est grand, le *théorème limite central* suggère que la loi de Y est proche d'une loi gaussienne centrée réduite. De plus, les forces aléatoires exercées sur les intervalles de temps

disjoints sont dûs à des molécules d'eau différentes (elles se déplacent beaucoup moins que la particule). On peut donc les supposer indépendantes. En conclusion, il paraît naturel de modéliser l'action des chocs par un processus à accroissements indépendants dont la loi d'un accroissement entre deux instants t et $t+\alpha$ est de loi gaussienne centrée de variance $\sigma^2\alpha$. Ce processus est le mouvement brownien $(\sigma B_t)_{t\geqslant 0}$. On retrouve ici l'idée du *théorème de Donsker* (théorème 26.8). Comme la vitesse est la dérivée de la position, o obtient le système :

$$\begin{cases} dX_t = V_t\,dt \\ dV_t = -\lambda V_t\,dt + \sigma dB_t. \end{cases}$$

On dit qu'il s'agit d'une équation cinétique car elle fait intervenir position et vitesse.

Théorème 25.1 (Processus de Langevin)**.** *Conditionnellement à $\{X_0 = x_0\}$, le processus $(X_t)_{t\geqslant 0}$ est un processus gaussien de moyenne et de covariance données par*

$$x_t = x_0 + \frac{1-e^{-\lambda t}}{\lambda} v_0 \quad \textit{et} \quad \rho(t,s) = \frac{\sigma^2}{\lambda^2} + \frac{\sigma^2}{2\lambda^3}\Big(-2 + 2e^{-\lambda t} + 2e^{-\lambda s} - e^{-\lambda|t-s|} + e^{-\lambda(t+s)}\Big).$$

En particulier, la variance est donnée par

$$\mathbb{V}(X_t) = \frac{\sigma^2}{\lambda^2} + \frac{\sigma^2}{2\lambda^3}\Big(-3 + 4e^{-\lambda t} - e^{-2\lambda t}\Big).$$

L'évolution de la vitesse ne dépend pas de la position.

Démonstration. Le processus $(V_t)_{t\geqslant 0}$ est un *processus d'Ornstein-Uhlenbeck*, étudié dans le chapitre 24. Si le processus $(U_t)_{t\geqslant 0}$ est solution de

$$U_t = U_0 - \int_0^t U_s\,ds + \sqrt{2}B_t.$$

alors le processus $(V_t)_{t\geqslant 0}$ défini par $V_t = (\sigma/\sqrt{2\lambda})U_{\lambda t}$ est solution de

$$V_t = V_0 - \lambda\int_0^t V_s\,ds + \sigma B_t.$$

On en déduit que $(V_t)_{t\geqslant 0}$ est un processus gaussien de moyenne $v_t = e^{-\lambda t}v_0$ et de covariance

$$r(t,s) = \frac{\sigma^2}{2\lambda}\Big(e^{-\lambda|t-s|} - e^{-\lambda(t+s)}\Big).$$

À présent le processus X est gaussien car image du processus V par une application linéaire : $X_t = X_0 + \int_0^t V_s\,ds$ (il s'agit d'une application immédiate de la formule d'Itô). □

Remarque 25.2 (Ordres de grandeurs)**.** *La constante $1/\lambda$, homogène à un temps, est appelée* temps de relaxation. *Elle est en pratique de l'ordre de 10^{-8} seconde. Ceci suggère qu'il est imposible de distinguer le processus $(X_t)_{t\geqslant 0}$ d'un mouvement brownien de variance σ^2/λ^2.*

25.2 Processus de Langevin confiné

On modifie à présent légèrement le modèle de la section précédente. Si la particule coule au fond d'un récipient dont la forme est donnée par la fonction $x \mapsto \mu x^2/2$ alors la particule est soumise à une force de plus qui dérive de l'énergie potentielle et qui est de la forme $-\mu x$. L'évolution du couple vitesse/position est donnée par le système suivant :

$$\begin{cases} dX_t = V_t\,dt \\ dV_t = -\lambda V_t\,dt - \mu X_t\,dt + \sigma dB_t. \end{cases}$$

Recueil de modèles stochastiques – Page 236/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

On dit que X est un *processus de Langevin confiné*[1]. Pour alléger les notations, on pose $Z_t = (X_t, V_t)$. On peut réécrire le système sous la forme matricielle suivante :

$$Z_t = Z_0 + \int_0^t M Z_s \, ds + \sigma W_t \quad \text{où} \quad M = \begin{pmatrix} 0 & 1 \\ -\mu & -\lambda \end{pmatrix} \quad \text{et} \quad W_t = \begin{pmatrix} 0 \\ B_t \end{pmatrix}.$$

La solution de cette équation linéaire est de la forme

$$Z_t = e^{tM} Z_0 + \sigma \int_0^t e^{(t-s)M} \, dW_s.$$

Lemme 25.3 (Loi au temps t)**.** *Si (X_0, V_0) est un vecteur gaussien, alors il en est de même pour (X_t, V_t) pour tout $t \geqslant 0$. Si on note, pour $t \geqslant 0$,*

$$x_t = \mathbb{E}(X_t), \quad v_t = \mathbb{E}(V_t), \quad a_t = \mathbb{E}(X_t^2), \quad b_t = \mathbb{E}(X_t Y_t) \quad \textit{et} \quad c_t = \mathbb{E}(V_t^2).$$

alors $(x_t)_{t\geqslant 0}$, $(v_t)_{t\geqslant 0}$, $(a_t)_{t\geqslant 0}$, $(b_t)_{t\geqslant 0}$ et $(c_t)_{t\geqslant 0}$ sont solutions des équations différentielles linéaires :

$$\frac{d}{dt}\begin{pmatrix} x_t \\ v_t \end{pmatrix} = M \begin{pmatrix} x_t \\ v_t \end{pmatrix} \quad \textit{et} \quad \frac{d}{dt}\begin{pmatrix} a_t \\ b_t \\ c_t \end{pmatrix} = N \begin{pmatrix} a_t \\ b_t \\ c_t \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ \sigma^2 \end{pmatrix}$$

où

$$M = \begin{pmatrix} 0 & 1 \\ -\mu & -\lambda \end{pmatrix} \quad \textit{et} \quad N = \begin{pmatrix} 0 & 2 & 0 \\ -\mu & -\lambda & 1 \\ 0 & -2\mu & -2\lambda \end{pmatrix}.$$

Démonstration. Il s'agit d'une conséquence immédiate de la formule d'Itô. □

Remarque 25.4 (Formules explicites)**.** *À propos de la matrice e^{tM}, on distingue trois cas :*

— *si $\lambda^2 > 4\mu$, alors, en notant $\omega = \sqrt{\lambda^2 - 4\mu}$, $\alpha_1 = (-\lambda + \omega)/2$, et $\alpha_2 = (-\lambda - \omega)/2$,*

$$e^{tM} = \frac{e^{\alpha_2 t}}{\omega}\begin{pmatrix} -\alpha_1 & 1 \\ -\mu & \alpha_2 \end{pmatrix} + \frac{e^{\alpha_1 t}}{\omega}\begin{pmatrix} \alpha_2 & -1 \\ \mu & -\alpha_1 \end{pmatrix};$$

— *si $\lambda^2 = 4\mu$, alors*

$$e^{tM} = e^{-\lambda t/2}\begin{pmatrix} 1 - \lambda s/2 & t \\ -\lambda^2 t/4 & 1 + \lambda t/2 \end{pmatrix};$$

— *si $\lambda^2 - 4\mu < 0$, alors en notant $\omega = \sqrt{\mu - \lambda^2/4}$,*

$$e^{tM} = e^{-\lambda t/2}\begin{pmatrix} \cos(\omega t) + \frac{\lambda \sin(\omega t)}{2\omega} & \frac{\sin(\omega t)}{\omega} \\ -\frac{\mu \sin(\omega t)}{\omega} & \cos(\omega t) - \frac{\lambda \sin(\omega t)}{2\omega} \end{pmatrix}.$$

L'expression explicite de e^{tN} est beaucoup plus pénible à obtenir. On peut toutefois remarquer que les valeurs propres de N ont toutes une partie réelle strictement négative.

Les équations différentielles linéaires du lemme 25.3 sont simples à résoudre numériquement (illustration donnée par la figure 25.1).

Lemme 25.5 (Densité de la loi à tout instant)**.** *Pour toute loi initiale et tout instant $t > 0$, la loi du processus de Langevin confiné (X_t, V_t) admet une densité strictement positive sur $\mathbb{R}^2$.*

1. «Damped Langevin process» en anglais.

Recueil de modèles stochastiques – Page 237/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

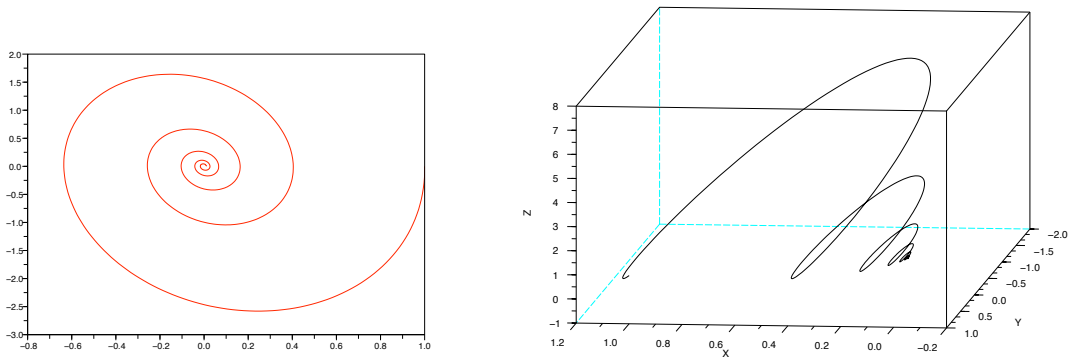


Figure 25.1 – Évolution des vecteurs moyenne (x_t, v_t) (à gauche) et covariance (a_t, b_t, c_t) (à droite) pour $\sigma = 1$, $\lambda = 1$ et $\mu = 10$.

Démonstration. Il suffit d'établir le résultat lorsque la loi initiale est une masse de Dirac en un point (x, v), car obtient le cas général en conditionnant par la valeur initiale. Si la loi initiale est une masse de Dirac, le couple (X_t, V_t) est un vecteur gaussien. Sa loi admet une densité strictement positive sur $\mathbb{R}^2$ si et seulement si le déterminant de sa matrice de covariance est strictement positif. Sans perte de généralité, on peut supposer que $(X_0, V_0) = (0, 0)$. Le déterminant de la matrice de covariance alors est donné par $d_t = a_t c_t - b_t^2$ avec les notations du lemme 25.3. Donc

$$d_t' = -2\lambda(a_t c_t - b_t^2) + \sigma^2 a_t = -2\lambda d_t + \sigma^2 a_t.$$

Comme $d_0 = a_0 = 0$, la dérivée de $t \mapsto d_t$ est nulle en 0. En itérant, on obtient

$$\begin{aligned} d_t'' &= -2\lambda d_t' + 2\sigma^2 b_t \\ d_t^{(3)} &= -2\lambda d_t'' + 2\sigma^2(-\mu a_t - \lambda b_t + c_t) \\ d_t^{(4)} &= -2\lambda d_t^{(3)} + 2\sigma^2(\lambda\mu a_t + (\lambda^2 - 4\mu) b_t - 3\lambda c_t) + 2\sigma^4. \end{aligned}$$

En particulier, les trois premières dérivées de $t \mapsto d_t$ sont nulles en 0 et $d_0^{(4)} = 2\sigma^4$. Donc le déterminant devient strictement positif pour $t > 0$ s'il est nul à l'instant initial. □

Remarque 25.6 (Hypoellipticité)**.** *Dans la preuve du lemme 25.5, on peut noter que*

$$d_t \underset{t \sim 0}{\sim} \frac{2\sigma^4}{4!} t^4.$$

Pour une diffusion sur $\mathbb{R}^2$ avec un bruit brownien non dégénéré, le déterminant serait de l'ordre de t^2. Ce retard à l'allumage est dû au fait que seule la composante des vitesses est perturbée par un mouvement brownien. Cependant, la structure subtile du coefficient de dérive dans l'équation du processus de Langevin confiné assure que ce bruit se propage sur les deux composantes et que la loi du couple (X_t, V_t) a une densité dès que $t > 0$. Il s'agit d'un cas particulier du phénomène de l'hypoellipticité, étudié dans un cadre général par Hörmander puis par Malliavin notamment.

La figure 25.1 illustre l'apparition de densité lorsque la loi initiale est une mesure sans densité sur $\mathbb{R}^2$ au travers de l'évolution du déterminant de la matrice de covariance. Sa valeur converge vers celui de la mesure invariante qui vaut $\sigma^4/(4\lambda^2\mu)$.

Recueil de modèles stochastiques – Page 238/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

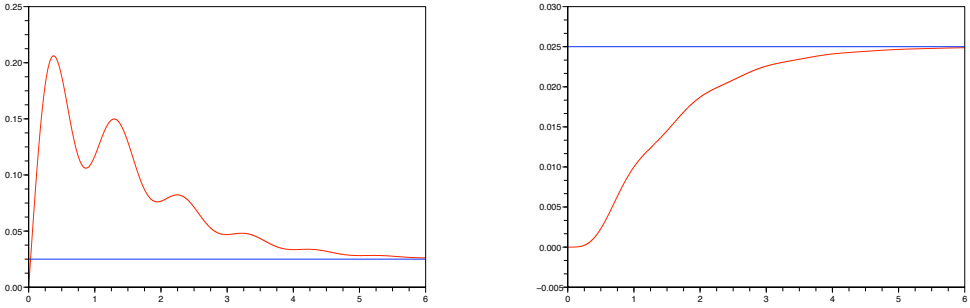


FIGURE 25.2 – Évolution en fonction du temps du déterminant de la matrice de covariance issu de $(x_0, v_0, a_0, b_0, c_0) = (0, 0, 1, 1, 1)$ à gauche et $(0, 0, 0, 0, 0)$ à droite pour $\sigma = 1$, $\lambda = 1$ et $\mu = 10$.

Comme pour le processus d'Ornstein-Uhlenbeck, on associe au processus de Langevin confiné $Z = (X, V)$ un semi-groupe $(Q_t)_{t \geqslant 0}$, donné pour $t \geqslant 0$, $z \in \mathbb{R}^2$, et f régulière par

$$Q_t f(z) = \mathbb{E}(f(Z_t) | Z_0 = z).$$

Lemme 25.7 (Générateur infinitésimal)**.** *Si f est une fonction de classe $\mathcal{C}^\infty$ sur $\mathbb{R}^2$ et à support compact, alors pour tout $t > 0$ et $z \in \mathbb{R}^2$,*

$$\partial_t Q_t f(z) = Q_t(Af)(z)$$

où, avec la notation $z = (x, v)$,

$$Af(x, v) = v \cdot \nabla_x f(x, v) - (\mu x + \lambda v) \cdot \nabla_v f(x, v) + \Delta_v f.$$

Démonstration. On applique la formule d'Itô à f avec $Z_0 = z$:

$$f(Z_t) = f(Z_0) + \int_0^t Af(Z_s)\, ds + M_t,$$

où $(M_t)_{t \geqslant 0}$ est une martingale. Ainsi,

$$\lim_{t \to 0} \frac{Q_t f(z) - f(z)}{t} = Af(z).$$

Pour obtenir la relation à tout temps $t \geqslant 0$, on applique la propriété de Markov. □

Théorème 25.8 (Mesure invariante)**.** *La mesure de probabilité gaussienne produit*

$$\pi = \mathcal{N}\left(0, \frac{\sigma^2}{2\lambda\mu}\right) \otimes \mathcal{N}\left(0, \frac{\sigma^2}{2\lambda}\right)$$

est une mesure invariante du processus (X, V). Elle vérifie, pour toute fonction f régulière,

$$\int Af\, d\pi = 0.$$

Démonstration. Puisque (X, V) est un processus gaussien, sa probabilité invariante doit être une loi normale. De plus, le point $\sigma^2/(2\lambda\mu)(0, 0, 1, 0, \mu)$ est l'unique point fixe des équations différentielles du lemme 25.3. On remarque enfin que

$$(x_t, v_t, a_t, b_t, c_t) \xrightarrow[t \to \infty]{} \frac{\sigma^2}{2\lambda\mu}(0, 0, 1, 0, \mu)$$

pour toutes conditions initiales finies. □

Recueil de modèles stochastiques – Page 239/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Remarque 25.9 (Irréversibilité)**.** *Contrairement au cas du processus d'Ornstein-Uhlenbeck, on peut vérifier que A n'est pas auto-adjoint dans $L^2(\pi)$. En d'autres termes, le semi-groupe $(Q_t)_{t\geqslant 0}$ n'est pas réversible pour la mesure invariante π.*

Dans la suite, on cherche à quantifier la convergence de la loi de X_t vers la mesure invariante π lorsque t tend vers $+\infty$ en termes de distance de Wasserstein ou d'entropie relative comme cela a été fait pour le processus d'Ornstein-Uhlenbeck dans le chapitre 24. Il s'avère cependant que l'étude du comportement en temps long de l'entropie relative est plus délicate à mener pour le processus (X, V), en partie à cause de l'absence de composante brownienne sur la coordonnée de position. On peut toutefois obtenir sans trop d'efforts le résultat de monotonie suivant.

Théorème 25.10 (Décroissance de l'entropie)**.** *Si π_t est la loi de (X_t, V_t) et h_t est sa densité par rapport à la mesure π, alors*

$$\frac{d}{dt}\mathrm{Ent}(\pi_t|\pi) = -\int \frac{|\nabla_v h_t|^2}{h_t}\, d\pi.$$

Démonstration. Pour alléger les calculs, on suppose sans perte de généralité que σ est égal à $\sqrt{2}$. En vertu du lemme 25.5, la loi π_t de (X_t, V_t) admet une densité u_t par rapport à la mesure de Lebesgue. Elle est solution de l'équation aux dérivées partielles suivante :

$$\partial_t u_t + v \cdot \nabla_x u_t - \mu x \cdot \nabla_v u_t = \Delta_v u_t + \nabla_v \cdot (\lambda v u_t).$$

La structure de cette équation est intéressante : le membre de gauche est conservatif et décrit les trajectoires d'un système dynamique dans $\mathbb{R}^d \times \mathbb{R}^d$ associé à l'hamiltonien

$$(x, v) \mapsto \mu \frac{|x|^2}{2} + \lambda \frac{|v|^2}{2},$$

tandis que le membre de droite est dissipatif mais n'agit que sur la variable de vitesse. Notons $h_t(x, v) = u_t(x, v)/\pi(x, v)$ où $\pi(x, v)$ désigne par abus de notation la densité de π au point (x, v). On vérifie que la fonction h est alors solution de l'équation aux dérivées partielles (ÉDP) suivante :

$$\partial_t h_t + v \cdot \nabla_x h_t - \mu x \cdot \nabla_v h_t = \Delta_v h_t - \lambda v \cdot \nabla_v h_t.$$

La variation d'entropie relative est donnée par

$$\frac{d}{dt}\mathrm{Ent}(\pi_t|\pi) = \frac{d}{dt}\int h_t \log h_t d\pi = \int \partial_t h_t (1 + \log h_t)\, d\pi.$$

L'expression de $\partial_t h_t$ est donnée par l'ÉDP ci-dessus. Les intégrations par parties suivantes

$$-\int (v \cdot \nabla_x h_t)(1 + \log h_t)\, d\pi = -\int v \cdot \nabla_x (h_t \log h_t)\, d\pi = -\lambda\mu \int (x \cdot v) h_t \log h_t\, d\pi$$
$$\int (\mu x \cdot \nabla_v h_t)(1 + \log h_t)\, d\pi = \int \mu x \cdot \nabla_v (h_t \log h_t)\, d\pi = \lambda\mu \int (x \cdot v) h_t \log h_t d\pi$$
$$\int (\Delta_v h_t)(1 + \log h_t)\, d\pi = -\int \frac{\nabla_v h_t \cdot \nabla_v h_t}{h_t}\, d\pi + \lambda \int (1 + \log h_t) \nabla_v h_t \cdot v\, d\pi$$

fournissent le résultat annoncé. □

Il est plus facile d'obtenir un résultat de convergence pour la distance de Wasserstein d'ordre 2.

Recueil de modèles stochastiques – Page 240/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Théorème 25.11 (Convergence en distance de Wasserstein)**.** *Soient ν et $\tilde{\nu}$ deux mesures de probabilité sur $\mathbb{R}^2$ possédant un moment d'ordre 2 fini. Soient ν_t (respectivement $\tilde{\nu}_t$) la loi de Z_t lorsque Z_0 suit la loi ν (respectivement $\tilde{\nu}$). Alors il existe c et d telles que*

$$W_2(\nu_t, \tilde{\nu}_t) \leqslant (c + dt)e^{-\rho t} W_2(\nu, \tilde{\nu}),$$

avec

$$\begin{cases} \rho = \dfrac{\lambda - \sqrt{\lambda^2 - 4\mu}}{2} \quad \text{et} \quad d = 0 & \text{si } \lambda > 2\sqrt{\mu}, \\ \rho = \dfrac{\lambda}{2} & \text{si } \lambda \leqslant 2\sqrt{\mu}. \end{cases}$$

Démonstration. On procède par couplage. On suppose tout d'abord que ν et $\tilde{\nu}$ sont des masses de Dirac aux points $z = (x, y)$ et $\tilde{z} = (\tilde{x}, \tilde{y})$. Notons $Z = (X_t, V_t)_{t \geqslant 0}$ (respectivement $\tilde{Z} = (\tilde{X}_t, \tilde{V}_t)_{t \geqslant 0}$) le processus de Langevin confiné issu de z (respectivement de $\tilde{z}$), tous les deux dirigés par le même mouvement brownien $(B_t)_{t \geqslant 0}$. Notons $z_t = Z_t - \tilde{Z}_t$. La fonction $t \mapsto z_t$ est déterministe. En effet, elle est solution de l'équation différentielle

$$\begin{cases} z_t' = M z_t \\ z_0 = z - \tilde{z} \end{cases} \qquad \text{avec} \quad M = \begin{pmatrix} 0 & 1 \\ -\mu & -\lambda \end{pmatrix}.$$

Ainsi, $z_t = e^{tM} z_0$. Le comportement en temps long de la fonction $t \mapsto z_t$ dépend des valeurs propres de la matrice M, qui sont données par :

$$\begin{aligned} -\frac{\lambda \pm \sqrt{\lambda^2 - 4\mu}}{2} \quad & \text{si} \quad \lambda \geqslant 2\sqrt{\mu}, \\ -\frac{\lambda}{2} \pm i\frac{\sqrt{4\mu - \lambda^2}}{2} \quad & \text{si} \quad \lambda \leqslant 2\sqrt{\mu}. \end{aligned}$$

Supposons que $\lambda > 2\sqrt{\mu}$. Puisque les deux valeurs propres de M sont réelles et distinctes, il existe une matrice inversible P à coefficients réels et une matrice diagonale Δ telles que

$$M = P^{-1} \Delta P.$$

Par conséquence, $P z_t = e^{t\Delta} P z_0$ et

$$|P z_t|_2 \leqslant e^{-\rho t} |P z_0|_2 \quad \text{où} \quad \rho = \frac{\lambda - \sqrt{\lambda^2 - 4\mu}}{2}.$$

Puisque P est inversible, l'application $z \mapsto \|Pz\|_2$ est une norme équivalente à $z \mapsto \|z\|_2$. Il existe donc une constance c telle que, pour tout $z_0 \in \mathbb{R}^2$,

$$|z_t|_2 \leqslant c e^{-\rho t} |z_0|_2.$$

En particulier, on a

$$W_2(\delta_z Q_t, \delta_{\tilde{z}} Q_t)^2 \leqslant \mathbb{E}\Big(|Z_t|^2\Big) \leqslant c e^{-2\rho t} |z - \tilde{z}|^2.$$

Supposons à présent que $\lambda \leqslant 2\sqrt{\mu}$. Alors la matrice M a deux valeurs propres qui sont ou confondues et égales à $-\lambda/2$, ou complexes conjuguées et de partie réelle égale à $-\lambda/2$. Dans ce cas, on peut établir qu'il existe deux constantes c et d telles que

$$\left\|e^{tM}\right\| \leqslant (c + dt) e^{-\lambda t/2} \quad \text{avec} \quad \|A\| = \sum_{i,j=1,2} |a_{ij}|.$$

On en déduit que

$$\|z_t\|_2 \leqslant (c + dt) e^{-\lambda t/2} \|z_0\|_2.$$

On conclut dans les deux cas en prenant l'infimum sur les couplages de ν et $\tilde{\nu}$. □

Recueil de modèles stochastiques – Page 241/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Grâce au théorème 24.12, on peut estimer la distance de Wasserstein entre la loi de (X_t, V_t) et la mesure invariante lorsque la loi initiale est gaussienne. La figure 25.3 représente l'évolution de l'entropie relative de la loi au temps t par rapport à la mesure invariante et du carré de la distance de Wasserstein entre ces deux lois. La figure 25.3 suggère que l'entropie relative de la loi au temps t par rapport à la mesure invariante décroît par paliers. Ceci est à relier à l'expression de sa dérivée qui ne fait intervenir que des dérivées partielles de h_t en la variable vitesse, ce qui est une conséquence de l'absence de composante brownienne sur la position. Contrairement à l'entropie relative, la distance W_2 n'est pas du tout une fonction décroissante du temps !

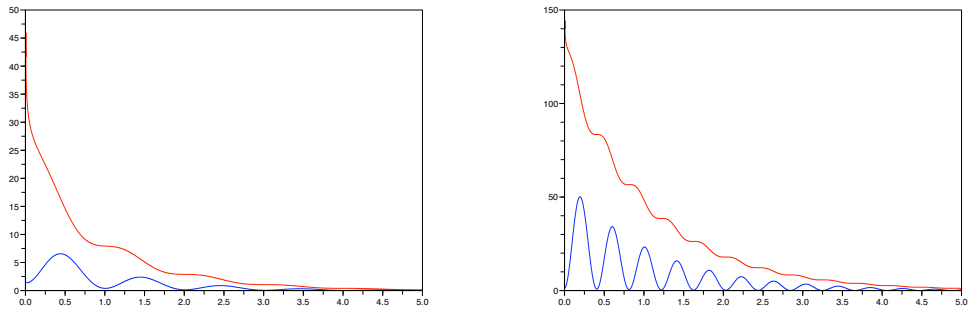


Figure 25.3 – Évolutions comparées de l'entropie relative et de la distance de Wasserstein d'ordre 2 entre la loi au temps t et la mesure invariante $(x_0, v_0, a_0, b_0, c_0) = (1, 0, 0, 0, 0)$ pour $\sigma = 1$, $\lambda = 1$ et $\mu = 10$ (à gauche) et $\mu = 60$ (à droite).

25.3 Notes et commentaires

En 1828, un botaniste du nom de Robert Brown [2] rend compte de ses observations : une particule de pollen, observable au microscope, plongée dans l'eau, contient des grains animés de mouvements aussi rapides qu'irréguliers. La thèse qui s'est imposée est la suivante : le mouvement serait lié aux chocs de la particule de pollen avec les molécules d'eau. À la fin du 19e siecle, les points suivants avaient été établis empiriquement :

1. le mouvement est très irrégulier et la trajectoire semble ne pas admettre de tangente ;
2. deux particules de pollen se déplacent indépendamment (si elles ne se choquent pas), même si la distance les séparant ne dépasse pas leur diamètre ;
3. le mouvement est d'autant plus rapide que la particule est petite ;
4. la composition et la densité de la particule n'ont aucun effet ;
5. le mouvement est d'autant plus lent que le fluide est visqueux ;
6. le mouvement est d'autant plus rapide que la température est élevée ;
7. le mouvement est incessant.

Le dernier point fut admis après une expérience de trente ans et l'observation de liquides prisonniers de quartz vieux de plusieurs milliers d'années. Il a alors été admis que le mouvement ne pouvait pas s'expliquer par une force interne à la particule de pollen. Les travaux de Albert Einstein, en 1905, sont souvent présentés comme les premiers pas mathématiques vers ce qui va devenir la théorie du mouvement brownien. Norbert Wiener fournira l'essentiel du travail de la mathématisation de la théorie du mouvement brownien et on parle aussi de processus de

2. ...resté dans l'histoire pour cette découverte, anecdotique dans son oeuvre, par ailleurs oubliée.

Recueil de modèles stochastiques – Page 242/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Wiener. Cependant, un peu avant les travaux d'Einstein, et dans un tout autre contexte, Louis Bachelier avait déjà obtenu, en 1900, la loi du mouvement brownien dans sa thèse intitulée «La théorie de la spéculation», qui n'a pas eu d'impact à l'époque. Il y a également les travaux de Marian Smoluchowski, contemporains de ceux d'Einstein. Le modèle étudié dans ce chapitre a été introduit par Paul Langevin et perfectionné par Leonard Ornstein et George Uhlenbeck dans les années 1930. On trouvera dans l'ouvrage de Edward Nelson [Nel67] la modélisation du début du chapitre, des perspectives historiques et des prolongements mathématiques.

On peut généraliser le processus de Langevin, en sortant du monde confortable des processus gaussiens, en considérant la solution de l'équation différentielle stochastique suivante :

$$\begin{cases} dX_t = V_t dt \\ dV_t = -\lambda V_t dt - \nabla U(X_t) dt + \sigma dB_t, \end{cases}$$

où la fonction U joue le rôle d'un potentiel de confinement (penser à une fonction convexe). Dans ce cas, la mesure invariante est encore explicite et s'écrit comme le produit de deux mesures sur les espaces de position et de vitesse. C'est un modèle simple de processus hypoelliptique ergodique. Lorsque la diffusion agit sur toutes les coordonnées, on parle de processus de diffusion non dégénéré, l'équation aux dérivées partielles de Kolmogorov décrivant l'évolution de la loi du processus au temps t est dite elliptique et admet une solution de classe $\mathcal{C}^\infty$. Cependant, beaucoup d'équations d'évolution, comme les équations cinétiques décrivant l'évolution de la vitesse et de la position d'une particule, sortent de ce cadre. Lars Hörmander dans [Hör67] classifie les diffusions dégénérées dont la loi au temps t est absolument continue et régulière.

Les résultats précis de convergence à l'équilibre pour ces processus sont assez récents. L'article de Denis Talay [Tal02] fournit des estimations pour la convergence de la densité de la loi au temps vers la densité invariante dans des espaces L^p à poids. Les travaux de Cédric Villani [Vil06, Vil09] proposent une autre approche via les inégalités fonctionnelles (inégalité de Poincaré, inégalité de Sobolev logarithmique) et la théorie de Lars Hörmander sur l'hypoellipticité.

Le théorème 25.10 qui établit que l'entropie relative par rapport à la mesure invariante est une fonction décroissante du temps est un résultat qui est vrai dans de très nombreux modèles. Il peut toutefois être très difficile à établir. Par exemple, pour la célèbre équation de Boltzmann, qui régit l'évolution de la densité de particules de gaz en interaction, il a fallu attendre les travaux de Cédric Villani [Vil03a], une centaine d'années après les travaux précurseurs de Ludwig Boltzmann.

Recueil de modèles stochastiques – Page 243/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Chapitre 26

Des chaînes de Markov aux processus de diffusion

Recueil de modèles stochastiques – Page 245/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Mots-clés. Convergence de processus ; approximation par une diffusion ; chaîne d'Ehrenfest ; chaîne de Fisher-Wright ; file d'attente M/M/∞ ; processus de Yule ; processus d'Orstein-Uhlenbeck ; mouvement brownien.

26.1 Paresse et échantillonnage

Soit $\mathbf{P}$ une matrice de transition sur un espace d'états fini E et I la matrice identité sur E. Pour tout $N \geqslant 1$,

$$\mathbf{P}_N = \left(1 - \frac{1}{N}\right) I + \frac{1}{N}\mathbf{P} = I + \frac{\mathbf{P} - I}{N}$$

est une matrice de transition. La chaîne X^N associée reste en un état $i \in E$ pendant un temps aléatoire de loi géométrique de paramètre $1 - (1 - \mathbf{P}(i,i))/N$ puis, si cette probabilité n'est pas nulle, saute sur $E \backslash \{i\}$ avec la mesure de probabilité proportionnelle à $(\mathbf{P}(i,j))_{j \neq i}$. On dit que X^N pour $N \geqslant 2$ est une *chaîne paresseuse*. Posons $Y_t^N = X_{\lfloor Nt \rfloor}$ pour $t \in \mathbb{R}_+$. La suite de processus $(Y^N)_{N \geqslant 1}$ converge vers le processus de Markov à temps continu sur E de générateur $A = \mathbf{P} - I$ et de semi-groupe $(P_t)_{t \geqslant 0}$. La convergence pour un temps donné t s'écrit simplement :

$$P_t = e^{tA} = \lim_{N \to \infty} \mathbf{P}_N^{\lfloor Nt \rfloor} = \lim_{N \to \infty} \left(I - \frac{A}{N}\right)^{\lfloor Nt \rfloor}.$$

La propriété de semi-groupe permet d'étendre cette relation à tout n-uplet de temps $t_1 < t_2 < \cdots < t_n$.

Réciproquement, si X est une chaîne de Markov à temps continu sur E de générateur A alors $(Y_n^\delta)_{n \geqslant 0} = (X_{\delta n})_{n \geqslant 0}$ est une chaîne de Markov à temps discret sur E de matrice de transition $\mathbf{P}_\delta = e^{\alpha A}$. Elle est appelée *chaîne échantillonnée*.

Les processus de Markov $(X_n^1)_{n \in \mathbb{N}}$, $(Y_n^\delta)_{n \in \mathbb{N}}$ et $(X_t)_{t \in \mathbb{R}_+}$ sont très liés : les propriétés de récurrence et transience coïncident et ces processus ont les mêmes mesures invariantes. Cependant même si la chaîne X^1 est périodique, ce n'est pas le cas pour Y^δ et X. En effet, on montre facilement par récurrence sur n que si $\mathbb{P}_{x_0}(X_n^1 = x_0) > 0$ alors $\mathbb{P}_x(X_t = x_n) > 0$ pour tout $t > 0$.

Remarque 26.1 (Bernoulli vers Poisson)**.** *Un processus de Bernoulli rendu de plus en plus paresseux convenablement renormalisé en temps converge vers un processus de Poisson. Ceci entraine notamment la* loi des petits nombres *déjà croisée au chapitre 1.*

Remarque 26.2 (Généalogie et paresse)**.** *Dans le chapitre 10, la chaîne à temps discret décrivant le nombre de lignées à suivre pour trouver l'ancêtre commun le plus récent dans une population de taille N converge, quand $N \to \infty$ vers un processus de mort pur à temps continu de manière semblable à la convergence des chaînes paresseuses.*

26.2 Diffusion sur un intervalle

On considère le processus $(X_t)_{t\geqslant 0}$ solution de l'équation différentielle stochastique

$$dX_t = \sigma(X_t)\, dB_t + b(X_t)\, dt.$$

où σ et b sont des fonction régulières sur l'intervalle $I =]l, r[$ à valeurs dans $]0, \infty[$ et $\mathbb{R}$ respectivement. Le générateur infinitésimal de X est l'opérateur agissant sur les fonctions f de classe $\mathcal{C}^2$ sur I donné par

$$Lf(x) = \frac{1}{2}\sigma^2(x)f''(x) + b(x)f'(x).$$

Soit $c \in I$. Introduisons

— la *fonction d'échelle* p :

$$p(x) = \int_c^x \exp\left(-2\int_c^y \frac{b(z)\, dz}{\sigma^2(z)}\right) dy,$$

— la *mesure de vitesse* m est définie par

$$m(dx) = \mathbf{1}_I(x)\frac{2}{p'(x)\sigma^2(x)}\, dx = \mathbf{1}_I(x)\frac{2}{\sigma^2(x)} \exp\left(2\int_c^x \frac{b(z)\, dz}{\sigma^2(z)}\right) dx,$$

— et pour $l < a < b < r$ la fonction sur $]a, b[$,

$$M_{a,b}(x) = -\int_a^x (p(x) - p(y))m(dy) + \frac{p(x) - p(a)}{p(b) - p(a)} \int_a^b (p(b) - p(y))m(dy).$$

Le choix des fonctions p et $M_{a,b}$ ne doit rien au hasard ! Elles vérifient les relations cruciales suivantes :

$$\begin{cases} Lp(x) = 0 & \text{pour } x \in I, \\ p(c) = 0, \\ p'(c) = 1, \end{cases} \quad \text{et} \quad \begin{cases} LM_{a,b}(x) = 1 & \text{pour } x \in]a, b[, \\ M_{a,b}(a) = 0, \\ M_{a,b}(b) = 0. \end{cases}$$

Théorème 26.3 (Sortie d'un intervalle)**.** *Soit $l < a < b < r$ et $T_{a,b} = \inf\{t \geqslant 0 \,:\, X_t \notin [a, b]\}$ le temps de sortie de l'intervalle $[a, b]$ pour X. Alors*

$$\mathbb{P}_x\big(X_{T_{a,b}} = a\big) = 1 - \mathbb{P}_x\big(X_{T_{a,b}} = b\big) = \frac{p(b) - p(x)}{p(b) - p(a)},$$

et

$$\mathbb{E}_x(T_{a,b}) = M_{a,b}(x).$$

Remarque 26.4 (Mouvement brownien)**.** *Ce résultat généralise le cas bien connu du mouvement brownien pour lequel la probabilité de sortir en a est égale à $(b - x)/(b - a)$.*

Recueil de modèles stochastiques – Page 246/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Démonstration. La fonction p est croissante par définition. Ceci assure que, pour $a < y < b$, les fonctions

$$x \in [y,b] \mapsto \frac{p(x)-p(y)}{p(x)-p(a)} \quad \text{et} \quad x \in [a,b] \mapsto \int_a^x \frac{p(x)-p(y)}{p(x)-p(a)} m(dy)$$

sont croissantes. En conséquence, la fonction $M_{a,b}$ est positive $[a,b]$. Posons, pour $n \in \mathbb{N}^*$,

$$\tau_n = \inf\left\{ t \geqslant 0\,;\, \int_0^t \sigma^2(X_s)\,ds \geqslant n \right\}.$$

Puisque $LM_{a,b}(x) = 1$ pour $x \in \,]a,b[$, la formule d'Itô appliquée à $M_{a,b}(X_t)$ assure que

$$M_{a,b}(X_{t\wedge\tau_n\wedge T_{a,b}}) = M_{a,b}(X_0) - (t \wedge \tau_n \wedge T_{a,b}) + \int_0^{t\wedge\tau_n\wedge T_{a,b}} M'_{a,b}(X_s)\sigma(X_s)dB_s.$$

Grâce à l'arrêt au temps $\tau_n \wedge T_{a,b}$ l'intégrale stochastique ci-dessus est une martingale. En prenant l'espérance avec $X_0 = x$, on obtient grâce à la positivité de $M_{a,b}$,

$$\mathbb{E}_x(t \wedge \tau_n \wedge T_{a,b}) = M_{a,b}(x) - \mathbb{E}_x\big[M_{a,b}(X_{t\wedge\tau_n\wedge T_{a,b}})\big] \leqslant M_{a,b}(x) < \infty.$$

Il reste à faire tendre n puis t vers ∞ pour obtenir, par convergence monotone, que $\mathbb{E}_x(T_{a,b})$ est fini. Le temps $T_{a,b}$ est donc fini p.s. et, par convergence dominée, on obtient

$$\lim_{t\to\infty} \mathbb{E}\big[M_{a,b}(X_{t\wedge T_{a,b}})\big] = \mathbb{E}\big[M_{a,b}(X_{T_{a,b}})\big] = 0,$$

puisque $M_{a,b}$ est bornée et nulle en a et b. On obtient ainsi l'expression de $\mathbb{E}_x(T_{a,b})$.

De même, la formule d'Itô appliquée à $p(X_t)$ donne

$$p(x) = \mathbb{E}_x\big[p(X_{T_{a,b}})\big] = p(a)\mathbb{P}_x\big(X_{T_{a,b}} = a\big) + p(b)\mathbb{P}_x\big(X_{T_{a,b}} = b\big).$$

On conclut en utilisant la relation $\mathbb{P}_x\big(X_{T_{a,b}} = a\big) + \mathbb{P}_x\big(X_{T_{a,b}} = b\big) = 1$. □

26.3 Convergence d'une chaîne de Markov vers une diffusion

L'objet de cette section est d'énoncer un théorème assurant la convergence en loi d'une suite de chaînes de Markov convenablement renormalisées en temps et en espace vers un processus de diffusion dont les coefficients de diffusion et de dérive sont explicites.

Théorème 26.5 (Approximation diffusive)**.** *Soit $(\mathbf{P}_N)_{N\geqslant 1}$ une suite de noyaux de transition sur $\mathbb{R}$ et posons*

$$b_N(x) = N\int_{|y-x|\leqslant 1} (y-x)\mathbf{P}_N(x,dy) \quad \textit{et} \quad a_N(x) = N\int_{|y-x|\leqslant 1} (y-x)^2\mathbf{P}_N(x,dy).$$

Supposons qu'il existe deux fonction régulières a et b à valeurs dans $]0,\infty[$ et $\mathbb{R}$ respectivement telles que pour tous $r > 0$ et $\varepsilon > 0$,

$$\sup_{|x|\leqslant r} |b_N(x) - b(x)| \xrightarrow[N\to\infty]{} 0, \quad \textit{et} \quad \sup_{|x|\leqslant r} |a_N(x) - a(x)| \xrightarrow[N\to\infty]{} 0$$

et

$$\sup_{|x|\leqslant r} N\mathbf{P}_N(x,[x-\varepsilon,x+\varepsilon]) \xrightarrow[N\to\infty]{} 0.$$

Soit $(Y_n^N)_{n\in\mathbb{N}}$ la chaîne de Markov de noyau de transition $\mathbf{P}_N$ et $(Z_t^N)_{t\in\mathbb{R}_+}$ le processus défini par $Z_t^N = Y_{\lfloor Nt\rfloor}^N$. Si $(Y_0^N)_{N\geqslant 1}$ converge en loi vers ν alors, pour tout $T > 0$, la suite

Recueil de modèles stochastiques – Page 247/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

de processus $((Z_t^N)_{t\in[0,T]})_{N\geqslant 1}$ *converge en loi vers le processus de diffusion* $(X_t)_{t\in[0,T]}$ *de loi initiale* ν *de générateur* A *donné par*

$$Af(x) = \frac{1}{2}a(x)f''(x) + b(x)f'(x),$$

pour toute fonction régulière f.

Idée de preuve. Le semi-groupe $(P_t)_{t\geqslant 0}$ associé au générateur infinitésimal A vérifie, pour toute fonction régulière f et $t > 0$,

$$P_t f = f + \int_0^t P_s A f \, ds.$$

Pour tout $N \geqslant 1$, notons $A_N = \mathbf{P}_N - I$. Soit $t > 0$. Alors

$$\begin{aligned}\mathbf{P}_N^{\lfloor Nt\rfloor} f &= f + \sum_{i=0}^{\lfloor (N-1)t\rfloor} \mathbf{P}_N^i A_N f = f + \frac{1}{N}\sum_{i=0}^{\lfloor (N-1)t\rfloor} \mathbf{P}_N^i (NA_N) f \\ &= f + \sum_{i=0}^{\lfloor (N-1)t\rfloor} \int_{i/N}^{(i+1)/N} \mathbf{P}_N^i (NA_N) f \, ds.\end{aligned}$$

On obtient ainsi que, pour tout $t \geqslant 0$,

$$\mathbf{P}_N^{\lfloor Nt\rfloor} f = f + \int_0^t \mathbf{P}_N^{\lfloor Ns\rfloor} (NA_N) f \, ds.$$

Ainsi, la famille de mesures de probabilité $(\mathbf{P}_N^{\lfloor Nt\rfloor})_{t\geqslant 0}$ est-elle liée à NA_N comme $(P_t)_{t\geqslant 0}$ l'est à A. De plus, si f est de classe $\mathcal{C}^2$ à dérivées bornées,

$$\begin{aligned}NA_N f(x) &= N\int (f(y) - f(x))\mathbf{P}_N(x, dy) \\ &= b_N(x)f'(x) + \frac{1}{2}a_N(x)f''(x) + \mathcal{O}(N\mathbf{P}_N(x, [x-1, x+1])).\end{aligned}$$

Les hypothèses du théorème assurent que $NA_N f$ converge vers Af. Reste à en déduire la convergence en loi des processus associés. Cette étape dépasse le cadre de cet ouvrage. On trouvera dans la section Notes et commentaires des références pour la preuve complète de ce résultat. □

Remarque 26.6 (Rien n'est simple)**.** *En pratique, la chaîne de Markov* Y^N *est en général à valeurs dans un sous-ensemble* I_N *discret et souvent fini de* $\mathbb{R}$. *Ainsi, son noyau de transition* $\mathbf{P}_N$ *n'est-il pas défini sur tout* $\mathbb{R}$. *Il faut alors étendre sa définition à l'extérieur de* I_N *de manière un peu artificielle.*

Exemple 26.7 (Marche aléatoire et mouvement brownien)**.** *Soit* S^N *la marche aléatoire sur* $\mathbb{Z}$ *définie par*

$$S_n^N = \sum_{i=1}^n \varepsilon_i$$

où $(\varepsilon_i^N)_{i\geqslant 1}$ *est une suite de v.a. i.i.d. de loi* $p_N\delta_1 + (1-p_N)\delta_{-1}$.

Si $p_N = 1/2$, $(S_{\lfloor N\cdot\rfloor}^N/\sqrt{N})_{N\geqslant 1}$ *converge en loi vers le mouvement brownien issu de 0.*

Si $p_N = p \neq 1/2$, $(S_{\lfloor N\cdot\rfloor}/N)_{N\geqslant 1}$ *converge en loi vers le processus déterministe* $t \mapsto (2p-1)t$.

Si $p_N = 1/2 + \alpha/(2\sqrt{N})$, $(S_{\lfloor N\cdot\rfloor}/N)_{N\geqslant 1}$ *converge en loi vers le mouvement brownien avec dérive* α *issu de 0.*

Recueil de modèles stochastiques – Page 248/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Le premier cas de l'exemple ci-dessous est une instance du *théorème de Donsker* [1], analogue du théorème limite central pour la convergence des processus.

Théorème 26.8 (Théorème de Donsker)**.** *Soit* $(Y_n)_{n\geqslant 1}$ *une suite de variables aléatoires i.i.d. centrées et de variance 1. Notons*

$$X_t^N = \frac{1}{\sqrt{N}}\left(\sum_{k=1}^{\lfloor Nt\rfloor} Y_k - (Nt - \lfloor Nt\rfloor)Y_{\lfloor Nt\rfloor+1}\right).$$

La suite de processus $(X^N)_{N\geqslant 1}$ *converge en loi vers le* mouvement brownien.

Remarque 26.9 (Interpolation linéaire)**.** *Dans l'énoncé du théorème de Donsker, les trajectoires de* X^N *sont construites par interpolation linéaire entre les points d'abscisses multiples de* $1/N$*. Cette précaution permet de considérer la convergence de* $(X_t^N)_{0\leqslant 1\leqslant T}$ *comme une variable aléatoire à valeurs dans* $\mathcal{C}([0,T])$ *muni de la norme infinie et de la tribu borélienne.*

Exemple 26.10 (Chaîne d'Ehrenfest et processus d'Ornstein-Uhlenbeck)**.** *Pour* $N \geqslant 1$*, soit* Z^N *la chaîne d'Ehrenfest (voir chapitre 3) à valeurs dans* $\{0,\dots,N\}$ *dont la matrice de transition est donnée par*

$$\mathbb{P}(Z_1^N = j | Z_0^N = i) = \begin{cases} i/N & \text{si } j = i-1, \\ 1 - i/N & \text{si } j = i+1, \\ 0 & \text{sinon.} \end{cases}$$

On définit la chaîne $(Y_n^N)_{n\geqslant 0}$ *à valeurs dans* $E_N = \left\{-(1/2)\sqrt{N} + k/\sqrt{N} \,:\, 0 \leqslant k \leqslant N\right\}$ *en posant*

$$Y_n^N = \sqrt{N}\left(\frac{Z_n^N}{N} - \frac{1}{2}\right).$$

Son noyau de transition est donné par

$$\mathbf{P}_N(x,\cdot) = \left(\frac{1}{2} + \frac{x}{\sqrt{N}}\right)\delta_{x-1/\sqrt{N}}(\cdot) + \left(\frac{1}{2} - \frac{x}{\sqrt{N}}\right)\delta_{x+1/\sqrt{N}}(\cdot),$$

pour $x \in E_N$*. Après avoir étendu* $\mathbf{P}_N$ *raisonnablement sur* $\mathbb{R}$*, on obtient*

$$b_N(x) = N\left[\left(\frac{1}{2} + \frac{x}{\sqrt{N}}\right)\left(x - \frac{1}{\sqrt{N}}\right) + \left(\frac{1}{2} - \frac{x}{\sqrt{N}}\right)\left(x + \frac{1}{\sqrt{N}}\right)\right] = -2x.$$

De même, $a_N(x) = 1$*. L'hypothèse de concentration du noyau* $\mathbb{P}_N$ *est clairement satisfaite puisque son support converge vers* $\{x\}$*. Ainsi, si* $(Y_0^N)_{N\geqslant 1}$ *converge en loi vers* ν*, alors la suite de processus* $(Y_{\lfloor N\cdot\rfloor}^N)_{N\geqslant 1}$ *converge en loi vers le processus d'Ornstein-Uhlenbeck de loi initiale* ν *solution de l'équation différentielle stochastique*

$$dX_t = dB_t - 2X_t\,dt.$$

26.4 Application aux processus de Fisher-Wright

Considérons le modèle de Fisher-Wright introduit au chapitre 4. Il s'agit de la chaîne de Markov $(X_n^N)_{n\in\mathbb{N}}$ sur $\{0,\dots,N\}$ de transition

$$\mathcal{L}(X_{n+1}^N | X_n^N = x) = \mathrm{Bin}(N, \eta_x)$$

1. Ce résultat est également appelé *principe d'invariance*.

Recueil de modèles stochastiques – Page 249/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

avec

$$\eta_x = \psi_x(1-u_A) + (1-\psi_x)u_B \quad \text{et} \quad \psi_x = \frac{(1+s)x}{(1+s)x+N-x}.$$

Le réel s représente l'avantage sélectif de A sur B et u_A est la probabilité de mutation d'un allèle A vers un allèle B. Notons $Y_n^N = X_n^N/N$ la proportion d'allèles A dans une population de taille N à la génération n et posons $Z_t^N = Y_{\lfloor Nt \rfloor}^N$ pour $t \in \mathbb{R}_+$.

Théorème 26.11 (Limite diffusive de Fisher-Wright)**.** *Supposons que les taux de sélection s et de mutation u_A et u_B (qui dépendent de N) soient tels que*

$$Ns \to \alpha, \quad Nu_A \to \beta_A \quad \textit{et} \quad Nu_B \to \beta_B.$$

Alors la suite de processus $(Z_\cdot^N)_{N\geqslant 1}$ converge en loi vers le processus de diffusion $(Z_t)_{t\in\mathbb{R}_+}$ solution de l'équation différentielle stochastique suivante :

$$dZ_t = \sigma(Z_t)\,dB_t + b(Z_t)\,dt.$$

où

$$\sigma^2(z) = z(1-z) \quad \textit{et} \quad b(x) = \alpha z(1-z) - \beta_A z + \beta_B(1-z).$$

Démonstration. Soit $z \in E_N = \{0, 1/N, \dots, 1\}$. Alors

$$\begin{aligned}\eta_{Nz} &= \psi_{Nz}(1-u_A) + (1-\psi_{Nz})u_B \\ &= z + \frac{-\beta_A z + \beta_B(1-z) + \alpha z(1-z)}{N} + \mathcal{O}(1/N).\end{aligned}$$

Puisque la loi de NZ_1^N sachant que $Z_0^N = z$ est la loi $\mathrm{Bin}(N, \eta_{Nz})$, on a

$$\begin{aligned}N\mathbb{E}\big[Z_1^N - z | Z_0^N = z\big] &= N(\eta_{Nz} - z) \\ &= -\beta_A z + \beta_B(1-z) + \alpha z(1-z) + \mathcal{O}(1/N).\end{aligned}$$

De même

$$\begin{aligned}N\mathbb{E}\Big[\big(Z_1^N - z\big)^2 | Z_0^N = z\Big] &= \eta_{Nz}(1-\eta_{Nz}) \\ &= z(1-z) + \mathcal{O}(1/N).\end{aligned}$$

Enfin,

$$N\mathbb{E}\Big[\big(Z_1^N - z\big)^4 | Z_0^N = z\Big] = \mathcal{O}\big(N^{-1}\big),$$

puisque, si X suit la loi $\mathrm{Bin}(N,p)$, son moment centré d'ordre 4 est tel que

$$\mathbb{E}\Big[(X-Np)^4\Big] = (3Np(1-p) + 1 - 6p(1-p))Np(1-p) = \mathcal{O}(N^2).$$

Toutes ces estimations sont uniformes en $z \in E_N$. Le théorème 26.5 permet de conclure. □

Remarque 26.12 (Échelles de temps)**.** *Pour transférer les estimations de temps d'absorption de la diffusion à la chaîne de Markov, il faudra prendre garde à les multiplier par N. Les probabilités d'atteinte ou les profils de mesures invariantes sont pour leurs parts insensibles à ce changement d'échelle temporelle.*

Exprimons la fonction d'échelle : pour $c, x \in (0,1)$

$$S(x) = \int_c^x y^{-2\beta_B}(1-y)^{-2\beta_A} e^{2\alpha y}\,dy,$$

et la mesure de vitesse :

$$m(x) = 2x^{2\beta_B - 1}(1-x)^{2\beta_A - 1} e^{-2\alpha x}\,dx.$$

Recueil de modèles stochastiques – Page 250/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Exemple 26.13 (Ni mutation, ni sélection)**.** *On suppose ici que* $s = \beta_A = \beta_B = 0$*. Dans ce cas, la dérive de la diffusion est nulle : la diffusion est une martingale et* $\mathbb{P}_x(Z_{T_{0,1}} = 1) = x$*. On a aussi l'expression de l'espérance et de la variance du temps de fixation.*

$$\mathbb{E}_x(T_{0,1}) = -2[x \log x + (1-x)\log(1-x)].$$

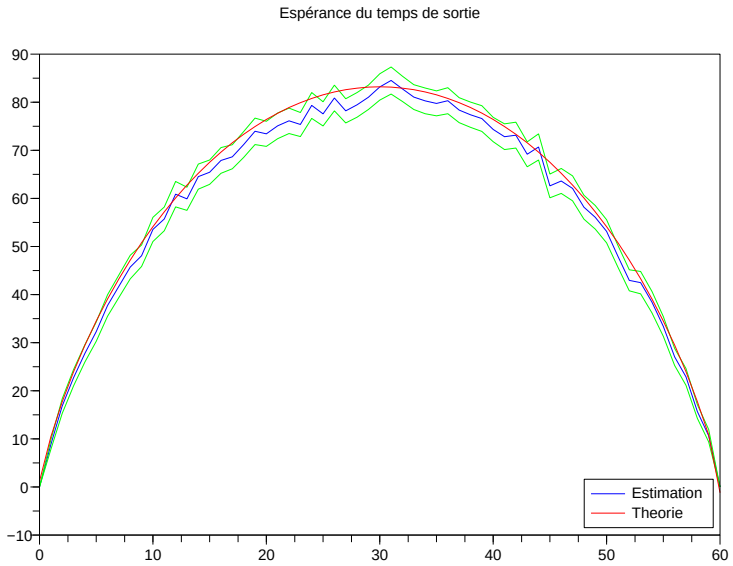


FIGURE 26.1 – Estimation du temps de fixation pour $N = 40$ en l'absence de sélection et de mutation.

Exemple 26.14 (Sélection sans mutation)**.** *On choisit ici* $\alpha > 0$ *et* $\beta_A = \beta_B = 0$*. Alors* $T_{0,1}$ *est fini p.s. et la probabilité de disparition de l'allèle* B

$$\mathbb{P}_x(Z_{T_{0,1}} = 1) = \frac{1 - e^{-2\alpha x}}{1 - e^{-2\alpha}}.$$

La figure 26.2 propose une illustration de l'estimation de la probabilité de fixation de l'allèle A *pour* $N = 30$ *et* $\alpha = 2$*. Supposons que* $N = 10^5$*,* $s = 10^{-4}$ *et* $x = 0,5$*. Alors* $\alpha = 20$ *et la probabilité de fixation de l'allèle* A *vaut environ* $0,999955$*. Ce faible avantage sélectif, inobservable en laboratoire ou par des mesures statistiques, est pourtant suffisant pour avoir un effet déterminant sur la fixation des allèles. Le calcul de l'espérance du temps de fixation est plus délicat. L'équation différentielle suivante :*

$$z''(x) + 2\alpha z'(x) = -\frac{1}{x(1-x)}, \quad \text{avec } z(0) = 0 \text{ et } z(1) = 0,$$

n'a pas de solution explicite. La figure 26.3 propose une illustration de l'estimation du temps d'absorption pour $N = 30$ *et* $\alpha = 10$*.*

Exemple 26.15 (Mutation sans sélection)**.** *On suppose ici que* α *est nul. La diffusion est solution de*

$$dX_t = \sqrt{X_t(1-X_t)}\, dB_t - \beta_A X_t\, dt + \beta_B(1-X_t)\, dt.$$

Recueil de modèles stochastiques – Page 251/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

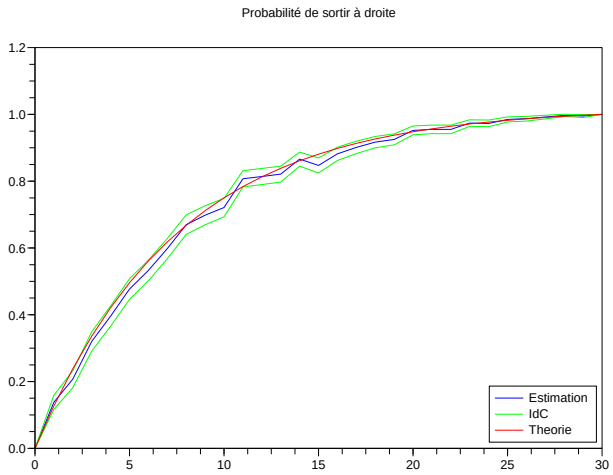


Figure 26.2 – Probabilité de fixation de A pour $N = 30$ et $\alpha = 2$.

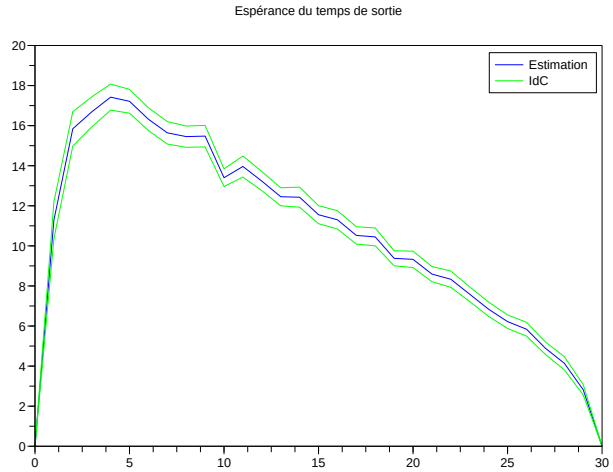


Figure 26.3 – Espérance du temps de fixation pour $N = 30$ et $\alpha = 10$.

Cette diffusion reste dans l'intervalle $]0,1[$*, c'est-à-dire que* $T_{0,1}$ *est infini p.s. De plus, elle admet pour mesure invariante* ν *la loi Beta de paramètres* $2\beta_B$ *et* $2\beta_A$*, c'est-à-dire que*

$$\nu(y) = \frac{\Gamma(2(\beta_A + \beta_B))}{\Gamma(2\beta_A)\Gamma(2\beta_B)} y^{2\beta_B - 1}(1-y)^{2\beta_A - 1}.$$

En particulier, si Y *suit la loi* ν*,*

$$\mathbb{E}(Y) = \frac{\beta_B}{\beta_A + \beta_B} \quad \text{et} \quad \mathbb{V}(Y) = \frac{\beta_A \beta_B}{(\beta_A + \beta_B)^2 (2(\beta_A + \beta_B) + 1)}.$$

Remarque 26.16 (Modèle multi-allèles)**.** *Dans le modèle à* k *allèles distincts avec un taux de mutation total* $\theta/2$ *et des probabilités respectives de mutation* $(\pi_i)_{1 \leqslant i \leqslant k}$*, le processus des proportions des allèles est à valeurs dans le simplexe* Δ *de* $\mathbb{R}^k$*, de dimension* $k-1$*, et la*

Recueil de modèles stochastiques – Page 252/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

densité de sa mesure invariante par rapport à la mesure de Lebesgue sur Δ est

$$\nu(y_1,\dots,y_K) = \frac{\Gamma(\theta)}{\Gamma(\theta\pi_1)\cdots\Gamma(\theta\pi_K)} y_1^{\Gamma(\theta\pi_1)-1}\cdots y_K^{\Gamma(\theta\pi_K)-1}\mathbf{1}_\Delta(y_1,\dots,y_k).$$

26.5 Schéma d'Euler

Considérons le processus de diffusion X solution de l'équation différentielle stochastique suivante :

$$dX_t = \sigma(X_t)\,dB_t + b(X_t)dt.$$

Sauf cas particulier, cette équation n'admet de solution explicite. On cherche donc à l'approcher par une chaîne de Markov facile à simuler. L'idée la plus simple, et la plus robuste en pratique, est de procéder comme pour les équations différentielles ordinaires en introduisant le schéma d'Euler associé à une discrétisation du temps. Soit $\alpha > 0$, le *schéma d'Euler* de pas α est la chaîne de Markov $(Y_n^\alpha)_{n\geqslant 0}$ à valeurs dans $\mathbb{R}$ associée au noyau

$$\mathbf{P}_\alpha(x,\cdot) = \mathcal{N}\left(x+\alpha b(x), \alpha\frac{\sigma(x)^2}{2}\right).$$

On peut l'étendre à un processus de diffusion $(X_t^\alpha)_{t\geqslant 0}$ dont les coefficients sont constants entre de temps multiples de α : pour $t\in[\alpha n,\alpha(n+1)[$,

$$\begin{aligned} X_t^\alpha &= X_{n\alpha}^\alpha + \int_{n\alpha}^t b(X_{n\alpha}^\alpha)\,ds + \int_{t_\alpha}^t \sigma(X_{n\alpha}^\alpha)dB_s \\ &= X_{n\alpha}^\alpha + (t-n\alpha)b(X_{n\alpha}^\alpha) + \sigma(X_{n\alpha}^\alpha)(B_t - B_{n\alpha}). \end{aligned}$$

On a alors $Y_n^\alpha = X_{n\alpha}^\alpha$ pour tout $n\in\mathbb{N}$.

Théorème 26.17 (Erreur faible)**.** *Si les coefficients b et σ sont des fonctions de classe $\mathcal{C}^4$ à dérivées bornées, alors, pour toute fonction f mesurable bornée et tout $t>0$, il existe $C>0$ tel que*

$$|\mathbb{E}_x f(X_t) - \mathbb{E}_x f(X_t^\alpha)| \leqslant C\alpha.$$

Démonstration. Esquissons ici une idée de la preuve. On écrit

$$\mathbf{P}_\alpha^n f - P_{\alpha n} f = \sum_{k=1}^n \mathbf{P}_\alpha^{k-1}(\mathbf{P}_\alpha - P_\alpha)P_{\alpha(n-k)}f.$$

Or les développements limités des deux noyaux de probabilité P_α et $\mathbf{P}_\alpha$ coïncident jusqu'à l'ordre 2. En effet, puisque $\partial_t P_t = LP_t$,

$$P_\alpha g(x) = g(x) + \alpha Lg(x) + \mathcal{O}(\alpha^2)$$

Soit Y de loi $\mathcal{N}(0,1)$. En posant $\delta(x) = \sqrt{\alpha}\sigma(x)Y + \alpha b(x)$, on a

$$\begin{aligned} \mathbf{P}_\alpha g(x) &= \mathbb{E}g(x+\alpha b(x) + \sqrt{\alpha}\sigma(x)Y) \\ &= \mathbb{E}\left(g(x) + \delta(x)g'(x) + \frac{\delta(x)^2}{2}g''(x) + \frac{\delta(x)^3}{6}g'''(x) + \mathcal{O}(\delta(x)^3)\right) \\ &= g(x) + \alpha Lg(x) + \mathcal{O}(\alpha^2) \end{aligned}$$

puisque les moments d'ordres 1 et 3 de Y sont nuls. L'erreur entre diffusion et schéma d'Euler sur un pas de temps α est donc de l'ordre de α^2. Au bout de n itérations, l'erreur sera donc d'ordre α (puisque αn est de l'ordre de t). Reste une difficulté de taille : montrer que l'on contrôle les dérivées de $P_{\alpha(n-k)}f$ pour que le développement limité ci-dessus soit valide. Ceci découle de propriétés de régularité de la solution de l'équation aux dérivées partielles parabolique $\partial_t u = Lu$. □

Recueil de modèles stochastiques – Page 253/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Il est également possible de contrôler l'écart trajectoriel entre les processus X et X^α.

Théorème 26.18 (Erreur forte)**.** *S'il existe $C > 0$ tel que pour tous $x, y \in \mathbb{R}$,*

$$|b(x) - b(y)| + |\sigma(x) - \sigma(y)| \leqslant C|x - y|,$$

alors, pour tout $p \geqslant 1$ et tout $T > 0$, il existe $K_p(T)$ tel que

$$\mathbb{E}\left(\sup_{0 \leqslant t \leqslant T} |X_t - X_t^\alpha|^p\right) \leqslant K_p(T)\alpha^{p/2}.$$

La démonstration de ce résultat est technique mais assez directe. Elle repose essentiellement sur la formule d'Itô, le lemme de Gronwall et l'inégalité de Burkholder-Davis-Gundy.

26.6 Notes et commentaires

De nombreux ouvrages sont consacrés au calcul stochastique et aux processus de diffusion. On trouvera notamment dans [GS72] l'étude des diffusions sur un intervalle de $\mathbb{R}$ et leurs comportements aux bords du domaine. Les différentes variantes et subtilités des résultats de convergence de processus sont rassemblées dans le livre [EK86] de Stewart Ethier et Thomas Kurtz. La section 7.4 de cet ouvrage très dense concerne notamment l'approximation diffusive et propose une démonstration du théorème 26.5.

Citons encore deux exemples classiques de convergence pour des processus de Markov à temps continu et espaces d'états discrets vers une diffusion.

Pour $N \in \mathbb{N}^*$, soit $Z^{N\lambda,\mu}$ la *file d'attente $M/M/\infty$* [2] de taux d'arrivée $N\lambda$ et de taux de service μ. Pour tous $N \geqslant 1$ et $t \geqslant 0$, posons

$$X_t^N = \sqrt{N}\left(\frac{Z_t^{N\lambda,\mu}}{N} - \frac{\lambda}{\mu}\right) \quad \text{avec} \quad Z_0^{N\lambda,\mu} = \lfloor N\lambda/\mu + \sqrt{N}x \rfloor.$$

La suite $(X_{\lfloor N\cdot \rfloor}^N)_{N \geqslant 1}$ converge en loi vers le processus d'Ornstein-Uhlenbeck solution de

$$dX_t = \sqrt{2\lambda}\, dB_t - \mu X_t\, dt \quad \text{avec} \quad X_0 = x,$$

De même, soit une suite $(Y^N)_{N \geqslant 1}$ de *processus de Yule* de paramètres respectifs (λ_N, μ_N), c'est-à-dire que Y^N est le processus de naissance et mort de taux respectifs $(\lambda_N n)_{n \in \mathbb{N}}$ et $(\mu_N n)_{n \in \mathbb{N}}$. On suppose que

$$N(\lambda_N - \mu_N) = b + \mathcal{O}(1/N) \quad \text{et} \quad \lambda_N + \mu_N = 2a + \mathcal{O}(1/N).$$

Alors la suite de processus $(Y_{\lfloor N\cdot \rfloor}^N/N)_{N \geqslant 1}$ converge vers la diffusion X solution de l'équation différentielle suivante :

$$dX_t = \sqrt{2aX_t}\, dB_t + bX_t\, dt.$$

Les schémas d'Euler associés à des processus de diffusion ont été très largement étudiés. Citons les ouvrages de synthèse de Peter Kloeden et Eckhard Kloeden [KP95], Nicolas Bouleau et Dominique Lépingle [BL94] pour la preuve du théorème 26.18. La contribution de Denis Talay [Tal96] propose également une synthèse sur les estimations de l'erreur faible. Il est possible d'améliorer le théorème 26.17 en obtenant un développement limité de l'erreur faible en puissances de α. Ceci permet d'accélérer la convergence : en combinant des schémas de pas 2α et α on obtient une erreur de l'ordre de α^2 ! Cette méthode, appelé extrapolation de Romberg, est également utilisée en analyse numérique des équations différentielles ordinaires. Les deux travaux fondateurs dans cette directions sont dus à Denis Talay et Luciano Tubaro [TT90] qui supposent que la fonction test est très régulière puis à Vlad Bally et Denis Talay [BT95] qui, grâce au calcul de Malliavin obtiennent le résultat pour des fonctions mesurables bornées si les coefficients σ et b sont réguliers.

2. Ce processus de Markov est étudié dans le chapitre 18.

Recueil de modèles stochastiques – Page 254/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Table des figures

1.1 Lois binomiales de même taille, pour trois valeurs du paramètre, illustrant la pertinence de l'approximation par une loi normale et par une loi de Poisson selon les cas. 3
1.2 Approximations de la queue de T par celle de $r\log(r) + rG$ où G suit la loi de Gumbel en vertu du théorème 1.15, ainsi que par une méthode de Monte-Carlo avec N tirages en utilisant la représentation de T en somme de v.a. géométriques indépendantes. 9

2.1 Trajectoires issues de 0 de la marche aléatoire simple sur $\mathbb{Z}$. 13
2.2 Réflexion de Désiré André (preuve du théorème 2.5). 16
2.3 Réflexion de Désiré André (preuve du théorème 2.6, $n = 10, k = 2$). 17
2.4 Début de trajectoire de la marche aléatoire simple sur $\mathbb{Z}^2$. 18
2.5 La carte en position k se déplace en position $\sigma(k)$. 23

3.1 Répartition des particules dans les deux urnes. 27
3.2 Évolution de la proportion de particules dans l'urne A pour différentes valeurs initiales 0, 1000, et 4000, pour $a = 4000$ particules. 33

4.1 Quelques trajectoires, jusqu'au temps $t = 200$, de la chaîne de Fisher-Wright pour une population de taille $N = 100$, toutes issues de $X_0 = N/4 = 25$. 39
4.2 Loi invariante, et histogramme d'un échantillon de taille 10000 de X_{100} avec $X_0 = 50$, pour une population de taille $N = 100$, avec taux de mutation $u = v = 0.05$. 43

5.1 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. 52
5.2 Arbre et marche aléatoire associée. 57
5.3 Une marche parmi n reste positive avant d'atteindre -1. 57

6.1 Réalisation du graphe aléatoire (V, F) du graphe euclidien $\mathcal{E}_2$ sous $\mathbb{P}_{1/2}$. 62
6.2 Les premières branches de $\mathcal{B}_4$ avec des arêtes supprimées par $\mathbb{P}_p$ (en pointillé). . 63
6.3 Percolation pour $\mathcal{B}_r$ avec $r = 3$ (trait plein) et $r = 4$ (trait en pointillé). 64
6.4 À gauche, le graphe euclidien $\mathcal{E}_2$ (trait plein) et son dual $\mathcal{E}_2^*$ (trait pointillé). À droite, une composante finie et sa boucle ceinturante. 67

7.1 Quelques trajectoires de l'urne de Pólya. 73
7.2 Réalisation d'un graphe à attachement préférentiel de Barabási-Albert. 74
7.3 Occupation d'un chemin pour la marche renforcée géométriquement. 76

Recueil de modèles stochastiques – Page 255/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

8.1 Quelques étapes de l'algorithme du recuit pour le voyageur de commerce ($n = 10$). 91

9.1 Configuration avec $n = 6$ clients répartis sur 3 tables dans le restaurant chinois. . 96
9.2 Histogramme d'un échantillon de taille 5000 de la loi d'Ewens de taille $n = 1000$ et de paramètre $\theta = 1$. 98

10.1 Dans le modèle de Fisher-Wright, tout se passe comme si pour chaque génération, chaque individu, indépendemment de tous les autres, choisissait uniformément un père dans la génération précédente et en héritait. Le graphique ci-dessus illustre par exemple l'évolution de ces relations de filiation sur quelques générations pour une population de taille $N = 5$. Ce point de vue généalogique est cœur de ce chapitre. 106

11.1 Une réalisation de $A(1000)$ en dimension 2. 120

12.1 Trajectoire de U (courbe continue) et probabilité de lissage pour $v = 0$ (courbe en pointillés) avec $l = 200$ et les transitions de l'exemple 12.1. 126

13.1 Estimation du mélange de trois gaussiennes par l'algorithme EM. 136

14.1 Convergence p.s. de la suite $(\overline{Y}_n - \log n)_n$. 140
14.2 Histogramme de la limite p.s. de la suite $(\overline{Y}_n - \log n)_n$ et densité de la loi de Gumbel. 141
14.3 Convergence de la probabilité optimale du meilleur choix (croix) et de la proportion idéale (trait continu) vers $1/e$. 144

16.1 Illustration du théorème 16.1 : densité de la loi du demi-cercle μ_1^{DC} et histogramme des valeurs propres d'une matrice symétrique de taille $n = 1000$ avec la loi uniforme sur $[-\sqrt{3/n}, \sqrt{3/n}]$ (dont la variance vaut $1/n$). 156
16.2 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 $t = 3$. 161
16.3 Illustration du théorème 16.7 : histogramme des valeurs propres non nulles d'une matrice de covariance empirique avec $n = 800$ et $d = 1000$ et loi de Marchenko-Pastur. 162
16.4 Partie à densité de μ_ρ^{MP} pour différentes valeurs de ρ. 162

17.1 Trajet le plus court pour $n = 20$ points uniformément répartis dans le carré unité obtenu par l'algorithme stochastique du recuit simulé (chapitre 8). 168
17.2 Tracé d'une approximation de la courbe $n \mapsto n^{-1/2}\mathbb{E}(L_n)$ dans le cas uniforme, avec une méthode de Monte-Carlo et l'algorithme stochastique du recuit simulé (chapitre 8). 168

18.1 Un début de trajectoire de file d'attente M/M/1 de paramètres 1 et 3, et de file d'attente M/M/∞ de paramètres 1 et .5. Les deux trajectoires sont issues de 100. Le premier processus a pour loi invariante la loi géométrique sur $\mathbb{N}$ de paramètre 1/3, et le second la loi de Poisson de paramètre $1/.5 = 2$. Dans les deux cas, la loi invariante a pour moyenne 2. La première trajectoire à un comportement linéaire et la seconde un comportement exponentiel. 184

19.1 Une trajectoire menant à la ruine au temps $\tau(c)$ avec un découvert $Y(c)$. 188

21.1 Allure d'une trajectoire typique du processus X. 200
21.2 Histogramme d'un échantillon de X_t pour $t = 100$, $\lambda = 1$, $\mathcal{Q} = \delta_{1/2}$. 202
21.3 Début de l'arbre binaire définissant le processus N. 205

22.1 Une trajectoire. 213
22.2 Décomposition d'une excursion. Les points signalent les sauts de la vitesse. . . . 214

Recueil de modèles stochastiques – Page 256/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

23.1 Découpage d'une trajectoire sortant de D après être sortie de $B(x,r)$. 219
23.2 Une trajectoire brownienne issue de x arrêtée hors d'une couronne. 223

25.1 Évolution des vecteurs moyenne (x_t, v_t) (à gauche) et covariance (a_t, b_t, c_t) (à droite) pour $\sigma = 1$, $\lambda = 1$ et $\mu = 10$. 238
25.2 Évolution en fonction du temps du déterminant de la matrice de covariance issu de $(x_0, v_0, a_0, b_0, c_0) = (0, 0, 1, 1, 1)$ à gauche et $(0, 0, 0, 0, 0)$ à droite pour $\sigma = 1$, $\lambda = 1$ et $\mu = 10$. 239
25.3 Évolutions comparées de l'entropie relative et de la distance de Wasserstein d'ordre 2 entre la loi au temps t et la mesure invariante $(x_0, v_0, a_0, b_0, c_0) = (1, 0, 0, 0, 0)$ pour $\sigma = 1$, $\lambda = 1$ et $\mu = 10$ (à gauche) et $\mu = 60$ (à droite). 242

26.1 Estimation du temps de fixation pour $N = 40$ en l'absence de sélection et de mutation. 251
26.2 Probabilité de fixation de A pour $N = 30$ et $\alpha = 2$. 252
26.3 Espérance du temps de fixation pour $N = 30$ et $\alpha = 10$. 252

Recueil de modèles stochastiques – Page 257/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Recueil de modèles stochastiques – Page 258/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Bibliographie

[AB99] R. Albert et A.-L. Barabási – « Emergence of scaling in random networks », *Science* (1999), no. 286, p. 509–512.

[AB02] — , « Statistical mechanics of complex networks », *Rev. Modern Phys.* **74** (2002), no. 1, p. 47–97.

[ABC+00] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto et G. Scheffer – *Sur les inégalités de Sobolev logarithmiques*, Panoramas et Synthèses [Panoramas and Syntheses], vol. 10, Société Mathématique de France, Paris, 2000, With a preface by Dominique Bakry and Michel Ledoux.

[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.

[ACG+99] G. Ausiello, P. Crescenzi, G. Gambosi, V. Kann, A. Marchetti-Spaccamela et M. Protasi – *Complexity and approximation*, Springer-Verlag, Berlin, 1999, Combinatorial optimization problems and their approximability properties, With 1 CD-ROM (Windows and UNIX).

[AF01] D. J. Aldous et J. A. Fill – « Reversible Markov Chains and Random Walks on Graphs », cours non publié disponible sous forme électronique sur Internet, 1994-2001.

[AG13a] A. Asselah et A. Gaudillière – « From logarithmic to subdiffusive polynomial fluctuations for internal dla and related growth models », *The Annals of Probability* **41** (2013), no. 3A, p. 1115–1159.

[AG13b] — , « Sublogarithmic fluctuations for internal dla », *The Annals of Probability* **41** (2013), no. 3A, p. 1160–1179.

[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.

[AHU75] A. V. Aho, J. E. Hopcroft et J. D. Ullman – *The design and analysis of computer algorithms*, Addison-Wesley Publishing Co., Reading, Mass.-London-Amsterdam, 1975, Second printing, Addison-Wesley Series in Computer Science and Information Processing.

Recueil de modèles stochastiques – Page 259/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

[AK90] D. Aldous et W. B. Krebs – « The “birth-and-assassination” process », *Statist. Probab. Lett.* **10** (1990), no. 5, p. 427–430.

[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].

[And87] D. André – « Solution directe du problème résolu par M. Bertrand », *Comptes Rendus Acad. Sci. Paris* **105** (1887), p. 436–437.

[Asm00] S. Asmussen – *Ruin probabilities*, Advanced Series on Statistical Science & Applied Probability, vol. 2, World Scientific Publishing Co. Inc., River Edge, NJ, 2000.

[Bak94] D. Bakry – « L’hypercontractivité et son utilisation en théorie des semigroupes », Lectures on probability theory (Saint-Flour, 1992), Lecture Notes in Math., vol. 1581, Springer, Berlin, 1994, p. 1–114.

[Bas95] R. F. Bass – *Probabilistic techniques in analysis*, Probability and its Applications (New York), Springer-Verlag, New York, 1995.

[BC07] B. Bercu et D. Chafaï – *Modélisation stochastique et simulation, cours et applications*, Collection Sciences Sup, Dunod, 2007, Mathématiques appliquées/SMAI.

[BDMT11] V. Bansaye, J.-F. Delmas, L. Marsalle et V. C. Tran – « Limit theorems for Markov processes indexed by continuous time Galton-Watson trees », *Ann. Appl. Probab.* **21** (2011), no. 6, p. 2263–2314.

[BÉ85] D. Bakry et M. Émery – « Diffusions hypercontractives », Séminaire de probabilités, XIX, 1983/84, Lecture Notes in Math., vol. 1123, Springer, Berlin, 1985, p. 177–206.

[Ber76] C. Berge – *Graphs and hypergraphs*, revised éd., North-Holland Publishing Co., Amsterdam-London ; American Elsevier Publishing Co., Inc., New York., 1976, Translated from the French by Edward Minieka, North-Holland Mathematical Library, Vol. 6.

[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.

[BGL14] D. Bakry, I. Gentil et M. Ledoux – *Analysis and geometry of Markov diffusion operators*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 348, Springer, Cham, 2014.

[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.

[BK05] M. Benaïm et N. Karoui – *Promenade aléatoire : chaînes de markov et simulations ; martingales et stratégies*, Mathématiques appliquées, Les éd. de l’École polytechnique, 2005.

Recueil de modèles stochastiques – Page 260/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

[BL94] N. Bouleau et D. Lépingle – *Numerical methods for stochastic processes*, Wiley Series in Probability and Mathematical Statistics : Applied Probability and Statistics, John Wiley & Sons, Inc., New York, 1994, A Wiley-Interscience Publication.

[BMR14] S. Bubeck, E. Mossel et M. Z. Rácz – « On the influence of the seed graph in the preferential attachment model », arXiv :1401.4849, 2014.

[Bod14] T. Bodineau – « Promenade aléatoire : chaînes de Markov et martingales », Cours de l'École Polytechnique, Palaiseau, France, 2014.

[Bol80] B. Bollobás – « A probabilistic proof of an asymptotic formula for the number of labelled regular graphs », *European J. Combin.* **1** (1980), no. 4, p. 311–316.

[Bol89] E. Bolthausen – « A note on the diffusion of directed polymers in a random environment », *Comm. Math. Phys.* **123** (1989), no. 4, p. 529–534.

[Bor08] C. Bordenave – « On the birth-and-assassination process, with an application to scotching a rumor in a network », *Electron. J. Probab.* **13** (2008), p. no. 66, 2014–2030.

[Bor14a] C. Bordenave – « Lecture notes on random graphs and probabilistic combinatorial optimization », Lecture notes, 2014.

[Bor14b] — , « A short course on random matrices », Lecture notes, 2014.

[BPPC99] C. Bouzitat, G. Pagès, F. Petit et F. Carrance – *En passant par hasard…les probabilités de tous les jours*, Vuibert, 1999.

[BR11] M. Benaïm et O. Raimond – « Self-interacting diffusions IV : Rate of convergence », *Electron. J. Probab.* **16** (2011), p. no. 66, 1815–1843.

[BRST01] B. Bollobás, O. Riordan, J. Spencer et G. Tusnády – « The degree sequence of a scale-free random graph process », *Random Structures Algorithms* **18** (2001), no. 3, p. 279–290.

[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.

[BT95] V. Bally et D. Talay – « The Euler scheme for stochastic differential equations : error analysis with Malliavin calculus », *Math. Comput. Simulation* **38** (1995), no. 1-3, p. 35–41, Probabilités numériques (Paris, 1992).

[Cap12] M. Capitaine – « Introduction aux grandes matrices aléatoires », Notes de cours, 2012.

[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.

[CD04] A. Charpentier et M. Denuit – *Mathématiques de l'assurance non-vie - concepts fondamentaux de théorie du risque, tome 1*, Economica, 2004.

[CDM13] D. Chafaï, Y. Doumerc et F. Malrieu – « Processus des restaurants chinois et loi d'ewens », *RMS* **123** (2013), no. 3, p. 56–74.

Recueil de modèles stochastiques – Page 261/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

[CDPP09] P. Caputo, P. Dai Pra et G. Posta – « Convex entropy decay via the Bochner-Bakry-Emery approach », *Ann. Inst. Henri Poincaré Probab. Stat.* **45** (2009), no. 3, p. 734–753.

[CGLP12] D. Chafaï, O. Guédon, G. Lecué et A. Pajor – *Interactions between compressed sensing random matrices and high dimensional geometry*, Panoramas et Synthèses [Panoramas and Syntheses], vol. 37, Société Mathématique de France, Paris, 2012.

[CH02] P. Carmona et Y. Hu – « On the partition function of a directed polymer in a Gaussian random environment », *Probab. Theory Related Fields* **124** (2002), no. 3, p. 431–457.

[Cha06] D. Chafaï – « Binomial-Poisson entropic inequalities and the $M/M/\infty$ queue », *ESAIM Probab. Stat.* **10** (2006), p. 317–339 (electronic).

[Cha13] — , « Introduction aux matrices aléatoires », Aléatoire, Ed. Éc. Polytech., Palaiseau, 2013, p. 87–122.

[Clo13] B. Cloez – « Comportement asymptotique de processus avec sauts et applications pour des modèles avec branchement », Thèse, Université Paris-Est, 2013.

[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.

[CSY03] F. Comets, T. Shiga et N. Yoshida – « Directed polymers in a random environment : path localization and strong disorder », *Bernoulli* **9** (2003), no. 4, p. 705–723.

[CV06] F. Comets et V. Vargas – « Majorizing multiplicative cascades for directed polymers in random media », *ALEA Lat. Am. J. Probab. Math. Stat.* **2** (2006), p. 267–277.

[CY06] F. Comets et N. Yoshida – « Directed polymers in random environment are diffusive at weak disorder », *Ann. Probab.* **34** (2006), no. 5, p. 1746–1770.

[Dav90] B. Davis – « Reinforced random walk », *Probab. Theory Related Fields* **84** (1990), no. 2, p. 203–229.

[DCD83] D. Dacunha-Castelle et M. Duflo – *Probabilités et statistiques. Tome 2*, Collection Mathématiques Appliquées pour la Maîtrise. [Collection of Applied Mathematics for the Master's Degree], Masson, Paris, 1983, Problèmes à temps mobile. [Movable-time problems].

[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.

[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].

[DLR77] A. P. Dempster, N. M. Laird et D. B. Rubin – « Maximum likelihood from incomplete data via the EM algorithm », *J. Roy. Statist. Soc. Ser. B* **39** (1977), no. 1, p. 1–38, With discussion.

Recueil de modèles stochastiques – Page 262/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

[DM47] A. Dvoretzky et T. Motzkin – « A problem of arrangements », *Duke Math. J.* **14** (1947), p. 305–313.

[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. Doumas 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.

[DS87] P. Diaconis et M. Shahshahani – « Time to reach stationarity in the Bernoulli-Laplace diffusion model », *SIAM J. Math. Anal.* **18** (1987), no. 1, p. 208–218.

[DST14] M. Disertori, C. Sabot et P. Tarrès – « Transience of Edge-Reinforced Random Walk », arXiv :1403.6079, 2014.

[Dur08] R. Durrett – *Probability models for DNA sequence evolution*, second éd., Probability and its Applications (New York), Springer, New York, 2008.

[Dur10] R. Durrett – *Random graph dynamics*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, 2010.

[Dwa69] M. Dwass – « The total progeny in a branching process and a related random walk. », *J. Appl. Probability* **6** (1969), p. 682–686.

[DZ10] A. Dembo et O. Zeitouni – *Large deviations techniques and applications*, Stochastic Modelling and Applied Probability, vol. 38, Springer-Verlag, Berlin, 2010, Corrected reprint of the second (1998) edition.

[EG60] P. Erdős et T. Gallai – « Gráfok előírt fokszámú pontokkal », *Matematikai Lapok* (1960), no. 11, p. 264–274.

[EK86] S. N. Ethier et T. G. Kurtz – *Markov processes*, Wiley Series in Probability and Mathematical Statistics : Probability and Mathematical Statistics, John Wiley & Sons Inc., New York, 1986, Characterization and convergence.

[EO05] R. Erban et H. G. Othmer – « From individual to collective behavior in bacterial chemotaxis », *SIAM J. Appl. Math.* **65** (2004/05), no. 2, p. 361–391 (electronic).

[ER61] P. Erdős et A. Rényi – « On the evolution of random graphs », *Bull. Inst. Internat. Statist.* **38** (1961), p. 343–347.

[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.

Recueil de modèles stochastiques – Page 263/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

[Fer89] T. S. Ferguson – « Who solved the secretary problem ? », *Statist. Sci.* **4** (1989), no. 3, p. 282–296, With comments and a rejoinder by the author.

[FF04] D. Foata et A. Fuchs – *Processus stochastiques : processus de poisson, chaînes de markov et martingales*, Collection Sciences Sup, Dunod, 2004.

[Fil98] J. A. Fill – « An interruptible algorithm for perfect sampling via Markov chains », *Ann. Appl. Probab.* **8** (1998), no. 1, p. 131–162.

[Fre65] D. A. Freedman – « Bernard Friedman's urn », *Ann. Math. Statist* **36** (1965), p. 956–970.

[FY48] R. A. Fisher et F. Yates – *Statistical Tables for Biological, Agricultural and Medical Research*, Oliver and Boyd, London, 1948, 3d ed.

[GJ87] J. Glimm et A. Jaffe – *Quantum physics*, second éd., Springer-Verlag, New York, 1987, A functional integral point of view.

[GR10] C. Graham et P. Robert – « A multi-class mean-field model with graph structure for TCP flows », Progress in industrial mathematics at ECMI 2008, Math. Ind., vol. 15, Springer, Heidelberg, 2010, p. 125–131.

[Gri99] G. Grimmett – *Percolation*, second éd., Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 321, Springer-Verlag, Berlin, 1999.

[Gri10] — , *Probability on graphs*, Institute of Mathematical Statistics Textbooks, vol. 1, Cambridge University Press, Cambridge, 2010, Random processes on graphs and lattices.

[GS72] Ĭ. Ī. Gīhman et A. V. Skorohod – *Stochastic differential equations*, Springer-Verlag, New York, 1972, Translated from the Russian by Kenneth Wickwire, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 72.

[GS84] C. R. Givens et R. M. Shortt – « A class of Wasserstein metrics for probability distributions », *Michigan Math. J.* **31** (1984), no. 2, p. 231–240.

[GS01] G. R. Grimmett et D. R. Stirzaker – *Probability and random processes*, third éd., Oxford University Press, New York, 2001.

[Häg02] O. Häggström – *Finite markov chains and algorithmic applications*, London Mathematical Society student texts, Cambridge University Press, 2002.

[Har60] T. E. Harris – « A lower bound for the critical probability in a certain percolation process », *Proc. Cambridge Philos. Soc.* **56** (1960), p. 13–20.

[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.

[HJ06] O. Häggström et J. Jonasson – « Uniqueness and non-uniqueness in percolation theory », *Probab. Surv.* **3** (2006), p. 289–344.

[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.

[Hol01] L. Holst – « Extreme value distributions for random coupon collector and birthday problems », *Extremes* **4** (2001), no. 2, p. 129–145 (2002).

Recueil de modèles stochastiques – Page 264/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

[Hör67] L. Hörmander – « Hypoelliptic second order differential equations », *Acta Math.* **119** (1967), p. 147–171.

[HV10] S. Herrmann et P. Vallois – « From persistent random walk to the telegraph noise », *Stoch. Dyn.* **10** (2010), no. 2, p. 161–196.

[IS88] J. Z. Imbrie et T. Spencer – « Diffusion of directed polymers in a random environment », *J. Statist. Phys.* **52** (1988), no. 3-4, p. 609–626.

[JK77] N. L. Johnson et S. Kotz – *Urn models and their application*, John Wiley & Sons, New York-London-Sydney, 1977, An approach to modern discrete probability theory, Wiley Series in Probability and Mathematical Statistics.

[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.

[JLS12] D. Jerison, L. Levine et S. Sheffield – « Logarithmic fluctuations for internal DLA », *J. Amer. Math. Soc.* **25** (2012), no. 1, p. 271–301.

[JLS13] D. Jerison, L. Levine et S. Sheffield – « Internal dla in higher dimensions », *Electron. J. Probab.* **18** (2013), p. no. 98, 1–14.

[Kac47] M. Kac – « Random walk and the theory of Brownian motion », *Amer. Math. Monthly* **54** (1947), p. 369–391.

[Kac74] M. Kac – « A stochastic model related to the telegrapher's equation », *Rocky Mountain J. Math.* **4** (1974), p. 497–509, Reprinting of an article published in 1956, Papers arising from a Conference on Stochastic Differential Equations (Univ. Alberta, Edmonton, Alta., 1972).

[KB97] S. Kotz et N. Balakrishnan – « Advances in urn models during the past two decades », Advances in combinatorial methods and applications to probability and statistics, Stat. Ind. Technol., Birkhäuser Boston, Boston, MA, 1997, p. 203–257.

[Kes80] H. Kesten – « The critical probability of bond percolation on the square lattice equals $\frac{1}{2}$ », *Comm. Math. Phys.* **74** (1980), no. 1, p. 41–59.

[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.

[KP95] P. E. Kloeden et E. Platen – « Numerical methods for stochastic differential equations », Nonlinear dynamics and stochastic mechanics, CRC Math. Model. Ser., CRC, Boca Raton, FL, 1995, p. 437–461.

[KPSC08] A. M. Krieger, M. Pollak et E. Samuel-Cahn – « Beat the mean : sequential selection by better than average rules », *J. Appl. Probab.* **45** (2008), no. 1, p. 244–259.

[KS76] J. G. Kemeny et J. L. Snell – *Finite Markov chains*, Springer-Verlag, New York, 1976, Reprinting of the 1960 original, Undergraduate Texts in Mathematics.

Recueil de modèles stochastiques – Page 265/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

[Lac10] H. Lacoin – « New bounds for the free energy of directed polymers in dimension $1+1$ and $1+2$ », *Comm. Math. Phys.* **294** (2010), no. 2, p. 471–503.

[Law13] G. F. Lawler – *Intersections of random walks*, Modern Birkhäuser Classics, Birkhäuser/Springer, New York, 2013, Reprint of the 1996 edition.

[LBG92] G. F. Lawler, M. Bramson et D. Griffeath – « Internal diffusion limited aggregation », *Ann. Probab.* **20** (1992), no. 4, p. 2117–2140.

[LL10] G. F. Lawler et V. Limic – *Random walk : a modern introduction*, Cambridge Studies in Advanced Mathematics, vol. 123, Cambridge University Press, Cambridge, 2010.

[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.

[LW09] Q. Liu et F. Watbled – « Exponential inequalities for martingales and asymptotic properties of the free energy of directed polymers in a random environment », *Stochastic Process. Appl.* **119** (2009), no. 10, p. 3101–3132.

[Lé08] T. Lévy – « Probabilistic methods in statistical physics », Lecture notes - Tsinghua University, October 2008.

[Mah09] H. M. Mahmoud – *Pólya urn models*, Texts in Statistical Science Series, CRC Press, Boca Raton, FL, 2009.

[Mar04] J.-F. Marckert – « Structures arborescentes, algorithmes », Habilitation à diriger des recherches, 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.

[MM09] M. Mézard et A. Montanari – *Information, physics, and computation*, Oxford Graduate Texts, Oxford University Press, Oxford, 2009.

[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.

[MW90] B. D. McKay et N. C. Wormald – « Uniform generation of random regular graphs of moderate degree », *J. Algorithms* **11** (1990), no. 1, p. 52–67.

[Nel67] E. Nelson – *Dynamical theories of Brownian motion*, Princeton University Press, Princeton, N.J., 1967.

[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.

Recueil de modèles stochastiques – Page 266/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

[ODA88] H. G. Othmer, S. R. Dunbar et W. Alt – « Models of dispersal in biological systems », *J. Math. Biol.* **26** (1988), no. 3, p. 263–298.

[OK08] T. J. Ott et J. H. B. Kemperman – « Transient behavior of processes in the TCP paradigm », *Probab. Engrg. Inform. Sci.* **22** (2008), no. 3, p. 431–471.

[Par07] E. Pardoux – *Processus de Markov et applications*, Collection Sciences Sup, Dunod, 2007, Mathématiques appliquées/SMAI.

[Pem07] R. Pemantle – « A survey of random processes with reinforcement », *Probab. Surv.* **4** (2007), p. 1–79.

[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.

[Pre00] J. Preater – « Sequential selection with a better-than-average rule », *Statist. Probab. Lett.* **50** (2000), no. 2, p. 187–191.

[PW96] J. D. Propp et D. B. Wilson – « Exact sampling with coupled Markov chains and applications to statistical mechanics », *Random Structures Algorithms* **9** (1996), no. 1-2, 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.

[RC04] C. P. Robert et G. Casella – *Monte Carlo statistical methods*, second éd., Springer Texts in Statistics, Springer-Verlag, New York, 2004.

[Rob00] P. Robert – *Réseaux et files d'attente : méthodes probabilistes*, Mathématiques & Applications (Berlin) [Mathematics & Applications], vol. 35, Springer-Verlag, Berlin, 2000.

[Rob05] P. Robert – « Réseaux de communication, algorithmes et probabilités », Cours de l'École Polytechnique, Palaiseau, France, 2005.

[Rob10] — , « Mathématiques et réseau de communication », Dossiers pour la science (2010) hal :inria-00472217, 2010.

[Roy99] G. Royer – *Une initiation aux inégalités de Sobolev logarithmiques*, Cours Spécialisés [Specialized Courses], vol. 5, Société Mathématique de France, Paris, 1999.

[RRS05] S. Robin, F. Rodolphe et S. Schbath – *DNA, words and models*, Cambridge University Press, Cambridge, 2005, Translated from the 2003 French original.

[Rud87] W. Rudin – *Real and complex analysis*, third éd., McGraw-Hill Book Co., New York, 1987.

[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].

[She07] S. Sheffield – « Gaussian free fields for mathematicians », *Probab. Theory Related Fields* **139** (2007), no. 3-4, p. 521–541.

Recueil de modèles stochastiques – Page 267/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

[Spi70] F. Spitzer – *Principes des cheminements aléatoires*, Traduit de l'anglais par E. Baverez et J.-L. Guignard. Centre Interarmés de Recherche Opérationnelle, vol. 2, Dunod, Paris, 1970.

[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.

[Tal96] D. Talay – « Probabilistic numerical methods for partial differential equations : elements of analysis », Probabilistic models for nonlinear partial differential equations (Montecatini Terme, 1995), Lecture Notes in Math., vol. 1627, Springer, Berlin, 1996, p. 148–196.

[Tal02] D. Talay – « Stochastic Hamiltonian systems : exponential convergence to the invariant measure, and discretization by the implicit Euler scheme », *Markov Process. Related Fields* **8** (2002), no. 2, p. 163–198, Inhomogeneous random systems (Cergy-Pontoise, 2001).

[Tar11] P. Tarrès – « Localization of reinforced random walks », arXiv :1103.5536, 2011.

[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.

[TPH86] J. N. Tsitsiklis, C. H. Papadimitriou et P. Humblet – « The performance of a precedence-based queueing discipline », *J. Assoc. Comput. Mach.* **33** (1986), no. 3, p. 593–602.

[TT90] D. Talay et L. Tubaro – « Expansion of the global error for numerical schemes solving stochastic differential equations », *Stochastic Anal. Appl.* **8** (1990), no. 4, p. 483–509 (1991).

[TVW10] A. Tripathi, S. Venugopalan et D. B. West – « A short constructive proof of the Erdős-Gallai characterization of graphic lists », *Discrete Math.* **310** (2010), no. 4, p. 843–844.

[vdH14] R. van der Hofstad – « Random Graphs and Complex Networks », Lecture notes available on Internet, 2014.

[Vil03a] C. Villani – « Cercignani's conjecture is sometimes true and always almost true », *Comm. Math. Phys.* **234** (2003), no. 3, p. 455–490.

[Vil03b] — , *Topics in optimal transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003.

[Vil06] C. Villani – « Hypocoercive diffusion operators », International Congress of Mathematicians. Vol. III, Eur. Math. Soc., Zürich, 2006, p. 473–498.

[Vil09] — , « Hypocoercivity », *Mem. Amer. Math. Soc.* **202** (2009), no. 950, p. iv+141.

[Wil91] D. Williams – *Probability with martingales*, Cambridge Mathematical Textbooks, Cambridge University Press, Cambridge, 1991.

Recueil de modèles stochastiques – Page 268/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

[WL01] F. Wang et D. P. Landau – « Efficient, Multiple-Range Random Walk Algorithm to Calculate the Density of States », *Physical Review Letters* **86** (2001), p. 2050.

[Yca02] B. Ycart – *Modèles et algorithmes markoviens*, Mathématiques & Applications (Berlin) [Mathematics & Applications], vol. 39, Springer-Verlag, Berlin, 2002.

[Yuk98] J. E. Yukich – *Probability theory of classical Euclidean optimization problems*, Lecture Notes in Mathematics, vol. 1675, Springer-Verlag, Berlin, 1998.

Recueil de modèles stochastiques – Page 269/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Index

A
absence de mémoire, 80
agrégation limitée par diffusion interne, 115
algorithme de Fisher-Yates, 84
algorithme de Knuth, 84
algorithme de parcours en largeur, 68
algorithme de tri rapide, 84
algorithme de tri rapide randomisé, 92
algorithme du recuit simulé, 89
ancêtre commun le plus récent, 106
appariement, 85, 87
arbre, 61
arbre de Bethe, 63
arbre de Galton-Watson, 63
arêtes, 61
arrêt optimal, 143

B
battage de cartes, 23
bistochastique, 24

C
chaîne échantillonnée, 245
chaîne paresseuse, 245
champ libre gaussien, 22
chemin, 61
chemin auto-évitant, 61
chemins, 145
circuit hamiltonien, 167
coalescence, 91
coalescent de Kingman, 113
collectionneur de coupons, 1, 7, 18
connexe par arcs, 116
constante de connectivité, 65
couplage par le passé, 92
crible de Poincaré, 85
cycle, 61

D
degré d'un sommet, 87
dépoissonisation, 172
dérangement, 85
dimension temporelle, 145
distance de Wasserstein, 182
distance en variation totale, 4, 182
doublement stochastique, 24, 28

E
échantillonneur de Gibbs, 89
énergie libre, 146
entropie, 27
équation de Fokker-Planck, 228
équation de la chaleur avec dérive, 226
équation du télégraphe, 212
équation maître, 75
excursion, 213

F
file d'attente M/M/∞, 183, 254
fixation, 38
fonction contractante, 90
fonction d'acceptation-rejet, 89
fonction de Green, 22
fonction de partition, 146
fonction de partition normalisée, 146
fonction de renforcement, 77
fonction d'échelle, 246
fonction harmonique, 217
formule de la moyenne, 217
formule de Mehler, 226

G
générateur infinitésimal, 181, 226
graphe aléatoire, 62
graphe aléatoire à attachement préférentiel de Barabási-Albert, 71, 74
graphe aléatoire de Erdős-Rényi, 67, 86

Recueil de modèles stochastiques – Page 271/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

graphe complet infini, 67
graphe d-régulier, 87
graphe fini, 86
groupe symétrique, 23

I
inégalité de concentration de Azuma-Hoeffding, 169
irréversiblité macroscopique, 27

J
jeu de pile ou face, 1

L
lemme de Borel-Cantelli, 2, 9, 91, 99, 102, 165, 174
lemme de Slutsky, 9, 100
log-vraisemblance conditionnelle des observations, 133
log-vraisemblance des observations, 133
loi de Hardy-Weinberg, 38
loi de Pareto, 142
loi des événements rares, 184
loi des grands nombres, 127, 155, 156, 189
loi des grands nombres pour les chaînes de Markov, 42
loi des petits nombres, 6, 184, 199, 245
loi du demi-cercle, 157
loi du quart de cercle, 163
loi du zéro-un pour les sommes de v.a.r. exponentielles indépendantes, 80
loi forte des grands nombres, 2, 130, 189
loi multinomiale, 18

M
marche aléatoire, 68
marche aléatoire renforcée, 71
marche aléatoire renforcée par sites, 77
marche aléatoire sur l'hypercube discret, 28
matrice aléatoire, 155
matrice d'adjacence, 86
matrice de covariance, 155
matrice de covariance empirique, 155
matrice de remise, 73
mélange fini, 131
mesure de vitesse, 246
mesure harmonique, 21
mesure spectrale empirique, 156
modèle de Cannings, 44
modèle de Fisher-Wright, 38, 105
modèle de Moran, 46
modèle d'Ehrenfest, 27
modèle des configurations, 87
monomères, 145
mouvement brownien, 219, 249
multigraphe, 87

N
nombre de Bell, 102
nombre de Stirling de première espèce, 102
nombre de Stirling de seconde espèce, 7, 102
nombres de Bell, 86
noyau de Poisson, 219
noyau de proposition ou d'exploration, 89

O
optimisation combinatoire randomisée, 167

P
percolation, 61
phénomène de seuil, 63
phylogénie, 114
poissonisation, 172
polymère, 145
polymère dirigé, 145
principe de Laplace, 89
principe des tiroirs, 171
principe d'inclusion-exclusion, 85
principe d'invariance, 249
probabilité critique, 62
problème de Dirichlet, 20, 217
problème des secrétaires, 143
problème du mariage, 143
problème du voyageur de commerce, 90, 167
processus de Bernoulli, 2
processus de branchement de Galton-Watson, 68, 205
processus de Langevin confiné, 237
processus de Markov déterministe par morceaux, 188, 199
processus de Poisson, 2, 30, 111, 177, 188, 193, 199, 210
processus de Poisson composé, 188
processus de taille de fenêtre TCP, 199
processus de vie et de mort, 183
processus de Yule, 110, 183, 205, 254
processus d'Ornstein-Uhlenbeck, 36, 184, 225, 236

R
racine, 62, 63
recuit simulé, 167
renforcement, 71
renforcement géométrique, 77
renforcement linéaire, 77

Recueil de modèles stochastiques – Page 272/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

réseau, 69
réversibilité microscopique, 27
ruine d'un joueur, 14
runs, 209

S
schéma d'Euler, 253
second principe de la thermodynamique, 27
sommets, 61
spin, 22
statistique d'ordre, 110
suite de degrés, 87
suite échangeable, 72

T
temps de fixation, 38
temps de recouvrement, 11
temps de retour, 34
tension d'une suite de mesures de probabilités, 157
théorème de Donsker, 236, 249
théorème de Erdős-Gallai, 87
théorème de Marchenko-Pastur, 162
théorème de Prohorov, 157
théorème de Rubin, 71
théorème de Weierstrass, 159
théorème de Wigner, 156
théorème des moments, 159
théorème ergodique, 127
théorème limite central, 2, 3, 16, 25, 26, 33, 100, 165, 235
théorème limite central de Lindeberg-Lévy, 138
théorème limite central pour les martingales, 121, 130
théorème limite central pour les martingales de carré intégrable, 118, 129
théorème limite central quantitatif de Berry-Esseen, 100
théorie cinétique des gaz de Boltzmann, 27
théorie de Mendel, 37
théorie des probabilités libres, 166

U
urne de Pólya, 71, 121
urne de Pólya généralisée, 79

Recueil de modèles stochastiques – Page 273/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.

Recueil de modèles stochastiques – Page 274/274 – Copyright © Djalil Chafaï and Florent Malrieu, 2014 – Compilé le 29 octobre 2014.