

FACULTÉ DES SCIENCES D'ORSAY

LDD2 Mathématiques / OLMA255

Modélisation aléatoire

Notes de cours

Modèles probabilistes et simulations

Arnaud Durand

Département de Mathématiques d'Orsay

arnaud.durand@universite-paris-saclay.fr

Avant-propos

L'unité d'enseignement (UE) *Modélisation* a pour objectif de vous immerger dans l'univers fascinant de la modélisation mathématique, en vous montrant comment les outils théoriques acquis dans d'autres UE peuvent être appliqués à des problèmes concrets issus de domaines variés tels que la biologie, l'économie, l'informatique et bien d'autres. Cette transversalité entre disciplines est une source inépuisable d'enrichissement, car elle révèle des idées et des techniques mathématiques communes, tout en mettant en lumière des applications surprenantes et souvent inattendues.

Ces notes de cours sont consacrées à l'étude de *modèles aléatoires*, qui constituent un terrain d'exploration idéal pour mettre en pratique les concepts abordés dans l'UE *Probabilités et statistiques*. Ces modèles, bien que simples, sont d'une richesse insoupçonnée et permettent d'illustrer des phénomènes complexes par des exemples intuitifs. Voici un aperçu des thèmes abordés :

- **Pile ou face : le modèle de Bernoulli.** Nous commençons par l'étude du jeu de *pile ou face*, un modèle fondamental en probabilités. Vous découvrirez les concepts d'*épreuve de Bernoulli*, de *processus de Bernoulli*, et vous explorerez les *instants de succès*, c'est-à-dire les moments où un événement particulier se produit. Nous approfondirons également le *comportement asymptotique* de ces processus, en étudiant des résultats clés tels que la *loi des grands nombres*, le *lemme de Borel-Cantelli*, et les *théorèmes limites* (approximation normale et poissonienne). Enfin, nous verrons comment ces processus peuvent être interprétés sous l'angle des *chaînes de Markov*, et nous explorerons les *processus de Bernoulli modulo un entier*, qui ouvrent la voie à des applications plus avancées.
- **Le problème du collectionneur de coupons.** Qui n'a jamais rêvé de compléter une collection de coupons, de cartes ou d'objets rares ? Ce problème, en apparence ludique, est en réalité un excellent prétexte pour étudier des concepts probabilistes tels que la *loi du temps de recouvrement*, qui permet de déterminer combien de temps il faut pour collecter tous les éléments d'une série. Nous analyserons également le *comportement asymptotique* de ce modèle et nous verrons comment il peut être interprété à travers le prisme des *chaînes de Markov*.
- **Marche aléatoire : du joueur ruiné au scrutin électoral.** La *marche aléatoire* est un modèle central en probabilités, avec des applications allant de la physique à la finance. Nous commencerons par étudier la *marche aléatoire unidimensionnelle*, en nous intéressant notamment au *problème de la ruine du joueur*, qui modélise la fortune d'un joueur au fil de ses paris. Nous explorerons également les *retours au point de départ*, ainsi que le *problème du scrutin*, qui établit un lien surprenant entre les marches aléatoires et les élections. Enfin, nous aborderons les *marches aléatoires multidimensionnelles*, qui offrent des perspectives encore plus vastes.
- **Branchement et dynamique des populations.** Comment une population évolue-t-elle au fil du temps ? Pour répondre à cette question, nous étudierons les *processus de Galton-Watson*, qui modélisent la croissance d'une population à travers des générations successives. Nous verrons comment les *fonctions génératrices* permettent de calculer des espérances et des variances, et d'analyser le *comportement d'extinction* de ces processus. Enfin, nous nous intéresserons à la *taille totale de la population*, un indicateur clé pour en comprendre la dynamique.
- **Génétique : des modèles classiques aux processus stochastiques.** La génétique est un domaine où les modèles probabilistes jouent un rôle crucial. Nous commencerons par la *loi de Hardy-Weinberg*, qui décrit la distribution des génotypes dans une population. Nous explorerons ensuite deux modèles classiques : le *modèle de Moran*, qui modélise l'évolution des fréquences alléliques dans une population de taille fixe, et le *modèle de Wright-Fisher*, qui introduit une dimension probabiliste pour étudier la dérive génétique.
- **Annexes : outils mathématiques et scripts Python.** Pour vous aider à maîtriser les concepts abordés, nous avons inclus des annexes détaillées sur les *fonctions génératrices* et les *chaînes de Markov*, deux outils essentiels en probabilités. Enfin, pour ceux qui souhaitent mettre en pratique ces notions, nous proposons une série de *scripts Python* qui vous permettront de simuler les modèles étudiés dans chaque chapitre. Ces scripts sont conçus pour être à la fois pédagogiques et pratiques, vous offrant ainsi une passerelle entre la théorie et l'application.

Ce cours est conçu pour vous offrir une vision à la fois large et précise de la modélisation probabiliste, telle qu'elle est accessible en premier cycle universitaire. À travers des exemples concrets et des applications variées, vous découvrirez comment les mathématiques éclairent des phénomènes complexes et fournissent des outils puissants pour comprendre le monde qui vous entoure. Que vous soyez intéressé par la biologie, l'économie, l'informatique ou simplement curieux des liens entre ces disciplines, ce cours vise à développer vos compétences en modélisation tout en explorant des idées stimulantes.

Table des matières

Avant-propos	i
1 Pile ou face	1
1.1 Épreuve, schéma et processus de Bernoulli	1
1.2 Instants de succès	3
1.2.1 Instant de premier succès	3
1.2.2 Nombre d'échecs avant le premier succès	5
1.2.3 Suite des instants de succès	5
1.3 Comportement asymptotique	8
1.3.1 Lois des grands nombres	8
1.3.2 Lemme de Borel-Cantelli	9
1.3.3 Approximation normale et théorème limite central	11
1.3.4 Approximation poissonnienne et loi des petits nombres	17
1.4 Aspect markovien	23
1.5 Processus de Bernoulli modulo un entier	24
2 Collection de coupons	27
2.1 Loi du temps de recouvrement	27
2.2 Comportement asymptotique	32
2.3 Aspect markovien	34
3 Marche aléatoire	39
3.1 Marche aléatoire unidimensionnelle	39
3.2 Ruine du joueur	41
3.3 Premier retour au point de départ	45
3.4 Problème du scrutin	49
3.5 Marche aléatoire symétrique multidimensionnelle	50
4 Branchement	55
4.1 Processus et arbres de Galton-Watson	55
4.2 Quelques lois de reproduction élémentaires	58
4.3 Fonctions génératrices, espérances et variances	59
4.4 Comportement à l'infini	62
4.5 Taille totale de la population	65
5 Génétique	68
5.1 Loi de Hardy-Weinberg	68
5.2 Modèle de Moran	70
5.3 Modèle de Wright-Fisher	74
A Fonctions génératrices	79

B	Chaînes de Markov	83
B.1	Vecteurs et matrices indexés par un ensemble dénombrable	83
B.2	Définitions et premières propriétés des chaînes de Markov	86
B.3	Visites successives d'un état	89
B.3.1	Accessibilité d'un état	89
B.3.2	Graphe d'accessibilité	91
B.3.3	Retour à un état	91
B.3.4	Nombre de visites d'un état	92
B.4	Réurrence et transience	96
B.5	Probabilité invariante	99
B.6	Réurrence positive	102
B.7	Apériodicité et convergence	105
C	Scripts Python	110
C.1	Scripts utilisés dans le chapitre 1	110
C.2	Scripts utilisés dans le chapitre 2	113
C.3	Scripts utilisés dans le chapitre 3	114
C.4	Scripts utilisés dans le chapitre 4	118
C.5	Scripts utilisés dans le chapitre 5	120

Pile ou face

Le jeu de pile ou face consiste à lancer un certain nombre de fois une pièce de monnaie et à noter le résultat, à savoir « pile » ou « face », de chacun des lancers. Ces lancers sont supposés indépendants entre eux, c'est-à-dire que le résultat d'un lancer n'a pas d'influence sur les autres.

1.1 Épreuve, schéma et processus de Bernoulli

Le jeu de pile ou face est un cas particulier d'expérience aléatoire ne pouvant avoir que deux issues possibles, appelées usuellement « succès » et « échec ». Une telle expérience aléatoire est appelée *épreuve de Bernoulli* de paramètre $p \in [0, 1]$, la probabilité de succès. Cette appellation fait référence au mathématicien suisse Jacques Bernoulli (1654–1705), qui aborde ce sujet dans le livre posthume *Ars Conjectandi* (1713). Bernoulli et son livre sont représentés sur la figure 1.1. Jouer à pile ou face avec une pièce équilibrée correspond au cas où les deux issues ont la même probabilité $p = 1/2$.

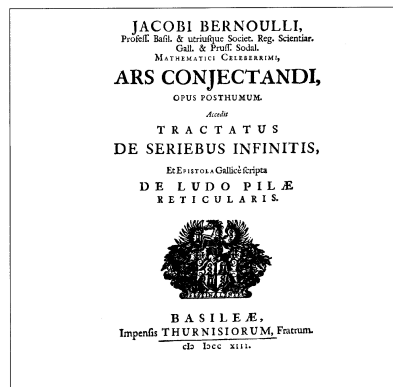


FIGURE 1.1 : Jacques Bernoulli (1654–1705) et son ouvrage *Ars Conjectandi* (1713)

En attribuant la valeur 1 au succès et la valeur 0 à l'échec, on peut coder le résultat d'une épreuve de Bernoulli par une variable aléatoire qui répond à la définition suivante.

Définition 1.1 (Loi de Bernoulli). Soit $p \in [0, 1]$. On dit qu'une variable aléatoire X suit la *loi de Bernoulli* de paramètre p , notée $\text{Ber}(p)$, si X est presque sûrement à valeurs dans $\{0, 1\}$ et vérifie

$$\mathbb{P}(X = 1) = p \quad \text{et} \quad \mathbb{P}(X = 0) = 1 - p.$$

L'espérance et la variance d'une variable aléatoire X qui suit la loi $\text{Ber}(p)$ sont données par

$$\mathbb{E}[X] = p \quad \text{et} \quad \text{Var}(X) = p(1 - p). \quad (1.1)$$

On appelle ensuite *schéma de Bernoulli* de paramètres n et p l'expérience aléatoire qui consiste à répéter n fois de façon indépendante une épreuve de Bernoulli de paramètre p . On code l'issue de la m -ième épreuve par une variable aléatoire X_m . Les variables aléatoires X_m suivent toutes la loi $\text{Ber}(p)$, donc pour tout entier $m \in \{1, \dots, n\}$, on a

$$\mathbb{P}(X_m = 1) = p \quad \text{et} \quad \mathbb{P}(X_m = 0) = 1 - p.$$

De plus, comme les épreuves sont indépendantes, on peut supposer que les variables aléatoires X_m sont *indépendantes*. Ces variables sont dites *indépendantes et identiquement distribuées* (i.i.d.). Le nombre total de succès parmi les n répétitions de l'expérience est donné par la somme

$$S_n = X_1 + \dots + X_n. \quad (1.2)$$

Il s'agit d'une variable aléatoire qui répond à la définition suivante.

Définition 1.2 (Loi binomiale). Soient $n \in \mathbb{N}$ et $p \in [0, 1]$. On dit qu'une variable aléatoire S suit la *loi binomiale* de paramètres n et p , notée $\text{Bin}(n, p)$, si S est presque sûrement à valeurs dans $\{0, \dots, n\}$ et vérifie

$$\forall k \in \{0, \dots, n\} \quad \mathbb{P}(S = k) = \binom{n}{k} p^k (1-p)^{n-k}. \quad (1.3)$$

On adopte la convention $0^0 = 1$. Pour $n = 0$, cette loi est concentrée en 0.

Le coefficient binomial $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ donne le nombre de positions possibles de k succès, et donc de $n - k$ échecs, parmi n épreuves de Bernoulli, et la probabilité d'avoir k succès, et donc $n - k$ échecs, à des positions fixées est égale à $p^k (1-p)^{n-k}$. Cela prouve que S_n suit la loi $\text{Bin}(n, p)$.

On rappelle que l'espérance est linéaire et que la variance d'une somme de variables aléatoires indépendantes est égale à la somme des variances. Ainsi, l'espérance et la variance de S_n sont

$$\mathbb{E}[S_n] = np \quad \text{et} \quad \text{Var}(S_n) = np(1-p). \quad (1.4)$$

Ces valeurs correspondent donc à l'espérance et la variance de toute variable qui suit la loi $\text{Bin}(n, p)$.

On peut retrouver les résultats énoncés ci-dessus en utilisant la notion de fonction génératrice, abordée dans l'annexe A. Soient $n \in \mathbb{N}^*$ et $p \in [0, 1]$. Si des variables aléatoires X et S suivent respectivement les lois $\text{Ber}(p)$ et $\text{Bin}(n, p)$, alors leurs fonctions génératrices sont données par

$$G_X(z) = (1-p)z^0 + pz^1 = pz + 1 - p \quad \text{et} \quad G_S(z) = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} z^k = (pz + 1 - p)^n.$$

La dernière égalité résulte de la formule du binôme de Newton. Dans ces deux cas, les séries entières sont des fonctions polynomiales. De plus, la proposition A.3 implique que

$$\mathbb{E}[X] = \lim_{\substack{z \rightarrow 1 \\ z < 1}} G'_X(z) = p \quad \text{et} \quad \mathbb{E}[X(X-1)] = \lim_{\substack{z \rightarrow 1 \\ z < 1}} G''_X(z) = 0,$$

ce qui permet de retrouver que $\text{Var}(X) = p(1-p)$. De même, on retrouve que $\text{Var}(S) = np(1-p)$, car

$$\mathbb{E}[S] = \lim_{\substack{z \rightarrow 1 \\ z < 1}} G'_S(z) = np \quad \text{et} \quad \mathbb{E}[S(S-1)] = \lim_{\substack{z \rightarrow 1 \\ z < 1}} G''_S(z) = n(n-1)p^2.$$

Enfin, on déduit de l'expression de G_X et de la proposition A.5 que si S_n désigne la somme (1.2) de n variables indépendantes $X_1, \dots, X_n$ qui suivent la loi $\text{Ber}(p)$, alors pour tous $z \in \mathbb{C}$ et $m \in \{1, \dots, n\}$,

$$G_{X_m}(z) = pz + 1 - p \quad \text{et} \quad G_{S_n}(z) = G_{X_1}(z) \cdots G_{X_n}(z) = (pz + 1 - p)^n.$$

La proposition A.4 permet de retrouver que la variable aléatoire S_n suit la loi $\text{Bin}(n, p)$.

On appelle *processus aléatoire à temps discret*, ou plus simplement *processus*, toute suite de variables aléatoires. Si on répète indéfiniment de façon indépendante une même épreuve de Bernoulli, le nombre de succès au cours du temps constitue un processus qui répond à la définition suivante.

Définition 1.3 (Processus de Bernoulli). Soit $p \in [0, 1]$. On appelle *processus de Bernoulli* de paramètre p toute suite $(S_n)_{n \geq 0}$ de variables aléatoires pour laquelle il existe une suite $(X_m)_{m \geq 1}$ de variables aléatoires indépendantes et identiquement distribuées selon la loi $\text{Ber}(p)$ vérifiant, presque sûrement, pour tout $n \in \mathbb{N}$,

$$S_n = \sum_{m=1}^n X_m. \quad (1.5)$$

Rappel 1.1 (Somme et produit indexés par l'ensemble vide). Par convention, une somme indexée par l'ensemble vide est égale à 0. Ainsi, un processus de Bernoulli $(S_n)_{n \geq 0}$ a pour valeur initiale $S_0 = 0$ avec probabilité 1. De plus, par convention, un produit indexé par l'ensemble vide est égal à 1.

Rappel 1.2 (Indépendance d'une famille infinie dénombrable de variables aléatoires). Des variables aléatoires indexées par un ensemble infini dénombrable sont indépendantes si pour tout choix fini d'indices, les variables aléatoires correspondantes sont indépendantes.

1.2 Instants de succès

Désignons par $(S_n)_{n \geq 0}$ le processus de Bernoulli de paramètre $p \in [0, 1]$ obtenu via (1.5) à partir d'une suite $(X_m)_{m \geq 1}$ de variables aléatoires indépendantes et identiquement distribuées selon la loi $\text{Ber}(p)$. Il s'agit d'un processus de comptage qui donne le nombre S_n de succès jusqu'à un instant n donné, quand on répète de façon indépendante la même épreuve de Bernoulli de paramètre p . Sa valeur est inchangée quand on rencontre un échec, et augmente de 1 à chaque succès. On s'intéresse dans cette section aux propriétés probabilistes des instants où les succès ont lieu.

1.2.1 Instant de premier succès

L'instant T du premier succès au cours de la répétition de l'épreuve de Bernoulli vérifie les relations

$$T = \inf\{m \geq 1 \mid X_m = 1\} = \inf\{n \geq 1 \mid S_n = 1\},$$

avec la convention que les infima sont égaux à ∞ si les ensembles sont vides. Pour tout entier $k \geq 1$, on a $T \geq k$ si et seulement si $X_1 = \dots = X_{k-1} = 0$. Par indépendance, on a donc

$$\mathbb{P}(T \geq k) = \mathbb{P}(X_1 = 0) \cdots \mathbb{P}(X_{k-1} = 0) = (1 - p)^{k-1}.$$

On en déduit directement que

$$\mathbb{P}(T = k) = \mathbb{P}(T \geq k) - \mathbb{P}(T \geq k + 1) = (1 - p)^{k-1} - (1 - p)^k = (1 - p)^{k-1} p. \quad (1.6)$$

Par conséquent, la variable aléatoire T vaut ∞ avec probabilité 1 si $p = 0$. Si $p > 0$, elle est finie avec probabilité 1 et répond à la définition suivante.

Définition 1.4 (Loi géométrique sur $\mathbb{N}^*$). Soit $p \in]0, 1]$. On dit qu'une variable aléatoire T suit la loi géométrique sur $\mathbb{N}^*$ de paramètre p , notée $\text{Géo}_{\mathbb{N}^*}(p)$, si T est presque sûrement à valeurs dans $\mathbb{N}^*$ et vérifie

$$\forall k \in \mathbb{N}^* \quad \mathbb{P}(T = k) = (1 - p)^{k-1} p.$$

Pour calculer l'espérance de l'instant de premier succès T , donc plus généralement de toute variable aléatoire suivant la loi géométrique sur $\mathbb{N}^*$ de paramètre p , on peut utiliser le résultat suivant. On désigne par $\overline{\mathbb{N}}$ l'ensemble des entiers auquel on adjoint l'infini, c'est-à-dire que $\overline{\mathbb{N}} = \mathbb{N} \cup \{\infty\}$.

Lemme 1.1 (Espérance comme somme de la queue de distribution). Si Y désigne une variable aléatoire à valeurs dans l'ensemble $\overline{\mathbb{N}}$, alors son espérance vérifie

$$\mathbb{E}[Y] = \sum_{k=1}^{\infty} \mathbb{P}(Y \geq k).$$

Démonstration. Supposons d'abord que la variable aléatoire Y est à valeurs dans $\mathbb{N}$. On établit le résultat à l'aide d'une transformation d'Abel. On observe que pour tout $k \in \mathbb{N}^*$,

$$k \mathbb{P}(Y = k) = k (\mathbb{P}(Y \geq k) - \mathbb{P}(Y \geq k + 1)) = k \mathbb{P}(Y \geq k) - (k + 1) \mathbb{P}(Y \geq k + 1) + \mathbb{P}(Y \geq k + 1)$$

En sommant ces identités jusqu'à un entier quelconque $n \in \mathbb{N}^*$, on en déduit que

$$\begin{aligned} \sum_{k=1}^n k \mathbb{P}(Y = k) &= \sum_{k=0}^n k \mathbb{P}(Y = k) = \sum_{k=0}^n (k \mathbb{P}(Y \geq k) - (k + 1) \mathbb{P}(Y \geq k + 1)) + \sum_{k=0}^n \mathbb{P}(Y \geq k + 1) \\ &= -(n + 1) \mathbb{P}(Y \geq n + 1) + \sum_{k=1}^{n+1} \mathbb{P}(Y \geq k) \\ &= -n \mathbb{P}(Y \geq n + 1) + \sum_{k=1}^n \mathbb{P}(Y \geq k) \leq \sum_{k=1}^n \mathbb{P}(Y \geq k). \end{aligned}$$

Si la série $\sum \mathbb{P}(Y \geq k)$ converge, la série $\sum k \mathbb{P}(Y = k)$ converge aussi, par comparaison entre séries à termes positifs. De plus, si la série $\sum k \mathbb{P}(Y = k)$ converge, on a

$$0 \leq (n+1) \mathbb{P}(Y \geq n+1) = (n+1) \sum_{k=n+1}^{\infty} \mathbb{P}(Y = k) \leq \sum_{k=n+1}^{\infty} k \mathbb{P}(Y = k) \xrightarrow{n \rightarrow \infty} 0,$$

donc l'identité précédente implique que la série $\sum \mathbb{P}(Y \geq k)$ converge aussi, et que les sommes vérifient

$$\sum_{k=1}^{\infty} k \mathbb{P}(Y = k) = \sum_{k=1}^{\infty} \mathbb{P}(Y \geq k).$$

Le membre de gauche ci-dessus est par définition égal à $\mathbb{E}[Y]$. L'égalité de l'énoncé est donc vraie, avec éventuellement ses deux membres égaux à ∞ .

Revenons au cas général où la variable aléatoire Y est à valeurs dans $\overline{\mathbb{N}}$. Si la variable Y vaut l'infini avec probabilité strictement positive, on a

$$\mathbb{E}[Y] \geq \infty \cdot \mathbb{P}(Y = \infty) = \infty \quad \text{et} \quad \sum_{k=1}^{\infty} \mathbb{P}(Y \geq k) \geq \sum_{k=1}^{\infty} \mathbb{P}(Y = \infty) = \infty,$$

donc l'égalité de l'énoncé est vraie avec les deux membres égaux à ∞ . Si la variable Y vaut l'infini avec probabilité nulle, elle coïncide presque sûrement avec $Y \mathbf{1}_{\{Y < \infty\}}$, qui est à valeurs dans $\mathbb{N}$. Ainsi,

$$\mathbb{E}[Y] = \mathbb{E}[Y \mathbf{1}_{\{Y < \infty\}}] = \sum_{k=1}^{\infty} \mathbb{P}(Y \mathbf{1}_{\{Y < \infty\}} \geq k) = \sum_{k=1}^{\infty} \mathbb{P}(\infty > Y \geq k) = \sum_{k=1}^{\infty} \mathbb{P}(Y \geq k),$$

par le cas traité initialement. On obtient donc l'égalité annoncée dans tous les cas possibles. $\square$

Remarque 1.1 (Espérance conditionnellement à un événement). Si Y désigne une variable aléatoire à valeurs dans $\overline{\mathbb{N}}$ et si A désigne un événement de probabilité non nulle, on rappelle que l'espérance de Y conditionnellement à A est

$$\mathbb{E}[Y | A] = \frac{\mathbb{E}[\mathbf{1}_A Y]}{\mathbb{P}(A)},$$

c'est-à-dire l'espérance de la variable Y sous la probabilité $\mathbb{P}(\cdot | A)$. Le lemme 1.1 implique alors

$$\mathbb{E}[Y | A] = \frac{\mathbb{E}[\mathbf{1}_A Y]}{\mathbb{P}(A)} = \sum_{k=1}^{\infty} \frac{\mathbb{P}(\mathbf{1}_A Y \geq k)}{\mathbb{P}(A)} = \sum_{k=1}^{\infty} \frac{\mathbb{P}(A \text{ et } Y \geq k)}{\mathbb{P}(A)} = \sum_{k=1}^{\infty} \mathbb{P}(Y \geq k | A). \quad (1.7)$$

En appliquant le lemme 1.1 à une variable aléatoire T qui, comme l'instant de premier succès, suit la loi géométrique sur $\mathbb{N}^*$ de paramètre $p \in]0, 1]$, on obtient

$$\mathbb{E}[T] = \sum_{k=1}^{\infty} \mathbb{P}(T \geq k) = \sum_{k=1}^{\infty} (1-p)^{k-1} = \sum_{k=0}^{\infty} (1-p)^k = \frac{1}{1-(1-p)} = \frac{1}{p}.$$

On peut retrouver ce résultat à partir de la fonction génératrice de T , qui vérifie pour $(1-p)|z| < 1$,

$$G_T(z) = \sum_{k=0}^{\infty} \mathbb{P}(T = k) z^k = \sum_{k=1}^{\infty} (1-p)^{k-1} p z^k = pz \sum_{k=0}^{\infty} ((1-p)z)^k = \frac{pz}{1-(1-p)z}. \quad (1.8)$$

On montre ensuite par une récurrence élémentaire que pour $r \in \mathbb{N}^*$ et $(1-p)|z| < 1$,

$$G_T^{(r)}(z) = \frac{r! (1-p)^{r-1} p}{(1-(1-p)z)^{r+1}} \xrightarrow[z < 1]{z \rightarrow 1} \frac{r!}{p} \left(\frac{1}{p} - 1 \right)^{r-1} = \mathbb{E}[(T)_r], \quad (1.9)$$

où la dernière égalité découle de la proposition A.3. On en déduit que

$$\mathbb{E}[T] = \mathbb{E}[(T)_1] = \frac{1}{p} \quad \text{et} \quad \text{Var}(T) = \mathbb{E}[(T)_2] + \mathbb{E}[(T)_1] - \mathbb{E}[(T)_1]^2 = \frac{2}{p} \left(\frac{1}{p} - 1 \right) + \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2}. \quad (1.10)$$

On retrouve ainsi l'espérance obtenue précédemment, et on obtient en outre la variance de T .

1.2.2 Nombre d'échecs avant le premier succès

Considérons maintenant le nombre N d'échecs avant le premier succès au cours des épreuves de Bernoulli. Celui-ci vérifie les relations

$$N = \sup\{m \geq 1 \mid X_1 = \dots = X_m = 0\} = \sup\{n \geq 1 \mid S_n = 0\} = T - 1,$$

avec la convention que les suprema sont égaux à 0 si les ensembles sont vides. Tout comme l'instant T de premier succès, la variable aléatoire $N = T - 1$ vaut ∞ avec probabilité 1 si $p = 0$. Si $p > 0$, elle est finie avec probabilité 1 et relève de la définition suivante.

Définition 1.5 (Loi géométrique sur $\mathbb{N}$). Soit $p \in]0, 1]$. On dit qu'une variable aléatoire N suit la loi géométrique sur $\mathbb{N}$ de paramètre p , notée $\text{Géo}_{\mathbb{N}}(p)$, si N est presque sûrement à valeurs dans $\mathbb{N}$ et vérifie

$$\forall k \in \mathbb{N} \quad \mathbb{P}(N = k) = (1 - p)^k p.$$

Comme $N = T - 1$, on peut déduire la fonction génératrice de N de celle de T , qui est donnée par (1.8). Ainsi, pour $z \neq 0$ et $(1 - p)|z| < 1$, on a

$$G_N(z) = \mathbb{E}[z^N] = \mathbb{E}[z^{T-1}] = \frac{G_T(z)}{z} = \frac{p}{1 - (1 - p)z}. \quad (1.11)$$

Le quotient se prolonge en $z = 0$ par $G_N(0) = p$. Cette fonction est la fonction génératrice de n'importe quelle variable aléatoire qui suit la loi $\text{Géo}_{\mathbb{N}}(p)$. Par une récurrence élémentaire et la proposition A.3, on a donc pour $r \in \mathbb{N}^*$ et $(1 - p)|z| < 1$,

$$G_N^{(r)}(z) = \frac{r! (1 - p)^r p}{(1 - (1 - p)z)^{r+1}} \xrightarrow[z < 1]{z \rightarrow 1} r! \left(\frac{1}{p} - 1\right)^r = \mathbb{E}[(N)_r].$$

Les deux moments factoriels $\mathbb{E}[(N)_1]$ et $\mathbb{E}[(N)_2]$ permettent de déterminer l'espérance et la variance de N , comme en (1.10). On peut aussi s'appuyer sur (1.10) et la relation $N = T - 1$ pour obtenir

$$\mathbb{E}[N] = \mathbb{E}[T] - 1 = \frac{1}{p} - 1 \quad \text{et} \quad \text{Var}(N) = \mathbb{E}[(T - 1) - (\mathbb{E}[T] - 1)]^2 = \mathbb{E}[(T - \mathbb{E}[T])^2] = \text{Var}(T) = \frac{1 - p}{p^2}. \quad (1.12)$$

Remarque 1.2 (Espérance et variance d'une variable décalée). Le raisonnement ci-dessus montre que pour toute variable aléatoire réelle Y admettant une espérance finie et pour tout réel λ , on a

$$\mathbb{E}[Y - \lambda] = \mathbb{E}[Y] - \lambda \quad \text{et} \quad \text{Var}(Y - \lambda) = \text{Var}(Y).$$

1.2.3 Suite des instants de succès

Considérons la suite des variables aléatoires T_l définie par les conditions suivantes :

$$T_1 = T \quad \text{et} \quad T_{l+1} = \inf\{m \geq T_l + 1 \mid X_m = 1\}, \quad (1.13)$$

toujours avec la convention que l'infimum est égal à ∞ si l'ensemble est vide. Si une variable aléatoire T_l est finie, elle donne l'instant où se produit le l -ième succès. Si la variable T_l est infinie, toutes les variables $T_{l+1}, T_{l+2}, \dots$ le sont aussi. On a plus précisément le résultat suivant.

Proposition 1.2 (Durée entre succès consécutifs). Les variables aléatoires T_l vérifient les propriétés suivantes :

- (i). Si $p = 0$, alors presque sûrement, $T_l = \infty$ pour tout entier $l \in \mathbb{N}^*$;
- (ii). Si $p > 0$, alors presque sûrement, $T_l < \infty$ pour tout entier $l \in \mathbb{N}^*$, et de plus les variables $T_1, T_2 - T_1, T_3 - T_2, \dots$ sont indépendantes et identiquement distribuées selon la loi $\text{Géo}_{\mathbb{N}^*}(p)$.

Démonstration. Si $p = 0$, on sait par (1.6) que $T_1 = T = \infty$ avec probabilité 1. De plus, la relation de récurrence (1.13) montre que T_{l+1} est infini si T_l l'est. Ainsi, par une récurrence immédiate, toutes les variables T_l sont infinies avec probabilité 1.

Supposons $p > 0$. Avant de considérer les différences entre les instants de succès, montrons qu'ils sont presque sûrement finis. Pour $l, m \geq 1$, si l'un des m blocs disjoints de l épreuves est constitué uniquement de succès, alors $T_l \leq ml$. L'indépendance donne donc

$$\mathbb{P}(T_l > ml) \leq (1 - p^l)^m \xrightarrow{m \rightarrow \infty} 0.$$

Ainsi, $T_l < \infty$ presque sûrement pour chaque l , puis simultanément pour tous les l par intersection dénombrable. Les différences $T_l - T_{l-1}$ sont donc bien définies presque sûrement; on peut les fixer arbitrairement sur l'événement négligeable restant.

Fixons des entiers $l \in \mathbb{N}^*$ et $k_1, \dots, k_l \in \mathbb{N}^*$, puis notons $k'_j = k_1 + \dots + k_j$. En posant $T_0 = 0$, on a les égalités suivantes entre événements :

$$\bigcap_{j=1}^l \{T_j - T_{j-1} = k_j\} = \bigcap_{j=1}^l \{T_j = k'_j\} = \bigcap_{m \in \{1, \dots, k'_l\} \setminus \{k'_1, \dots, k'_l\}} \{X_m = 0\} \cap \bigcap_{m \in \{k'_1, \dots, k'_l\}} \{X_m = 1\}.$$

Comme les variables aléatoires $X_1, \dots, X_{k'_l}$ sont indépendantes et suivent la loi $\text{Ber}(p)$, on obtient

$$\mathbb{P}\left(\bigcap_{j=1}^l \{T_j - T_{j-1} = k_j\}\right) = (1 - p)^{k'_l - l} p^l = \prod_{j=1}^l ((1 - p)^{k_j - 1} p). \quad (1.14)$$

Chaque événement de la forme $\{T_j - T_{j-1} < \infty\}$ est l'union disjointe des événements $\{T_j - T_{j-1} = k_j\}$ pour $k_j \in \mathbb{N}^*$. En sommant (1.14) sur tous les entiers $k_1, \dots, k_l \in \mathbb{N}^*$, on a donc

$$\mathbb{P}\left(\bigcap_{j=1}^l \{T_j - T_{j-1} < \infty\}\right) = \sum_{k_1, \dots, k_l \geq 1} \prod_{j=1}^l ((1 - p)^{k_j - 1} p) = \prod_{j=1}^l \left(\sum_{k_j=1}^{\infty} (1 - p)^{k_j - 1} p \right) = \left(\frac{p}{1 - (1 - p)} \right)^l = 1,$$

ce qui prouve que toutes les variables aléatoires $T_1, \dots, T_l$ sont finies avec probabilité 1. En choisissant un entier $j_0 \in \{1, \dots, l\}$ et en sommant (1.14) sur les entiers $k_1, \dots, k_{j_0-1}, k_{j_0+1}, \dots, k_l \in \mathbb{N}^*$, on a de plus

$$\begin{aligned} \mathbb{P}(T_{j_0} - T_{j_0-1} = k_{j_0}) &= \mathbb{P}\left(\{T_{j_0} - T_{j_0-1} = k_{j_0}\} \cap \bigcap_{j \in \{1, \dots, l\} \setminus \{j_0\}} \{T_j - T_{j-1} < \infty\}\right) \\ &= \sum_{k_1, \dots, k_{j_0-1}, k_{j_0+1}, \dots, k_l \geq 1} \prod_{j=1}^l ((1 - p)^{k_j - 1} p) \\ &= ((1 - p)^{k_{j_0} - 1} p) \prod_{j \in \{1, \dots, l\} \setminus \{j_0\}} \left(\sum_{k_j=1}^{\infty} (1 - p)^{k_j - 1} p \right) \\ &= ((1 - p)^{k_{j_0} - 1} p) \left(\frac{p}{1 - (1 - p)} \right)^{l-1} = (1 - p)^{k_{j_0} - 1} p. \end{aligned}$$

Ainsi, les variables aléatoires $T_j - T_{j-1}$ suivent toutes la loi $\text{Géo}_{\mathbb{N}^*}(p)$. En combinant le calcul précédent avec (1.14), on obtient de plus

$$\mathbb{P}\left(\bigcap_{j=1}^l \{T_j - T_{j-1} = k_j\}\right) = \prod_{j=1}^l \mathbb{P}(T_j - T_{j-1} = k_j),$$

donc les variables aléatoires $T_j - T_{j-1}$ sont indépendantes. $\square$

Remarque 1.3 (Interprétation comme processus de comptage). La proposition 1.2 implique que les nombres d'échecs entre deux succès consécutifs constituent des variables aléatoires indépendantes et identiquement distribuées selon la loi $\text{Géo}_{\mathbb{N}}(p)$. On peut interpréter le processus de Bernoulli $(S_n)_{n \geq 0}$ comme un processus de comptage d'impulsions qui sont espacées par des durées indépendantes et de loi géométrique $\text{Géo}_{\mathbb{N}^*}(p)$.

Finalement, dans le cas où $p > 0$, les variables aléatoires T_l relèvent de la définition suivante.

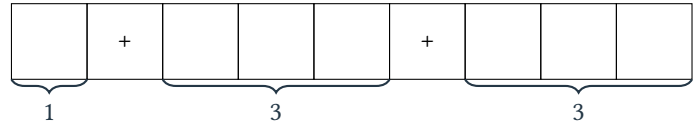
Définition 1.6 (Loi de Pascal). Soient $l \in \mathbb{N}^*$ et $p \in]0, 1]$. On dit qu'une variable aléatoire U suit la loi de Pascal de paramètres l et p , notée $\text{Géo}_{\mathbb{N}^*}(p)^{*l}$, si U est presque sûrement à valeurs dans $\{l, l+1, \dots\}$ et vérifie

$$\forall k \geq l \quad \mathbb{P}(U = k) = \binom{k-1}{l-1} (1-p)^{k-l} p^l.$$

Supposons $p > 0$ et montrons que pour tout entier $l \in \mathbb{N}^*$, la variable aléatoire T_l suit la loi $\text{Géo}_{\mathbb{N}^*}(p)^{*l}$. On déduit directement de (1.13) et de la proposition 1.2 que T_l est presque sûrement à valeurs dans l'ensemble $\{l, l+1, \dots\}$. De plus, on observe à l'aide de (1.14) que pour tout entier $k \geq l$,

$$\mathbb{P}(T_l = k) = \sum_{\substack{k_1, \dots, k_l \geq 1 \\ k_1 + \dots + k_l = k}} \mathbb{P}\left(\bigcap_{j=1}^l \{T_j - T_{j-1} = k_j\}\right) = (1-p)^{k-l} p^l \# \{k_1, \dots, k_l \in \mathbb{N}^* \mid k_1 + \dots + k_l = k\},$$

où $\#$ désigne le cardinal. Le nombre de façons de décomposer k en une somme de l entiers supérieurs ou égaux à 1 est égal au nombre de façons de décomposer $k-l$ en une somme de l entiers supérieurs ou égaux à 0, ce qui revient à choisir la position de $l-1$ symboles $+$ parmi $(k-l) + (l-1) = k-1$ positions possibles, le nombre de positions séparées par les $+$ donnant les entiers de la décomposition, voir la figure 1.2. Le cardinal dans la formule ci-dessus est donc égal à $\binom{k-1}{l-1}$. Cela confirme que T_l suit la loi $\text{Géo}_{\mathbb{N}^*}(p)^{*l}$.



Les cases vides entre les $+$ indiquent la décomposition $k-l = 7 = 1 + 3 + 3$. Elle correspond à $k = 10 = 2 + 4 + 4$, avec $l = 3$.

FIGURE 1.2 : Blaise Pascal (1623–1662) et exemple de décomposition

On peut facilement calculer l'espérance et la variance de la variable aléatoire T_l , et donc de toute variable aléatoire qui suit la loi $\text{Géo}_{\mathbb{N}^*}(p)^{*l}$, en écrivant

$$T_l = T_1 + (T_2 - T_1) + \dots + (T_l - T_{l-1}) \quad (1.15)$$

et en observant, à l'aide de la proposition 1.2, que les termes de cette décomposition sont des variables aléatoires indépendantes et identiquement distribuées selon la loi $\text{Géo}_{\mathbb{N}^*}(p)$. (Par conséquent, la loi de Pascal $\text{Géo}_{\mathbb{N}^*}(p)^{*l}$ est la puissance l -ième de convolution, définie par la proposition 4.1, de la loi géométrique $\text{Géo}_{\mathbb{N}^*}(p)$, ce qui justifie la notation de la définition 1.6.) Ainsi, pour $l \in \mathbb{N}^*$,

$$\mathbb{E}[T_l] = l \mathbb{E}[T] = \frac{l}{p} \quad \text{et} \quad \text{Var}(T_l) = l \text{Var}(T) = \frac{(1-p)l}{p^2}. \quad (1.16)$$

On a utilisé (1.10) pour exprimer l'espérance et la variance de la variable T , qui suit la loi $\text{Géo}_{\mathbb{N}^*}(p)$.

On peut exprimer la fonction génératrice de T_l , et donc de toute variable aléatoire qui suit la loi $\text{Géo}_{\mathbb{N}^*}(p)^{*l}$, en utilisant la proposition 1.2 ainsi que (1.8) et (1.15). Pour $(1-p)|z| < 1$, on obtient

$$\begin{aligned} G_{T_l}(z) &= \mathbb{E}[z^{T_l}] = \mathbb{E}[z^{T_1} z^{T_2 - T_1} \dots z^{T_l - T_{l-1}}] = \mathbb{E}[z^{T_1}] \mathbb{E}[z^{T_2 - T_1}] \dots \mathbb{E}[z^{T_l - T_{l-1}}] \\ &= G_{T_1}(z) G_{T_2 - T_1}(z) \dots G_{T_l - T_{l-1}}(z) = \left(\frac{pz}{1 - (1-p)z} \right)^l. \end{aligned} \quad (1.17)$$

On en déduit au passage le développement en série entière

$$\left(\frac{pz}{1-(1-p)z}\right)^l = \sum_{k=l}^{\infty} \binom{k-1}{l-1} (1-p)^{k-l} p^l z^k. \quad (1.18)$$

Enfin, en calculant les dérivées successives des fonctions génératrices G_{T_l} , on peut déterminer par récurrence tous les moments factoriels $\mathbb{E}[(T_l)_r]$. Plus précisément, en posant encore $T_0 = 0$ et dérivant la relation $G_{T_{l+1}} = G_T G_{T_l}$ à l'aide de la formule de Leibniz, on obtient pour $l, r \in \mathbb{N}$ et $(1-p)|z| < 1$,

$$G_{T_{l+1}}^{(r)}(z) = \sum_{s=0}^r \binom{r}{s} G_T^{(r-s)}(z) G_{T_l}^{(s)}(z) = \frac{pz}{1-(1-p)z} G_{T_l}^{(r)}(z) + \sum_{s=0}^{r-1} \frac{r! (1-p)^{r-s-1} p}{s! (1-(1-p)z)^{r-s+1}} G_{T_l}^{(s)}(z),$$

à l'aide de (1.8) et (1.9). La proposition A.3 permet d'en déduire que pour $l, r \in \mathbb{N}$,

$$\mathbb{E}[(T_{l+1})_r] = \mathbb{E}[(T_l)_r] + \sum_{s=0}^{r-1} \frac{r!}{s!} p \left(\frac{1}{p} - 1\right)^{r-s-1} \mathbb{E}[(T_l)_s].$$

Pour $r = 1$ et $r = 2$, cette formule conduit respectivement aux relations

$$\mathbb{E}[T_{l+1}] = \mathbb{E}[T_l] + \frac{1}{p} \quad \text{et} \quad \mathbb{E}[T_{l+1}(T_{l+1} - 1)] = \mathbb{E}[T_l(T_l - 1)] + \frac{2}{p} \mathbb{E}[T_l] + 2 \frac{1-p}{p^2}.$$

On en déduit par une récurrence élémentaire que pour tout $l \in \mathbb{N}$,

$$\mathbb{E}[T_l] = \frac{l}{p} \quad \text{et} \quad \mathbb{E}[T_l(T_l - 1)] = \sum_{j=0}^{l-1} \frac{2}{p^2} (j+1-p) = \frac{l}{p^2} (l+1-2p).$$

Cela permet de retrouver directement l'espérance et la variance de T_l calculées dans (1.16).

1.3 Comportement asymptotique

Soit $(S_n)_{n \geq 0}$ un processus de Bernoulli de paramètre $p \in [0, 1]$. On s'intéresse dans cette section au comportement du nombre de succès S_n au bout d'un grand nombre n d'épreuves.

1.3.1 Loïs des grands nombres

Intuitivement, on s'attend à ce que la fréquence de succès S_n/n converge vers la probabilité p de succès. Cette intuition est confirmée par le résultat de *convergence en probabilité* suivant.

Proposition 1.3 (Loi faible des grands nombres pour le processus de Bernoulli). Soit $(S_n)_{n \geq 0}$ un processus de Bernoulli de paramètre $p \in [0, 1]$. Alors S_n/n converge en probabilité vers p quand $n \rightarrow \infty$, ce que l'on note $S_n/n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} p$ et ce qui signifie que pour tout $\varepsilon > 0$,

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \geq \varepsilon\right) \xrightarrow[n \rightarrow \infty]{} 0.$$

Démonstration. À l'aide de (1.4), l'inégalité de Bienaymé-Tchebychev appliquée à S_n donne

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \geq \varepsilon\right) = \mathbb{P}(|S_n - \mathbb{E}[S_n]| \geq \varepsilon n) \leq \frac{\text{Var}(S_n)}{\varepsilon^2 n^2} = \frac{np(1-p)}{\varepsilon^2 n^2} \leq \frac{1}{4\varepsilon^2 n} \xrightarrow[n \rightarrow \infty]{} 0,$$

ce qui conclut la preuve. □

On peut renforcer la proposition 1.3 par le résultat de *convergence presque sûre* suivant.

Proposition 1.4 (Loi forte des grands nombres pour le processus de Bernoulli). Soit $(S_n)_{n \geq 0}$ un processus de Bernoulli de paramètre $p \in [0, 1]$. On a

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} \frac{S_n}{n} = p\right) = 1.$$

Démonstration. Pour tout $\varepsilon > 0$, l'inégalité de Markov appliquée à $(S_n - np)^4$ donne

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \geq \varepsilon\right) = \mathbb{P}(|S_n - np| \geq \varepsilon n) = \mathbb{P}((S_n - np)^4 \geq \varepsilon^4 n^4) \leq \frac{\mathbb{E}[(S_n - np)^4]}{\varepsilon^4 n^4}.$$

Notons X_m les variables aléatoires indépendantes et identiquement distribuées selon la loi $\text{Ber}(p)$ qui apparaissent dans la formule (1.5) définissant le processus de Bernoulli. On a

$$(S_n - np)^4 = \left(\sum_{m=1}^n (X_m - p)\right)^4 = \sum_{1 \leq m_1, m_2, m_3, m_4 \leq n} (X_{m_1} - p)(X_{m_2} - p)(X_{m_3} - p)(X_{m_4} - p).$$

Comme les variables $X_m - p$ sont indépendantes, les termes dans le membre de droite ont une espérance nulle sauf si les quatre indices m_1, m_2, m_3, m_4 sont les mêmes, ou si deux indices sont les mêmes et les deux autres sont les mêmes. Comme les variables $X_m - p$ sont en outre identiquement distribuées, on en déduit que la somme dans le membre de droite a pour espérance

$$n \mathbb{E}[(X_1 - p)^4] + \binom{4}{2} \binom{n}{2} \mathbb{E}[(X_1 - p)^2]^2 = n \mathbb{E}[(X_1 - p)^4] + 3n(n-1) \text{Var}(X_1)^2.$$

Le facteur $\binom{4}{2}$ correspond aux façons de choisir deux indices parmi les quatre qui vont prendre une valeur identique, et donc les deux autres indices vont prendre une autre valeur identique, tandis que le facteur $\binom{n}{2}$ correspond aux façons de choisir les deux valeurs entre 1 et n des indices en question. En utilisant l'expression (1.1) pour la variance de X_1 et en calculant $\mathbb{E}[(X_1 - p)^4]$, on obtient

$$\mathbb{E}[(S_n - np)^4] = np(1-p)((1-p)^3 + p^3) + 3n(n-1)p^2(1-p)^2 = \mathcal{O}_{n \rightarrow \infty}(n^2).$$

On en déduit que

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \geq \varepsilon\right) \leq \sum_{n=1}^{\infty} \frac{\mathbb{E}[(S_n - np)^4]}{\varepsilon^4 n^4} < \infty.$$

Le cas de convergence du lemme 1.5 ci-dessous implique alors qu'avec une probabilité égale à 1, il existe un entier $n_0 \in \mathbb{N}$ tel que pour tout $n \geq n_0$, on a $|S_n - np| < \varepsilon n$. Ainsi, avec probabilité 1, on a

$$\lim_{n_0 \rightarrow \infty} \sup_{n \geq n_0} \left|\frac{S_n}{n} - p\right| \leq \varepsilon.$$

La limite dans le membre de gauche existe toujours par monotonie. En prenant $\varepsilon = 1/k$, on en déduit qu'avec probabilité 1, la limite est majorée par $1/k$ pour tout $k \geq 1$. Elle est donc nulle. $\square$

Algorithme 1.1 (Illustration de la loi des grands nombres pour le processus de Bernoulli). Pour illustrer les propositions 1.3 et 1.4, on simule un processus de Bernoulli $(S_n)_{n \geq 0}$ de paramètre $p \in [0, 1]$, puis on trace le nombre de succès S_n en fonction de n pour le comparer avec son espérance np . On trace aussi la fréquence de succès S_n/n pour la comparer avec la probabilité p . On utilise le script C.1, pour produire les graphiques de la figure 1.3.

1.3.2 Lemme de Borel-Cantelli

La preuve de la proposition 1.4 ci-dessus repose sur le lemme de Borel-Cantelli suivant.

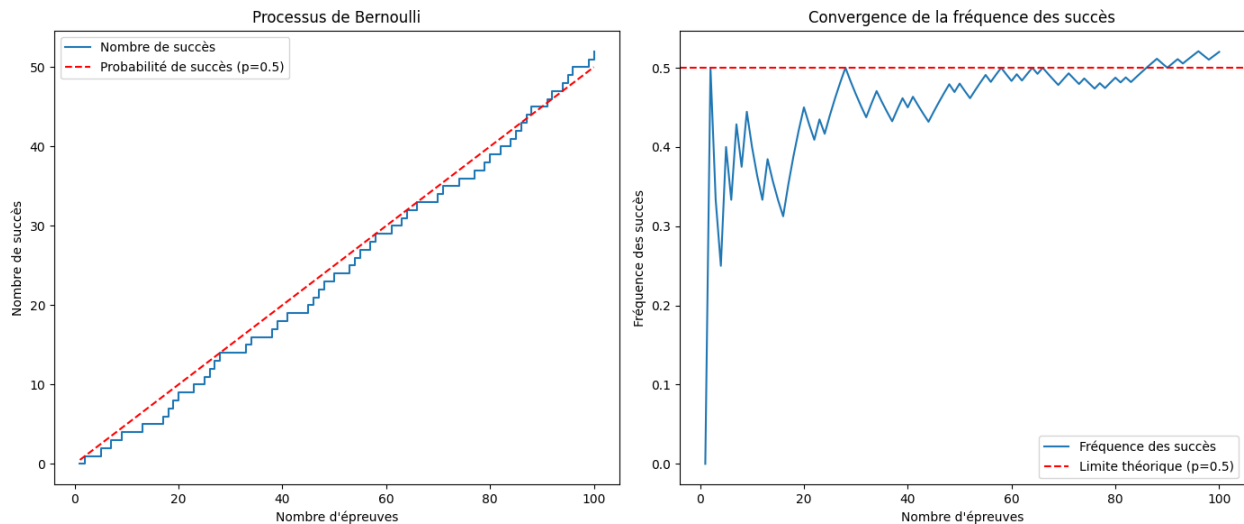


FIGURE 1.3 : Simulation du processus de Bernoulli pour $p = 0.5$ jusqu'à l'instant $n = 100$

Lemme 1.5 (Lemme de Borel-Cantelli). Soit $(A_n)_{n \geq 0}$ une suite d'événements. L'événement $A_\infty = \bigcap_{n_0=0}^{\infty} \bigcup_{n=n_0}^{\infty} A_n$ se réalise si et seulement si les événements A_n se réalisent pour une infinité d'indices n . On a aussi les propriétés suivantes :

- (i). Si la série $\sum_n \mathbb{P}(A_n)$ converge, alors $\mathbb{P}(A_\infty) = 0$;
- (ii). Si les événements A_n sont indépendants et la série $\sum_n \mathbb{P}(A_n)$ diverge, alors $\mathbb{P}(A_\infty) = 1$.

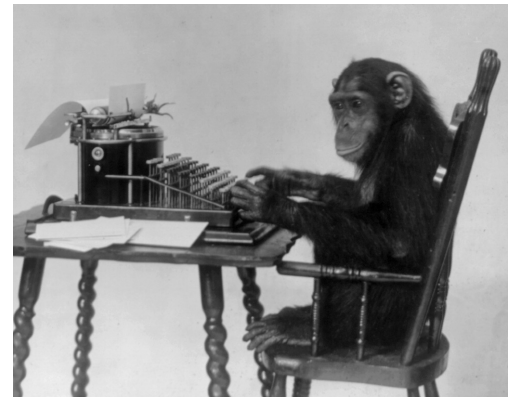
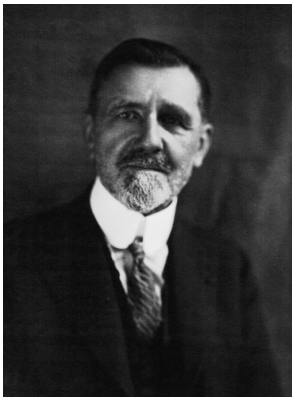


FIGURE 1.4 : Emile Borel (1871–1956), Francesco Paolo Cantelli (1875–1966) et un chimpanzé tapant peut-être les œuvres complètes de William Shakespeare

Démonstration. L'événement A_∞ se réalise si et seulement si pour tout entier $n_0 \in \mathbb{N}$, l'événement A_n se réalise pour au moins un indice $n \geq n_0$, ce qui signifie que les événements A_n se réalisent pour une infinité d'indices n . Supposons de plus que la série $\sum_n \mathbb{P}(A_n)$ converge. On a alors

$$\mathbb{P}(A_\infty) \leq \mathbb{P}\left(\bigcup_{n=n_0}^{\infty} A_n\right) \leq \sum_{n=n_0}^{\infty} \mathbb{P}(A_n) \xrightarrow{n_0 \rightarrow \infty} 0.$$

Supposons maintenant que les événements A_n sont indépendants et que la série $\sum_n \mathbb{P}(A_n)$ diverge. Le complémentaire A_∞^c de l'événement A_∞ est l'union sur l'entier $n_0 \in \mathbb{N}$ des intersections sur l'entier $n \geq n_0$ des complémentaires des événements A_n . Comme les événements A_n sont indépendants, leurs complémentaires sont

indépendants. On a alors pour tous entiers $m \geq n_0 \geq 0$,

$$\mathbb{P}\left(\bigcap_{n=n_0}^{\infty} A_n^c\right) \leq \mathbb{P}\left(\bigcap_{n=n_0}^m A_n^c\right) = \prod_{n=n_0}^m \mathbb{P}(A_n^c) = \prod_{n=n_0}^m (1 - \mathbb{P}(A_n)) \leq \exp\left(-\sum_{n=n_0}^m \mathbb{P}(A_n)\right) \xrightarrow{m \rightarrow \infty} 0.$$

On a utilisé ici d'une part le fait que la probabilité d'une intersection infinie d'événements indépendants est le produit infini des probabilités des événements, et d'autre part l'encadrement $0 \leq 1 - x \leq e^{-x}$ valable pour tout $x \leq 1$. On en déduit que $\mathbb{P}(A_{\infty}^c) = 0$, donc que $\mathbb{P}(A_{\infty}) = 1$. $\square$

Remarque 1.4 (Paradoxe du singe savant). Le cas de divergence du lemme 1.5 est souvent illustré par l'expérience de pensée du singe savant : si un singe tape indéfiniment et au hasard sur le clavier d'une machine à écrire, il pourra presque sûrement écrire un texte donné, par exemple les œuvres complètes de William Shakespeare. On modélise l'expérience en supposant que les touches $X_1, X_2, \dots$ tapées par le singe sont des variables aléatoires indépendantes et identiquement distribuées à valeurs dans l'ensemble fini $\mathcal{X}$ des touches de la machine à écrire. On suppose aussi que chaque touche peut être tapée par le singe avec une probabilité non nulle, ce qui revient à supposer que $\varepsilon = \inf_{k \in \mathcal{X}} \mathbb{P}(X_1 = k) > 0$. Notons l la longueur des œuvres complètes de Shakespeare. Pour tout entier $n \in \mathbb{N}$, notons A_n l'événement correspondant au fait que les touches tapées par le singe entre les instants $nl + 1$ et $(n + 1)l$, c'est-à-dire $X_{nl+1}, \dots, X_{(n+1)l}$, forment les œuvres complètes de Shakespeare. Les événements A_n sont indépendants, et on a $\mathbb{P}(A_n) \geq \varepsilon^l$ pour tout $n \in \mathbb{N}$. La série $\sum_n \mathbb{P}(A_n)$ diverge, donc le lemme de Borel-Cantelli implique qu'avec une probabilité égale à 1, le singe pourra écrire infiniment souvent les œuvres complètes de Shakespeare.

1.3.3 Approximation normale et théorème limite central

On peut compléter les résultats ci-dessus par l'étude des déviations de la fréquence de succès S_n/n par rapport à la probabilité $p \in]0, 1[$ de succès. Pour cela, il est naturel de *normaliser* le nombre de succès S_n en lui soustrayant son espérance et en le divisant par son écart-type. Au vu de (1.1), on introduit le nombre *réduit* de succès

$$S_n^* = \frac{S_n - np}{\sqrt{np(1-p)}}, \quad (1.19)$$

qui est donc une variable aléatoire centrée réduite, c'est-à-dire d'espérance égale à 0 et de variance égale à 1. On peut alors montrer que, pour n grand, la variable aléatoire S_n^* a une loi approchée par la *loi normale (ou gaussienne) centrée réduite*, dont la densité est la fonction

$$\varphi(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}. \quad (1.20)$$

Plus précisément, on a le résultat de *convergence en loi* suivant.

Théorème 1.6 (Théorème limite de Moivre-Laplace). Soit $(S_n)_{n \geq 0}$ un processus de Bernoulli de paramètre $p \in]0, 1[$. Pour tous réels $a < b$, on a

$$\mathbb{P}\left(a < \frac{S_n - np}{\sqrt{np(1-p)}} \leq b\right) \xrightarrow{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx.$$

Le reste de cette section est consacré à la preuve du théorème 1.6, qui repose sur plusieurs lemmes. Le premier lemme donne une approximation classique du logarithme.

Lemme 1.7 (Approximation du logarithme). Pour tout nombre réel $x \in \mathbb{R}$ tel que $|x| \leq 2/3$, on peut écrire

$$\ln(1+x) = x - \frac{x^2}{2} + \theta(x) \quad \text{avec} \quad |\theta(x)| \leq |x|^3.$$

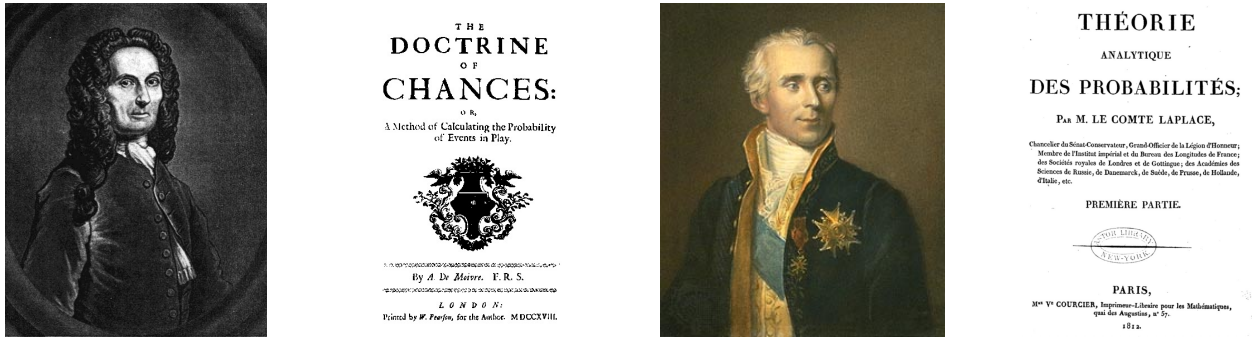


FIGURE 1.5 : Abraham de Moivre (1667–1754) et Pierre-Simon de Laplace (1749–1827), et leurs ouvrages respectifs *Doctrine of chances* (1718) et *Théorie analytique des probabilités* (1812)

Démonstration. On peut établir la borne sur $\theta(x) = \ln(1+x) - x + x^2/2$ en dérivant successivement. On peut aussi partir du développement en série entière du logarithme, qui s'écrit pour $|x| < 1$,

$$\ln(1+x) = \sum_{m=1}^{\infty} (-1)^{m-1} \frac{x^m}{m} = x - \frac{x^2}{2} + \theta(x) \quad \text{avec} \quad \theta(x) = \sum_{m=3}^{\infty} (-1)^{m-1} \frac{x^m}{m}.$$

Pour $|x| \leq 2/3$, on a de plus

$$|\theta(x)| \leq \sum_{m=3}^{\infty} \frac{|x|^m}{3} = \frac{|x|^3}{3(1-|x|)} \leq |x|^3,$$

ce qui conclut la preuve. □

Le deuxième lemme donne l'approximation classique de la factorielle attribuée à Stirling.

Lemme 1.8 (Formule de Stirling). *On a l'équivalent suivant de la factorielle au voisinage de l'infini, appelé formule de Stirling :*

$$n! \underset{n \rightarrow \infty}{\sim} \sqrt{2\pi n} \left(\frac{n}{e}\right)^n.$$

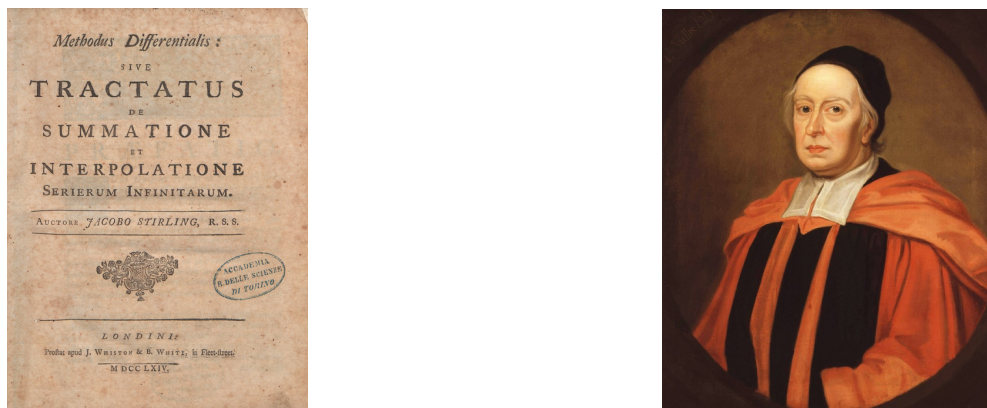


FIGURE 1.6 : *Methodus differentialis* (1764) de James Stirling (1692–1770), et John Wallis (1616–1703)

Démonstration. Pour tout entier $n \in \mathbb{N}^*$, considérons ensuite l'expression

$$d_n = \ln(n!) - \left(n + \frac{1}{2}\right) \ln n + n.$$

En considérant la fonction θ définie par le lemme 1.7, on a pour $n \geq 2$

$$d_{n+1} - d_n = \left(\ln(n+1) + \ln(n!) - \left(n+1 + \frac{1}{2}\right) \ln(n+1) + n+1\right) - \left(\ln(n!) - \left(n + \frac{1}{2}\right) \ln n + n\right)$$

$$= 1 - \left(n + \frac{1}{2}\right) \ln\left(1 + \frac{1}{n}\right) = 1 - \left(n + \frac{1}{2}\right) \left(\frac{1}{n} - \frac{1}{2n^2} + \theta\left(\frac{1}{n}\right)\right) = \frac{1}{4n^2} - \left(n + \frac{1}{2}\right) \theta\left(\frac{1}{n}\right).$$

La borne sur θ donnée par le lemme 1.7 implique que la série $\sum_n (d_{n+1} - d_n)$ converge absolument. Par télescopage, la suite $(d_n)_{n \geq 1}$ converge donc vers une limite notée l . En posant $\kappa = e^l$, on obtient

$$n! \underset{n \rightarrow \infty}{\sim} \kappa \sqrt{n} \left(\frac{n}{e}\right)^n. \quad (1.21)$$

Pour déterminer la valeur de la constante κ , on considère les *intégrales de Wallis* définies par

$$W_n = \int_0^{\pi/2} \cos^n x \, dx.$$

Comme $0 < \cos x \leq 1$ pour tout $x \in [0, \pi/2[$, on a pour tout $n \in \mathbb{N}$ les inégalités $0 < W_{n+1} \leq W_n \leq \pi/2$. De plus, en intégrant par parties, on trouve la relation

$$\begin{aligned} W_{n+2} &= \int_0^{\pi/2} (\cos x)(\cos x)^{n+1} \, dx = [(\sin x)(\cos x)^{n+1}]_0^{\pi/2} + (n+1) \int_0^{\pi/2} (\sin x)^2 (\cos x)^n \, dx \\ &= (n+1) \int_0^{\pi/2} (1 - (\cos x)^2)(\cos x)^n \, dx = (n+1)(W_n - W_{n+2}), \end{aligned}$$

ce qui conduit à la relation de récurrence

$$W_{n+2} = \frac{n+1}{n+2} W_n. \quad (1.22)$$

En combinant cette égalité avec la décroissance de la suite $(W_n)_{n \geq 0}$, on obtient

$$\frac{n+1}{n+2} = \frac{W_{n+2}}{W_n} \leq \frac{W_{n+1}}{W_n} \leq 1 \quad \text{et donc} \quad W_{n+1} \underset{n \rightarrow \infty}{\sim} W_n.$$

Posons $C_n = (n+1)W_n W_{n+1}$. La relation de récurrence (1.22) donne aussi

$$C_{n+1} = (n+2)W_{n+1}W_{n+2} = (n+2)W_{n+1} \frac{n+1}{n+2} W_n = (n+1)W_n W_{n+1} = C_n.$$

On en déduit que la suite $(C_n)_{n \geq 0}$ est constante, égale à

$$C_0 = W_0 W_1 = \left(\int_0^{\pi/2} dx\right) \left(\int_0^{\pi/2} \cos x \, dx\right) = \frac{\pi}{2} [\sin x]_0^{\pi/2} = \frac{\pi}{2}.$$

Sachant que $W_{n+1} \underset{n \rightarrow \infty}{\sim} W_n$, on observe que $C_n \underset{n \rightarrow \infty}{\sim} n W_n^2$, ce qui donne l'équivalent $W_n \underset{n \rightarrow \infty}{\sim} \sqrt{\frac{\pi}{2n}}$. Enfin, la relation de récurrence (1.22) permet de montrer que pour tout entier $p \in \mathbb{N}$,

$$W_{2p} = \left(\prod_{j=1}^p \frac{2j-1}{2j}\right) W_0 = \frac{(2p)!}{(2^p p!)^2} \frac{\pi}{2}.$$

L'équivalent ci-dessus et l'équivalent (1.21) donnent alors

$$\frac{\sqrt{\pi}}{2\sqrt{p}} \underset{p \rightarrow \infty}{\sim} W_{2p} \underset{p \rightarrow \infty}{\sim} \frac{\kappa \sqrt{2p} (2p/e)^{2p}}{2^{2p} \kappa^2 p (p/e)^{2p}} \frac{\pi}{2} = \frac{\pi}{\kappa \sqrt{2p}}.$$

On en conclut que $\kappa = \sqrt{2\pi}$, ce qui avec (1.21) donne la formule de Stirling. □

Le troisième lemme donne une approximation de la loi binomiale $\text{Bin}(n, p)$ quand n est grand qui fait appel à la fonction φ donnée par (1.20).

Lemme 1.9 (Approximation normale de la loi binomiale). Soient deux réels $p \in]0, 1[$ et $\alpha > 0$. Pour tout entier $n \in \mathbb{N}^*$, soit S_n une variable aléatoire qui suit la loi binomiale $\text{Bin}(n, p)$ et soit

$$\mathcal{K}_{n,\alpha} = \{k \in \{0, \dots, n\} \mid |x_{n,k}| \leq \alpha\} \quad \text{avec} \quad x_{n,k} = \frac{k - np}{\sqrt{np(1-p)}}.$$

Pour tout réel $\varepsilon > 0$, il existe un entier $n_0 \in \mathbb{N}^*$ tel que pour tous entiers $n \geq n_0$ et $k \in \mathcal{K}_{n,\alpha}$,

$$1 - \varepsilon \leq \frac{\mathbb{P}(S_n = k)}{(x_{n,k+1} - x_{n,k}) \varphi(x_{n,k})} \leq 1 + \varepsilon.$$

Démonstration. D'après l'expression (1.3), en posant $q = 1 - p$, on a pour tout $k \in \{0, \dots, n\}$,

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

Fixons un réel $\varepsilon \in]0, 2/5]$. D'après le lemme 1.8, il existe un entier $n_0 \in \mathbb{N}^*$ tel que pour tout $n \geq n_0$,

$$(1 - \varepsilon) \sqrt{2\pi n} \left(\frac{n}{e}\right)^n \leq n! \leq (1 + \varepsilon) \sqrt{2\pi n} \left(\frac{n}{e}\right)^n.$$

On a de plus $\sqrt{n} = o(n)$, donc il existe un entier $n_1 \in \mathbb{N}^*$ tel que pour tout $n \geq n_1$, on a simultanément les inégalités $|\alpha \sqrt{npq}| \leq \min\{p, q\} \varepsilon n$ et $\min\{p, q\}(1 - \varepsilon)n \geq n_0$. Ainsi, pour tous $n \geq n_1$ et $k \in \mathcal{K}_{n,\alpha}$,

$$\begin{cases} n_0 \leq (1 - \varepsilon)np \leq np - \alpha \sqrt{npq} \leq k \leq np + \alpha \sqrt{npq} \leq (1 + \varepsilon)np \\ n_0 \leq (1 - \varepsilon)nq \leq nq - \alpha \sqrt{npq} \leq n - k \leq nq + \alpha \sqrt{npq} \leq (1 + \varepsilon)nq. \end{cases}$$

Cela permet d'écrire que

$$\frac{(1 - \varepsilon) \sqrt{2\pi n} \left(\frac{n}{e}\right)^n}{2\pi(1 + \varepsilon)^2 \sqrt{k} \left(\frac{k}{e}\right)^k \sqrt{(n-k)} \left(\frac{n-k}{e}\right)^{n-k}} \leq \frac{n!}{k!(n-k)!} \leq \frac{(1 + \varepsilon) \sqrt{2\pi n} \left(\frac{n}{e}\right)^n}{2\pi(1 - \varepsilon)^2 \sqrt{k} \left(\frac{k}{e}\right)^k \sqrt{(n-k)} \left(\frac{n-k}{e}\right)^{n-k}},$$

et ensuite

$$\frac{1 - \varepsilon}{(1 + \varepsilon)^3} \frac{1}{\sqrt{2\pi npq}} \left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k} \leq \mathbb{P}(S_n = k) \leq \frac{1 + \varepsilon}{(1 - \varepsilon)^3} \frac{1}{\sqrt{2\pi npq}} \left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k}. \quad (1.23)$$

Comme $\varepsilon \leq 2/5$, on observe que

$$|x_{n,k}| \sqrt{npq} \leq \alpha \sqrt{npq} \leq \min\{p, q\} \varepsilon n \leq \frac{2}{3} \min\{p, q\} (1 - \varepsilon)n \leq \frac{2}{3} \min\{k, n - k\}.$$

On peut donc utiliser le lemme 1.7 afin d'écrire que

$$\begin{cases} \ln\left(\left(\frac{np}{k}\right)^k\right) = k \ln\left(1 - \frac{x_{n,k} \sqrt{npq}}{k}\right) = -x_{n,k} \sqrt{npq} - \frac{x_{n,k}^2 npq}{2k} + k\theta\left(-\frac{x_{n,k} \sqrt{npq}}{k}\right) \\ \ln\left(\left(\frac{nq}{n-k}\right)^{n-k}\right) = (n-k) \ln\left(1 + \frac{x_{n,k} \sqrt{npq}}{n-k}\right) = x_{n,k} \sqrt{npq} - \frac{x_{n,k}^2 npq}{2(n-k)} + (n-k)\theta\left(\frac{x_{n,k} \sqrt{npq}}{n-k}\right). \end{cases}$$

La somme de ces deux égalités conduit à

$$\ln\left(\left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k}\right) + \frac{x_{n,k}^2}{2} = \left(1 - \frac{n^2 pq}{k(n-k)}\right) \frac{x_{n,k}^2}{2} + \left(k\theta\left(-\frac{x_{n,k} \sqrt{npq}}{k}\right) + (n-k)\theta\left(\frac{x_{n,k} \sqrt{npq}}{n-k}\right)\right).$$

Comme $\varepsilon \leq 1/2$, on encadre ainsi le premier facteur entre parenthèses dans le membre de droite :

$$-8\varepsilon \leq \frac{-(2 - \varepsilon)\varepsilon}{(1 - \varepsilon)^2} = 1 - \frac{1}{(1 - \varepsilon)^2} \leq 1 - \frac{n^2 pq}{k(n-k)} \leq 1 - \frac{1}{(1 + \varepsilon)^2} = \frac{(2 + \varepsilon)\varepsilon}{(1 + \varepsilon)^2} \leq 3\varepsilon.$$

La borne fournie par le lemme 1.7 permet de majorer la valeur absolue du deuxième terme entre parenthèses dans le membre de droite par

$$\frac{|x_{n,k}|^3 (npq)^{3/2}}{k^2} + \frac{|x_{n,k}|^3 (npq)^{3/2}}{(n-k)^2} \leq \alpha^3 (npq)^{3/2} \left(\frac{1}{k^2} + \frac{1}{(n-k)^2} \right) \leq \frac{\alpha^3 (pq)^{3/2}}{(1-\varepsilon)^2 \sqrt{n}} \left(\frac{1}{p^2} + \frac{1}{q^2} \right) \leq \varepsilon,$$

où la dernière inégalité a lieu pour tout entier $n \geq n_2$, pour un certain entier $n_2 \geq n_1$. On en déduit que

$$\left| \ln \left(\left(\frac{np}{k} \right)^k \left(\frac{nq}{n-k} \right)^{n-k} \right) + \frac{x_{n,k}^2}{2} \right| \leq (4\alpha^2 + 1)\varepsilon.$$

En combinant avec (1.23) l'encadrement qui se déduit de cette borne en prenant l'exponentielle, on a

$$\frac{1-\varepsilon}{(1+\varepsilon)^3} e^{-(4\alpha^2+1)\varepsilon} \frac{e^{-x_{n,k}^2/2}}{\sqrt{2\pi npq}} \leq \mathbb{P}(S_n = k) \leq \frac{1+\varepsilon}{(1-\varepsilon)^3} e^{(4\alpha^2+1)\varepsilon} \frac{e^{-x_{n,k}^2/2}}{\sqrt{2\pi npq}}.$$

On conclut la preuve du lemme en observant que $x_{n,k+1} - x_{n,k} = 1/\sqrt{npq}$ et que les facteurs dépendant de ε sont arbitrairement proches de 1 pour ε suffisamment petit. $\square$

Algorithme 1.2 (Illustration de l'approximation normale de la loi binomiale). Le lemme 1.9 donne, pour une variable S_n qui suit la loi binomiale $\text{Bin}(n, p)$ et pour n grand,

$$\mathbb{P}(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k} \approx \frac{\exp\left(-\frac{(k-np)^2}{2np(1-p)}\right)}{\sqrt{2\pi np(1-p)}} = \frac{1}{\sigma} \varphi\left(\frac{k-\mu}{\sigma}\right) \quad \text{avec} \quad \begin{cases} \mu = np \\ \sigma = \sqrt{np(1-p)}. \end{cases}$$

On illustre cette approximation en comparant les tracés des deux applications $k \mapsto \mathbb{P}(S_n = k)$ et $x \mapsto \frac{1}{\sigma} \varphi\left(\frac{x-\mu}{\sigma}\right)$. La seconde est la densité de la *loi normale* (ou *gaussienne*) de paramètres μ et σ^2 . Cette loi est notée $N(\mu, \sigma^2)$. On utilise pour cela le script C.2. Les graphes produits pour diverses valeurs de n et p sont représentés sur la figure 1.7. On observe que l'approximation est d'autant meilleure que n est grand et que p est proche de $1/2$.

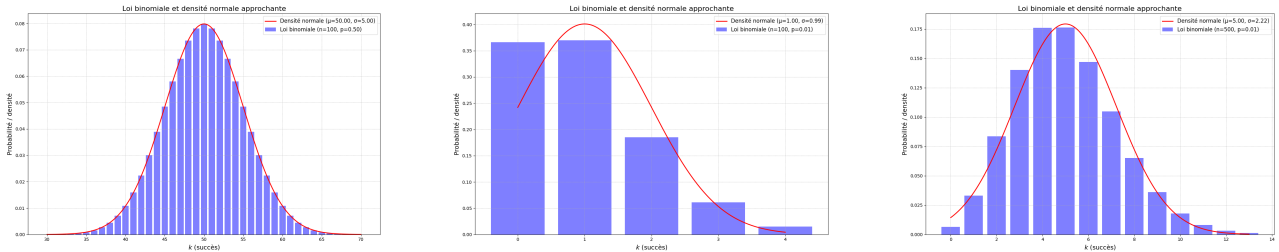


FIGURE 1.7 : Approximation de la loi binomiale par la loi normale

On peut maintenant prouver le théorème 1.6. Considérons un processus de Bernoulli $(S_n)_{n \geq 0}$ de paramètre $p \in]0, 1[$. On sait que, pour tout entier $n \in \mathbb{N}^*$, la variable aléatoire S_n suit la loi binomiale $\text{Bin}(n, p)$. Fixons $a < b$ et $\varepsilon > 0$, ainsi que $\alpha \geq \max\{|a|, |b|\}$, et considérons un entier $n_0 \in \mathbb{N}$ fourni par le lemme 1.9. Pour tous entiers $n \geq n_0$ et $k \in \{0, \dots, n\}$ tels que $a < x_{n,k} \leq b$, on a $k \in \mathcal{K}_{n,\alpha}$, donc

$$(1-\varepsilon)I_n \leq \mathbb{P}\left(a < \frac{S_n - np}{\sqrt{npq}} \leq b\right) = \sum_{\substack{k \in \{0, \dots, n\} \\ a < x_{n,k} \leq b}} \mathbb{P}(S_n = k) \leq (1+\varepsilon)I_n, \quad (1.24)$$

où l'on a posé $q = 1 - p$ et

$$I_n = \sum_{\substack{k \in \{0, \dots, n\} \\ a < x_{n,k} \leq b}} (x_{n,k+1} - x_{n,k}) \varphi(x_{n,k}).$$

On utilise maintenant l'uniforme continuité de la fonction φ sur le segment $[a, b+1]$, qui donne un réel $\delta > 0$ tel que pour tous réels $x, y \in [a, b+1]$ tels que $|x - y| \leq \delta$, on a $|\varphi(x) - \varphi(y)| \leq \varepsilon$. On choisit alors un entier $n_1 \geq n_0$ tel que pour tous entiers $n \geq n_1$ et $k \in \{0, \dots, n\}$,

$$-\sqrt{np/q} = x_{n,0} \leq a < b \leq x_{n,n} = \sqrt{nq/p} \quad \text{et} \quad x_{n,k+1} - x_{n,k} = \frac{1}{\sqrt{npq}} \leq \min\{1, \delta, b - a\}.$$

Pour $n \geq n_1$ fixé, notons x_{n,k_1} et x_{n,k_2} respectivement le plus petit et le plus grand terme de la forme $x_{n,k}$ dans l'intervalle $]a, b]$. Pour $k \in \{k_1, \dots, k_2\}$, les points $x_{n,k}$ et $x_{n,k+1}$ sont dans l'intervalle $[a, b + 1]$ et sont distants d'au plus δ , donc

$$\left| \int_{x_{n,k}}^{x_{n,k+1}} \varphi(x) dx - (x_{n,k+1} - x_{n,k})\varphi(x_{n,k}) \right| = \left| \int_{x_{n,k}}^{x_{n,k+1}} (\varphi(x) - \varphi(x_{n,k})) dx \right| \leq (x_{n,k+1} - x_{n,k})\varepsilon.$$

La relation de Chasles et l'inégalité triangulaire donnent alors

$$\left| \int_{x_{n,k_1}}^{x_{n,k_2+1}} \varphi(x) dx - I_n \right| \leq \varepsilon \sum_{k=k_1}^{k_2} (x_{n,k+1} - x_{n,k}) = (x_{n,k_2+1} - x_{n,k_1})\varepsilon \leq \left(b - a + \frac{2}{\sqrt{npq}} \right) \varepsilon.$$

La dernière inégalité vient du fait que les points a et b sont distants d'au plus $1/\sqrt{npq}$ des points x_{n,k_1} et x_{n,k_2+1} respectivement. De plus, comme la fonction φ est bornée par $1/\sqrt{2\pi}$, on a aussi

$$\left| \int_a^b \varphi(x) dx - \int_{x_{n,k_1}}^{x_{n,k_2+1}} \varphi(x) dx \right| \leq \int_a^{x_{n,k_1}} \varphi(x) dx + \int_b^{x_{n,k_2+1}} \varphi(x) dx \leq \frac{2}{\sqrt{2\pi npq}}.$$

On en déduit que pour tout entier $n \geq n_1$,

$$\left| \int_a^b \varphi(x) dx - I_n \right| \leq \frac{2}{\sqrt{2\pi npq}} + \left(b - a + \frac{2}{\sqrt{npq}} \right) \varepsilon.$$

Ainsi, I_n converge vers $\int_a^b \varphi(x) dx$ quand $n \rightarrow \infty$. On conclut en combinant cela avec (1.24).

Algorithme 1.3 (Illustration du théorème de Moivre-Laplace par une planche de Galton). Une planche de Galton est un dispositif permettant d'illustrer expérimentalement le théorème 1.6. Elle consiste à faire descendre des jetons à travers une série de bifurcations où, à chaque niveau, le jeton peut aller à droite avec probabilité p , en général égale à $1/2$, ou à gauche avec probabilité $1 - p$. Au bout de n bifurcations, chaque jeton atterrit dans l'une des $n + 1$ cases finales. Un tel dispositif est visible par exemple à la Cité des Sciences et de l'Industrie à Paris, ou dans les jeux télévisés *The Wall* et *Le Juste Prix* (sous le nom de jeu du fakir), voir la figure 1.8. Dans la forme simple du jeu, le joueur remporte une somme d'argent d'autant plus grande que le jeton atterrit dans une case tout à gauche ou tout à droite, ce qui arrive avec les probabilités les plus faibles.

On peut modéliser la position finale S_n d'un jeton par une somme de n variables indépendantes de loi $\text{Ber}(p)$, de sorte que S_n suit la loi binomiale $\text{Bin}(n, p)$. Pour n assez grand, le nombre réduit de succès S_n^* , défini par (1.19), se comporte alors approximativement selon la loi normale centrée réduite de densité φ donnée par (1.20). Pour illustrer ce phénomène, on simule la trajectoire de m jetons sur une planche de Galton, avec n bifurcations et une probabilité p d'aller à droite. Les résultats sont présentés sous forme de deux graphiques : un histogramme des positions finales S_n des m jetons ; un histogramme des quantités réduites S_n^* correspondantes, superposé au graphe de la fonction φ . On utilise le script C.3, et le résultat est sur la figure 1.9.

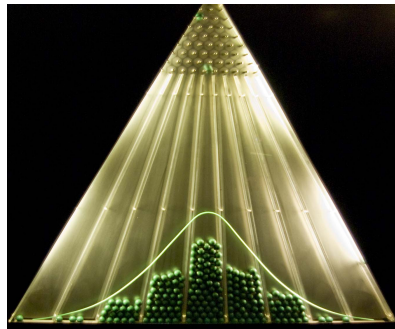
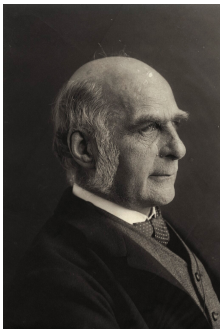


FIGURE 1.8 : Sir Francis Galton (1822–1911), planche de Galton exposée à la Cité des Sciences et de l'Industrie, et jeu télévisé *The Wall* (2017)

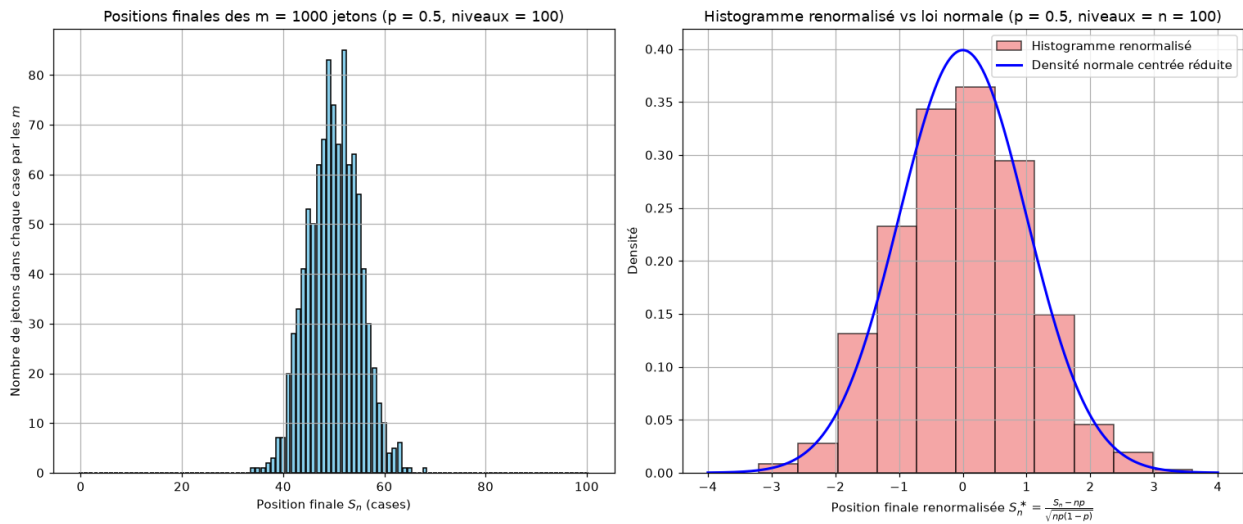


FIGURE 1.9 : Simulation d'une planche de Galton

Remarque 1.5 (Théorème limite central). Le théorème 1.6 est un cas particulier du *théorème limite central*, qui affirme que pour toute suite $(X_n)_{n \geq 1}$ de variables aléatoires i.i.d. de moyenne $\mu \in \mathbb{R}$ et de variance $\sigma^2 \in]0, \infty[$, et tous $a < b$,

$$\mathbb{P}\left(a < \frac{(X_1 + \dots + X_n) - n\mu}{\sigma\sqrt{n}} \leq b\right) \xrightarrow{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx.$$

Il s'agit d'un résultat de convergence en loi qui justifie le rôle fondamental de la loi normale dans la modélisation des phénomènes aléatoires.

1.3.4 Approximation poissonnienne et loi des petits nombres

Comme on peut le voir sur la figure 1.7, l'approximation normale de la loi binomiale est assez mauvaise quand p est proche de 0 et n n'est pas suffisamment grand. Par symétrie, on pourrait observer le même phénomène pour p proche de 1. Ces deux cas extrêmes correspondent à des succès très rares ou au contraire très fréquents. Lorsque p est petit et que np^2 est petit, on peut approcher le nombre de succès S_n par une *loi de Poisson* de paramètre np . Lorsque p est proche de 1, c'est le nombre d'échecs $n - S_n$, de loi $\text{Bin}(n, 1 - p)$, qu'on approche par une loi de Poisson de paramètre $n(1 - p)$, avec une erreur contrôlée par $n(1 - p)^2$. Le nombre de succès est alors approché par $n - N$, où N suit cette loi de Poisson. Commençons par rappeler la définition de cette loi.

Définition 1.7 (Loi de Poisson). Soit $\lambda \in [0, \infty[$. On dit qu'une variable aléatoire N suit la loi de Poisson de paramètre λ , notée $\text{Poi}(\lambda)$, si N est presque sûrement à valeurs dans $\mathbb{N}$ et vérifie

$$\forall k \in \mathbb{N} \quad \mathbb{P}(N = k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

On adopte la convention $0^0 = 1$; pour $\lambda = 0$, la loi est concentrée en 0.

La fonction génératrice d'une variable aléatoire N qui suit la loi de Poisson de paramètre $\lambda \in]0, \infty[$ est donnée, pour tout nombre complexe $z \in \mathbb{C}$, par la formule

$$G_N(z) = \sum_{k=0}^{\infty} \mathbb{P}(N = k) z^k = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} z^k = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda z)^k}{k!} = e^{-\lambda} e^{\lambda z} = e^{\lambda(z-1)}. \quad (1.25)$$

Les dérivées successives de la fonction génératrice de N vérifient alors, pour tout $r \in \mathbb{N}$,

$$G_N^{(r)}(z) = \lambda^r e^{\lambda(z-1)} \xrightarrow[z < 1]{z \rightarrow 1} \lambda^r = \mathbb{E}[(N)_r], \quad (1.26)$$

où la dernière égalité découle de la proposition A.3. On en déduit que l'espérance et la variance de la variable aléatoire N sont toutes deux égales au paramètre λ . En effet, on a

$$\mathbb{E}[N] = \mathbb{E}[(N)_1] = \lambda \quad \text{et} \quad \text{Var}(N) = \mathbb{E}[(N)_2] + \mathbb{E}[(N)_1] - \mathbb{E}[(N)_1]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda. \quad (1.27)$$

La fonction génératrice (1.25) permet de montrer simplement qu'une somme de variables aléatoires indépendantes qui suivent des lois de Poisson suit aussi une loi de Poisson.

Proposition 1.10 (Sommes de variables de Poisson indépendantes). *Pour $n \in \mathbb{N}^*$, soient $\lambda_1, \dots, \lambda_n \in [0, \infty[$ et $N_1, \dots, N_n$ des variables aléatoires indépendantes de lois respectives $\text{Poi}(\lambda_1), \dots, \text{Poi}(\lambda_n)$. Leur somme $N_1 + \dots + N_n$ suit la loi de Poisson $\text{Poi}(\lambda_1 + \dots + \lambda_n)$.*

Démonstration. D'après (1.25) et la proposition A.5 sur les fonctions génératrices de sommes de variables aléatoires indépendantes, la fonction génératrice de $N_1 + \dots + N_n$ vérifie pour tout $z \in \mathbb{C}$,

$$G_{N_1 + \dots + N_n}(z) = G_{N_1}(z) \cdots G_{N_n}(z) = e^{\lambda_1(z-1)} \cdots e^{\lambda_n(z-1)} = e^{(\lambda_1 + \dots + \lambda_n)(z-1)},$$

qui est la fonction génératrice de la loi $\text{Poi}(\lambda_1 + \dots + \lambda_n)$. On conclut avec la proposition A.4. $\square$

Afin de donner un sens au fait que des lois de variables aléatoires sont proches, on peut utiliser la notion de *variation totale*. Considérons un ensemble dénombrable $\mathcal{S}$, fini ou infini. Conformément à la définition B.1, on appelle *mesure de probabilité*, ou plus simplement *probabilité*, sur $\mathcal{S}$ tout vecteur ligne $\mu = (\mu_s)_{s \in \mathcal{S}}$ à coefficients dans $[0, 1]$ tel que $\sum_{s \in \mathcal{S}} \mu_s = 1$. Pour tout sous-ensemble $A \subseteq \mathcal{S}$, on note $\mu(A) = \sum_{s \in A} \mu_s$, et on appelle cette quantité la *mesure* de l'ensemble A .

Exemple 1.1 (Loi d'une variable aléatoire discrète). Soit X une variable aléatoire *discrète*, c'est-à-dire à valeurs dans un ensemble dénombrable $\mathcal{S}$. En posant $\mu_s = \mathbb{P}(X = s)$ pour tout $s \in \mathcal{S}$, on obtient une mesure de probabilité $\mu = (\mu_s)_{s \in \mathcal{S}}$ sur $\mathcal{S}$ qui vérifie $\mu(A) = \mathbb{P}(X \in A)$ pour tout sous-ensemble $A \subseteq \mathcal{S}$. Cette mesure de probabilité μ est appelée la *loi* de la variable aléatoire X .

On peut alors comparer deux mesures de probabilités sur un ensemble dénombrable $\mathcal{S}$ à l'aide de la notion définie ci-dessous.

Définition 1.8 (Distance en variation totale). Soit $\mathcal{S}$ un ensemble dénombrable. La *distance en variation totale* entre deux mesures de probabilité μ et ν sur $\mathcal{S}$ est définie par la formule

$$d_{\text{TV}}(\mu, \nu) = \sup_{A \subseteq \mathcal{S}} |\mu(A) - \nu(A)|.$$

La distance en variation totale entre deux mesures de probabilité est donc l'écart maximal entre les mesures qu'elles attribuent à un même sous-ensemble de $\mathcal{S}$. Il s'agit bien d'une distance au sens topologique sur l'ensemble des mesures de probabilité sur $\mathcal{S}$. En particulier, la distance en variation totale entre deux mesures de probabilité est nulle si et seulement si ces mesures sont égales, et l'inégalité triangulaire est clairement vérifiée : pour toutes mesures de probabilité μ, ν et ρ sur $\mathcal{S}$,

$$d_{\text{TV}}(\mu, \nu) = \sup_{A \subseteq \mathcal{S}} |\mu(A) - \nu(A)| \leq \sup_{A \subseteq \mathcal{S}} |\mu(A) - \rho(A)| + \sup_{A \subseteq \mathcal{S}} |\rho(A) - \nu(A)| = d_{\text{TV}}(\mu, \rho) + d_{\text{TV}}(\rho, \nu). \quad (1.28)$$

Le résultat suivant donne une autre formulation de la distance en variation totale. Pour simplifier, on suppose que $\mathcal{S} = \mathbb{N}$, puisque nous voulons comparer les lois binomiales et de Poisson, à valeurs entières.

Proposition 1.11 (Formulation équivalente sur $\mathbb{N}$ de la distance en variation totale). *Soient μ et ν deux mesures de probabilité sur $\mathbb{N}$. On a*

$$d_{\text{TV}}(\mu, \nu) = \frac{1}{2} \sum_{k=0}^{\infty} |\mu_k - \nu_k|.$$

Démonstration. Considérons un sous-ensemble $A \subseteq \mathbb{N}$ et rappelons que l'indicatrice de A , notée 1_A , est la fonction qui prend la valeur 1 sur l'ensemble A et la valeur 0 ailleurs. Observons que

$$\mu(A) + \mu(\mathbb{N} \setminus A) = \sum_{k \in A} \mu_k + \sum_{k \in \mathbb{N} \setminus A} \mu_k = \sum_{k=0}^{\infty} 1_A(k) \mu_k + \sum_{k=0}^{\infty} 1_{\mathbb{N} \setminus A}(k) \mu_k = \sum_{k=0}^{\infty} (1_A(k) + 1_{\mathbb{N} \setminus A}(k)) \mu_k = \sum_{k=0}^{\infty} \mu_k = 1,$$

et de même pour ν . Considérons alors la fonction $f = 1_A - 1_{\mathbb{N} \setminus A}$, qui vaut 1 sur l'ensemble A et -1 sur son complémentaire. On a, avec convergence absolue de toutes les séries considérées,

$$\begin{aligned} \sum_{k=0}^{\infty} f(k)(\mu_k - \nu_k) &= \sum_{k=0}^{\infty} (1_A(k) - 1_{\mathbb{N} \setminus A}(k))(\mu_k - \nu_k) \\ &= \sum_{k=0}^{\infty} 1_A(k) \mu_k - \sum_{k=0}^{\infty} 1_A(k) \nu_k - \sum_{k=0}^{\infty} 1_{\mathbb{N} \setminus A}(k) \mu_k + \sum_{k=0}^{\infty} 1_{\mathbb{N} \setminus A}(k) \nu_k \\ &= \mu(A) - \nu(A) - \mu(\mathbb{N} \setminus A) + \nu(\mathbb{N} \setminus A) \\ &= (\mu(A) - \nu(A)) - (1 - \mu(A)) + (1 - \nu(A)) = 2(\mu(A) - \nu(A)). \end{aligned}$$

L'inégalité triangulaire pour les séries et le fait que la fonction $|f|$ est constante égale à 1 conduisent à

$$2|\mu(A) - \nu(A)| = \left| \sum_{k=0}^{\infty} f(k)(\mu_k - \nu_k) \right| \leq \sum_{k=0}^{\infty} |f(k)| |\mu_k - \nu_k| \leq \sum_{k=0}^{\infty} |\mu_k - \nu_k|.$$

On en déduit une première inégalité en prenant le supremum sur $A \subseteq \mathbb{N}$. Les inégalités ci-dessus deviennent des égalités si $A = \{k \in \mathbb{N} \mid \mu_k \geq \nu_k\}$, car on a alors, quel que soit $k \in \mathbb{N}$,

$$f(k)(\mu_k - \nu_k) = |\mu_k - \nu_k| \geq 0.$$

On en déduit directement l'égalité annoncée. $\square$

Le résultat ci-dessous concerne la loi d'une somme de variables aléatoires indépendantes à valeurs dans $\mathbb{N}$, et son comportement vis-à-vis de la distance en variation totale.

Proposition 1.12 (Produit de convolution et distance en variation totale). Soient μ et ν deux mesures de probabilité sur $\mathbb{N}$. On considère le vecteur $\mu * \nu$ dont les composantes sont données, pour $k \in \mathbb{N}$, par la formule

$$(\mu * \nu)_k = \sum_{j=0}^k \mu_j \nu_{k-j}.$$

On a alors les propriétés suivantes :

- (i). Le vecteur $\mu * \nu$ est une probabilité sur $\mathbb{N}$, appelée le produit de convolution de μ et ν ;
- (ii). Si X et Y sont deux variables aléatoires indépendantes à valeurs dans $\mathbb{N}$ de lois respectives μ et ν , alors la variable aléatoire $X + Y$ a pour loi le produit de convolution $\mu * \nu$;
- (iii). Si μ' et ν' sont deux mesures de probabilité sur $\mathbb{N}$, alors on a la majoration

$$d_{TV}(\mu * \nu, \mu' * \nu') \leq d_{TV}(\mu, \mu') + d_{TV}(\nu, \nu').$$

Démonstration. Pour prouver (i), on commence par observer que $(\mu * \nu)_k \in [0, \infty]$ pour tout $k \in \mathbb{N}$. Ensuite, on utilise d'une part la proposition B.2 avec la partition de $\mathbb{N}^2$ donnée par les ensembles finis $\{(j, k-j), j \in \{0, \dots, k\}\}$, pour $k \in \mathbb{N}$, et d'autre part le théorème B.3. On obtient alors

$$\sum_{(m,n) \in \mathbb{N}^2} \mu_m \nu_n = \sum_{k=0}^{\infty} \sum_{j=0}^k \mu_j \nu_{k-j} = \sum_{k=0}^{\infty} (\mu * \nu)_k \quad \text{et} \quad \sum_{(m,n) \in \mathbb{N}^2} \mu_m \nu_n = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \mu_m \nu_n = 1.$$

Pour établir (ii), considérons deux variables aléatoires X et Y indépendantes à valeurs dans $\mathbb{N}$, de lois respectives μ et ν . On a alors, pour tout $k \in \mathbb{N}$,

$$\mathbb{P}(X + Y = k) = \mathbb{P}\left(\bigsqcup_{j=0}^k \{X = j \text{ et } Y = k - j\}\right) = \sum_{j=0}^k \mathbb{P}(X = j)\mathbb{P}(Y = k - j) = (\mu * \nu)_k,$$

Le symbole $\bigsqcup$ indique que l'union est constituée d'événements disjoints, voir la remarque B.1.

Pour démontrer (iii), considérons deux mesures de probabilité μ' et ν' sur $\mathbb{N}$. Pour tout $k \in \mathbb{N}$,

$$(\mu * \nu)_k - (\mu' * \nu')_k = \sum_{j=0}^k \mu_j \nu_{k-j} - \sum_{j=0}^k \mu'_j \nu'_{k-j} = \sum_{j=0}^k (\mu_j - \mu'_j) \nu_{k-j} + \sum_{j=0}^k \mu'_j (\nu_{k-j} - \nu'_{k-j}).$$

En utilisant l'inégalité triangulaire, en sommant sur $k \in \mathbb{N}$, en procédant comme dans la preuve de (i) pour réécrire les sommes doubles, on obtient

$$\begin{aligned} \sum_{k=0}^{\infty} |(\mu * \nu)_k - (\mu' * \nu')_k| &\leq \sum_{k=0}^{\infty} \sum_{j=0}^k |\mu_j - \mu'_j| \nu_{k-j} + \sum_{k=0}^{\infty} \sum_{j=0}^k \mu'_j |\nu_{k-j} - \nu'_{k-j}| \\ &= \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} |\mu_m - \mu'_m| \nu_n + \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \mu'_m |\nu_n - \nu'_n| \\ &= \sum_{m=0}^{\infty} |\mu_m - \mu'_m| + \sum_{n=0}^{\infty} |\nu_n - \nu'_n|. \end{aligned}$$

On conclut à l'aide de l'expression de la distance en variation totale donnée par la proposition 1.11. $\square$

Le lien entre la loi binomiale et la loi de Poisson sera une conséquence du théorème suivant, qui majore la distance en variation totale entre la loi d'une somme de variables aléatoires indépendantes qui suivent des lois de Bernoulli et une loi de Poisson de paramètre approprié.

Théorème 1.13 (Inégalité de Le Cam). *Pour $n \in \mathbb{N}^*$, soient $p_1, \dots, p_n \in [0, 1]$ et $X_1, \dots, X_n$ des variables aléatoires indépendantes, de lois respectives $\text{Ber}(p_1), \dots, \text{Ber}(p_n)$. Notons $\mu^{(n)}$ la loi de la variable aléatoire $X_1 + \dots + X_n$, et $\nu^{(n)}$ la mesure de probabilité correspondant à la loi $\text{Poi}(p_1 + \dots + p_n)$, c'est-à-dire que pour tout $k \in \mathbb{N}$,*

$$\mu_k^{(n)} = \mathbb{P}(X_1 + \dots + X_n = k) \quad \text{et} \quad \nu_k^{(n)} = e^{-(p_1 + \dots + p_n)} \frac{(p_1 + \dots + p_n)^k}{k!}.$$

On a alors la majoration

$$d_{\text{TV}}(\mu^{(n)}, \nu^{(n)}) \leq p_1^2 + \dots + p_n^2.$$



AN APPROXIMATION THEOREM FOR THE POISSON BINOMIAL DISTRIBUTION

LUCIEN LE CAM

1. Introduction. Let $x, j = 1, 2, \dots$ be independent random variables such that $\text{Prob}\{X = j\} = 1 - \text{Prob}\{X = 0\} = p_j$. Let $Q = \sum_{j=1}^{\infty} p_j$ be the distribution of their sum. This kind of distribution is often referred to as a Poisson binomial distribution. For any finite measure μ on the real line let $\|\mu\|$ be the norm defined by

$$\|\mu\| = \sup \left\{ \left| \int f d\mu \right| \right\},$$

the supremum being taken over all measurable functions f such that $|f| \leq 1$. Let $\lambda = \sum_{j=1}^{\infty} p_j$, let $2\lambda = \lambda + \lambda$ and let $\alpha = \sup_j p_j$. Finally let P be the Poisson distribution whose expectation is equal to λ . The purpose of the present paper is to show that there exist absolute constants D and δ , such that $\|Q - P\| \leq D\alpha$ for all values of the p_j 's and $\|Q - P\| \leq \delta \lambda$ if $\lambda \leq 1$ and $\alpha \leq \delta$. The constant D is not larger than 9 and the constant δ is not larger than 10.

Such a result can be considered a generalization of a theorem of Yu. V. Prohorov [6] according to which such constants exist when all the probabilities p_j are equal.

The norm $\|Q - P\|$ is always larger than the maximum distance $\sup_j |p_j - P_j|$ between the cumulative distributions. For the distance ρ a very general theorem of A. N. Kolmogorov [4] implies that $\rho(P, Q)$ is at most of order $\alpha^{1/2}$. The improvement obtained here is made possible by the smaller scope of our assumptions.

The method of proof used in the present paper is not quite elementary, since it uses both operator theoretic methods and characteristic functions. The relevant concepts are described in §2.

A completely elementary approach, described in [4] leads to bounds of the order of $\lambda^{1/2}$ for the distance ρ . Unfortunately, the elementary method does not seem to be able to provide the more precise result of the present paper.

The developments given here were prompted by discussions with J. H. Hoeffding, Jr. in connection with the writing of [6].

2. Measures as operators. Let $(\mathcal{H}, \mathcal{U})$ be a measurable Abelian group, that is, an Abelian group on which a σ -field $\mathcal{H}$ has been selected. Received December 9, 1959. The author is a research fellow of the A.P. Sloan Foundation.

FIGURE 1.10 : Lucien Le Cam (1924–2000) et son article *An approximation theorem for the Poisson binomial distribution* (1960) dans lequel est prouvée l'inégalité du théorème 1.13

Démonstration. Le résultat se prouve par récurrence sur $n \in \mathbb{N}^*$. Pour $n = 1$, les mesures de probabilité $\mu^{(1)}$ et $\nu^{(1)}$ correspondent respectivement aux lois $\text{Ber}(p_1)$ et $\text{Poi}(p_1)$, donc d'après la proposition 1.11,

$$\begin{aligned} 2 d_{\text{TV}}(\mu^{(1)}, \nu^{(1)}) &= \sum_{k=0}^{\infty} |\mu_k^{(1)} - \nu_k^{(1)}| = |(1 - p_1) - e^{-p_1}| + |p_1 - p_1 e^{-p_1}| + \sum_{k=2}^{\infty} \left| e^{-p_1} \frac{p_1^k}{k!} \right| \\ &= (e^{-p_1} - 1 + p_1) + (p_1 - p_1 e^{-p_1}) + (1 - e^{-p_1} - p_1 e^{-p_1}) \\ &= 2p_1(1 - e^{-p_1}) \leq 2p_1^2. \end{aligned}$$

On observe ensuite à l'aide des propositions 1.10 et 1.12 que $\mu^{(n)} = \mu^{(n-1)} * \tilde{\mu}^{(n)}$ et $\nu^{(n)} = \nu^{(n-1)} * \tilde{\nu}^{(n)}$, où $\tilde{\mu}^{(n)}$ et $\tilde{\nu}^{(n)}$ désignent respectivement les mesures de probabilité correspondant aux lois $\text{Ber}(p_n)$ et $\text{Poi}(p_n)$. On a alors, à nouveau à l'aide de la proposition 1.12,

$$d_{\text{TV}}(\mu^{(n)}, \nu^{(n)}) \leq d_{\text{TV}}(\mu^{(n-1)}, \nu^{(n-1)}) + d_{\text{TV}}(\tilde{\mu}^{(n)}, \tilde{\nu}^{(n)}) \leq d_{\text{TV}}(\mu^{(n-1)}, \nu^{(n-1)}) + p_n^2,$$

où la dernière inégalité résulte du cas $n = 1$. On conclut donc par une récurrence élémentaire. $\square$

Le théorème 1.13 permet d'estimer la proximité entre la loi binomiale et la loi de Poisson. Plus précisément, si les probabilités $p_1, \dots, p_n$ sont toutes égales à une même valeur p , la somme $X_1 + \dots + X_n$ est constituée de n variables aléatoires i.i.d. de loi $\text{Ber}(p)$, donc la mesure de probabilité $\mu^{(n)}$ correspond à la loi $\text{Bin}(n, p)$. De plus, la mesure de probabilité $\nu^{(n)}$ correspond à la loi $\text{Poi}(np)$. À l'aide de la proposition 1.11, on en déduit le résultat ci-dessous.

Corollaire 1.14 (Approximation poissonnienne de la loi binomiale). *Pour tous $n \in \mathbb{N}^*$ et $p \in [0, 1]$, si S_n est une variable aléatoire qui suit la loi $\text{Bin}(n, p)$, alors*

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

Ce résultat est à comparer avec le théorème 1.6 et le lemme 1.9, qui donnent une approximation de la loi binomiale par une loi normale. L'approximation poissonnienne est à privilégier dans le cas où p est proche de 0 et n est grand, de sorte que la borne np^2 sur la distance en variation totale entre les lois binomiale et de Poisson est petite. Par symétrie, si p est proche de 1 et n est grand, on utilise le fait que $n - S_n$ suit la loi $\text{Bin}(n, 1 - p)$, et on l'approche par la loi de Poisson de paramètre $n(1 - p)$. On se ramène à la situation précédente puisque $1 - p$ est alors proche de 0.

Algorithme 1.4 (Illustration de l'approximation poissonnienne de la loi binomiale). On illustre le corollaire 1.14 en comparant les tracés des applications $k \mapsto \mathbb{P}(S_n = k)$ et $k \mapsto e^{-np} \frac{(np)^k}{k!}$, où S_n suit la loi $\text{Bin}(n, p)$. On utilise pour cela le script C.4. Les graphes produits pour diverses valeurs de n et p sont représentés sur la figure 1.11. La borne du corollaire 1.14 garantit une bonne approximation lorsque np^2 est petit.

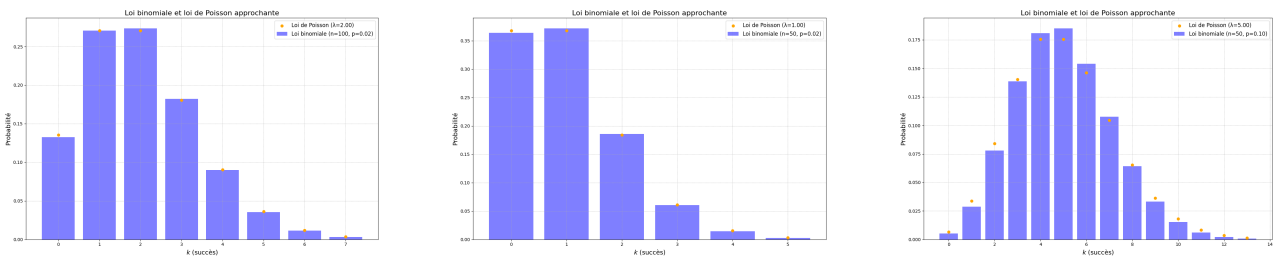


FIGURE 1.11 : Approximation de la loi binomiale par la loi de Poisson

Le théorème 1.13 admet aussi pour corollaire le résultat de convergence en loi suivant, connu sous le nom de *loi des petits nombres*.

Corollaire 1.15 (Loi des petits nombres). Soit $(X_{n,m})_{1 \leq m \leq n}$ un tableau triangulaire de variables aléatoires qui suivent des lois de Bernoulli, telles que $p_{n,m}$ désigne le paramètre de la variable $X_{n,m}$. On suppose que les variables de chaque ligne sont indépendantes. On suppose que, pour un nombre réel $\lambda \in]0, \infty[$, les paramètres de la n -ième ligne du tableau vérifient

$$p_{n,1} + \dots + p_{n,n} \xrightarrow{n \rightarrow \infty} \lambda \quad \text{et} \quad \max\{p_{n,1}, \dots, p_{n,n}\} \xrightarrow{n \rightarrow \infty} 0.$$

La loi de la somme $S_n = X_{n,1} + \dots + X_{n,n}$ sur la n -ième ligne du tableau converge vers la loi de Poisson de paramètre λ , au sens que

$$\sum_{k=0}^{\infty} \left| \mathbb{P}(S_n = k) - e^{-\lambda} \frac{\lambda^k}{k!} \right| \xrightarrow{n \rightarrow \infty} 0.$$

Démonstration. Notons respectivement $\mu^{(n)}$ et $\nu^{(n)}$ la loi de la variable aléatoire S_n et la loi de Poisson de paramètre $p_{n,1} + \dots + p_{n,n}$. D'après le théorème 1.13, on a

$$d_{TV}(\mu^{(n)}, \nu^{(n)}) \leq p_{n,1}^2 + \dots + p_{n,n}^2 \leq (p_{n,1} + \dots + p_{n,n}) \max\{p_{n,1}, \dots, p_{n,n}\} \xrightarrow{n \rightarrow \infty} 0.$$

Notons de plus ν la loi de Poisson de paramètre λ . Pour tout $k \in \mathbb{N}$, on observe que

$$\nu_k^{(n)} = e^{-(p_{n,1} + \dots + p_{n,n})} \frac{(p_{n,1} + \dots + p_{n,n})^k}{k!} \xrightarrow{n \rightarrow \infty} e^{-\lambda} \frac{\lambda^k}{k!} = \nu_k.$$

La proposition 1.16 ci-après montre alors que $d_{TV}(\nu^{(n)}, \nu) \xrightarrow{n \rightarrow \infty} 0$. L'inégalité triangulaire (1.28) pour la distance en variation totale implique que $d_{TV}(\mu^{(n)}, \nu) \xrightarrow{n \rightarrow \infty} 0$. On conclut avec la proposition 1.11. $\square$

Proposition 1.16 (Caractérisation de la convergence en variation totale sur $\mathbb{N}$). Soient $\mu^{(n)}$, pour $n \in \mathbb{N}$, et μ des mesures de probabilité sur $\mathbb{N}$. Alors,

$$d_{TV}(\mu^{(n)}, \mu) \xrightarrow{n \rightarrow \infty} 0 \quad \Longleftrightarrow \quad [\forall k \in \mathbb{N} \quad \mu_k^{(n)} \xrightarrow{n \rightarrow \infty} \mu_k].$$

Démonstration. L'implication directe est claire, par définition de la distance en variation totale. Pour établir l'implication réciproque, observons que pour tous entiers $n, K \in \mathbb{N}$,

$$\begin{aligned} \sum_{k=0}^{\infty} |\mu_k^{(n)} - \mu_k| &\leq \sum_{k=0}^K |\mu_k^{(n)} - \mu_k| + \sum_{k=K+1}^{\infty} \mu_k^{(n)} + \sum_{k=K+1}^{\infty} \mu_k \\ &= \sum_{k=0}^K |\mu_k^{(n)} - \mu_k| + \left(\left(\sum_{k=K+1}^{\infty} \mu_k + \sum_{k=0}^K \mu_k \right) - \sum_{k=0}^K \mu_k^{(n)} \right) + \sum_{k=K+1}^{\infty} \mu_k \\ &= \sum_{k=0}^K |\mu_k^{(n)} - \mu_k| + \sum_{k=0}^K (\mu_k - \mu_k^{(n)}) + 2 \sum_{k=K+1}^{\infty} \mu_k. \end{aligned}$$

On a utilisé l'inégalité triangulaire et le fait que $\mu^{(n)}$ et μ sont des mesures de probabilité. Grâce à la proposition 1.11, on en déduit que

$$d_{TV}(\mu^{(n)}, \mu) \leq \sum_{k=0}^K |\mu_k^{(n)} - \mu_k| + \sum_{k=K+1}^{\infty} \mu_k.$$

Supposons que $\mu_k^{(n)} \xrightarrow{n \rightarrow \infty} \mu_k$ pour tout $k \in \mathbb{N}$, et considérons un réel $\varepsilon > 0$. Comme μ est une mesure de probabilité, il existe un entier $K \in \mathbb{N}$ tel que $\sum_{k=K+1}^{\infty} \mu_k \leq \varepsilon$. De plus, pour tout $k \in \{0, \dots, K\}$, il existe un entier $n_k \in \mathbb{N}$ tel que pour tout entier $n \geq n_k$, on a $|\mu_k^{(n)} - \mu_k| \leq \varepsilon/(K+1)$. Par conséquent, pour tout entier $n \geq \max\{n_0, \dots, n_K\}$, on a la majoration $d_{TV}(\mu^{(n)}, \mu) \leq 2\varepsilon$, ce qui permet de conclure. $\square$

1.4 Aspect markovien

Le résultat ci-dessous montre que les processus de Bernoulli sont des exemples de chaînes de Markov. Une présentation générale de cette dernière notion est donnée dans l'annexe B.

Proposition 1.17 (Aspect markovien du processus de Bernoulli). Soit $(S_n)_{n \geq 0}$ un processus de Bernoulli de paramètre $p \in [0, 1]$. Le processus $(S_n)_{n \geq 0}$ est une chaîne de Markov homogène à espace d'états $\mathbb{N}$, dont la loi initiale est le vecteur $(1, 0, \dots)$, et la matrice de transition $P = (p_{k,l})_{k,l \in \mathbb{N}}$ a des coefficients donnés, pour tous $k, l \in \mathbb{N}$, par la formule

$$p_{k,l} = \begin{cases} 1 - p & \text{si } l = k \\ p & \text{si } l = k + 1 \\ 0 & \text{sinon.} \end{cases}$$

De plus, pour tout $m \in \mathbb{N}$, les coefficients $p_{k,l}^{(m)}$ de la matrice P^m vérifient, pour tous $k, l \in \mathbb{N}$,

$$p_{k,l}^{(m)} = \begin{cases} \binom{m}{l-k} p^{l-k} (1-p)^{m-l+k} & \text{si } 0 \leq l - k \leq m \\ 0 & \text{sinon.} \end{cases}$$

Démonstration. On note $(X_m)_{m \geq 1}$ une suite de variables indépendantes et identiquement distribuées selon la loi $\text{Ber}(p)$, qui conduit au processus (S_n) via (1.5). Pour tout $n \in \mathbb{N}$, on a $S_{n+1} = S_n + X_{n+1}$, et la variable aléatoire X_{n+1} est indépendante des variables aléatoires $S_0, \dots, S_n$, qui sont des fonctions linéaires de $X_1, \dots, X_n$. Ainsi, pour tous $s_0, \dots, s_{n-1}, k, l \in \mathbb{N}$, on a

$$\begin{aligned} \mathbb{P}(S_0 = s_0, \dots, S_{n-1} = s_{n-1}, S_n = k, S_{n+1} = l) &= \mathbb{P}(S_0 = s_0, \dots, S_{n-1} = s_{n-1}, S_n = k, X_{n+1} = l - k) \\ &= \mathbb{P}(S_0 = s_0, \dots, S_{n-1} = s_{n-1}, S_n = k) \mathbb{P}(X_{n+1} = l - k) \\ &= \mathbb{P}(S_0 = s_0, \dots, S_{n-1} = s_{n-1}, S_n = k) p_{k,l}, \end{aligned}$$

où $p_{k,l}$ est défini comme dans l'énoncé de la proposition. En sommant ces identités sur tous les entiers $s_0, \dots, s_{n-1} \in \mathbb{N}$, on obtient aussi

$$\mathbb{P}(S_n = k \text{ et } S_{n+1} = l) = \mathbb{P}(S_n = k) p_{k,l}.$$

Si la probabilité $\mathbb{P}(S_0 = s_0, \dots, S_{n-1} = s_{n-1}, S_n = k)$ est strictement positive, on en déduit que

$$\mathbb{P}(S_{n+1} = l \mid S_0 = s_0, \dots, S_{n-1} = s_{n-1}, S_n = k) = p_{k,l} = \mathbb{P}(S_{n+1} = l \mid S_n = k).$$

Le processus $(S_n)_{n \geq 0}$ est donc bien une chaîne de Markov homogène à espace d'états $\mathbb{N}$, de matrice de transition $P = (p_{k,l})_{k,l \in \mathbb{N}}$. Son état initial est clairement $S_0 = 0$ d'après (1.5), donc sa loi initiale est bien le vecteur $(1, 0, \dots)$. Enfin, notons I la matrice identité et Q la matrice de décalage, de coefficients $Q_{k,l} = \mathbf{1}_{\{l=k+1\}}$. On a $P = (1-p)I + pQ$ et, pour tout entier $j \geq 0$, $(Q^j)_{k,l} = \mathbf{1}_{\{l=k+j\}}$. Comme I et Q commutent, la formule du binôme donne, pour tout $m \in \mathbb{N}$,

$$P^m = \sum_{j=0}^m \binom{m}{j} p^j (1-p)^{m-j} Q^j.$$

En prenant le coefficient d'indices k, l , on retrouve la formule annoncée, y compris pour les états jamais visités depuis l'état initial et pour $p = 0$ ou $p = 1$, avec la convention $0^0 = 1$. $\square$

D'après la proposition 1.17, le graphe associé à un processus de Bernoulli $(S_n)_{n \geq 0}$ de paramètre p , vu comme une chaîne de Markov homogène, admet $\mathbb{N}$ pour ensemble de sommets, et fait apparaître un arc entre un entier et son successeur si $p > 0$, et entre un entier et lui-même si $p < 1$. On représente ce graphe sur la figure 1.12. Si $p > 0$, tous les états $l \in \mathbb{N}$ sont accessibles à partir des états $k \leq l$, mais pas accessibles à partir des états $k > l$. Chaque état ne communique donc qu'avec lui-même. Ainsi, les classes d'équivalence sont des singletons, et la chaîne n'est clairement pas irréductible. De plus, toutes les classes sont transientes. Si $p = 0$, un état n'est accessible qu'à partir de lui-même. Toutes les classes sont des singletons et sont finales. Autrement dit, tous les états sont absorbants.

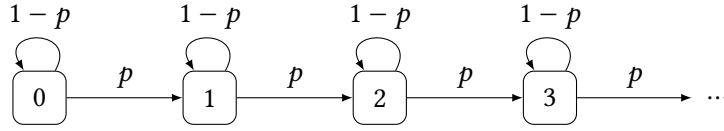


FIGURE 1.12 : Graphe du processus de Bernoulli vu comme une chaîne de Markov

Considérons, outre un processus de Bernoulli $(S_n)_{n \geq 0}$ de paramètre p , une variable aléatoire K à valeurs dans $\mathbb{N}$ indépendante du processus. On peut montrer à l'aide de la proposition 1.17 que le processus défini, pour tout entier $n \geq 0$, par la formule

$$Y_n = K + S_n \quad (1.29)$$

est une chaîne de Markov homogène à espace d'états $\mathbb{N}$, de matrice de transition celle de $(S_n)_{n \geq 0}$.

Les instants de visite $T_1^{(k)}$ et le nombre total de visites N_k d'un entier $k \in \mathbb{N}$ par la chaîne $(Y_n)_{n \geq 0}$ sont définis par (B.3) et (B.4). Si $\mathbb{P}(Y_0 = k) > 0$, le premier instant de retour en k vérifie

$$\begin{cases} \mathbb{P}(T_1^{(k)} = 1 \mid Y_0 = k) = \mathbb{P}(Y_1 = k \mid Y_0 = k) = 1 - p \\ \mathbb{P}(T_1^{(k)} = \infty \mid Y_0 = k) = \mathbb{P}(\forall n \geq 1, Y_n \neq k \mid Y_0 = k) = \mathbb{P}(Y_1 = k + 1 \mid Y_0 = k) = p. \end{cases}$$

À l'aide de la proposition B.14, qui relie la probabilité de finitude du premier instant de retour en k et l'espérance du nombre total de visites de k , on en conclut que

$$\mathbb{P}(T_1^{(k)} < \infty \mid Y_0 = k) = 1 - p, \quad \mathbb{E}[T_1^{(k)} \mid Y_0 = k] = \begin{cases} \infty & \text{si } p > 0 \\ 1 & \text{si } p = 0, \end{cases} \quad \mathbb{E}[N_k \mid Y_0 = k] = \begin{cases} 1/p & \text{si } p > 0 \\ \infty & \text{si } p = 0. \end{cases}$$

Ainsi, les états sont tous transients pour la chaîne $(Y_n)_{n \geq 0}$ si $p > 0$, et sont tous récurrents positifs si $p = 0$.

1.5 Processus de Bernoulli modulo un entier

On obtient à partir d'un processus de Bernoulli une nouvelle chaîne de Markov en ne conservant que le chiffre des unités de son écriture en base $b \geq 2$. Plus précisément, étant donné un processus de Bernoulli $(S_n)_{n \geq 0}$, définissons pour tout entier $n \geq 0$,

$$S_n^{[b]} = S_n \mod b.$$

On obtient ainsi une suite de variables aléatoires à valeurs dans l'ensemble $\mathbb{Z}/b\mathbb{Z}$, identifié à l'ensemble $\{0, \dots, b-1\}$. Cette suite hérite du caractère markovien du processus de Bernoulli de départ.

Proposition 1.18 (Aspect markovien du processus de Bernoulli modulo b). *Soit $(S_n)_{n \geq 0}$ un processus de Bernoulli de paramètre $p \in [0, 1]$. Pour tout entier $b \geq 2$, le processus $(S_n^{[b]})_{n \geq 0}$ est une chaîne de Markov homogène à espace d'états $\mathbb{Z}/b\mathbb{Z}$. Sa loi initiale est le vecteur $(1, 0, \dots, 0)$. Sa matrice de transition est la matrice de taille $b \times b$ dont les coefficients sur la diagonale valent $1 - p$, les coefficients juste au-dessus de la diagonale et le coefficient en bas à gauche valent p , et les autres coefficients sont nuls, c'est-à-dire la matrice*

$$P^{[b]} = \begin{pmatrix} 1-p & p & & \\ & 1-p & \ddots & \\ & & \ddots & p \\ p & & & 1-p \end{pmatrix}.$$

Démonstration. Pour tout $n \in \mathbb{N}$ et tout $k \in \mathbb{Z}/b\mathbb{Z}$, on a $S_n^{[b]} = k$ si et seulement si $S_n = k + br$ pour un entier $r \in \mathbb{N}$. De plus, la proposition 1.17 montre que le processus de Bernoulli $(S_n)_{n \geq 0}$ est une chaîne de Markov.

Notons $p_{k,l}$, pour $k, l \in \mathbb{N}$, les coefficients de sa matrice de transition. Pour des entiers $n, r_0, \dots, r_{n+1} \in \mathbb{N}$ et $s_0, \dots, s_{n-1}, k, l \in \mathbb{Z}/b\mathbb{Z}$ quelconques, on a

$$\begin{aligned} & \mathbb{P}(S_0 = s_0 + br_0, \dots, S_{n-1} = s_{n-1} + br_{n-1}, S_n = k + br_n, S_{n+1} = l + br_{n+1}) \\ &= p_{k+br_n, l+br_{n+1}} \mathbb{P}(S_0 = s_0 + br_0, \dots, S_{n-1} = s_{n-1} + br_{n-1}, S_n = k + br_n). \end{aligned}$$

En notant $p_{k,l}^{[b]}$, pour $k, l \in \mathbb{Z}/b\mathbb{Z}$, les coefficients de la matrice $P^{[b]}$ de l'énoncé, et en sommant l'identité précédente sur tous les entiers $r_0, \dots, r_{n+1} \in \mathbb{N}$, on obtient

$$\begin{aligned} & \mathbb{P}(S_0^{[b]} = s_0, \dots, S_{n-1}^{[b]} = s_{n-1}, S_n^{[b]} = k, S_{n+1}^{[b]} = l) \\ &= \sum_{r_0, \dots, r_n \in \mathbb{N}} \mathbb{P}(S_0 = s_0 + br_0, \dots, S_{n-1} = s_{n-1} + br_{n-1}, S_n = k + br_n) \sum_{r_{n+1} \in \mathbb{N}} p_{k+br_n, l+br_{n+1}} \\ &= p_{k,l}^{[b]} \mathbb{P}(S_0^{[b]} = s_0, \dots, S_{n-1}^{[b]} = s_{n-1}, S_n^{[b]} = k), \end{aligned}$$

car la somme indexée par r_{n+1} qui apparaît à la deuxième ligne est égale à $p_{k,l}^{[b]}$. En sommant ensuite sur tous les entiers $s_0, \dots, s_{n-1} \in \mathbb{Z}/b\mathbb{Z}$, on obtient aussi

$$\mathbb{P}(S_n^{[b]} = k \text{ et } S_{n+1}^{[b]} = l) = p_{k,l}^{[b]} \mathbb{P}(S_n^{[b]} = k).$$

Finalement, si la probabilité $\mathbb{P}(S_0^{[b]} = s_0, \dots, S_{n-1}^{[b]} = s_{n-1}, S_n^{[b]} = k)$ est strictement positive, on déduit

$$\mathbb{P}(S_{n+1}^{[b]} = l \mid S_0^{[b]} = s_0, \dots, S_{n-1}^{[b]} = s_{n-1}, S_n^{[b]} = k) = \mathbb{P}(S_{n+1}^{[b]} = l \mid S_n^{[b]} = k) = p_{k,l}^{[b]}.$$

On conclut la preuve en observant que $S_0^{[b]} = 0$, puisque $S_0 = 0$. □

D'après la proposition 1.18, pour tout entier $b \geq 2$, le graphe associé à la chaîne de Markov $(S_n^{[b]})_{n \geq 0}$ admet $0, \dots, b-1$ pour sommets, et fait apparaître un arc entre un entier et son successeur modulo b si $p > 0$, et entre un entier et lui-même si $p < 1$. On représente ce graphe sur la figure 1.13. Tous les états communiquent si $p > 0$. Cette chaîne de Markov est donc irréductible, et son unique classe est finale. Si $p = 0$, toutes les classes, réduites à des singletons, sont finales.

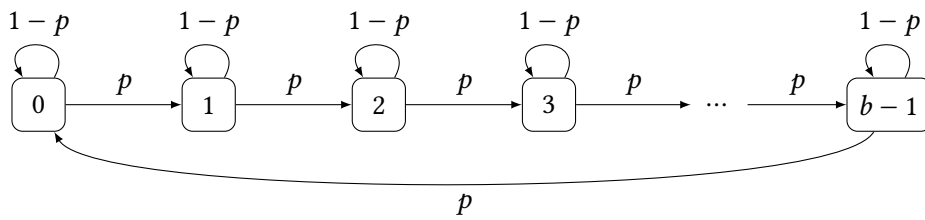


FIGURE 1.13 : Graphe du processus de Bernoulli modulo b vu comme une chaîne de Markov

Considérons ensuite le processus obtenu en ne gardant que les chiffres des unités dans le processus défini par (1.29). Soient K une variable aléatoire à valeurs dans $\mathbb{N}$, et $(S_n)_{n \geq 0}$ un processus de Bernoulli de paramètre p et indépendant de la variable K . Étant donné un entier $b \geq 2$, définissons

$$Y_n^{[b]} = K + S_n \mod b$$

pour tout entier $n \geq 0$. On peut montrer en utilisant la proposition 1.18 qu'il s'agit d'une chaîne de Markov homogène à espace d'états $\mathbb{Z}/b\mathbb{Z}$, de matrice de transition identique à celle de $(S_n^{[b]})_{n \geq 0}$.

Étudions enfin les instants de visite $T_l^{(k)}$ et le nombre total de visites N_k d'un entier $k \in \mathbb{Z}/b\mathbb{Z}$ par la chaîne $(Y_n^{[b]})_{n \geq 0}$, qui sont définis par (B.3) et (B.4). Si $Y_0^{[b]} = k$ avec probabilité non nulle, on peut alors avoir $Y_1^{[b]} = k$ avec probabilité $1 - p$, et le premier instant de retour en k vérifie $T_1^{(k)} = 1$. Ainsi,

$$\mathbb{P}(T_1^{(k)} = 1 \mid Y_0^{[b]} = k) = \mathbb{P}(Y_1^{[b]} = k \mid Y_0^{[b]} = k) = 1 - p,$$

Si $p > 0$, on peut aussi avoir $Y_1^{[b]} = k + 1$ avec probabilité p , et il faut obtenir $b - 1$ succès pour que la chaîne retourne en k , donc, conditionnellement à l'événement $\{Y_0^{[b]} = k, Y_1^{[b]} = k + 1\}$, la variable $T_1^{(k)} - 1$ suit la loi de Pascal $\text{Géo}_{\mathbb{N}^*}(p)^{*(b-1)}$. D'où, pour tout entier $u \geq b$,

$$\begin{aligned} \mathbb{P}(T_1^{(k)} = u \mid Y_0^{[b]} = k) &= \mathbb{P}(T_1^{(k)} - 1 = u - 1 \mid Y_1^{[b]} = k + 1 \text{ et } Y_0^{[b]} = k) \mathbb{P}(Y_1^{[b]} = k + 1 \mid Y_0^{[b]} = k) \\ &= \binom{u-2}{b-2} (1-p)^{u-b} p^{b-1} p = \binom{u-2}{b-2} (1-p)^{u-b} p^b. \end{aligned}$$

La relation donnée par la proposition B.14 entre la probabilité de finitude du premier instant de retour en k et l'espérance du nombre total de visites de k implique que

$$\mathbb{P}(T_1^{(k)} < \infty \mid Y_0^{[b]} = k) = 1 \quad \text{et} \quad \mathbb{E}[N_k \mid Y_0^{[b]} = k] = \infty.$$

En particulier, tous les états sont récurrents. De plus, si $p = 0$, le premier instant de retour $T_1^{(k)}$ vaut 1 presque sûrement, donc admet 1 pour espérance. Si $p > 0$, on peut calculer l'espérance de ce premier instant de retour à partir de sa loi exprimée ci-dessus, ou en utilisant l'espérance de la loi $\text{Géo}_{\mathbb{N}^*}(p)^{*(b-1)}$. En effet, en notant U une variable aléatoire qui suit cette dernière loi, on a

$$\begin{aligned} \mathbb{E}[T_1^{(k)} \mid Y_0^{[b]} = k] &= \mathbb{E}[T_1^{(k)} \mathbf{1}_{\{Y_1^{[b]} = k\}} \mid Y_0^{[b]} = k] + \mathbb{E}[T_1^{(k)} \mathbf{1}_{\{Y_1^{[b]} = k+1\}} \mid Y_0^{[b]} = k] \\ &= \mathbb{P}(Y_1^{[b]} = k \mid Y_0^{[b]} = k) + \mathbb{E}[T_1^{(k)} \mid Y_1^{[b]} = k + 1 \text{ et } Y_0^{[b]} = k] \mathbb{P}(Y_1^{[b]} = k + 1 \mid Y_0^{[b]} = k) \\ &= (1-p) + p \mathbb{E}[1 + U] = 1 + p \mathbb{E}[U] = 1 + p \frac{b-1}{p} = b. \end{aligned}$$

Finalement, l'espérance $\mathbb{E}[T_1^{(k)} \mid Y_0^{[b]} = k]$ est égale à 1 si $p = 0$, et elle est égale à b sinon. Cela signifie que tous les états sont récurrents positifs pour la chaîne. Enfin, si $p > 0$, la chaîne admet pour probabilité invariante la loi uniforme sur $\mathbb{Z}/b\mathbb{Z}$. Si $0 < p < 1$, le corollaire B.32 montre alors que pour tout $k \in \mathbb{Z}/b\mathbb{Z}$,

$$\mathbb{P}(Y_n^{[b]} = k) \xrightarrow{n \rightarrow \infty} \frac{1}{b}.$$

La chaîne est en effet clairement apériodique, puisque la probabilité de rester en un état vaut $1 - p > 0$. En revanche, si $p = 1$, on revient à un état au bout d'un temps multiple de b avec une probabilité 1, et au bout d'un temps non multiple de b avec une probabilité 0. Tous les états ont donc pour période b , de sorte que la chaîne n'est pas apériodique.

Collection de coupons

On s'intéresse au *problème du collectionneur de coupons* suivant. On considère un ensemble fini et non vide dont les éléments sont appelés *coupons*. On tire au hasard avec remise des coupons et on s'intéresse au nombre de tirages nécessaires pour obtenir au moins une fois chaque coupon. Pour modéliser ce problème, notons $r \geq 2$ le nombre total de coupons et considérons une suite $(X_n)_{n \geq 1}$ de variables aléatoires indépendantes qui suivent toutes la *loi uniforme* sur l'ensemble $\{1, \dots, r\}$ représentant l'ensemble des coupons. Pour tous entiers $n \geq 1$ et $c \in \{1, \dots, r\}$, on a $\mathbb{P}(X_n = c) = 1/r$. Chacune des variables aléatoires X_n indique quel coupon a été obtenu au n -ième tirage. Le premier instant où on a obtenu au moins une fois chaque coupon est donné par la variable aléatoire

$$T^{(r)} = \inf\{n \geq 1 \mid \mathcal{C}_n^{(r)} = \{1, \dots, r\}\} = \inf\{n \geq 1 \mid \#\mathcal{C}_n^{(r)} = r\} \quad \text{avec} \quad \mathcal{C}_n^{(r)} = \{X_1, \dots, X_n\}. \quad (2.1)$$

On adopte la convention usuelle que l'infimum de l'ensemble vide vaut ∞ , de sorte que $T^{(r)}$ pourrait *a priori* être infini. On verra cependant dans la proposition 2.3 que $T^{(r)}$ est fini presque sûrement. Nous appelons $T^{(r)}$ le *temps de recouvrement* de la collection. Le but principal de ce chapitre est d'étudier la loi de $T^{(r)}$ et son comportement asymptotique quand $r \rightarrow +\infty$ tend vers l'infini.

Le nom du problème vient de l'idée qu'il modélise le fait de collectionner des coupons pour obtenir une récompense, comme par exemple des vignettes pour compléter un album ou des magnets de départements pour former une carte de France. Dans la pratique, certains coupons sont souvent plus rares que d'autres et on peut alors envisager une variante où ils sont distribués avec des probabilités différentes. Cette variante plus complexe n'est pas étudiée ici.



FIGURE 2.1 : Album et vignettes Panini National Geographic, et magnets des départements Le Gaulois

Si $r = 2$, il y a deux coupons qui ont chacun une probabilité $1/2$ d'apparaître. Cela revient au jeu de pile ou face avec une pièce équilibrée étudié dans le chapitre 1. Le temps de recouvrement $T^{(2)}$ est le nombre de lancers nécessaires pour obtenir pile et face au moins une fois. De même, pour $r = 6$, le temps de recouvrement $T^{(6)}$ s'interprète comme le nombre de lancers d'un dé équilibré à six faces nécessaires pour obtenir chaque face au moins une fois.

2.1 Loi du temps de recouvrement

La loi du temps de recouvrement $T^{(r)}$ défini par (2.1) peut être décrite de manière combinatoire à l'aide des *nombre de Stirling de seconde espèce* dont les valeurs sont données dans le résultat suivant.

Proposition 2.1 (Nombres de Stirling de seconde espèce et nombres de surjections). *Pour tous entiers $1 \leq m \leq n$, le nombre de Stirling de seconde espèce $\left\{ \begin{smallmatrix} n \\ m \end{smallmatrix} \right\}$, défini comme le nombre de façons de partitionner un ensemble de n éléments en m sous-ensembles non vides, vérifie*

$$\left\{ \begin{smallmatrix} n \\ m \end{smallmatrix} \right\} = \frac{1}{m!} \sum_{j=0}^m (-1)^{m-j} \binom{m}{j} j^n = \sum_{j=0}^m \frac{(-1)^{m-j} j^n}{(m-j)! j!}.$$

De plus, le nombre de surjections de $\{1, \dots, n\}$ dans $\{1, \dots, m\}$ est donné par $S_{n,m} = m! \left\{ \begin{smallmatrix} n \\ m \end{smallmatrix} \right\}$.

Démonstration. L'égalité consistant à remplacer le coefficient binomial par des factorielles est claire. Ensuite, l'égalité $S_{n,m} = m! \left\{ \begin{smallmatrix} n \\ m \end{smallmatrix} \right\}$ découle de deux observations. D'une part, pour toute surjection de $\{1, \dots, n\}$ dans $\{1, \dots, m\}$, les images réciproques des éléments de $\{1, \dots, m\}$ forment une partition de $\{1, \dots, n\}$ en m sous-ensembles non vides. D'autre part, partant d'une telle partition, on peut construire $m!$ surjections distinctes en choisissant les images sur chaque sous-ensemble de la partition.

Pour établir la formule donnant $\left\{ \begin{smallmatrix} n \\ m \end{smallmatrix} \right\}$, on calcule $S_{n,m}$ d'une autre façon. Pour tout $i \in \{1, \dots, m\}$, notons A_i l'ensemble des applications de $\{1, \dots, n\}$ dans $\{1, \dots, m\}$ dont i n'appartient pas à l'image. L'ensemble des surjections est le complémentaire de l'union des A_i au sein de l'ensemble des applications de $\{1, \dots, n\}$ dans $\{1, \dots, m\}$. Il y a m^n applications de la sorte. De plus, pour tous indices distincts $i_1, \dots, i_j \in \{1, \dots, m\}$, l'intersection $A_{i_1} \cap \dots \cap A_{i_j}$ contient $(m-j)^n$ éléments, qui sont les applications dont l'image est incluse dans $\{1, \dots, m\} \setminus \{i_1, \dots, i_j\}$. On sait aussi qu'il y a $\binom{m}{j}$ façons de choisir ces indices. Enfin, le *principe d'inclusion-exclusion*, qui se prouve facilement par récurrence sur m , indique que pour tous ensembles finis $A_1, \dots, A_m$, on a

$$\#(A_1 \cup \dots \cup A_m) = \sum_{j=1}^m (-1)^{j-1} \sum_{1 \leq i_1 < \dots < i_j \leq m} \#(A_{i_1} \cap \dots \cap A_{i_j}). \quad (2.2)$$

En appliquant cette formule aux ensembles A_i définis ci-dessus, on en conclut que

$$S_{n,m} = m^n - \#(A_1 \cup \dots \cup A_m) = m^n - \sum_{j=1}^m (-1)^{j-1} \binom{m}{j} (m-j)^n = \sum_{j=0}^m (-1)^j \binom{m}{j} (m-j)^n.$$

On termine la preuve en faisant un changement d'indice dans la somme ci-dessus et en observant que $\binom{m}{j} = \binom{m}{m-j}$, puis en identifiant cette somme avec l'expression obtenue précédemment pour $S_{n,m}$. $\square$

La proposition ci-dessus nous permet maintenant de donner une formule combinatoire pour la loi du temps de recouvrement d'une collection de coupons.

Proposition 2.2 (Description combinatoire de la loi du temps de recouvrement). *Soit $T^{(r)}$ le temps de recouvrement défini par (2.1) d'une collection de r coupons. On a $T^{(r)} \geq r$ et, pour tout entier $n \geq r$,*

$$\mathbb{P}(T^{(r)} = n) = \frac{r!}{r^n} \left\{ \begin{smallmatrix} n-1 \\ r-1 \end{smallmatrix} \right\} = \sum_{j=1}^{r-1} (-1)^{r-1-j} \binom{r-1}{j} \left(\frac{j}{r} \right)^{n-1}.$$

Démonstration. Il est clair que $T^{(r)} \geq r$. De plus, pour tous entiers $n \geq r$, l'événement $\{T^{(r)} = n\}$ est réalisé si et seulement s'il existe un entier $c \in \{1, \dots, r\}$ et une surjection s de $\{1, \dots, n-1\}$ dans $\{1, \dots, r\} \setminus \{c\}$ tels que $X_1 = s(1), \dots, X_{n-1} = s(n-1), X_n = c$. À c et s fixés, cela se produit avec une probabilité égale à r^{-n} . On en déduit que $\mathbb{P}(T^{(r)} = n) = r S_{n-1, r-1} r^{-n}$ et on conclut grâce à la proposition 2.1. $\square$

La proposition 2.2 donne une formule explicite pour la loi du temps de recouvrement $T^{(r)}$ d'une collection de r coupons, mais celle-ci n'est pas très intuitive et n'est pas évidente à manipuler dans les calculs concrets. Pour cela, il est plus commode de faire appel au résultat suivant, qui montre que $T^{(r)}$ s'écrit comme une somme de variables aléatoires géométriques indépendantes.

Proposition 2.3 (Description probabiliste de la loi du temps de recouvrement). *Le temps de recouvrement $T^{(r)}$ défini par (2.1) d'une collection de r coupons peut s'écrire*

$$T^{(r)} = D_1^{(r)} + \dots + D_r^{(r)},$$

où $D_1^{(r)}, \dots, D_r^{(r)}$ sont des variables aléatoires indépendantes telles que, pour tout $l \in \{1, \dots, r\}$, la variable $D_l^{(r)}$ suit la loi $\text{Géo}_{\mathbb{N}^*}(p_{r,l})$ où $p_{r,l} = (r-l+1)/r$. En particulier, $\mathbb{P}(T^{(r)} < \infty) = 1$.

Démonstration. Pour tout entier $l \in \{1, \dots, r\}$, considérons la variable aléatoire

$$T_l^{(r)} = \inf\{n \geq 1 \mid \#\{X_1, \dots, X_n\} = l\},$$

de sorte que $T_1^{(r)} = 1$ et $T_r^{(r)} = T^{(r)}$. Avant de former les différences de ces temps, observons que, pour tout $n \geq 1$, l'inégalité de l'union et l'indépendance des tirages donnent

$$\mathbb{P}(T^{(r)} > n) \leq \sum_{c=1}^r \mathbb{P}(X_1 \neq c, \dots, X_n \neq c) = r \left(1 - \frac{1}{r}\right)^n \xrightarrow{n \rightarrow \infty} 0.$$

Ainsi, les temps $T_l^{(r)} \leq T^{(r)}$ sont tous presque sûrement finis. Leurs différences sont bien définies presque sûrement et peuvent être fixées arbitrairement sur l'événement négligeable restant. La preuve est similaire à celle de la proposition 1.2. En effet, si on interprète comme un succès le fait d'obtenir un coupon différent des précédents, les variables $T_l^{(r)}$ donnent les instants de succès. Cependant, la probabilité de succès n'est pas constante : elle diminue à mesure que l'on obtient des coupons. Fixons des entiers $k_2, \dots, k_r \in \mathbb{N}^*$, puis notons $k'_l = 1 + k_2 + \dots + k_l$. On a les égalités suivantes entre événements :

$$\begin{aligned} \bigcap_{l=2}^r \{T_l^{(r)} - T_{l-1}^{(r)} = k_l\} &= \bigcap_{l=2}^r \{T_l^{(r)} = k'_l\} \\ &= \bigsqcup_{\{c_1, \dots, c_r\} = \{1, \dots, r\}} \{X_1 = c_1; X_2, \dots, X_{k'_2-1} \in \{c_1\}, X_{k'_2} = c_2; X_{k'_2+1}, \dots, X_{k'_3-1} \in \{c_1, c_2\}, X_{k'_3} = c_3; \dots\}, \end{aligned}$$

où l'union disjointe est prise sur les $r!$ permutations de $\{1, \dots, r\}$. On en déduit que

$$\mathbb{P}\left(\bigcap_{l=2}^r \{T_l^{(r)} - T_{l-1}^{(r)} = k_l\}\right) = r! \frac{1}{r} \left(\frac{1}{r}\right)^{k_2-1} \frac{1}{r} \left(\frac{2}{r}\right)^{k_3-1} \frac{1}{r} \dots \left(\frac{r-1}{r}\right)^{k_r-1} \frac{1}{r} = \prod_{l=2}^r ((1 - p_{r,l})^{k_l-1} p_{r,l}).$$

En sommant sur tous les $k_2, \dots, k_r \in \mathbb{N}^*$, on obtient

$$\mathbb{P}\left(\bigcap_{l=2}^r \{T_l^{(r)} - T_{l-1}^{(r)} < \infty\}\right) = \sum_{k_2, \dots, k_r \geq 1} \prod_{l=2}^r ((1 - p_{r,l})^{k_l-1} p_{r,l}) = \prod_{l=2}^r \left(\sum_{k_l=1}^{\infty} (1 - p_{r,l})^{k_l-1} p_{r,l}\right) = 1,$$

ce qui prouve que les variables aléatoires $D_l^{(r)} = T_l^{(r)} - T_{l-1}^{(r)}$, pour $l \in \{2, \dots, r\}$, sont presque sûrement finies, tout comme $D_1^{(r)} = T_1^{(r)} = 1$ qui est même presque sûrement constante. Enfin, en choisissant un entier $l_0 \in \{2, \dots, r\}$ et en sommant sur les entiers $k_l \in \mathbb{N}^*$ pour $l \in \{2, \dots, r\} \setminus \{l_0\}$, on a de plus

$$\mathbb{P}(D_{l_0}^{(r)} = k_{l_0}) = ((1 - p_{r,l_0})^{k_{l_0}-1} p_{r,l_0}) \prod_{l \in \{2, \dots, r\} \setminus \{l_0\}} \left(\sum_{k_l=1}^{\infty} (1 - p_{r,l})^{k_l-1} p_{r,l}\right) = (1 - p_{r,l_0})^{k_{l_0}-1} p_{r,l_0}.$$

Ainsi, les variables aléatoires $D_l^{(r)}$, pour $l \in \{1, \dots, r\}$, sont indépendantes et de lois $\text{Géo}_{\mathbb{N}^*}(p_{r,l})$. Ce qui précède implique de plus que $T^{(r)}$ est presque sûrement fini. $\square$

Les propositions 2.2 et 2.3 conduisent à deux expressions pour la fonction génératrice du temps de recouvrement $T^{(r)}$ d'une collection de r coupons. Pour $|z| < r/(r-1)$, on a d'une part

$$G_{T^{(r)}}(z) = \sum_{n=r}^{\infty} \mathbb{P}(T^{(r)} = n) z^n = \sum_{n=r}^{\infty} \sum_{j=1}^{r-1} (-1)^{r-1-j} \binom{r-1}{j} \left(\frac{j}{r}\right)^{n-1} z^n = \frac{z^r}{r^{r-2}} \sum_{j=1}^{r-1} (-1)^{r-1-j} \binom{r-1}{j} \frac{j^{r-1}}{r-jz}.$$

D'autre part, comme dans (1.17), on peut utiliser (1.8) et la proposition A.5 pour écrire que

$$G_{T^{(r)}}(z) = G_{D_1^{(r)}}(z) \cdots G_{D_r^{(r)}}(z) = \prod_{l=1}^r \frac{p_{r,l} z}{1 - (1 - p_{r,l})z} = \prod_{l=1}^r \frac{(r-l+1)z}{r - (l-1)z} = \frac{r! z^r}{\prod_{l=0}^{r-1} (r-lz)}.$$

En identifiant les deux expressions ci-dessus pour la fonction génératrice, on obtient la décomposition en éléments simples suivante :

$$\frac{r^{r-3} r!}{\prod_{l=1}^{r-1} (r-lz)} = \sum_{j=1}^{r-1} (-1)^{r-1-j} \binom{r-1}{j} \frac{j^{r-1}}{r-jz}.$$

La proposition 2.3 permet aussi de calculer l'espérance et la variance de $T^{(r)}$. L'estimation de l'espérance fait apparaître la constante d'Euler-Mascheroni $\gamma \approx 0.577$, définie comme suit.

Définition 2.1 (Constante d'Euler-Mascheroni). La limite $\gamma = \lim_{r \rightarrow \infty} (-\ln r + \sum_{l=1}^r 1/l)$ existe, est finie et s'appelle *constante d'Euler-Mascheroni*.



FIGURE 2.2 : Lorenzo Mascheroni (1750–1800) et Leonhard Euler (1707–1783)

Démonstration. La suite de terme général $u_r = -\ln r + \sum_{l=1}^r 1/l$ est décroissante car pour tout $r \geq 2$,

$$u_r - u_{r-1} = \left(-\ln r + \sum_{l=1}^r \frac{1}{l} \right) - \left(-\ln(r-1) + \sum_{l=1}^{r-1} \frac{1}{l} \right) = \ln\left(1 - \frac{1}{r}\right) + \frac{1}{r} \leq 0,$$

puisque $\ln(1+x) \leq x$ pour tout réel $x > -1$. De plus, la décroissance de $x \mapsto 1/x$ conduit à

$$u_r = -\ln r + \sum_{l=1}^r \frac{1}{l} \geq -\ln r + \sum_{l=1}^r \int_l^{l+1} \frac{dx}{x} = -\ln r + \int_1^{r+1} \frac{dx}{x} = \ln(r+1) - \ln r \geq 0,$$

de sorte que la suite de terme général u_r est minorée. Ainsi, elle est convergente. $\square$

Indiquons maintenant les valeurs de l'espérance et de la variance du temps de recouvrement.

Corollaire 2.4 (Espérance et variance du temps de recouvrement). En désignant par γ la constante d'Euler-Mascheroni, on a les égalités suivantes pour l'espérance et la variance du temps de recouvrement $T^{(r)}$ défini par (2.1) d'une collection de r coupons :

$$\mathbb{E}[T^{(r)}] = r \sum_{l=1}^r \frac{1}{l} = r(\ln r + \gamma) + o_{r \rightarrow \infty}(r) \quad \text{et} \quad \text{Var}(T^{(r)}) = r \sum_{l=1}^r \frac{r-l}{l^2} = \frac{\pi^2}{6} r^2 + o_{r \rightarrow \infty}(r^2).$$

Démonstration. D'après la proposition 2.3, on a $T^{(r)} = D_1^{(r)} + \dots + D_r^{(r)}$, où les variables aléatoires $D_l^{(r)}$ sont indépendantes et de loi $\text{Géo}_{\mathbb{N}^*}(p_{r,l})$ avec $p_{r,l} = (r - l + 1)/r$. En utilisant les formules (1.10) pour l'espérance et la variance d'une variable aléatoire géométrique, on a d'une part

$$\mathbb{E}[T^{(r)}] = \sum_{l=1}^r \mathbb{E}[D_l^{(r)}] = \sum_{l=1}^r \frac{1}{p_{r,l}} = \sum_{l=1}^r \frac{r}{r-l+1} = r \sum_{l=1}^r \frac{1}{l} = r(\ln r + \gamma) + o_{r \rightarrow \infty}(r) = o_{r \rightarrow \infty}(r^2),$$

et d'autre part, en utilisant aussi l'indépendance des variables aléatoires constituant la somme,

$$\text{Var}(T^{(r)}) = \sum_{l=1}^r \text{Var}(D_l^{(r)}) = \sum_{l=1}^r \frac{1-p_{r,l}}{p_{r,l}^2} = r \sum_{l=1}^r \frac{l-1}{(r-l+1)^2} = r^2 \sum_{l=1}^r \frac{1}{l^2} - r \sum_{l=1}^r \frac{1}{l} = \frac{\pi^2}{6} r^2 + o_{r \rightarrow \infty}(r^2),$$

ce qui achève la preuve. $\square$

Le résultat suivant exprime la queue de distribution du temps de recouvrement.

Proposition 2.5 (Queue de distribution du temps de recouvrement). *La queue de distribution du temps de recouvrement $T^{(r)}$ défini par (2.1) d'une collection de r coupons est donnée, pour tout entier $n \geq 1$, par la formule*

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

Démonstration. Pour tout $n \geq 1$, l'événement $\{T^{(r)} > n\}$ est l'union pour $i \in \{1, \dots, r\}$ des événements $E_i = \{X_1 \neq i, \dots, X_n \neq i\}$. On peut donc déterminer sa probabilité à l'aide du principe d'inclusion-exclusion déjà utilisé pour écrire (2.2). En effet, on peut prouver par récurrence sur $m \in \mathbb{N}^*$ que pour tous événements $A_1, \dots, A_m$, on a

$$\mathbb{P}(A_1 \cup \dots \cup A_m) = \sum_{j=1}^m (-1)^{j-1} \sum_{1 \leq i_1 < \dots < i_j \leq m} \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_j}).$$

Pour appliquer cette formule aux événements E_i , on calcule, pour tous $1 \leq i_1 < \dots < i_j \leq r$,

$$\mathbb{P}(E_{i_1} \cap \dots \cap E_{i_j}) = \mathbb{P}(X_1, \dots, X_n \notin \{i_1, \dots, i_j\}) = \left(1 - \frac{j}{r}\right)^n.$$

On conclut en observant qu'il y a $\binom{r}{j}$ façons de choisir les indices $i_1, \dots, i_j$ ci-dessus. Pour $n \geq r - 1$, on peut aussi prouver le résultat en sommant la formule de la proposition 2.2 à partir du rang $n + 1$. On peut remarquer à l'issue de cette preuve que, comme $T^{(r)} \geq r$ presque sûrement, la somme dans le membre de droite de l'égalité prouvée vaut toujours 1 pour $n \in \{1, \dots, r - 1\}$. $\square$

D'après le corollaire 2.4, l'espérance du temps de recouvrement $T^{(r)}$ se comporte asymptotiquement comme $r \ln r + \gamma r$. Le résultat suivant majore la probabilité avec laquelle la variable $T^{(r)}$ peut dévier de cet ordre de grandeur. On désigne par $\{x\} = x - \lfloor x \rfloor$ la partie fractionnaire d'un réel x .

Corollaire 2.6 (Déviation du temps de recouvrement). *Soit $t > 0$. Le temps de recouvrement $T^{(r)}$ défini par (2.1) d'une collection de r coupons vérifie*

$$\mathbb{P}(T^{(r)} > r \ln r + tr) \leq \exp\left(-t + \frac{\{r \ln r + tr\}}{r}\right).$$

Démonstration. Pour tout entier $n \geq 1$, avec les notations de la proposition 2.5, on a la majoration

$$\mathbb{P}(T^{(r)} > n) \leq \sum_{i=1}^r \mathbb{P}(E_i) = r \left(1 - \frac{1}{r}\right)^n \leq r e^{-n/r},$$

car on a l'encadrement $0 \leq 1 - x \leq e^{-x}$ quand $x \leq 1$. On conclut en prenant $n = \lfloor r \ln r + tr \rfloor \geq 1$. $\square$

Le nombre r de coupons et un seuil $\alpha \in]0, 1[$ étant fixés, le majorant du corollaire 2.6 est au plus α si $t = -\ln \alpha + 1/r$. Ainsi, $T^{(r)} \in [r, r \ln r - r \ln \alpha + 1]$ avec une probabilité au moins $1 - \alpha$. On choisit usuellement α petit et on obtient un intervalle auquel appartient $T^{(r)}$ avec une probabilité $1 - \alpha \approx 1$.

2.2 Comportement asymptotique

On commence l'étude du comportement asymptotique pour $r \rightarrow \infty$ du temps de recouvrement d'une collection de r coupons par le résultat de convergence en probabilité suivant.

Proposition 2.7 (Équivalent en probabilité du temps de recouvrement). *Le temps de recouvrement $T^{(r)}$ défini par (2.1) d'une collection de r coupons vérifie*

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

Démonstration. Pour tout $\varepsilon > 0$, l'inégalité de Markov donne

$$\mathbb{P}\left(\left|\frac{T^{(r)}}{r \ln r} - 1\right| \geq \varepsilon\right) = \mathbb{P}\left((T^{(r)} - r \ln r)^2 \geq \varepsilon^2 r^2 (\ln r)^2\right) \leq \frac{\mathbb{E}[(T^{(r)} - r \ln r)^2]}{\varepsilon^2 r^2 (\ln r)^2} \underset{r \rightarrow \infty}{\sim} \left(\frac{\pi^2}{6} + \gamma^2\right) \frac{1}{\varepsilon^2 (\ln r)^2}. \quad (2.3)$$

Pour établir l'équivalent annoncé ci-dessus, on réécrit le terme dans l'espérance sous la forme

$$\begin{aligned} (T^{(r)} - r \ln r)^2 &= ((T^{(r)} - \mathbb{E}[T^{(r)}]) + (\mathbb{E}[T^{(r)}] - r \ln r))^2 \\ &= (T^{(r)} - \mathbb{E}[T^{(r)}])^2 + (\mathbb{E}[T^{(r)}] - r \ln r)^2 + 2(T^{(r)} - \mathbb{E}[T^{(r)}])(\mathbb{E}[T^{(r)}] - r \ln r). \end{aligned}$$

D'après le corollaire 2.4, l'espérance vérifie donc

$$\mathbb{E}[(T^{(r)} - r \ln r)^2] = \text{Var}(T^{(r)}) + (\mathbb{E}[T^{(r)}] - r \ln r)^2 = \left(\frac{\pi^2}{6} r^2 + o(r^2)\right) + \left(\gamma r + o(r)\right)^2 \underset{r \rightarrow \infty}{\sim} \left(\frac{\pi^2}{6} + \gamma^2\right) r^2.$$

Cela confirme l'équivalent dans (2.3), et la probabilité dans (2.3) tend vers 0 quand $r \rightarrow \infty$. $\square$

Algorithme 2.1 (Illustration de l'équivalent en probabilité du temps de recouvrement). On illustre la proposition 2.7 en simulant une collecte de r coupons. On compare le temps de recouvrement $T^{(r)}$ obtenu à son équivalent en probabilité $r \ln r$ en affichant une valeur approchée du rapport $T^{(r)}/(r \ln r)$. Le script C.5 permet de réaliser cette simulation. Les graphes produits pour diverses valeurs de r sont représentés sur la figure 2.3.

Pour compléter la proposition 2.7, on peut étudier la vitesse de convergence vers 1 du quotient $T^{(r)}/(r \ln r)$. On obtient le résultat de convergence en loi suivant.

Proposition 2.8 (Fluctuations en loi du temps de recouvrement). *Les fluctuations renormalisées du temps de recouvrement $T^{(r)}$ défini par (2.1) d'une collection de r coupons suivent asymptotiquement la loi de Gumbel, c'est-à-dire que pour tout $t \in \mathbb{R}$, on a*

$$\mathbb{P}\left(\left(\frac{T^{(r)}}{r \ln r} - 1\right) \ln r \leq t\right) \xrightarrow[r \rightarrow \infty]{} \exp(-e^{-t}).$$

Démonstration. Posons $n_{t,r} = \lfloor r \ln r + tr \rfloor$. Comme $T^{(r)}$ est à valeurs entières, pour r assez grand la proposition 2.5 donne

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

Pour $1 \leq j < r$, les inégalités $1 - x \leq e^{-x}$ et $n_{t,r} \geq r \ln r + tr - 1$ entraînent

$$0 \leq u_j^{(r)} \leq \frac{r^j}{j!} e^{-jn_{t,r}/r} \leq \frac{e^{-tj+j/r}}{j!} \leq \frac{e^{(1-t)j}}{j!} =: v_j.$$

Pour chaque entier $j \geq 1$ fixé, on a $\binom{r}{j} \sim r^j/j!$ et

$$n_{t,r} \ln\left(1 - \frac{j}{r}\right) = -j \ln r - tj + o(1),$$

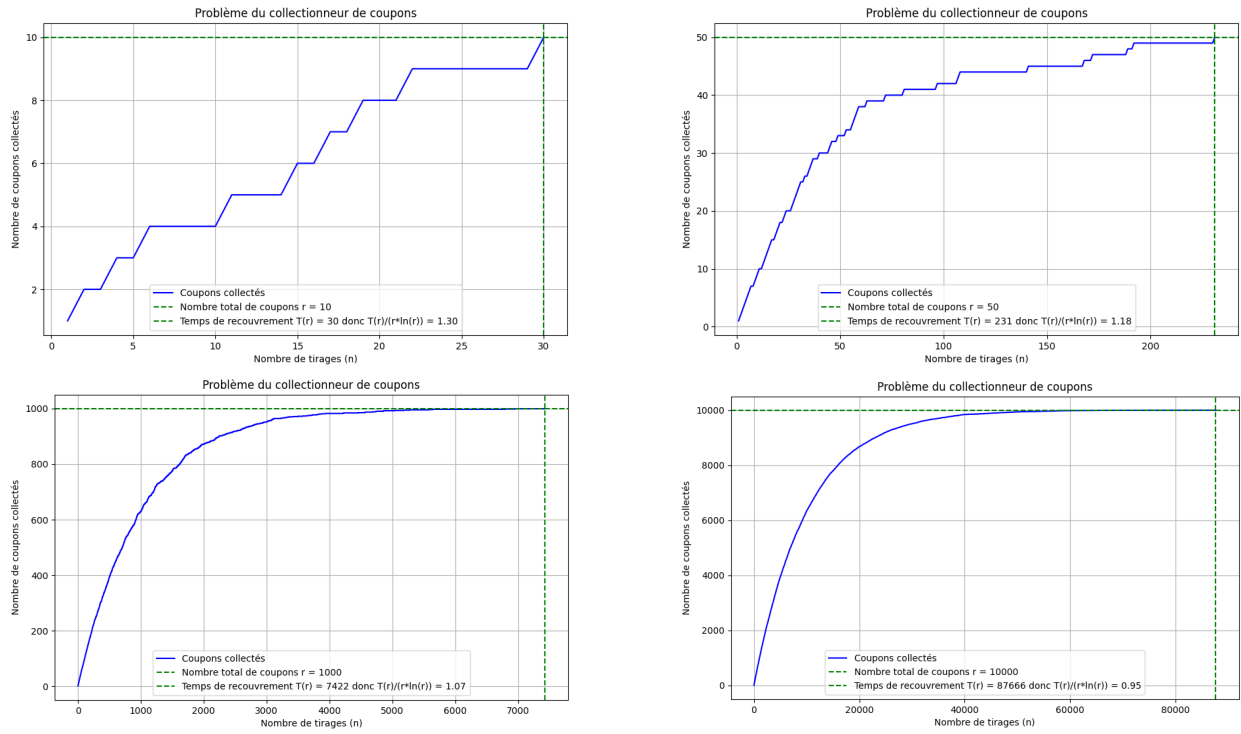


FIGURE 2.3 : Simulations de collectes de coupons

donc $u_j^{(r)} \rightarrow u_j := e^{-tj}/j!$. La série $\sum_{j \geq 1} v_j$ converge. Prolongeons $u_j^{(r)}$ par 0 pour $j \geq r$. Pour tout $J \geq 1$, on a alors

$$\left| \sum_{j=1}^{\infty} (-1)^{j-1} u_j^{(r)} - \sum_{j=1}^{\infty} (-1)^{j-1} u_j \right| \leq \sum_{j=1}^{J-1} |u_j^{(r)} - u_j| + 2 \sum_{j=J}^{\infty} v_j.$$

On fait d'abord tendre r vers l'infini, puis J vers l'infini. Il vient

$$\mathbb{P}(T^{(r)} > r \ln r + tr) \xrightarrow{r \rightarrow \infty} \sum_{j=1}^{\infty} (-1)^{j-1} \frac{e^{-tj}}{j!} = 1 - \exp(-e^{-t}),$$

ce qui prouve le résultat annoncé. □

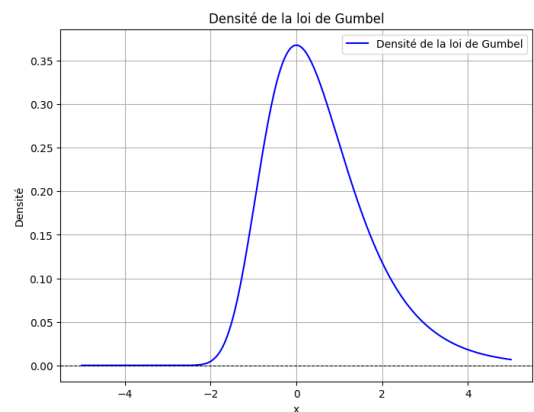
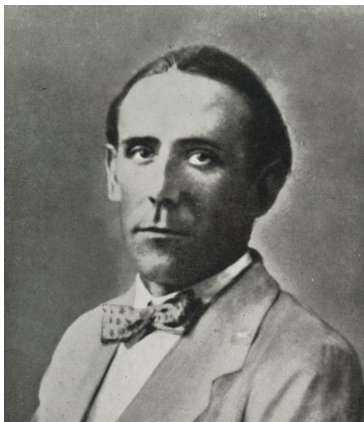


FIGURE 2.4 : Emil Julius Gumbel (1891–1966) et densité de la loi de Gumbel, tracée à l'aide du script C.6

La proposition 2.8 fait apparaître à la limite la fonction de répartition de la loi de Gumbel, c'est-à-dire la fonction $t \mapsto \exp(-e^{-t})$. En dérivant cette fonction, on obtient la densité de la loi de Gumbel. Celle-ci est donnée par $x \mapsto \exp(-x - e^{-x})$, et se voit représentée sur la figure 2.4.

Algorithme 2.2 (Illustration des fluctuations en loi du temps de recouvrement). On réalise m fois de façon indépendante l'expérience de collecte de r coupons, et on note $T_1^{(r)}, \dots, T_m^{(r)}$ les temps de recouvrement obtenus. La fonction de répartition empirique des fluctuations renormalisées est alors donnée par

$$t \mapsto \frac{1}{m} \# \left\{ j \in \{1, \dots, m\} \mid \left(\frac{T_j^{(r)}}{r \ln r} - 1 \right) \ln r \leq t \right\}.$$

D'après la proposition 2.8 et la loi des grands nombres (voir la section 1.3.1), on peut comparer cette fonction de répartition empirique à celle de la loi de Gumbel. Le script C.7 permet de réaliser cette simulation. Les graphes produits pour diverses valeurs de r et pour $m = 100$ répétitions indépendantes de la collecte sont représentés sur la figure 2.5.

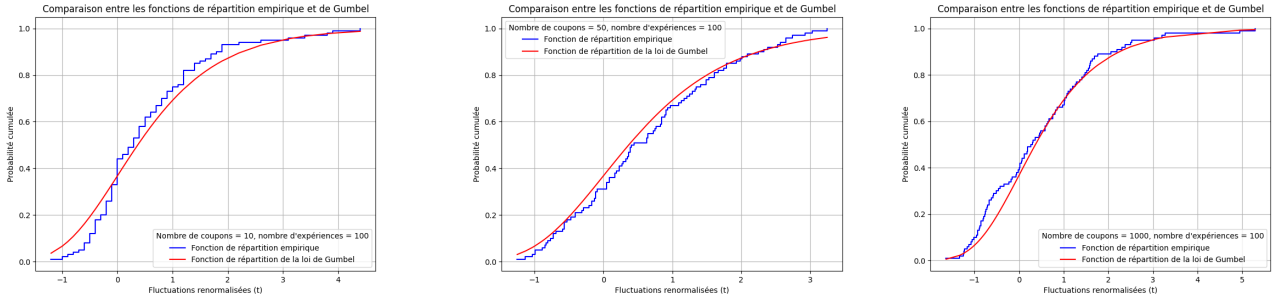


FIGURE 2.5 : Fonctions de répartition empirique des fluctuations renormalisées et de la loi de Gumbel

Les résultats précédents impliquent que la queue de distribution du temps de recouvrement $T^{(r)}$ d'une collection d'un très grand nombre r de coupons présente un phénomène de seuil autour de la valeur $r \ln r$. La queue de distribution passe en effet rapidement d'une valeur proche de 1 à une valeur proche de 0, comme l'indique plus précisément le résultat suivant.

Proposition 2.9 (Phénomène de seuil pour le temps de recouvrement). Soit $(c_r)_{r \geq 2}$ une suite divergeant vers l'infini. Le temps de recouvrement $T^{(r)}$ défini par (2.1) d'une collection de r coupons vérifie

$$\lim_{r \rightarrow \infty} \mathbb{P}(T^{(r)} > r \ln r - c_r) = 1 \quad \text{et} \quad \lim_{r \rightarrow \infty} \mathbb{P}(T^{(r)} > r \ln r + c_r) = 0.$$

Démonstration. Considérons un réel $\varepsilon \in]0, 1[$, et posons $t_\varepsilon = \ln \ln(1/\varepsilon)$. Il existe un entier $r_0 \geq 2$ tel que $c_r \geq t_\varepsilon$ pour tout entier $r \geq r_0$. D'après la proposition 2.8, on a alors

$$\mathbb{P}(T^{(r)} \leq r \ln r - c_r) = \mathbb{P}\left(\left(\frac{T^{(r)}}{r \ln r} - 1\right) \ln r \leq -c_r\right) \leq \mathbb{P}\left(\left(\frac{T^{(r)}}{r \ln r} - 1\right) \ln r \leq -t_\varepsilon\right) \xrightarrow{r \rightarrow \infty} \exp(-e^{t_\varepsilon}) = \varepsilon.$$

On en déduit que le membre de gauche est au plus 2ε pour r assez grand, ce qui prouve la première limite. La seconde limite découle du corollaire 2.6, appliqué avec $t = c_r > 0$. On a en effet

$$\mathbb{P}(T^{(r)} > r \ln r + c_r) \leq \exp\left(-c_r + \frac{\{r \ln r + c_r\}}{r}\right) \leq \exp\left(-c_r + \frac{1}{r}\right) \xrightarrow{r \rightarrow \infty} 0,$$

ce qui achève la preuve de la proposition. □

2.3 Aspect markovien

Les ensembles de coupons collectés au cours des tirages parmi un total de r coupons possibles constituent une chaîne de Markov dont les états sont les parties de $\{1, \dots, r\}$.

Proposition 2.10 (Aspect markovien du processus des ensembles de coupons collectés). *Pour tout entier $r \geq 2$, notons $\mathcal{S}_r$ l'ensemble des parties de $\{1, \dots, r\}$, et considérons la suite des ensembles $\mathcal{E}_n^{(r)}$, définis par (2.1), contenant les coupons collectés au bout de n tirages parmi un total de $r \geq 2$ coupons possibles. Le processus $(\mathcal{E}_n^{(r)})_{n \geq 0}$ est une chaîne de Markov homogène à espace d'états $\mathcal{S}_r$, dont la loi initiale est telle que $\mathcal{E}_0^{(r)} = \emptyset$ presque sûrement, et les coefficients de la matrice de transition sont donnés, pour tous $s, t \in \mathcal{S}_r$, par la formule*

$$p_{s,t} = \begin{cases} \#s/r & \text{si } t = s \\ 1/r & \text{si } t = s \sqcup \{c\} \text{ avec } c \in \{1, \dots, r\} \setminus s \\ 0 & \text{sinon.} \end{cases}$$

Démonstration. La preuve est assez proche de celle de la proposition 1.17. Notons $(X_n)_{n \geq 1}$ une suite de variables indépendantes et identiquement distribuées selon la loi uniforme sur l'ensemble $\{1, \dots, r\}$. Cette suite conduit via la formule (2.1) aux ensembles $\mathcal{E}_n^{(r)}$ de coupons collectés. Pour tout $n \in \mathbb{N}$, on a $\mathcal{E}_{n+1}^{(r)} = \mathcal{E}_n^{(r)} \cup \{X_{n+1}\}$, et la variable aléatoire X_{n+1} est indépendante des variables aléatoires $X_1, \dots, X_n$, qui déterminent les ensembles $\mathcal{E}_0^{(r)}, \dots, \mathcal{E}_n^{(r)}$. Ainsi, pour tous $s_0, \dots, s_{n-1}, s, t \in \mathcal{S}_r$, on a

$$\begin{aligned} & \mathbb{P}(\mathcal{E}_0^{(r)} = s_0, \dots, \mathcal{E}_{n-1}^{(r)} = s_{n-1}, \mathcal{E}_n^{(r)} = s, \mathcal{E}_{n+1}^{(r)} = t) \\ &= \mathbb{P}(\mathcal{E}_0^{(r)} = s_0, \dots, \mathcal{E}_{n-1}^{(r)} = s_{n-1}, \mathcal{E}_n^{(r)} = s, s \cup \{X_{n+1}\} = t) \\ &= \mathbb{P}(\mathcal{E}_0^{(r)} = s_0, \dots, \mathcal{E}_{n-1}^{(r)} = s_{n-1}, \mathcal{E}_n^{(r)} = s) \mathbb{P}(s \cup \{X_{n+1}\} = t). \end{aligned}$$

Si $t = s$, on a $s \cup \{X_{n+1}\} = t$ si et seulement si $X_{n+1} \in s$, ce qui se produit avec probabilité $\#s/r$. Si $t = s \sqcup \{c\}$ pour un certain $c \in \{1, \dots, r\} \setminus s$, on a $s \cup \{X_{n+1}\} = t$ si et seulement si $X_{n+1} = c$, ce qui se produit avec probabilité $1/r$. Enfin, si t n'est ni égal à s , ni de la forme $s \sqcup \{c\}$, on ne peut pas avoir $s \cup \{X_{n+1}\} = t$. Finalement, la probabilité de l'événement $\{s \cup \{X_{n+1}\} = t\}$ est égale à $p_{s,t}$, donc

$$\mathbb{P}(\mathcal{E}_0^{(r)} = s_0, \dots, \mathcal{E}_{n-1}^{(r)} = s_{n-1}, \mathcal{E}_n^{(r)} = s, \mathcal{E}_{n+1}^{(r)} = t) = \mathbb{P}(\mathcal{E}_0^{(r)} = s_0, \dots, \mathcal{E}_{n-1}^{(r)} = s_{n-1}, \mathcal{E}_n^{(r)} = s) p_{s,t}.$$

En sommant sur tous les ensembles $s_0, \dots, s_{n-1} \in \mathcal{S}_r$, on obtient aussi

$$\mathbb{P}(\mathcal{E}_n^{(r)} = s \text{ et } \mathcal{E}_{n+1}^{(r)} = t) = \mathbb{P}(\mathcal{E}_n^{(r)} = s) p_{s,t}.$$

On en déduit le résultat annoncé. □

On peut déduire du résultat ci-dessus que la suite donnée par le nombre de coupons collectés est une chaîne de Markov dont l'espace d'états est l'ensemble $\{0, \dots, r\}$.

Proposition 2.11 (Aspect markovien du processus des nombres de coupons collectés). *Pour tout entier $r \geq 2$, considérons la suite des ensembles $\mathcal{E}_n^{(r)}$, définis par (2.1), contenant les coupons collectés au bout de n tirages parmi un total de $r \geq 2$ coupons possibles. Le processus $(\#\mathcal{E}_n^{(r)})_{n \geq 0}$ est une chaîne de Markov homogène à espace d'états $\{0, \dots, r\}$, dont la loi initiale est telle que $\#\mathcal{E}_0^{(r)} = 0$ presque sûrement, et la matrice de transition est*

$$p^{[r]} = \begin{pmatrix} 0 & 1 & & & & \\ & \frac{1}{r} & \frac{r-1}{r} & & & \\ & & \frac{2}{r} & \frac{r-2}{r} & & \\ & & & \ddots & \ddots & \\ & & & & \frac{r-1}{r} & \frac{1}{r} \\ & & & & & 1 \end{pmatrix}.$$

Démonstration. Notons $\mathcal{S}_r$ l'ensemble des parties de $\{1, \dots, r\}$ et posons $Y_n = \#\mathcal{E}_n^{(r)}$ pour tout $n \in \mathbb{N}$. En notant $p_{s,t}$ les coefficients de la proposition 2.10, on a pour tous $y_0, \dots, y_{n-1}, k, l \in \{0, \dots, r\}$,

$$\mathbb{P}(Y_0 = y_0, \dots, Y_{n-1} = y_{n-1}, Y_n = k, Y_{n+1} = l)$$

$$\begin{aligned}
&= \sum_{\substack{s_0, \dots, s_{n-1}, s, t \in \mathcal{S}_r \\ \#s_0=y_0, \dots, \#s_{n-1}=y_{n-1}, \#s=k, \#t=l}} \mathbb{P}(\mathcal{C}_0^{(r)} = s_0, \dots, \mathcal{C}_{n-1}^{(r)} = s_{n-1}, \mathcal{C}_n^{(r)} = s, \mathcal{C}_{n+1}^{(r)} = t) \\
&= \sum_{\substack{s_0, \dots, s_{n-1}, s \in \mathcal{S}_r \\ \#s_0=y_0, \dots, \#s_{n-1}=y_{n-1}, \#s=k}} \mathbb{P}(\mathcal{C}_0^{(r)} = s_0, \dots, \mathcal{C}_{n-1}^{(r)} = s_{n-1}, \mathcal{C}_n^{(r)} = s) \sum_{\substack{t \in \mathcal{S}_r \\ \#t=l}} p_{s,t}.
\end{aligned}$$

Cependant, pour tout $s \in \mathcal{S}_r$ tel que $\#s = k$, la somme des coefficients $p_{s,t}$ sur les ensembles $t \in \mathcal{S}_r$ tels que $\#t = l$ est égale à k/r si $l = k$, égale à $(r - k)/r$ si $l = k + 1$, et nulle sinon. On en déduit que

$$\mathbb{P}(Y_0 = y_0, \dots, Y_{n-1} = y_{n-1}, Y_n = k, Y_{n+1} = l) = \mathbb{P}(Y_0 = y_0, \dots, Y_{n-1} = y_{n-1}, Y_n = k) \cdot \begin{cases} \frac{k}{r} & \text{si } l = k \\ \frac{r-k}{r} & \text{si } l = k + 1 \\ 0 & \text{sinon,} \end{cases}$$

ce qui achève la preuve de la proposition. $\square$

D'après la proposition 2.11, le graphe associé à la chaîne de Markov homogène $(\#\mathcal{C}_n^{(r)})_{n \geq 0}$ des nombres de coupons collectés admet $\{0, \dots, r\}$ pour ensemble de sommets, et fait apparaître un arc entre un sommet $k \geq 1$ et lui-même, et entre les sommets $k < r$ et $k + 1$. On représente ce graphe sur la figure 2.6. Tous les états $l \geq 1$ sont accessibles à partir d'un état $k \leq l$, mais pas accessibles à partir d'un état $k > l$. Ainsi, les classes d'équivalence sont des singletons, et la chaîne n'est pas irréductible. De plus, tous les états $k < r$ sont transients, et forment des classes transientes. L'état r est absorbant, donc récurrent, et forme une classe finale.

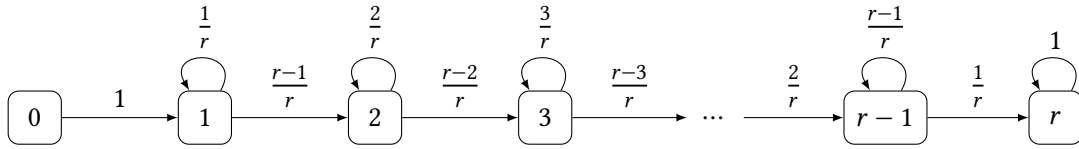


FIGURE 2.6 : Graphe de la chaîne de Markov des nombres de coupons collectés

De plus, d'après sa définition (2.1), le temps de recouvrement $T^{(r)}$ peut s'interpréter comme le premier instant de visite de l'état absorbant r par la chaîne $(\#\mathcal{C}_n^{(r)})_{n \geq 0}$. Pour exploiter cette observation, considérons une chaîne de Markov $(Y_n)_{n \geq 0}$ à espace d'états $\{0, \dots, r\}$, de matrice de transition la matrice $P^{[r]}$ donnée dans la proposition 2.11, et de loi initiale telle que $\mathbb{P}(Y_0 = k) > 0$ pour tout $k \in \{0, \dots, r\}$. Ainsi, conditionnellement à $Y_0 = 0$, la chaîne de Markov $(Y_n)_{n \geq 0}$ a les mêmes lois en dimension finie que la chaîne de Markov $(\#\mathcal{C}_n^{(r)})_{n \geq 0}$, d'après la proposition B.6. En définissant par (B.3) le premier instant de visite $T_0^{(r)}$ de l'état r par la chaîne $(Y_n)_{n \geq 0}$, on obtient que l'espérance du temps de recouvrement $T^{(r)}$ vérifie

$$\mathbb{E}[T^{(r)}] = \theta_0 \quad \text{avec} \quad \theta_k = \mathbb{E}[T_0^{(r)} \mid Y_0 = k].$$

Comme l'état r est absorbant, on déduit de (1.7) que pour tout $k \in \{0, \dots, r\}$, on a

$$\begin{aligned}
\theta_k &= \sum_{n=1}^{\infty} \mathbb{P}(T_0^{(r)} \geq n \mid Y_0 = k) = \sum_{n=0}^{\infty} \mathbb{P}(T_0^{(r)} \geq n+1 \mid Y_0 = k) = \sum_{n=0}^{\infty} \mathbb{P}(Y_n \neq r \mid Y_0 = k) \\
&= \sum_{n=0}^{\infty} \sum_{l=0}^{r-1} \mathbb{P}(Y_n = l \mid Y_0 = k) = \sum_{l=0}^{r-1} \sum_{n=0}^{\infty} \mathbb{P}(Y_n = l \mid Y_0 = k) = \sum_{l=0}^{r-1} p_{k,l}^{[r]*}.
\end{aligned} \tag{2.4}$$

Dans la dernière somme ci-dessus, $p_{k,l}^{[r]*}$ désigne le coefficient de la matrice potentielle $P^{[r]*}$ associée à la matrice $P^{[r]}$. Ce coefficient apparaît en conséquence du théorème B.7. Cependant, la définition B.4 montre que $P^{[r]*} = I + P^{[r]}P^{[r]*}$, où I est la matrice identité. On en déduit que

$$\theta_k = \sum_{l=0}^{r-1} \left(\mathbf{1}_{\{k=l\}} + \sum_{m=0}^r p_{k,m}^{[r]} p_{m,l}^{[r]*} \right) = \mathbf{1}_{\{k < r\}} + \sum_{m=0}^r p_{k,m}^{[r]} \sum_{l=0}^{r-1} p_{m,l}^{[r]*} = \mathbf{1}_{\{k < r\}} + \sum_{m=0}^r p_{k,m}^{[r]} \theta_m. \tag{2.5}$$

Avant de résoudre ces relations, justifions que les θ_k sont finis. Les états $l < r$ sont transients; la proposition B.19 et le corollaire B.18 donnent donc $p_{k,l}^{[r]*} \leq p_{l,l}^{[r]*} < \infty$. La somme finie dans (2.4) montre alors que $\theta_k < \infty$ pour tout $k \in \{0, \dots, r\}$. On peut ainsi soustraire les termes des relations précédentes.

En remplaçant les coefficients de la matrice $P^{[r]}$ par leur expression et en observant que $\theta_r = 0$, on obtient, pour tout $k \in \{0, \dots, r-1\}$,

$$\theta_k = 1 + \frac{k}{r}\theta_k + \frac{r-k}{r}\theta_{k+1} \iff \theta_k - \theta_{k+1} = \frac{r}{r-k}.$$

Comme $k < r$, le dénominateur est strictement positif, y compris lorsque $r = 2$. En sommant ces relations et en utilisant $\theta_r = 0$, on retrouve l'espérance du temps de recouvrement obtenue dans le corollaire 2.4 :

$$\mathbb{E}[T^{(r)}] = \theta_0 = \sum_{k=0}^{r-1} (\theta_k - \theta_{k+1}) = \sum_{k=0}^{r-1} \frac{r}{r-k} = r \sum_{l=1}^r \frac{1}{l}.$$

Enfin, d'après la proposition 2.10, le graphe associé à la chaîne de Markov homogène $(\mathcal{C}_n^{(r)})_{n \geq 0}$ admet l'ensemble des parties de $\{1, \dots, r\}$ pour ensemble de sommets, et fait apparaître un arc entre chaque sommet $s \neq \emptyset$ et lui-même, et entre chaque sommet $s \neq \{1, \dots, r\}$ et un sommet $s \sqcup \{c\}$. Tous les états $s \neq \emptyset$ sont accessibles à partir d'un état $t \subseteq s$, mais pas accessibles à partir d'un état $t \supset s$. Par conséquent, les classes d'équivalence sont des singletons, et la chaîne n'est pas irréductible. De plus, tous les états $s \neq \{1, \dots, r\}$ sont transients, et forment des classes transientes. L'état $\{1, \dots, r\}$ est absorbant, donc récurrent, et forme une classe finale. D'après (2.1), le temps de recouvrement $T^{(r)}$ peut s'interpréter comme le premier instant de visite de cet état absorbant. Le graphe associé de la chaîne de Markov homogène $(\mathcal{C}_n^{(r)})_{n \geq 0}$ est représenté sur les figures 2.7, 2.8 et 2.9, respectivement pour $r = 2, 3$ et 4. Sur ces figures, les sommets correspondant à des ensembles de même cardinal sont regroupés en blocs qui constituent les sommets du graphe de la chaîne de Markov homogène $(\#\mathcal{C}_n^{(r)})_{n \geq 0}$ des nombres de coupons collectés.

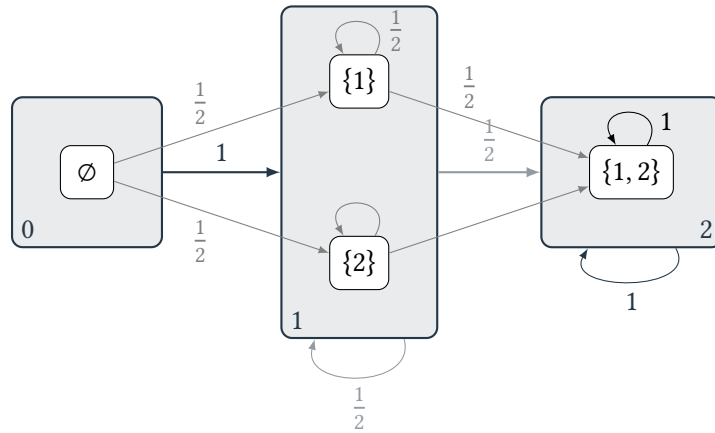


FIGURE 2.7 : Graphes des chaînes de Markov des ensembles et nombres de coupons collectés, $r = 2$

Le graphe devient exponentiellement plus complexe à mesure que le nombre de coupons r augmente, ce qui rend son analyse et sa visualisation de plus en plus difficiles. En effet, le nombre de sommets est 2^r , car chaque sommet représente un sous-ensemble de l'ensemble $\{1, \dots, r\}$. De plus, le nombre de boucles est $2^r - 1$, puisque chaque sommet sauf celui correspondant à l'ensemble vide $\emptyset$ est connecté à lui-même. Enfin, le nombre d'arêtes d'un sommet à un autre est $\sum_{k=0}^r \binom{r}{k} (r-k) = r2^{r-1}$, car chacun des $\binom{r}{k}$ sommets correspondant à un ensemble de cardinal k est connecté à $r-k$ autres sommets par des arêtes représentant l'ajout d'un coupon supplémentaire.

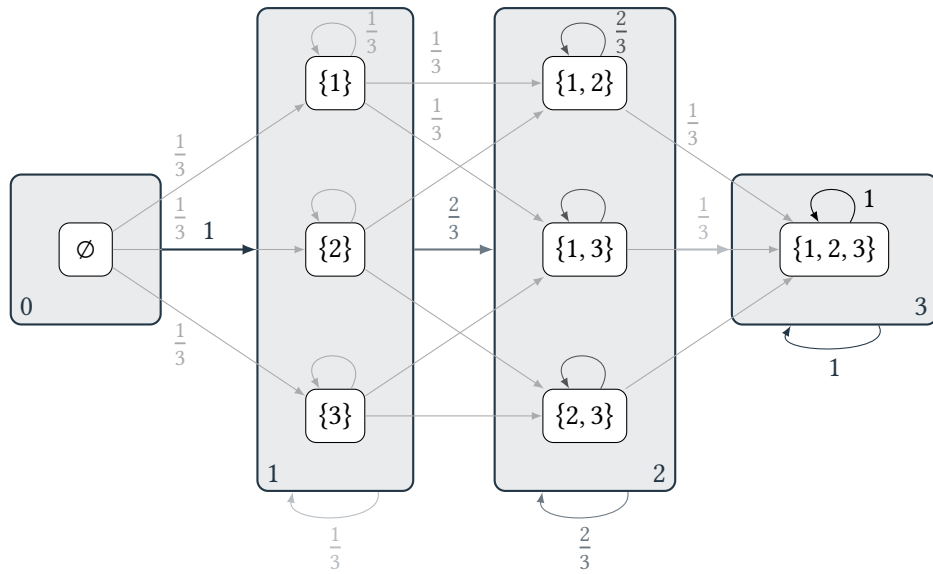


FIGURE 2.8 : Graphes des chaînes de Markov des ensembles et nombres de coupons collectés, $r = 3$

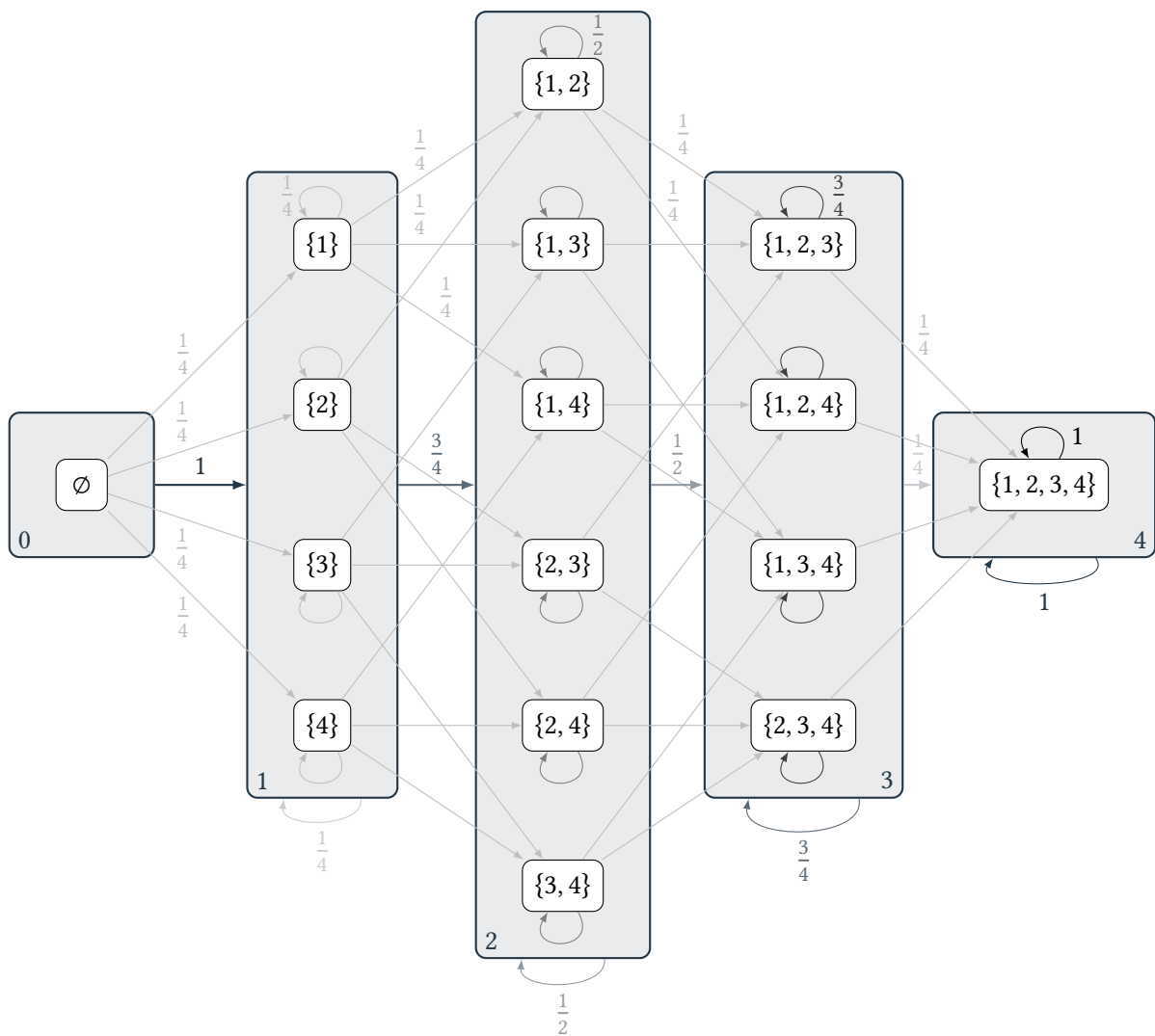


FIGURE 2.9 : Graphes des chaînes de Markov des ensembles et nombres de coupons collectés, $r = 4$

Marche aléatoire

Une marche aléatoire est un processus qui décrit le déplacement d'une particule dans un espace discret. À chaque pas, la particule se déplace d'une certaine distance dans une direction aléatoire. Les marches aléatoires sont utilisées pour modéliser un grand nombre de phénomènes physiques, biologiques et économiques. Dans ce chapitre, nous allons étudier les propriétés de base des marches aléatoires sur les réseaux $\mathbb{Z}^d$ de dimension $d \geq 1$.

3.1 Marche aléatoire unidimensionnelle

En dimension 1, une marche aléatoire est un processus défini sur l'ensemble $\mathbb{Z}$ des entiers relatifs. À chaque pas, la particule se déplace d'une unité au hasard vers la droite ou vers la gauche. L'évolution de la position de la particule au cours du temps est donnée par la somme des déplacements successifs et constitue, comme l'indique la définition ci-dessous, la marche aléatoire.

Définition 3.1 (Marche aléatoire simple unidimensionnelle). On appelle *marche aléatoire simple unidimensionnelle*, ou *marche aléatoire simple sur $\mathbb{Z}$* , toute suite $(W_n)_{n \geq 0}$ de variables aléatoires pour laquelle il existe une suite $(\xi_m)_{m \geq 1}$ de variables aléatoires indépendantes et identiquement distribuées à valeurs dans $\{-1, 1\}$, indépendantes de W_0 , et vérifiant presque sûrement, pour tout $n \in \mathbb{N}$,

$$W_n = W_0 + \sum_{m=1}^n \xi_m \in \mathbb{Z}.$$

Si $\mathbb{P}(\xi_1 = 1) = \mathbb{P}(\xi_1 = -1) = 1/2$, la marche aléatoire est dite *symétrique*. Sinon, elle est dite *biaisée* ou *asymétrique*. Si $W_0 = 0$ presque sûrement, la marche aléatoire est dite *issue de l'origine*.

Dans tout ce chapitre, l'adjectif « simple » sera omis car on ne considérera que ce type de marches aléatoires. On remarque que toute variable aléatoire ξ à valeurs dans $\{-1, 1\}$ peut s'écrire sous la forme $\xi = 2X - 1$, où $X = \mathbf{1}_{\{\xi=1\}}$ est une variable aléatoire de loi de Bernoulli de paramètre $p = \mathbb{P}(\xi = 1)$. Par conséquent, les marches aléatoires sur $\mathbb{Z}$ peuvent s'exprimer à partir des processus de Bernoulli de la définition 1.3. Plus précisément, une suite $(W_n)_{n \geq 0}$ de variables aléatoires est une marche aléatoire sur $\mathbb{Z}$ si et seulement si $W_0 \in \mathbb{Z}$ presque sûrement et il existe un processus de Bernoulli $(S_n)_{n \geq 0}$ indépendant de W_0 tel que presque sûrement, pour tout $n \in \mathbb{N}$,

$$W_n - W_0 = 2S_n - n. \quad (3.1)$$

Dans ce cas, le paramètre du processus de Bernoulli est donné par $p = \mathbb{P}(W_1 - W_0 = 1)$.

Algorithme 3.1 (Simulation d'une marche aléatoire unidimensionnelle). Dans le même esprit que pour l'algorithme 1.1, on simule m réalisations d'une marche aléatoire unidimensionnelle $(W_n)_{n \geq 0}$ issue de l'origine, puis on trace la position W_n en fonction de n . On utilise pour cela le script C.8. Pour $m = 10$ réalisations, une valeur finale de $n = 500$, et divers $p = \mathbb{P}(W_1 - W_0 = 1)$, on obtient les graphiques de la figure 3.1.

Du fait de la représentation (3.1), les marches aléatoires sur $\mathbb{Z}$ héritent de certaines propriétés des processus de Bernoulli obtenues dans le chapitre 1, notamment leur caractère markovien. En effet, la proposition 1.17 conduit au résultat suivant.

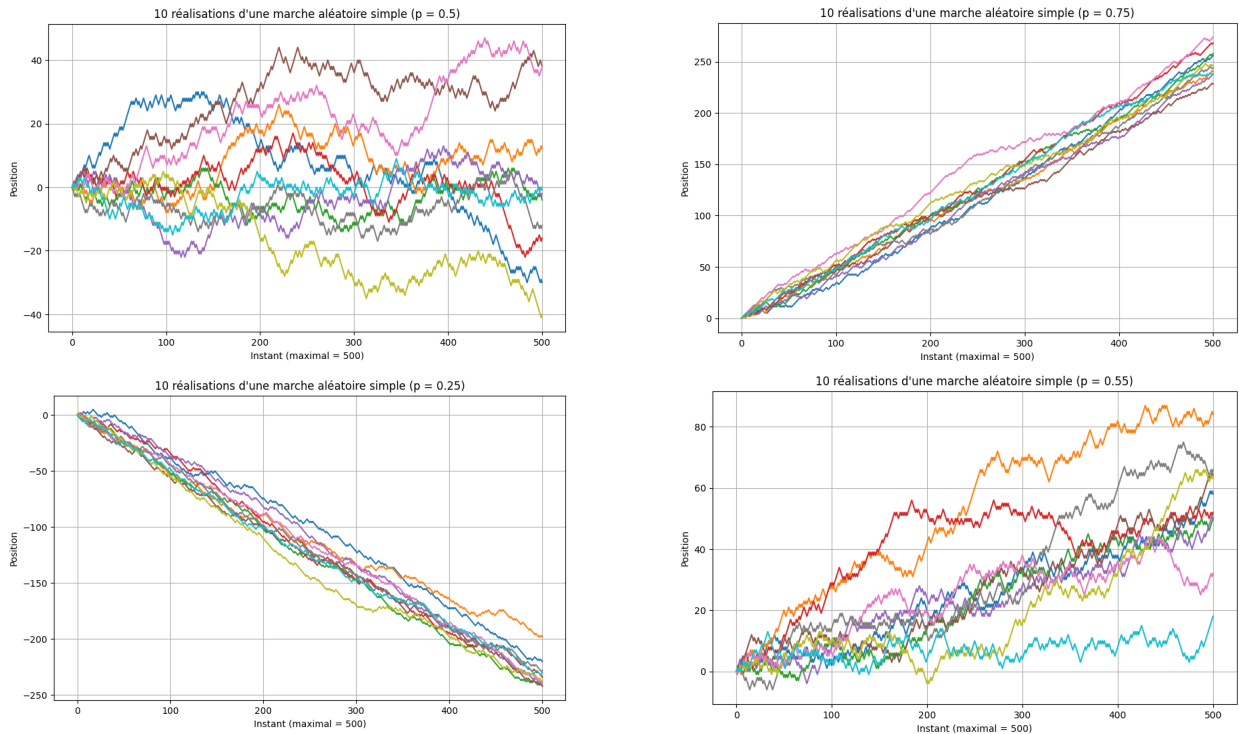


FIGURE 3.1 : $m = 10$ réalisations d'une marche aléatoire unidimensionnelle issue de l'origine et jusqu'à l'instant $n = 500$, avec une probabilité de faire un pas $+1$ égale à p

Proposition 3.1 (Aspect markovien des marches aléatoires unidimensionnelles). *Toute marche aléatoire unidimensionnelle $(W_n)_{n \geq 0}$ est une chaîne de Markov homogène à espace d'états $\mathbb{Z}$, dont les coefficients de la matrice de transition sont donnés, pour $k, l \in \mathbb{Z}$, par*

$$p_{k,l} = \begin{cases} 1-p & \text{si } l = k-1 \\ p & \text{si } l = k+1 \\ 0 & \text{sinon} \end{cases} \quad \text{avec} \quad p = \mathbb{P}(W_1 - W_0 = 1).$$

Démonstration. Considérons une marche aléatoire unidimensionnelle $(W_n)_{n \geq 0}$, et un processus de Bernoulli $(S_n)_{n \geq 0}$ indépendant de W_0 tel que (3.1) est vérifiée presque sûrement pour tout $n \in \mathbb{N}$. Pour tous entiers $n \in \mathbb{N}$ et $w_0, \dots, w_{n+1} \in \mathbb{Z}$, en posant $s_m = (w_m - w_0 + m)/2$ pour tout $m \in \{0, \dots, n+1\}$, on a

$$\begin{aligned} & \mathbb{P}(W_0 = w_0, \dots, W_n = w_n, W_{n+1} = w_{n+1}) \\ &= \mathbb{P}(W_0 = w_0) \mathbb{P}(S_0 = 0, S_1 = s_1, \dots, S_n = s_n, S_{n+1} = s_{n+1}) \\ &= \mathbb{P}(W_0 = w_0) \mathbb{P}(S_0 = 0, S_1 = s_1, \dots, S_n = s_n) ((1-p) \mathbf{1}_{\{s_{n+1}=s_n\}} + p \mathbf{1}_{\{s_{n+1}=s_n+1\}}) \\ &= \mathbb{P}(W_0 = w_0, \dots, W_n = w_n) p_{w_n, w_{n+1}}, \end{aligned}$$

où le coefficient $p_{w_n, w_{n+1}}$ est donné dans l'énoncé. La deuxième égalité découle de la proposition 1.17 si $s_1, \dots, s_n$ sont des entiers positifs, et est triviale ($0 = 0$) sinon. On en déduit ensuite que

$$\mathbb{P}(W_n = w_n \text{ et } W_{n+1} = w_{n+1}) = \mathbb{P}(W_n = w_n) p_{w_n, w_{n+1}},$$

en sommant sur tous les entiers $w_0, \dots, w_{n-1} \in \mathbb{Z}$. En divisant l'identité portant sur toute la trajectoire par $\mathbb{P}(W_0 = w_0, \dots, W_n = w_n)$ lorsque cette probabilité est non nulle, on obtient la propriété de Markov et l'homogénéité annoncées. $\square$

D'après la proposition 3.1, le graphe associé à une marche aléatoire unidimensionnelle $(W_n)_{n \geq 0}$, vue comme une chaîne de Markov homogène à espace d'états $\mathbb{Z}$, admet $\mathbb{Z}$ pour ensemble de sommets, et comporte un arc de k vers $k-1$ si $p < 1$, et de k vers $k+1$ si $p > 0$. On représente ce graphe sur la figure 3.2, en notant $p = \mathbb{P}(W_1 - W_0 = 1)$.

Si $0 < p < 1$, tous les états communiquent, de sorte que la chaîne de Markov est irréductible. D'après la proposition B.20, la chaîne est soit récurrente, soit transiente. Le résultat suivant montre qu'en dimension 1, seules les marches aléatoires symétriques sont récurrentes.

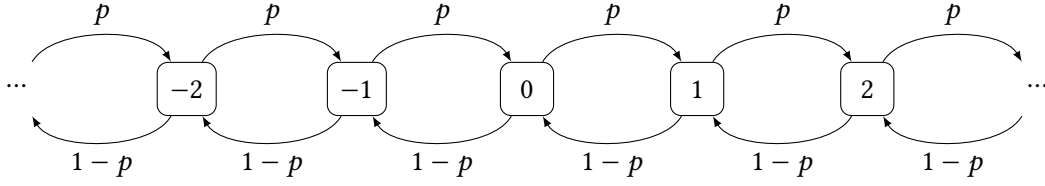


FIGURE 3.2 : Graphe de la chaîne de Markov associée à une marche aléatoire unidimensionnelle

Théorème 3.2 (Récurrence ou transience des marches aléatoires unidimensionnelles). *Toute marche aléatoire unidimensionnelle symétrique est récurrente nulle, tandis que toute marche aléatoire unidimensionnelle asymétrique est transiente.*

Démonstration. Si $p = \mathbb{P}(W_1 - W_0 = 1)$ vaut 0 ou 1, la marche se déplace toujours dans le même sens et tous les états sont transients. Supposons donc $0 < p < 1$. Soit $(W_n)_{n \geq 0}$ une marche aléatoire unidimensionnelle. D'après les propositions B.20 et B.28, comme la chaîne de Markov associée est irréductible, elle est soit récurrente positive, soit récurrente nulle, soit transiente. Il suffit donc de déterminer la nature d'un état quelconque $k \in \mathbb{Z}$ pour conclure. Notons $p_{k,l}^{(m)}$ et $p_{k,l}^*$ les coefficients des puissances de la matrice de transition de la chaîne, et ceux de sa matrice potentielle. Notons aussi $(S_n)_{n \geq 0}$ un processus de Bernoulli indépendant de W_0 tel que (3.1) est vérifiée presque sûrement pour tout $n \in \mathbb{N}$. En posant $p = \mathbb{P}(W_1 - W_0 = 1)$, on a

$$\mathbb{P}(W_0 = k) p_{k,k}^{(m)} = \mathbb{P}(W_m = k \text{ et } W_0 = k) = \mathbb{P}(W_0 = k) \mathbb{P}(S_m = m/2).$$

On choisit un état $k \in \mathbb{Z}$ tel que $\mathbb{P}(W_0 = k) > 0$, et on utilise le fait que S_m suit la loi $\text{Bin}(m, p)$. On observe alors que si m est impair, le coefficient $p_{k,k}^{(m)}$ est nul. De plus, si m est pair, on a

$$p_{k,k}^{(m)} = \binom{m}{m/2} p^{m/2} (1-p)^{m/2} = \frac{m!}{((m/2)!)^2} (p(1-p))^{m/2} \underset{m \rightarrow \infty}{\sim} \left(\frac{2}{\pi m} \right)^{1/2} (4p(1-p))^{m/2},$$

où l'équivalent découle de la formule de Stirling donnée dans le lemme 1.8. Par conséquent, le coefficient $p_{k,k}^* = \sum_{m=0}^{\infty} p_{k,k}^{(m)}$ est infini si et seulement si $p = 1/2$. La chaîne de Markov est donc récurrente si $p = 1/2$, et transiente sinon. Enfin, on observe que la mesure μ telle que $\mu_k = 1$ pour tout $k \in \mathbb{Z}$ est invariante pour la chaîne de Markov, qui est irréductible. Si $p = 1/2$, n'importe quel état $l \in \mathbb{Z}$ est récurrent, et la proposition B.24 implique que l'espérance τ_l du temps de retour en l vaut $\sum_{k \in \mathbb{Z}} \mu_k = \infty$. On en conclut que la chaîne est récurrente nulle.

Dans le cas asymétrique où $p \neq 1/2$, on peut aussi établir la transience à partir de l'identité (3.1) et de la loi forte des grands nombres pour le processus de Bernoulli donnée par la proposition 1.4. En effet, pour tout $k \in \mathbb{Z}$ tel que $\mathbb{P}(W_0 = k) > 0$, conditionnellement à $W_0 = k$, on a presque sûrement

$$W_n = W_0 + 2S_n - n = k + \left(2pn + o_{n \rightarrow \infty}(n) \right) - n \underset{n \rightarrow \infty}{\sim} (2p - 1)n,$$

ce qui montre que W_n est différent de k pour n assez grand. $\square$

3.2 Ruine du joueur

Les marches aléatoires peuvent s'utiliser pour étudier le problème dit de la *ruine du joueur*. Un joueur commence avec un capital de k euros, avec $k \in \{a, \dots, b\}$, pour deux entiers relatifs $a < b$. À chaque tour, le joueur gagne 1 euro avec une probabilité $p \in]0, 1[$, et perd 1 euro avec probabilité $1 - p$. Le joueur continue à jouer jusqu'à ce que son capital atteigne a ou b euros. On s'intéresse à la probabilité que son capital atteigne a euros en premier, et à la durée espérée de la partie. Si $a = 0$, le problème revient à déterminer la probabilité de ruine avant d'atteindre un capital de b euros.

En désignant par W_n le capital en euros du joueur après n tours, on peut modéliser le problème de la ruine du joueur par une marche aléatoire unidimensionnelle $(W_n)_{n \geq 0}$ telle que $\mathbb{P}(W_0 = k) = 1$ et $\mathbb{P}(W_1 - W_0 = 1) = p$. Si $p = 1/2$, la marche aléatoire est symétrique et le théorème 3.2 montre que la chaîne de Markov associée est récurrente. Ainsi, tous les états sont visités presque sûrement infiniment souvent, donc les deux états a et b sont forcément atteints. Si au contraire $p \neq 1/2$, la marche aléatoire est asymétrique et la chaîne de Markov est transiente. Dans ce cas, il est possible que l'état a ou l'état b ne soit pas atteint. Cependant, l'un des deux états est forcément atteint. On précise cela dans l'énoncé suivant, qui répond aussi aux problèmes soulevés au début de cette section.

Théorème 3.3 (Ruine du joueur). Soit $(W_n)_{n \geq 0}$ une marche aléatoire unidimensionnelle. Pour des entiers relatifs $a < b$, on introduit les premiers instants de passage de la marche aléatoire en a et en b , c'est-à-dire

$$T_a = \inf\{n \geq 0 \mid W_n = a\}, \quad T_b = \inf\{n \geq 0 \mid W_n = b\}, \quad T = \min\{T_a, T_b\}.$$

On suppose que $p = \mathbb{P}(W_1 - W_0 = 1) \in]0, 1[$ et on note $\rho = (1 - p)/p$. Pour tout entier $k \in \{a, \dots, b\}$ tel que $\mathbb{P}(W_0 = k) > 0$, on a les propriétés suivantes :

(i). Pour tout entier $n \in \mathbb{N}$, on a la majoration

$$\mathbb{P}(T \geq n \mid W_0 = k) \leq \delta^{\frac{n}{b-a}-1} \quad \text{avec} \quad \delta = 1 - \min\{p, 1 - p\}^{b-a} \in]0, 1[,$$

donc en particulier $\mathbb{E}[T \mid W_0 = k] < \infty$ et $\mathbb{P}(T < \infty \mid W_0 = k) = 1$;

(ii). La probabilité d'atteindre a avant b est donnée par

$$\mathbb{P}(T_a < T_b \mid W_0 = k) = \mathbb{P}(W_T = a \mid W_0 = k) = \begin{cases} \frac{\rho^b - \rho^k}{\rho^b - \rho^a} & \text{si } p \neq 1/2 \\ \frac{b - k}{b - a} & \text{si } p = 1/2; \end{cases}$$

(iii). La durée espérée de la partie est donnée par

$$\mathbb{E}[T \mid W_0 = k] = \begin{cases} \frac{k - a}{1 - 2p} - \frac{b - a}{1 - 2p} \cdot \frac{\rho^k - \rho^a}{\rho^b - \rho^a} & \text{si } p \neq 1/2 \\ (b - k)(k - a) & \text{si } p = 1/2. \end{cases}$$

Démonstration. Pour démontrer (i), considérons $r \in \mathbb{N}$ et notons $U_r = \inf\{m \geq 0 \mid W_{r(b-a)+m} \in \{a, b\}\}$. Pour tout entier $l \in \{a, \dots, b\}$, il existe un chemin à coordonnées entières de longueur minimale $m_l \in \{0, \dots, b - a\}$ reliant le point $(r(b - a), l)$ au point $(r(b - a) + m_l, a)$ ou au point $(r(b - a) + m_l, b)$. Ce chemin a des ordonnées $w_{r(b-a)}^* = l$, puis $w_{r(b-a)+1}^*, \dots, w_{r(b-a)+m_l-1}^*$, et enfin $w_{r(b-a)+m_l}^* \in \{a, b\}$, telles que pour tout $m \in \{0, \dots, m_l - 1\}$, on a $w_{r(b-a)+m+1}^* - w_{r(b-a)+m}^* \in \{-1, 1\}$. Si la marche aléatoire prend les valeurs $w_{r(b-a)}^*, \dots, w_{r(b-a)+m_l}^*$ aux instants $r(b - a), \dots, r(b - a) + m_l$, alors $U_r = m_l \leq b - a$. La proposition 3.1 montre que la marche aléatoire est une chaîne de Markov homogène, donc pour tous $w_0, \dots, w_{r(b-a)-1} \in \{a, \dots, b\}$, la proposition B.6 donne

$$\begin{aligned} & \mathbb{P}(W_0 = w_0, \dots, W_{r(b-a)-1} = w_{r(b-a)-1}, W_{r(b-a)} = l, U_r \leq b - a) \\ & \geq \mathbb{P}(W_0 = w_0, \dots, W_{r(b-a)-1} = w_{r(b-a)-1}, W_{r(b-a)} = w_{r(b-a)}^*, \dots, W_{r(b-a)+m_l} = w_{r(b-a)+m_l}^*) \\ & = \mathbb{P}(W_0 = w_0, \dots, W_{r(b-a)-1} = w_{r(b-a)-1}, W_{r(b-a)} = w_{r(b-a)}^*) \prod_{m=r(b-a)}^{r(b-a)+m_l-1} p_{w_m^*, w_{m+1}^*} \\ & \geq \mathbb{P}(W_0 = w_0, \dots, W_{r(b-a)-1} = w_{r(b-a)-1}, W_{r(b-a)} = l) \min\{p, 1 - p\}^{b-a}, \end{aligned}$$

où les coefficients $p_{w_m^*, w_{m+1}^*}$ sont ceux de la matrice de transition de la chaîne de Markov. En posant δ comme dans l'énoncé du théorème, on en déduit que pour tous $w_0, \dots, w_{r(b-a)} \in \{a, \dots, b\}$,

$$\mathbb{P}(W_0 = w_0, \dots, W_{r(b-a)-1} = w_{r(b-a)-1}, W_{r(b-a)} = w_{r(b-a)}, U_r > b - a)$$

$$\leq \delta \mathbb{P}(W_0 = w_0, \dots, W_{r(b-a)-1} = w_{r(b-a)-1}, W_{r(b-a)} = w_{r(b-a)}).$$

En prenant $w_0 = k \in \{a, \dots, b\}$ et en sommant sur tous les entiers $w_1, \dots, w_{r(b-a)} \in \{a+1, \dots, b-1\}$, on a

$$\mathbb{P}(W_0 = k \text{ et } T > (r+1)(b-a)) \leq \delta \mathbb{P}(W_0 = k \text{ et } T > r(b-a)).$$

Par une récurrence élémentaire sur l'entier $r \geq 0$, on a finalement

$$\mathbb{P}(W_0 = k \text{ et } T > r(b-a) \mid W_0 = k) \leq \delta^r.$$

On en déduit la borne annoncée sur $\mathbb{P}(T \geq n \mid W_0 = k)$ en considérant $r(b-a) < n \leq (r+1)(b-a)$. Ainsi, du fait de (1.7), la variable aléatoire T a une espérance finie, donc est nécessairement finie presque sûrement, conditionnellement à l'événement $\{W_0 = k\}$.

Pour démontrer (ii), on observe d'abord que les événements $\{T_a < T_b\}$ et $\{W_T = a\}$ sont égaux, donc ils ont la même probabilité conditionnellement à $\{W_0 = k\}$. Supposons dans un premier temps que $\mathbb{P}(W_0 = k) > 0$ pour tout $k \in \{a, \dots, b\}$, ce qui permet de définir

$$\theta(k) = \mathbb{P}(W_T = a \mid W_0 = k).$$

La propriété de Markov (B.2) pour la marche aléatoire donne pour tous entiers $n \geq 1$, $w_0, \dots, w_n \in \mathbb{Z}$,

$$\mathbb{P}(W_0 = w_0, \dots, W_n = w_n) = \mathbb{P}(W_0 = w_0 \text{ et } W_1 = w_1) \prod_{m=1}^{n-1} p_{w_m, w_{m+1}} = \mathbb{P}(W_0 = w_0) \prod_{m=0}^{n-1} p_{w_m, w_{m+1}}.$$

En sommant la première égalité sur tous les entiers $w_0 \in \mathbb{Z}$ et en utilisant la deuxième égalité en décalant les valeurs de w_m , on en déduit aussi que

$$\begin{cases} \mathbb{P}(W_1 = w_1, \dots, W_n = w_n) = \mathbb{P}(W_1 = w_1) \prod_{m=1}^{n-1} p_{w_m, w_{m+1}} \\ \mathbb{P}(W_0 = w_1, \dots, W_{n-1} = w_n) = \mathbb{P}(W_0 = w_1) \prod_{m=1}^{n-1} p_{w_m, w_{m+1}}. \end{cases}$$

En prenant $w_0 = k \in \{a+1, \dots, b-1\}$ et $w_1 = k+1$, on a par conséquent

$$\begin{aligned} \mathbb{P}(W_1 = k+1, W_2 = w_2, \dots, W_n = w_n \mid W_0 = k) &= p \prod_{m=1}^{n-1} p_{w_m, w_{m+1}} \\ &= p \mathbb{P}(W_1 = w_2, \dots, W_{n-1} = w_n \mid W_0 = k+1). \end{aligned}$$

Si $k \leq b-2$, en sommant sur les entiers $w_2, \dots, w_{n-1} \in \{a+1, \dots, b-1\}$ et en prenant $w_n = a$, on obtient

$$\mathbb{P}(W_1 = k+1, T = n, W_n = a \mid W_0 = k) = p \mathbb{P}(T = n-1 \text{ et } W_{n-1} = a \mid W_0 = k+1). \quad (3.2)$$

Cette identité reste vraie si $k = b-1$, les deux membres étant nuls. Conditionnellement à $\{W_0 = k\}$, on a nécessairement $T \geq 1$, donc en sommant sur tous les entiers $n \geq 1$, on en déduit que

$$\mathbb{P}(W_1 = k+1 \text{ et } W_T = a \mid W_0 = k) = p \mathbb{P}(W_T = a \mid W_0 = k+1).$$

Le même raisonnement est valable pour $k-1$ au lieu de $k+1$, en remplaçant p par $1-p$. Finalement,

$$\begin{aligned} \mathbb{P}(W_T = a \mid W_0 = k) &= \mathbb{P}(W_1 = k+1 \text{ et } W_T = a \mid W_0 = k) + \mathbb{P}(W_1 = k-1 \text{ et } W_T = a \mid W_0 = k) \\ &= p \mathbb{P}(W_T = a \mid W_0 = k+1) + (1-p) \mathbb{P}(W_T = a \mid W_0 = k-1), \end{aligned}$$

ce qui donne, pour tout $k \in \{a+1, \dots, b-1\}$, la relation de récurrence linéaire

$$\theta(k) = p \theta(k+1) + (1-p) \theta(k-1). \quad (3.3)$$

On en déduit la relation $\Theta_{k+1} = M\Theta_k$ pour tout $k \in \{a+1, \dots, b-1\}$, et donc par récurrence la relation $\Theta_k = M^{k-a-1}\Theta_{a+1}$ pour tout $k \in \{a+1, \dots, b\}$, où

$$M = \begin{pmatrix} 1/p & 1-1/p \\ 1 & 0 \end{pmatrix} \quad \text{et} \quad \Theta_k = \begin{pmatrix} \theta(k) \\ \theta(k-1) \end{pmatrix}.$$

La matrice M admet 1 et p pour valeurs propres. Si $p \neq 1/2$, ces deux valeurs propres sont distinctes et la matrice M^{k-a-1} est semblable à la matrice diagonale dont les coefficients diagonaux sont 1 et p^{k-a-1} . On en déduit que $\theta(k) = \alpha + \beta p^k$, pour deux constantes α et β indépendantes de k . Si $p = 1/2$, les deux valeurs propres sont identiques et la matrice M^{k-a-1} est semblable à la matrice $\begin{pmatrix} 1 & k-a-1 \\ 0 & 1 \end{pmatrix}$, donc on a $\theta(k) = \alpha + \beta k$ pour deux constantes α et β indépendantes de k . On peut alors déterminer α et β en observant que $\theta(a) = 1$ et $\theta(b) = 0$, et obtenir les formules annoncées pour $\theta(k)$.

Pour prouver (iii), toujours en supposant que $\mathbb{P}(W_0 = k) > 0$ pour tout $k \in \{a, \dots, b\}$, on définit

$$\eta(k) = \mathbb{E}[T \mid W_0 = k].$$

On a (3.2) et une identité analogue avec b à la place de a . En sommant ces deux identités, on a pour tous entiers $k \in \{a+1, \dots, b-1\}$ et $n \geq 1$,

$$\mathbb{P}(W_1 = k+1 \text{ et } T = n \mid W_0 = k) = p \mathbb{P}(T = n-1 \mid W_0 = k+1).$$

La même égalité est valable pour $k-1$ au lieu de $k+1$, en remplaçant p par $1-p$. On en déduit que

$$\begin{aligned} \mathbb{P}(T = n \mid W_0 = k) &= \mathbb{P}(W_1 = k+1 \text{ et } T = n \mid W_0 = k) + \mathbb{P}(W_1 = k-1 \text{ et } T = n \mid W_0 = k) \\ &= p \mathbb{P}(T = n-1 \mid W_0 = k+1) + (1-p) \mathbb{P}(T = n-1 \mid W_0 = k-1). \end{aligned}$$

Par linéarité de l'espérance et des sommes de séries, on obtient

$$\begin{aligned} \mathbb{E}[T \mid W_0 = k] &= 1 + \mathbb{E}[T-1 \mid W_0 = k] \\ &= 1 + \sum_{n=1}^{\infty} n \mathbb{P}(T-1 = n \mid W_0 = k) \\ &= 1 + p \sum_{n=1}^{\infty} n \mathbb{P}(T = n \mid W_0 = k+1) + (1-p) \sum_{n=1}^{\infty} n \mathbb{P}(T = n \mid W_0 = k-1) \\ &= 1 + p \mathbb{E}[T \mid W_0 = k+1] + (1-p) \mathbb{E}[T \mid W_0 = k-1]. \end{aligned}$$

Cela donne, pour tout $k \in \{a+1, \dots, b-1\}$, la relation de récurrence linéaire

$$\eta(k) = p\eta(k+1) + (1-p)\eta(k-1) + 1.$$

Si $p \neq 1/2$, on voit que $\eta_0(k) = k/(1-2p)$ vérifie cette relation. Par linéarité, η vérifie la relation si et seulement si $\eta - \eta_0$ vérifie (3.3). On a donc $\eta(k) = \eta_0(k) + \alpha + \beta p^k$, pour deux constantes α et β indépendantes de k . Si $p = 1/2$, on voit que $\eta_1(k) = -k^2$ vérifie la relation, donc $\eta(k) = \eta_1(k) + \alpha + \beta k$, pour deux constantes α et β indépendantes de k . On peut alors déterminer α et β en observant que $\eta(a) = \eta(b) = 0$, et finalement obtenir les formules annoncées pour $\eta(k)$.

Il reste à étendre (ii) et (iii) au cas où $\mathbb{P}(W_0 = k) > 0$, mais $\mathbb{P}(W_0 = l) = 0$ pour certains $l \in \{a, \dots, b\}$. Il suffit pour cela de construire une marche $(W'_n)_{n \geq 0}$ dont la position initiale est uniforme sur $\{a, \dots, b\}$ et dont les incréments sont indépendants, indépendants de W'_0 , avec probabilités p et $1-p$ de valoir respectivement 1 et -1 . Cette marche vérifie

$$\mathbb{P}(W'_1 = w_1, \dots, W'_n = w_n \mid W'_0 = k) = \mathbb{P}(W_1 = w_1, \dots, W_n = w_n \mid W_0 = k)$$

pour tous entiers $n \geq 1$ et $w_1, \dots, w_n \in \mathbb{Z}$, d'après la proposition B.6. Ainsi, en notant T' le premier instant de passage associé à la nouvelle marche aléatoire $(W'_n)_{n \geq 0}$, on en déduit que

$$\mathbb{P}(W_T = a \mid W_0 = k) = \mathbb{P}(W'_{T'} = a \mid W'_0 = k) \quad \text{et} \quad \mathbb{E}[T \mid W_0 = k] = \mathbb{E}[T' \mid W'_0 = k].$$

On peut alors appliquer ce qui précède à $(W'_n)_{n \geq 0}$, et transférer les résultats à $(W_n)_{n \geq 0}$. □

Algorithme 3.2 (Simulation du problème de ruine du joueur). Pour illustrer le problème de ruine du joueur, on simule m réalisations d'une marche aléatoire unidimensionnelle $(W_n)_{n \geq 0}$ partant de l'état initial $k \in \{a, \dots, b\}$, et s'arrêtant dès que la marche atteint une des bornes a ou b . On utilise le script C.9. Pour chaque réalisation, on suit l'évolution de la position W_n en fonction de n , en attribuant une couleur spécifique en fonction de la première borne atteinte : bleu si la marche atteint a en premier, rouge si la marche atteint b en premier. La proportion des marches atteignant a et b est calculée, ainsi que la distribution de la durée T de la partie. Les simulations correspondantes, pour diverses valeurs des paramètres, sont présentées sur la figure 3.3. La probabilité d'atteindre a avant b semble d'autant plus forte que la probabilité p de gagner un tour de la partie est faible.

3.3 Premier retour au point de départ

On s'intéresse maintenant à la loi du premier instant de retour d'une marche aléatoire unidimensionnelle à son point de départ. Sans perte de généralité, on peut supposer que la marche est issue de l'origine. Jusqu'à son premier retour à l'origine, la marche aléatoire forme une trajectoire qui présente des incréments ± 1 entre deux instants consécutifs, et reste toujours strictement au-dessus de l'axe des abscisses, ou toujours strictement en-dessous. Ces trajectoires sont liées aux *chemins de Dyck*. On commence par définir ces objets, puis on cherche à les dénombrer.

Définition 3.2 (Chemins de Dyck). Soit $n \in \mathbb{N}$. Un *chemin de Dyck* de longueur $2n + 2$ est un $(2n + 3)$ -uplet d'entiers $(w_0, \dots, w_{2n+2}) \in \mathbb{Z}^{2n+3}$ tel que

$$\begin{cases} w_0 = w_{2n+2} = 0 \\ w_1, \dots, w_{2n+1} > 0 \\ w_1 - w_0, \dots, w_{2n+2} - w_{2n+1} \in \{-1, 1\}. \end{cases}$$

On note $\mathcal{D}_n$ l'ensemble des chemins de Dyck de longueur $2n + 2$, et on note D_n le cardinal de $\mathcal{D}_n$.

Les chemins de Dyck de longueur $2n + 2$ peuvent aussi se voir comme des représentations des *mots de Dyck*, c'est-à-dire les parenthésages bien formés, de longueur $2n$. En effet, on peut faire correspondre un chemin de Dyck $(w_0, \dots, w_{2n+2}) \in \mathcal{D}_n$ et un mot de Dyck $\epsilon_2 \dots \epsilon_{2n+1}$ en convenant que ϵ_m est une parenthèse ouvrante si $w_m - w_{m-1} = 1$, et une parenthèse fermante si $w_m - w_{m-1} = -1$. Par exemple, le chemin de Dyck $(0, 1, 2, 3, 2, 1, 2, 1, 2, 1, 2, 3, 2, 3, 2, 1, 0)$ représenté sur la figure 3.4 correspond au mot de Dyck $((()))()()((()))$. Les cardinaux des ensembles de chemins de Dyck vérifient la *relation de Segner*, ce qui permet de les calculer de manière récurrente.

Proposition 3.4 (Relation de Segner pour les nombres de chemins de Dyck). La suite $(D_n)_{n \geq 0}$ des nombres de chemins de Dyck vérifie, pour tout $n \geq 1$, la relation de récurrence de Segner :

$$D_n = \sum_{m=0}^{n-1} D_m D_{n-1-m} \quad \text{et} \quad D_0 = 1.$$

Démonstration. Considérons un entier $n \geq 1$ et un chemin de Dyck $(w_0, \dots, w_{2n+2}) \in \mathcal{D}_n$. Le plus grand indice inférieur à $2n$ où le chemin prend la valeur 1 est un entier impair $2m_0 + 1$, avec $m_0 \in \{0, \dots, n-1\}$. On peut alors construire deux chemins $(w_0, \dots, w_{2m_0+1}, 0) \in \mathcal{D}_{m_0}$ et $(w_{2m_0+1} - 1, \dots, w_{2n+1} - 1) \in \mathcal{D}_{n-1-m_0}$, comme l'illustre la figure 3.4. Réciproquement, à partir de deux chemins $(w'_0, \dots, w'_{2m_0+2}) \in \mathcal{D}_{m_0}$ et $(w''_0, \dots, w''_{2(n-m_0)}) \in \mathcal{D}_{n-1-m_0}$, on peut former le chemin $(w'_0, \dots, w'_{2m_0+1}, w''_1 + 1, \dots, w''_{2(n-m_0)} + 1, 0) \in \mathcal{D}_n$. On en déduit que l'ensemble $\mathcal{D}_n$ est en bijection avec l'union disjointe sur $m_0 \in \{0, \dots, n-1\}$ des produits cartésiens $\mathcal{D}_{m_0} \times \mathcal{D}_{n-1-m_0}$. On obtient la relation de Segner en passant au cardinal. $\square$

Il est ensuite possible de résoudre la récurrence de Segner par la méthode des fonctions génératrices, ce qui permet de donner une expression explicite en fonction de n pour les cardinaux D_n des ensembles de chemins de Dyck. Plus précisément, on peut montrer que ces cardinaux sont donnés par les *nombres de Catalan*. Cependant, on donne ici une méthode moins technique basée sur un *principe de réflexion* souvent attribué à André.

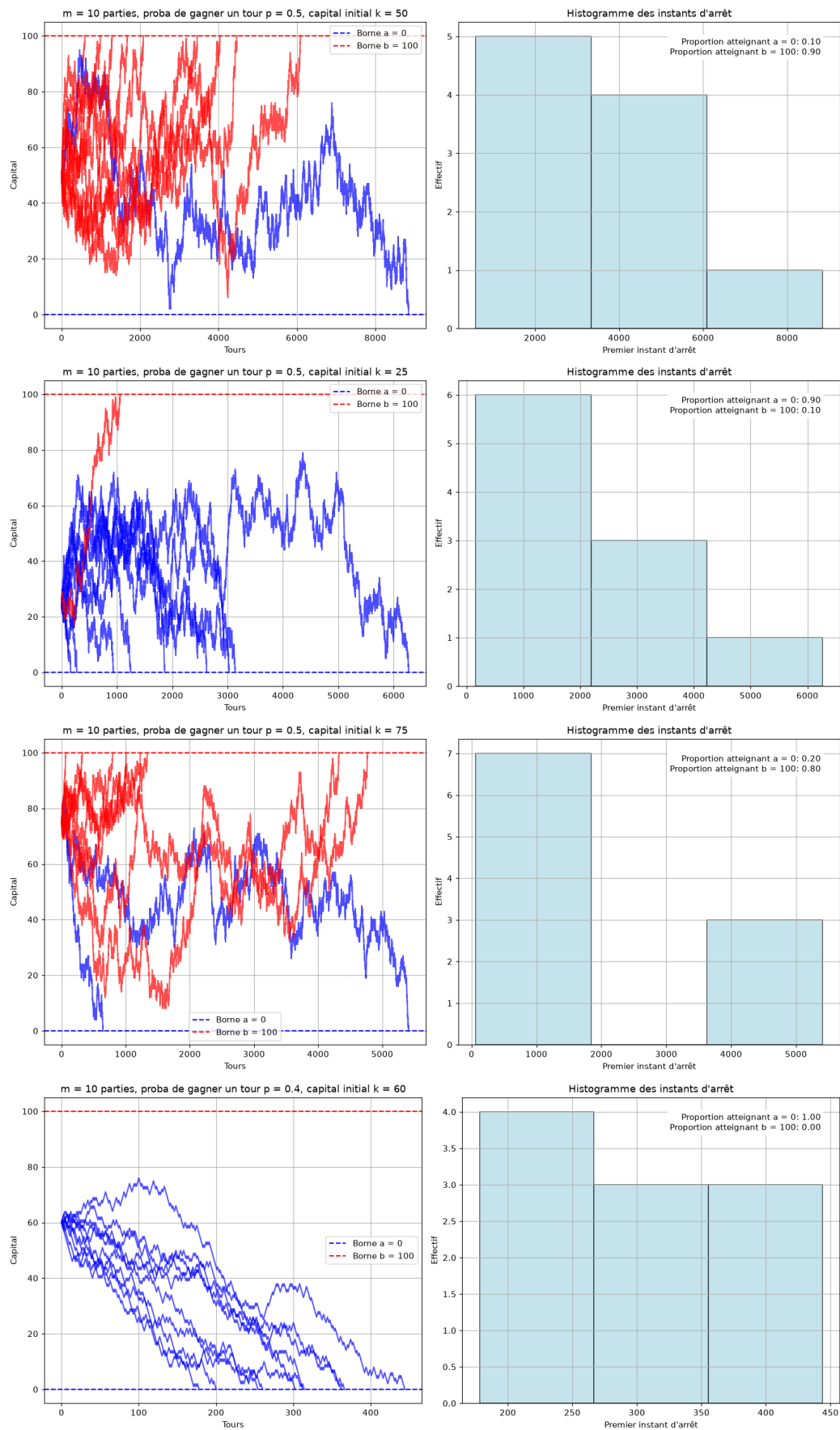


FIGURE 3.3 : Simulations du problème de la ruine du joueur

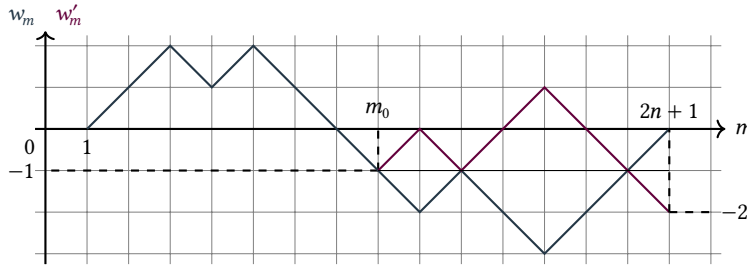


FIGURE 3.5 : Eugène Charles Catalan (1814–1894) et Désiré André (1840–1917) encadrant un principe de réflexion

Théorème 3.6 (Premier retour à l'origine pour une marche unidimensionnelle). Soit $(W_n)_{n \geq 0}$ une marche aléatoire unidimensionnelle issue de l'origine. On note $p = P(W_1 = 1)$, ainsi que T le premier instant de retour à l'origine, c'est-à-dire

$$T = \inf\{n \geq 1 \mid W_n = 0\}.$$

Si cet instant T est fini, il prend une valeur de la forme $2n$, pour $n \in \mathbb{N}^*$, et ce avec probabilité

$$P(T = 2n) = 2C_{n-1} p^n (1-p)^n.$$

Le premier instant de retour T a donc une probabilité d'être fini et une espérance donnés par

$$P(T < \infty) = 1 - |1 - 2p| \quad \text{et} \quad E[T] = \infty.$$

Démonstration. Si la marche aléatoire vérifie $T = m$, alors il y a autant d'entiers $n \in \{0, \dots, m-1\}$ tels que $W_{n+1} - W_n = 1$ que d'entiers tels que $W_{n+1} - W_n = -1$. On en déduit que m est pair, donc de la forme $2n$, avec $n \in \mathbb{N}^*$. De plus, comme la marche est issue de l'origine, l'événement $\{T = 2n\}$ correspond à des trajectoires reliant les points $(0, 0)$ et $(2n, 0)$, présentant des incréments ± 1 entre deux instants consécutifs, et restant toujours strictement au-dessus de l'axe des abscisses en dehors de leurs extrémités, ou toujours strictement en-dessous. Autrement dit, ces trajectoires sont des chemins de Dyck de longueur $2n$, ou leurs symétriques par rapport à l'axe des abscisses. On a donc

$$\{T = 2n\} = \{(W_0, \dots, W_{2n}) \in \mathcal{D}_{n-1}\} \sqcup \{(-W_0, \dots, -W_{2n}) \in \mathcal{D}_{n-1}\}.$$

Cependant, un chemin de Dyck quelconque $(w_0, \dots, w_{2n}) \in \mathcal{D}_{n-1}$ et son symétrique présentent tous les deux n incréments $+1$ et n incréments -1 , ce qui donne

$$P(W_0 = w_0, \dots, W_{2n} = w_{2n}) = P(-W_0 = w_0, \dots, -W_{2n} = w_{2n}) = p^n (1-p)^n.$$

On en déduit avec la proposition 3.5 la formule annoncée pour la probabilité de l'événement $\{T = 2n\}$. Ensuite, en sommant sur tous les entiers $n \in \mathbb{N}$, on obtient

$$P(T < \infty) = \sum_{n=1}^{\infty} P(T = 2n) = f(p(1-p)) \quad \text{avec} \quad f(x) = 2x \sum_{n=0}^{\infty} C_n x^n.$$

D'après la formule de Stirling donnée par le lemme 1.8, on a

$$C_n = \frac{1}{n+1} \binom{2n}{n} \underset{n \rightarrow \infty}{\sim} \frac{4^n}{(n+1)\sqrt{\pi n}} \underset{n \rightarrow \infty}{\sim} \frac{4^n}{n^{3/2}\sqrt{\pi}}, \quad (3.4)$$

de sorte que la série entière définissant f converge normalement sur l'intervalle $[-1/4, 1/4]$. De plus, les nombres de Catalan sont égaux aux nombres de chemins de Dyck, donc ils vérifient aussi la relation de récurrence de Segner donnée par la proposition 3.4. Ainsi, en utilisant la proposition B.2 et le théorème B.3, ou en effectuant un produit de Cauchy de séries entières, on a pour $x \neq 0$,

$$\frac{f(x)}{2x} = \sum_{n=0}^{\infty} C_n x^n = 1 + \sum_{n=1}^{\infty} \sum_{m=0}^{n-1} C_m C_{n-1-m} x^n = 1 + x \sum_{n=0}^{\infty} \sum_{m=0}^n C_m x^m C_{n-m} x^{n-m}$$

$$\begin{aligned}
&= 1 + x \sum_{n=0}^{\infty} \sum_{\substack{(m_1, m_2) \in \mathbb{N}^2 \\ m_1 + m_2 = n}} C_{m_1} x^{m_1} C_{m_2} x^{m_2} = 1 + x \sum_{(m_1, m_2) \in \mathbb{N}^2} C_{m_1} x^{m_1} C_{m_2} x^{m_2} \\
&= 1 + x \sum_{m_1=0}^{\infty} \sum_{m_2=0}^{\infty} C_{m_1} x^{m_1} C_{m_2} x^{m_2} = 1 + x \sum_{m_1=0}^{\infty} C_{m_1} x^{m_1} \frac{f(x)}{2x} = 1 + \frac{f(x)^2}{4x}.
\end{aligned}$$

On en tire l'équation fonctionnelle $2f(x) = 4x + f(x)^2$, vraie aussi pour $x = 0$. Il existe donc un réel $\varepsilon(x) \in \{-1, 1\}$ tel que $f(x) = 1 + \varepsilon(x)\sqrt{1 - 4x}$. Cependant, par convergence normale, la fonction f est continue sur l'intervalle $[-1/4, 1/4]$, donc la fonction ε est constante sur l'intervalle $[-1/4, 1/4[$. De plus, comme $0 = f(0) = 1 + \varepsilon(0)$, la fonction ε est égale à -1 sur ce dernier intervalle. Ainsi, $f(x) = 1 - \sqrt{1 - 4x}$ pour tout $x \in [-1/4, 1/4]$. Cela donne

$$\mathbb{P}(T < \infty) = f(p(1 - p)) = 1 - \sqrt{1 - 4p(1 - p)} = 1 - |1 - 2p|.$$

Si $p \neq 1/2$, on observe que $T = \infty$ avec probabilité non nulle, donc $\mathbb{E}[T] = \infty$. En revanche, si $p = 1/2$, on a $T < \infty$ presque sûrement, donc

$$\mathbb{E}[T] = \sum_{n=1}^{\infty} 2n \mathbb{P}(T = 2n) = \sum_{n=1}^{\infty} 4n C_{n-1} 4^{-n} = \sum_{n=0}^{\infty} (n+1) C_n 4^{-n} = \infty.$$

En effet, la série est divergente car son terme général est équivalent à $1/\sqrt{\pi n}$, d'après (3.4). $\square$

En combinant le théorème 3.6 avec la proposition B.19 et la définition B.11, on obtient que l'origine est un état récurrent nul pour la marche aléatoire si $p = 1/2$, et un état transient si $p \neq 1/2$. Pour $0 < p < 1$, la marche constitue une chaîne de Markov irréductible ; pour $p = 0$ ou $p = 1$, elle est déterministe et transiente. On retrouve ainsi le théorème 3.2.

3.4 Problème du scrutin

Le *problème du scrutin*, posé et résolu par Bertrand, consiste à déterminer la probabilité qu'à l'issue d'un scrutin avec deux candidats A et B , le candidat A ait toujours strictement plus de voix que le candidat B , tout au long du dépouillement. Ce problème peut être reformulé avec une marche aléatoire unidimensionnelle. En effet, supposons que les candidats A et B obtiennent respectivement a et b voix à l'issue du scrutin, avec $0 \leq b \leq a$, et désignons par W_m la différence de voix entre le candidat A et le candidat B après le dépouillement des m premiers bulletins. S'il y a $n = a + b$ bulletins au total, alors $(W_0, \dots, W_n)$ peut être vu comme une marche aléatoire issue de l'origine et arrivant à l'instant n à la valeur $k = a - b$, c'est-à-dire conditionnée à vérifier $W_0 = 0$ et $W_n = k$. Le problème du scrutin revient donc à déterminer la probabilité que $W_1, \dots, W_n > 0$. La solution est donnée par le résultat suivant. Il est remarquable que celle-ci ne dépende pas de $p = \mathbb{P}(W_1 - W_0 = 1)$.

Théorème 3.7 (Problème du scrutin). Soient des entiers $n \geq 1$ et $0 \leq k \leq n$, et soit $(W_m)_{m \geq 0}$ une marche aléatoire unidimensionnelle telle que $\mathbb{P}(W_0 = 0 \text{ et } W_n = k) > 0$. On a alors

$$\mathbb{P}(W_1 > 0, \dots, W_n > 0 \mid W_0 = 0 \text{ et } W_n = k) = \frac{k}{n}.$$

Démonstration. Désignons par $\mathcal{P}_{n,k}$ l'ensemble des $(n+1)$ -uplets d'entiers $(w_0, \dots, w_n) \in \mathbb{Z}^{n+1}$ tels que $w_0 = 0$, $w_n = k$ et $w_{m+1} - w_m \in \{-1, 1\}$ pour tout $m \in \{0, \dots, n-1\}$. Chaque élément de cet ensemble est uniquement déterminé par les $(n+k)/2$ indices $m \in \{0, \dots, n-1\}$ tels que $w_{m+1} - w_m = 1$, ou de manière équivalente par les $(n-k)/2$ indices m tels que $w_{m+1} - w_m = -1$. On en déduit que

$$\#\mathcal{P}_{n,k} = \binom{n}{\frac{n+k}{2}}.$$

Notons aussi $\mathcal{P}_{n,k}^+$ l'ensemble des chemins $(w_0, \dots, w_n) \in \mathcal{P}_{n,k}$ tels que $w_1, \dots, w_n > 0$. Conditionnellement à $\{W_0 = 0\}$, chaque chemin a une probabilité $p^{(n+k)/2}(1-p)^{(n-k)/2}$ d'apparaître. Le facteur $\mathbb{P}(W_0 = 0) > 0$ se

simplifie donc dans le quotient suivant :

$$\begin{aligned} \mathbb{P}(W_1 > 0, \dots, W_n > 0 \mid W_0 = 0 \text{ et } W_n = k) &= \frac{\mathbb{P}(W_0 = 0, W_1 > 0, \dots, W_n > 0, W_n = k)}{\mathbb{P}(W_0 = 0 \text{ et } W_n = k)} \\ &= \frac{p^{(n+k)/2}(1-p)^{(n-k)/2} \# \mathcal{P}_{n,k}^+}{p^{(n+k)/2}(1-p)^{(n-k)/2} \# \mathcal{P}_{n,k}^-} = 1 - \frac{\# \mathcal{P}_{n,k}^-}{\# \mathcal{P}_{n,k}^+}, \end{aligned}$$

où $\mathcal{P}_{n,k}^- = \mathcal{P}_{n,k} \setminus \mathcal{P}_{n,k}^+$. On utilise alors un principe de réflexion comme dans la preuve de la proposition 3.5. En effet, considérons l'application de $\mathcal{P}_{n,k}^-$ dans lui-même qui, à un chemin $(w_0, \dots, w_n)$, associe le chemin $(w_0, -w_1, \dots, -w_{m_0-1}, w_{m_0}, \dots, w_n)$, où m_0 est le plus petit indice dans $\{1, \dots, n\}$ tel que $w_{m_0} = 0$, voir la figure 3.6. Cette application est une involution qui échange les chemins de $\mathcal{P}_{n,k}^-$ tels que $w_1 = 1$ et ceux tels que $w_1 = -1$. De plus, ces derniers chemins sont en bijection avec les chemins de $\mathcal{P}_{n-1,k+1}$, via l'application $(w_0, \dots, w_n) \mapsto (w_1 + 1, w_2 + 1, \dots, w_n + 1)$. On en déduit que

$$\# \mathcal{P}_{n,k}^- = 2 \# \mathcal{P}_{n-1,k+1} = 2 \binom{n-1}{\frac{n+k}{2}}.$$

On conclut par de simples calculs sur les coefficients binomiaux. $\square$

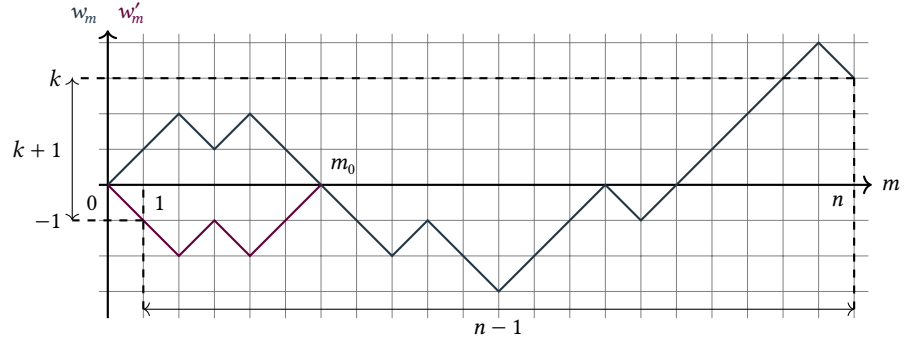


FIGURE 3.6 : Joseph Bertrand (1822–1900) et un principe de réflexion

3.5 Marche aléatoire symétrique multidimensionnelle

Les marches aléatoires peuvent se définir en n'importe quelle dimension $d \geq 1$, de la manière suivante. On note $(e_1, \dots, e_d)$ la base canonique de $\mathbb{R}^d$. On se restreint ici aux marches symétriques.

Définition 3.3 (Marche aléatoire simple symétrique en dimension d). Étant donné un entier $d \geq 1$, on appelle *marche aléatoire simple symétrique en dimension d* , ou *marche aléatoire simple symétrique sur $\mathbb{Z}^d$* , toute suite $(W_n)_{n \geq 0}$ de variables aléatoires pour laquelle il existe une suite $(\xi_m)_{m \geq 1}$ de variables aléatoires indépendantes et identiquement distribuées selon la loi uniforme dans l'ensemble $\{\pm e_1, \dots, \pm e_d\}$, indépendantes de W_0 , et vérifiant presque sûrement, pour tout $n \in \mathbb{N}$,

$$W_n = W_0 + \sum_{m=1}^n \xi_m \in \mathbb{Z}^d.$$

Si $W_0 = 0$ presque sûrement, la marche aléatoire est dite *issue de l'origine*.

À chaque instant, la marche aléatoire fait donc un saut sur un de ses plus proches voisins dans $\mathbb{Z}^d$ au sens de la norme définie par $\|x\|_1 = |x_1| + \dots + |x_d|$ pour tout $x = (x_1, \dots, x_d) \in \mathbb{R}^d$.

Algorithme 3.3 (Simulation d'une marche aléatoire symétrique dans le plan). Pour illustrer une marche aléatoire symétrique dans le plan, on simule une trajectoire $(W_n)_{n \geq 0}$ issue de l'origine. À chaque pas, la marche se déplace

d'une unité dans l'une des quatre directions possibles (haut, bas, gauche, droite) avec une probabilité égale à $1/4$. On utilise le script C.10. La trajectoire est représentée avec un dégradé de bleu pour indiquer l'évolution temporelle : les premiers pas sont en bleu clair et les derniers en bleu foncé. De plus, des flèches sont ajoutées au milieu de chaque segment pour indiquer la direction des pas. Les graphiques correspondants, générés pour diverses longueurs des marches, sont sur la figure 3.7.

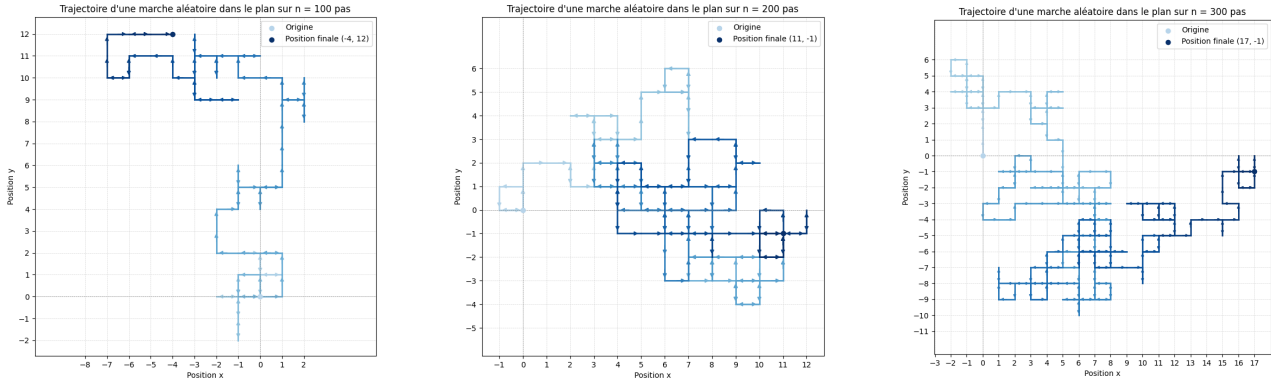


FIGURE 3.7 : Marches aléatoires symétriques issue de l'origine dans le plan

Tout comme dans le cas unidimensionnel, la marche aléatoire simple en dimension d est une chaîne de Markov. On a plus précisément le résultat suivant, à comparer avec la proposition 3.1.

Proposition 3.8 (Aspect markovien des marches aléatoires symétriques en dimension d). *Toute marche aléatoire symétrique $(W_n)_{n \geq 0}$ en dimension d est une chaîne de Markov homogène à espace d'états $\mathbb{Z}^d$, dont les coefficients de la matrice de transition vérifient, pour $k, l \in \mathbb{Z}^d$,*

$$p_{k,l} = \begin{cases} \frac{1}{2d} & \text{si } \|l - k\|_1 = 1 \\ 0 & \text{sinon.} \end{cases}$$

Démonstration. Considérons une marche aléatoire symétrique $(W_n)_{n \geq 0}$ en dimension d , et une suite $(\xi_m)_{m \geq 1}$ de variables aléatoires associées comme dans la définition 3.3. Pour tout entier $n \in \mathbb{N}$ et tous points $w_0, \dots, w_{n-1}, k, l \in \mathbb{Z}^d$, on a par indépendance

$$\begin{aligned} & \mathbb{P}(W_0 = w_0, \dots, W_{n-1} = w_{n-1}, W_n = k, W_{n+1} = l) \\ &= \mathbb{P}(W_0 = w_0, \xi_1 = w_1 - w_0, \dots, \xi_n = k - w_{n-1}, \xi_{n+1} = l - k) \\ &= \mathbb{P}(W_0 = w_0, \xi_1 = w_1 - w_0, \dots, \xi_n = k - w_{n-1}) \mathbb{P}(\xi_{n+1} = l - k) \\ &= \mathbb{P}(W_0 = w_0, \dots, W_{n-1} = w_{n-1}, W_n = k) p_{k,l}, \end{aligned}$$

où $p_{k,l}$ est le coefficient donné dans l'énoncé. On en déduit ensuite que

$$\mathbb{P}(W_n = k \text{ et } W_{n+1} = l) = \mathbb{P}(W_n = k) p_{k,l},$$

en sommant sur tous les points $w_0, \dots, w_{n-1} \in \mathbb{Z}^d$, ce qui achève la preuve. $\square$

D'après la proposition 3.8, le graphe associé à une marche aléatoire symétrique $(W_n)_{n \geq 0}$ en dimension d vue comme une chaîne de Markov homogène à espace d'états $\mathbb{Z}^d$ admet $\mathbb{Z}^d$ pour ensemble de sommets, et comporte des arcs de chaque sommet k vers ses voisins $k \pm e_i$. On représente ce graphe sur la figure 3.8 pour le cas de la dimension $d = 2$, auquel cas les probabilités de transition sont toutes égales à $1/4$. Tous les états communiquent, de sorte que la chaîne de Markov est irréductible. D'après la proposition B.20, la chaîne est soit récurrente, soit transiente. Le résultat suivant, dû à Pólya, indique dans quelle situation on se trouve selon la valeur de la dimension d .

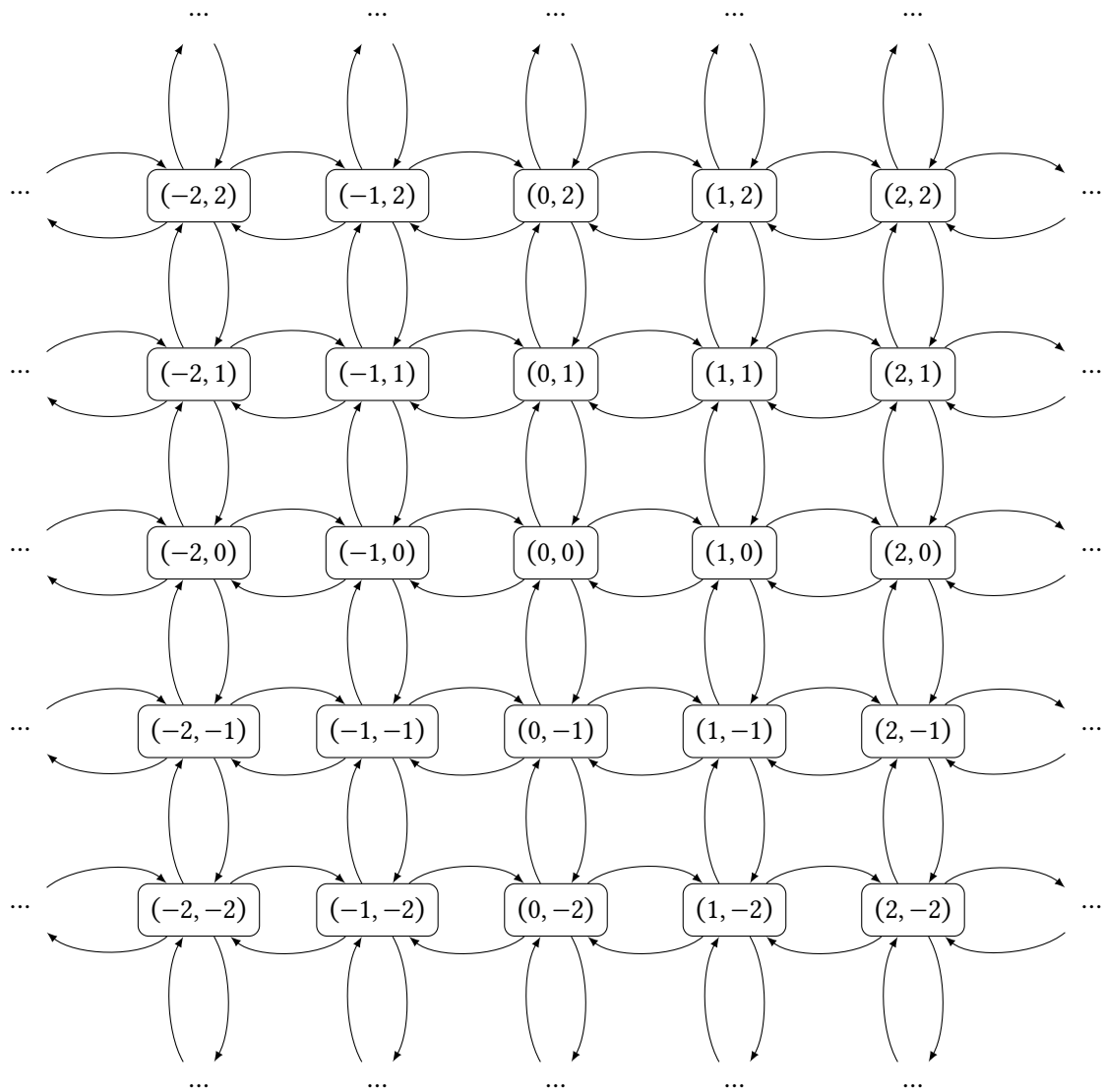


FIGURE 3.8 : Graphe de la chaîne de Markov associée à une marche aléatoire symétrique dans le plan, toutes les probabilités de transition étant égales à $1/4$

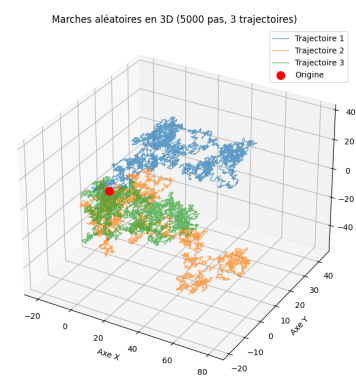
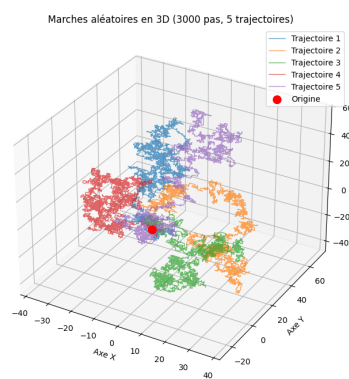


FIGURE 3.9 : Georg Pólya (1887–1985) et des marches aléatoires symétriques en dimension 3

Théorème 3.9 (Récurrence ou transience des marches symétriques en dimension d). *Toute marche aléatoire symétrique en dimension $d \in \{1, 2\}$ est récurrente nulle, tandis que toute marche aléatoire symétrique en dimension $d \geq 3$ est transiente.*

Démonstration. Considérons une marche aléatoire symétrique $(W_n)_{n \geq 0}$ en dimension $d \geq 1$, et une suite $(\xi_m)_{m \geq 1}$ de variables aléatoires associées comme dans la définition 3.3. Tout comme dans la preuve du théorème 3.2, il suffit de déterminer la nature d'un état quelconque $k \in \mathbb{Z}^d$. Notons $p_{k,l}^{(m)}$ et $p_{k,l}^*$ les coefficients des puissances de la matrice de transition de la chaîne, et ceux de sa matrice potentielle. Si $\mathbb{P}(W_0 = k) > 0$, le théorème B.7 montre que

$$p_{k,k}^{(m)} = \mathbb{P}(W_m = k \mid W_0 = k) = \mathbb{P}(\xi_1 + \dots + \xi_m = 0).$$

On utilise alors une technique d'analyse de Fourier. Pour tout point $x \in \mathbb{R}^d$, on a par indépendance

$$\begin{aligned} \varphi_m(x) &= \sum_{k \in \mathbb{Z}^d} e^{2i\pi k \cdot x} \mathbb{P}(\xi_1 + \dots + \xi_m = k) = \sum_{(k', k'') \in \mathbb{Z}^d \times \mathbb{Z}^d} e^{2i\pi(k' + k'') \cdot x} \mathbb{P}(\xi_1 + \dots + \xi_{m-1} = k') \mathbb{P}(\xi_m = k'') \\ &= \sum_{k' \in \mathbb{Z}^d} e^{2i\pi k' \cdot x} \mathbb{P}(\xi_1 + \dots + \xi_{m-1} = k') \sum_{k'' \in \mathbb{Z}^d} e^{2i\pi k'' \cdot x} \mathbb{P}(\xi_m = k'') = \varphi_{m-1}(x) \varphi_1(x). \end{aligned}$$

Toutes ces égalités sont valides car les sommes n'ont qu'un nombre fini de termes non nuls. En effet, presque sûrement, $\|\xi_1 + \dots + \xi_m\|_1 \leq m$. Pour la même raison, on peut échanger somme et intégrale :

$$\begin{aligned} \int_{[0,1]^d} \varphi_m(x) dx &= \int_{[0,1]^d} \sum_{k \in \mathbb{Z}^d} e^{2i\pi k \cdot x} \mathbb{P}(\xi_1 + \dots + \xi_m = k) dx \\ &= \sum_{k \in \mathbb{Z}^d} \mathbb{P}(\xi_1 + \dots + \xi_m = k) \underbrace{\int_{[0,1]^d} e^{2i\pi k \cdot x} dx}_{\mathbf{1}_{\{k=0\}}} = \mathbb{P}(\xi_1 + \dots + \xi_m = 0). \end{aligned}$$

On en déduit que pour tout point $k \in \mathbb{Z}^d$ tel que $\mathbb{P}(W_0 = k) > 0$, pour tout entier $m \in \mathbb{N}$, on a

$$p_{k,k}^{(m)} = \int_{[0,1]^d} (\varphi_1(x))^m dx.$$

Posons $r_n = 1 - 2^{-n}$ pour tout entier $n \in \mathbb{N}$. D'après le théorème A.2, on a

$$p_{k,k}^* = \sum_{m=0}^{\infty} p_{k,k}^{(m)} = \sum_{m=0}^{\infty} \sup_{n \in \mathbb{N}} r_n^m p_{k,k}^{(m)} = \sup_{n \in \mathbb{N}} \sum_{m=0}^{\infty} r_n^m p_{k,k}^{(m)} = \sup_{n \in \mathbb{N}} \sum_{m=0}^{\infty} \int_{[0,1]^d} (r_n \varphi_1(x))^m dx.$$

Les sommes partielles de la série dans le membre de droite vérifient, pour tout entier $M \in \mathbb{N}$,

$$\sum_{m=0}^M \int_{[0,1]^d} (r_n \varphi_1(x))^m dx = \int_{[0,1]^d} \sum_{m=0}^M (r_n \varphi_1(x))^m dx = \int_{[0,1]^d} \frac{1 - (r_n \varphi_1(x))^{M+1}}{1 - r_n \varphi_1(x)} dx.$$

Cependant, l'inégalité triangulaire pour les sommes implique que $|\varphi_1(x)| \leq 1$ pour tout $x \in \mathbb{R}^d$, et celle pour les intégrales permet d'en déduire que

$$\left| \int_{[0,1]^d} \frac{(r_n \varphi_1(x))^{M+1}}{1 - r_n \varphi_1(x)} dx \right| \leq \int_{[0,1]^d} \frac{|(r_n \varphi_1(x))^{M+1}|}{|1 - r_n \varphi_1(x)|} dx \leq \frac{r_n^{M+1}}{1 - r_n} \xrightarrow{M \rightarrow \infty} 0.$$

Il en résulte que

$$p_{k,k}^* = \lim_{n \rightarrow \infty} \int_{[0,1]^d} \frac{dx}{1 - r_n \varphi_1(x)} \quad \text{avec} \quad \varphi_1(x_1, \dots, x_d) = \frac{1}{d} \sum_{j=1}^d \cos(2\pi x_j).$$

Pour justifier le passage à la limite dans l'intégrale, il faut tenir compte du signe de φ_1 . Sur l'ensemble où $\varphi_1 \geq 0$, les intégrandes sont positives et croissantes en n : le théorème de convergence monotone pour les intégrales s'applique. Sur l'ensemble où $\varphi_1 < 0$, on a

$$\left| \frac{1}{1 - r_n \varphi_1(x)} - \frac{1}{1 - \varphi_1(x)} \right| = \frac{(1 - r_n)|\varphi_1(x)|}{(1 - r_n \varphi_1(x))(1 - \varphi_1(x))} \leq 1 - r_n = 2^{-n},$$

ce qui donne une convergence uniforme sur un ensemble de volume au plus 1. En additionnant les deux contributions, on obtient, avec une intégrale éventuellement infinie,

$$p_{k,k}^* = \int_{[0,1]^d} \frac{dx}{1 - \varphi_1(x)} = \int_{[0,1]^d} \frac{d}{\sum_{j=1}^d (1 - \cos(2\pi x_j))} dx. \quad (3.5)$$

Il reste à étudier la convergence de cette intégrale. Sur le cube fermé $[0, 1]^d$, le dénominateur s'annule aux 2^d sommets $v \in \{0, 1\}^d$. Au voisinage de chacun de ces sommets, il est équivalent à $2\pi^2 \|x - v\|_2^2$. Ici, $\|\cdot\|_2$ désigne la norme euclidienne sur $\mathbb{R}^d$. En passant aux coordonnées sphériques, on observe que l'intégrale de la fonction $x \mapsto \|x\|_2^{-2}$ sur la boule euclidienne centrée en 0 et de rayon δ est convergente si et seulement si l'intégrale de $r \mapsto r^{-2} r^{d-1}$ sur $[0, \delta]$ est convergente. Cela est le cas si et seulement si $d > 2$. On en déduit que la marche aléatoire est récurrente si $d \in \{1, 2\}$, et transiente si $d \geq 3$.

Finalement, on rappelle que le théorème 3.2 donne la récurrence nulle pour $d = 1$. Le cas où $d = 2$ se traite de même en observant que la mesure μ telle que $\mu_k = 1$ pour tout $k \in \mathbb{Z}^2$ est invariante pour la chaîne de Markov, qui est irréductible. Tout état $l \in \mathbb{Z}^2$ est récurrent, et la proposition B.24 implique que l'espérance τ_l du temps de retour en l vaut $\sum_{k \in \mathbb{Z}^2} \mu_k = \infty$. La chaîne est donc récurrente nulle aussi pour $d = 2$. $\square$

Algorithme 3.4 (Simulation d'une marche aléatoire symétrique dans l'espace). Dans la même veine que l'algorithme 3.3, on peut simuler une marche aléatoire symétrique issue de l'origine $(W_n)_{n \geq 0}$ en dimension $d = 3$. À chaque pas, la marche se déplace d'une unité dans l'une des six directions possibles (haut, bas, gauche, droite, avant, arrière) avec une probabilité $1/6$. On utilise le script C.11. On simule plusieurs trajectoires indépendantes, avec des couleurs différentes. Les graphiques correspondants, générés pour diverses longueurs des marches, sont représentés en figure 3.9.

Branchement

Le processus de branchement de Galton-Watson est un modèle probabiliste classique utilisé pour décrire l'évolution d'une population où chaque individu donne naissance à un nombre aléatoire d'enfants, selon une loi de probabilité fixée. Introduit par Francis Galton et Henry William Watson pour étudier la disparition des patronymes, ce processus trouve aujourd'hui des applications, notamment en biologie pour modéliser l'évolution d'une population asexuée au fil de générations qui ne se recoupent pas. On associe parfois à ce processus le nom de Bienaymé, qui a étudié un modèle similaire avant Galton et Watson. Ce chapitre explore les fondements théoriques du processus, ses propriétés asymptotiques, et les conditions menant à l'extinction ou à la survie de la population.

4.1 Processus et arbres de Galton-Watson

On rappelle que, suivant la définition B.1, une mesure de probabilité sur l'ensemble $\mathbb{N}$ des entiers naturels est un vecteur ligne $\mu = (\mu_l)_{l \in \mathbb{N}}$ à coefficients dans $[0, 1]$ dont la somme des coefficients $\sum_{l \in \mathbb{N}} \mu_l$ est égale à 1. En accord avec l'exemple 1.1, on dit qu'une variable aléatoire X à valeurs dans $\mathbb{N}$ est *distribuée selon la loi μ* , que X admet pour loi μ , ou encore que μ est la loi de X , si on a $\mathbb{P}(X = l) = \mu_l$ pour tout $l \in \mathbb{N}$. Le processus de Galton-Watson est alors défini comme suit.

Définition 4.1 (Processus de Galton-Watson). Soit μ une mesure de probabilité sur $\mathbb{N}$. On appelle *processus de Galton-Watson de loi de reproduction μ* toute suite $(Z_n)_{n \in \mathbb{N}}$ de variables aléatoires à valeurs dans $\mathbb{N}$ pour laquelle il existe une famille à deux indices $(X_{n,j})_{n,j \geq 1}$ de variables aléatoires indépendantes et identiquement distribuées selon la loi μ , indépendantes de Z_0 , et vérifiant presque sûrement, pour tout $n \in \mathbb{N}$,

$$Z_{n+1} = \sum_{j=1}^{Z_n} X_{n+1,j} \in \mathbb{N}.$$

On interprète la définition 4.1 en considérant que Z_n représente le nombre d'individus d'une population à la génération n . La taille de la population initiale est donnée par la variable aléatoire Z_0 , souvent supposée égale à 1 presque sûrement. Chaque individu de la génération n donne naissance, indépendamment des autres, à un nombre aléatoire d'enfants distribués selon la loi μ . La variable aléatoire $X_{n+1,j}$ représente le nombre d'enfants de l'individu j de la génération n . La taille Z_{n+1} de la population à la génération $n + 1$ est alors la somme des nombres d'enfants de chaque individu de génération n .

Si $Z_0 = 1$, les variables aléatoires $X_{n,j}$ permettent de construire un arbre aléatoire dont la racine est l'individu initial et dont les branches représentent les enfants des individus. Un tel arbre, appelé *arbre de Galton-Watson*, peut être codé par des mots finis sur l'alphabet $\mathbb{N}^*$ en suivant les branches de gauche à droite. La racine de l'arbre, correspondant au seul individu de génération 0, est codée par le mot vide $\emptyset$. Les sommets de profondeur n dans l'arbre représentent les individus de génération n et sont codés par des mots de longueur n . Le nombre de sommets de profondeur n est égal à Z_n . L'arbre de Galton-Watson permet donc de retrouver le processus de Galton-Watson. On donne un exemple d'arbre de Galton-Watson sur la figure 4.1. Si $Z_0 \geq 2$, on obtient une forêt d'arbres de Galton-Watson.

En vue de décrire la loi d'un processus de Galton-Watson, on introduit les puissances de convolution de sa loi de reproduction. Les puissances de convolution d'une mesure sont définies dans l'énoncé suivant, qui se déduit par une récurrence élémentaire de la proposition 1.12.

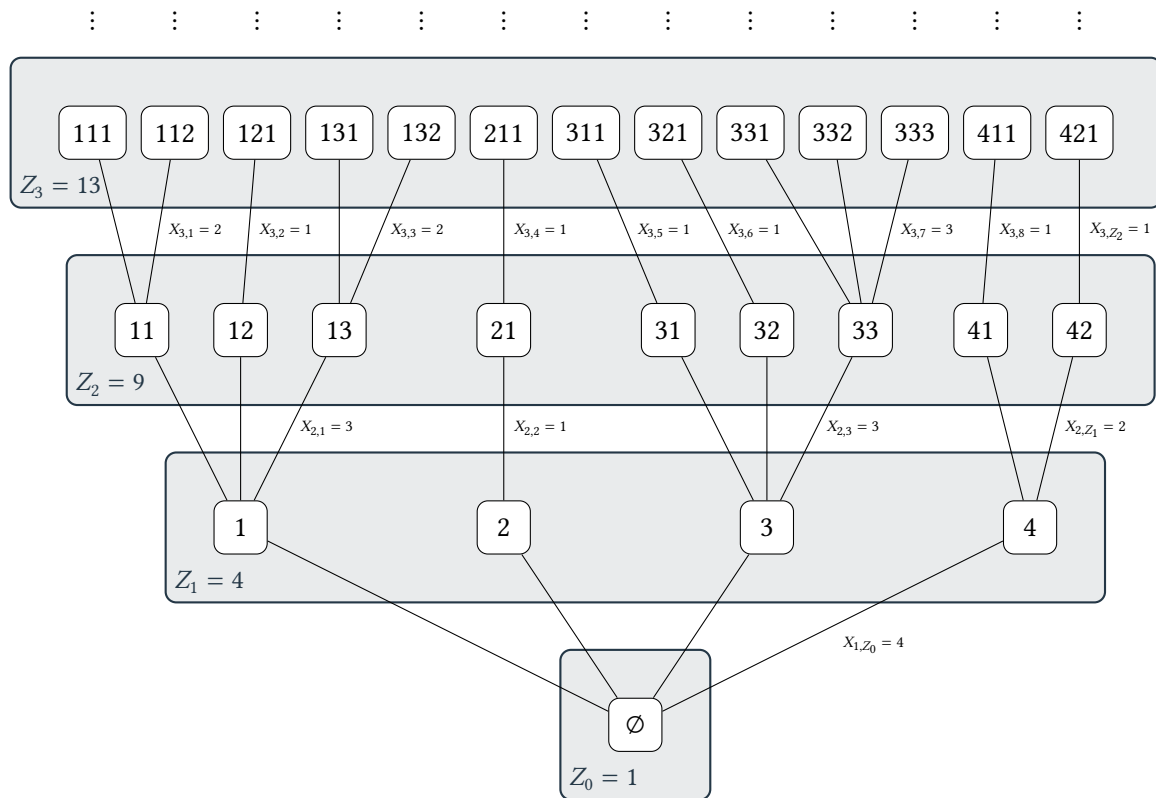


FIGURE 4.1 : Un arbre de Galton-Watson représentant la généalogie d'un processus de Galton-Watson

Proposition 4.1 (Puissances de convolution). Soit μ une mesure de probabilité sur $\mathbb{N}$. On définit de manière récurrente, pour $k \in \mathbb{N}$, les puissances k -ième de convolution de μ par les conditions

$$\mu_l^{*0} = 1_{\{l=0\}} \quad \text{et} \quad \mu^{*(k+1)} = \mu^{*k} * \mu.$$

Si $X_1, \dots, X_k$ désignent k variables aléatoires à valeurs dans $\mathbb{N}$ indépendantes et identiquement distribuées selon la loi μ , alors la variable aléatoire $X_1 + \dots + X_k$ est distribuée selon la loi μ^{*k} .

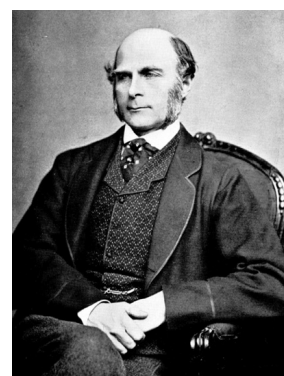


FIGURE 4.2 : Irénée-Jules Bienaymé (1796–1878), R.P. Henry William Watson (1827–1903) et Sir Francis Galton (1822–1911), voir aussi la figure 1.8

Comme le montre le résultat suivant, un processus de Galton-Watson peut être vu comme une chaîne de Markov homogène à espace d'états $\mathbb{N}$. On peut donc, grâce à la proposition B.6, exprimer ses lois en dimension finie à partir de sa loi de reproduction.

Proposition 4.2 (Aspect markovien des processus de Galton-Watson). Soit μ une mesure de probabilité sur $\mathbb{N}$. Tout processus de Galton-Watson $(Z_n)_{n \geq 0}$ de loi de reproduction μ est une chaîne de Markov homogène à espace d'états $\mathbb{N}$, dont les coefficients de la matrice de transition sont donnés, pour $k, l \in \mathbb{N}$, par

$$p_{k,l} = \mu_l^{*k}.$$

Démonstration. Considérons un processus de Galton-Watson $(Z_n)_{n \geq 0}$ de loi de reproduction μ , puis une famille à deux indices $(X_{n,j})_{n,j \geq 1}$ de variables aléatoires comme dans la définition 4.1. Pour tous entiers $n \in \mathbb{N}$ et $z_0, \dots, z_{n-1}, k, l \in \mathbb{N}$, en observant que le vecteur $(Z_0, \dots, Z_n)$ est indépendant de la famille $(X_{n+1,j})_{j \geq 1}$, on peut écrire que

$$\begin{aligned} \mathbb{P}(Z_0 = z_0, \dots, Z_{n-1} = z_{n-1}, Z_n = k, Z_{n+1} = l) &= \mathbb{P}\left(Z_0 = z_0, \dots, Z_{n-1} = z_{n-1}, Z_n = k, \sum_{j=1}^k X_{n+1,j} = l\right) \\ &= \mathbb{P}(Z_0 = z_0, \dots, Z_{n-1} = z_{n-1}, Z_n = k) \mathbb{P}\left(\sum_{j=1}^k X_{n+1,j} = l\right) \\ &= \mathbb{P}(Z_0 = z_0, \dots, Z_{n-1} = z_{n-1}, Z_n = k) \mu_l^{*k}. \end{aligned}$$

La dernière égalité découle de la proposition 4.1. On en déduit ensuite que

$$\mathbb{P}(Z_n = k \text{ et } Z_{n+1} = l) = \mathbb{P}(Z_n = k) \mu_l^{*k},$$

en sommant sur tous les entiers $z_0, \dots, z_{n-1} \in \mathbb{N}$, ce qui achève la preuve. $\square$

D'après la proposition 4.2, le graphe associé à un processus de Galton-Watson $(Z_n)_{n \geq 0}$ de loi de reproduction μ , vu comme une chaîne de Markov homogène à espace d'états $\mathbb{N}$, admet $\mathbb{N}$ pour ensemble de sommets, et des arêtes orientées de chaque sommet $k \in \mathbb{N}$ vers les sommets $l \in \mathbb{N}$ tels que $\mu_l^{*k} > 0$. L'aspect du graphe dépend fortement de la loi de reproduction μ et peut vite devenir complexe. On se contente de donner sur la figure 4.3 l'exemple du processus de Galton-Watson dont la loi de reproduction μ est la loi de Bernoulli $\text{Ber}(p)$ de paramètre $p \in]0, 1[$, c'est-à-dire que

$$\mu_l = \begin{cases} 1-p & \text{si } l = 0 \\ p & \text{si } l = 1 \\ 0 & \text{sinon,} \end{cases}$$

auquel cas la puissance m -ième de convolution μ^{*m} est la loi binomiale $\text{Bin}(m, p)$.

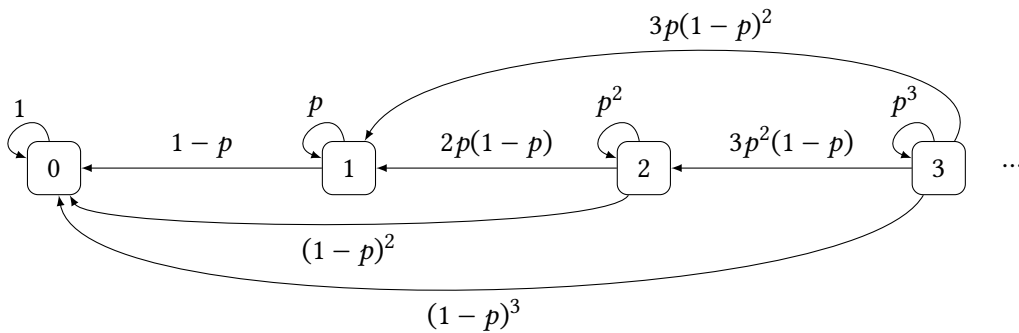


FIGURE 4.3 : Graphe de la chaîne de Markov associée à un processus de Galton-Watson de loi de reproduction $\text{Ber}(p)$

Cependant, il est toujours vrai que l'état 0, qui correspond à une *extinction* de la population, est absorbant. Cela découle de l'observation que $\{Z_n = 0\} \subseteq \{Z_{n+1} = 0\}$, ou de la proposition 4.2 puisque $\mu_0^{*0} = 1$. Conformément à (B.3), on définit le premier instant de visite de l'état 0 par la formule

$$T = \inf\{n \geq 0 \mid Z_n = 0\}, \quad (4.1)$$

avec la convention usuelle que l'infimum de l'ensemble vide vaut ∞ . On appelle T le *temps d'extinction* du processus de Galton-Watson. Sur l'événement $\{1 \leq T < \infty\}$, la dernière génération non vide est la génération $T - 1$. Si $T = 0$, la population est vide dès l'origine ; si $T = \infty$, chaque génération est non vide. La taille totale de la population est donnée par

$$N = \sum_{n=0}^{\infty} Z_n = \sum_{n=0}^{T-1} Z_n. \quad (4.2)$$

Une question principale dans l'étude des processus de Galton-Watson est de déterminer les conditions sur la loi de reproduction μ pour que le processus s'éteigne presque sûrement, c'est-à-dire pour que l'état 0 soit presque sûrement atteint, et que la *probabilité d'extinction* soit égale à 1. Celle-ci peut s'exprimer, à l'aide du lemme B.17 et de ses notations, des diverses manières suivantes :

$$\phi = \mathbb{P}(T < \infty) = \mathbb{P}(\exists n \geq 0 : Z_n = 0) = \mathbb{P}\left(\bigcup_{n=0}^{\infty} \uparrow \{Z_n = 0\}\right) = \lim_{n \rightarrow \infty} \uparrow \mathbb{P}(Z_n = 0) = \sup_{n \in \mathbb{N}} \mathbb{P}(Z_n = 0). \quad (4.3)$$

Remarque 4.1 (Propriété de branchement). On peut montrer qu'un processus de Galton-Watson $(Z_n)_{n \geq 0}$ de loi de reproduction μ et issu de $Z_0 = k$ individus équivaut à la somme de k processus de Galton-Watson indépendants, chacun ayant la même loi de reproduction μ et démarrant d'un seul individu. Pour avoir l'intuition de cette propriété, on peut imaginer que l'on regroupe tous les descendants d'un même individu initial dans la population. Étant donné que les nombres d'enfants des individus sont des variables aléatoires indépendantes et identiquement distribuées, les nombres de descendants des individus initiaux le sont également. En d'autres termes, dans la représentation de la généalogie du processus par une forêt d'arbres de Galton-Watson, les arbres issus de chaque individu initial évoluent indépendamment les uns des autres. Cette caractéristique du processus de Galton-Watson s'appelle la *propriété de branchement*. Pour cette raison, on peut voir tout processus de Galton-Watson comme un mélange de sommes de processus indépendants et vérifiant $Z_0 = 1$.

4.2 Quelques lois de reproduction élémentaires

Au vu de la remarque 4.1, on suppose dorénavant que les processus de Galton-Watson $(Z_n)_{n \geq 0}$ considérés sont issus d'un individu unique, donc vérifiant $Z_0 = 1$ presque sûrement. On commence par traiter plusieurs cas où le comportement limite d'un processus de Galton-Watson, en particulier ses propriétés d'extinction ou de survie, sont très faciles à déterminer.

Proposition 4.3 (Lois de reproduction élémentaires). Soit $(Z_n)_{n \geq 0}$ un processus de Galton-Watson de loi de reproduction $\mu = (\mu_l)_{l \in \mathbb{N}}$ et vérifiant $Z_0 = 1$ presque sûrement. On a les propriétés suivantes :

- (i). Si $\mu_0 = 1$, alors presque sûrement $Z_n = 0$ pour tout $n \in \mathbb{N}^*$ et $T = 1$;
- (ii). Si $\mu_l = 1$ pour un entier $l \in \mathbb{N}^*$, alors presque sûrement $Z_n = l^n$ pour tout $n \in \mathbb{N}$, et $T = \infty$;
- (iii). Si $\mu_0 = 0$ et $\mu_1 < 1$, alors presque sûrement $Z_n \xrightarrow[n \rightarrow \infty]{} \infty$ de façon monotone, et $T = \infty$;
- (iv). Si $\mu_0 = 1 - \mu_1 > 0$, alors presque sûrement $Z_n \xrightarrow[n \rightarrow \infty]{} 0$ de façon monotone, et T suit la loi géométrique $\text{Géo}_{\mathbb{N}^*}(\mu_0)$.

Démonstration. Si $\mu_0 = 1$, la proposition 4.2 implique que $\mathbb{P}(Z_1 = 0) = \mu_0 = 1$, donc $T = 1$ presque sûrement, et le processus $(Z_n)_{n \geq 0}$ vaut 0 à partir du rang 1, puisque 0 est un état absorbant.

Supposons que $\mu_l = 1$ pour un entier $l \in \mathbb{N}^*$, et notons $X_{n,j}$ les variables aléatoires introduites dans la définition 4.1. On a presque sûrement $X_{n,j} = l$ pour tous $n, j \geq 1$, donc

$$Z_{n+1} = \sum_{j=1}^{Z_n} X_{n+1,j} = l Z_n.$$

Une récurrence élémentaire implique que presque sûrement, $Z_n = l^n \geq 1$ pour tout $n \in \mathbb{N}$, et donc $T = \infty$. Le processus $(Z_n)_{n \geq 0}$ est croissant et diverge vers l'infini si $l > 1$, ou reste égal à 1 si $l = 1$.

Si maintenant $\mu_0 = 0$ et $\mu_1 < 1$, on a presque sûrement $X_{n,j} \geq 1 + \mathbf{1}_{\{X_{n,j} \geq 2\}}$ pour tous $n, j \geq 1$, donc

$$Z_{n+1} = \sum_{j=1}^{Z_n} X_{n+1,j} \geq Z_n + \sum_{j=1}^{Z_n} \mathbf{1}_{\{X_{n+1,j} \geq 2\}} \geq Z_n + \mathbf{1}_{\{X_{n+1,1} \geq 2\}} \mathbf{1}_{\{Z_n \geq 1\}}.$$

Ainsi, presque sûrement, le processus $(Z_n)_{n \geq 0}$ est croissant, et ne prend donc que des valeurs supérieures à $Z_0 = 1$. En particulier, $Z_n \neq 0$ pour tout $n \geq 0$, ce qui montre que $T = \infty$. De plus,

$$\lim_{n \rightarrow \infty} \uparrow Z_n \geq 1 + \sum_{n=0}^{\infty} \mathbf{1}_{\{X_{n+1,1} \geq 2\}} \quad \text{avec} \quad \mathbb{P}(X_{n+1,1} \geq 2) = 1 - \mu_1 > 0.$$

Les événements $\{X_{n+1,1} \geq 2\}$ sont indépendants et la série $\sum_n \mathbb{P}(X_{n+1,1} \geq 2)$ diverge. Le lemme 1.5 implique que, presque sûrement, pour une infinité d'indices n , l'événement $\{X_{n+1,1} \geq 2\}$ est réalisé, c'est-à-dire que $\mathbf{1}_{\{X_{n+1,1} \geq 2\}} = 1$. Ainsi, la série ci-dessus est infinie presque sûrement.

Enfin, si $\mu_0 = 1 - \mu_1 > 0$, on a alors presque sûrement $X_{n,j} \in \{0, 1\}$ pour tous $n, j \geq 1$, donc

$$Z_{n+1} = \sum_{j=1}^{Z_n} X_{n+1,j} \leq Z_n.$$

Le processus $(Z_n)_{n \geq 0}$ est donc presque sûrement décroissant. De plus, d'après les propositions 4.2 et B.6, on a pour tout entier $n \geq 0$,

$$\mathbb{P}(T > n) = \mathbb{P}(Z_0 = Z_1 = \dots = Z_n = 1) = \mu_1^n.$$

Cela implique que $T \geq 1$ presque sûrement et que pour tout entier $n \geq 1$,

$$\mathbb{P}(T = n) = \mathbb{P}(T > n-1) - \mathbb{P}(T > n) = \mu_1^{n-1} - \mu_1^n = (1 - \mu_0)^{n-1} \mu_0.$$

On en conclut que le temps d'extinction T suit la loi géométrique $\text{Géo}_{\mathbb{N}^*}(\mu_0)$. En particulier, on a $T < \infty$ avec probabilité 1, ce qui montre que Z_n est égal à 0 à partir d'un certain rang. $\square$

4.3 Fonctions génératrices, espérances et variances

Afin de décrire la loi de la taille d'un processus de Galton-Watson à un instant donné, on va utiliser comme outil les fonctions génératrices abordées dans l'annexe A. On définit pour cela, en complément de (A.1), la fonction génératrice d'une mesure de probabilité μ sur $\mathbb{N}$ par la formule

$$G_\mu(z) = \sum_{l=0}^{\infty} \mu_l z^l. \quad (4.4)$$

Si X est une variable aléatoire à valeurs dans $\mathbb{N}$ distribuée selon la loi μ , alors la fonction génératrice de X vérifie clairement $G_X = G_\mu$. De plus, l'espérance et la variance de X sont respectivement égales à

$$m_\mu = \sum_{l=0}^{\infty} l \mu_l \quad \text{et} \quad \sigma_\mu^2 = \sum_{l=0}^{\infty} (l - m_\mu)^2 \mu_l. \quad (4.5)$$

Ces deux quantités sont dans l'intervalle $[0, \infty]$, et la seconde est correctement définie seulement quand la première est finie. On appelle m_μ la *moyenne* de la mesure de probabilité μ et σ_μ^2 sa *variance*. On observe alors, à l'aide de la proposition A.3, que

$$G_\mu(z) \xrightarrow[z < 1]{z \rightarrow 1} 1, \quad G'_\mu(z) \xrightarrow[z < 1]{z \rightarrow 1} m_\mu, \quad G''_\mu(z) \xrightarrow[z < 1]{z \rightarrow 1} \sigma_\mu^2 + m_\mu^2 - m_\mu, \quad (4.6)$$

la dernière convergence ayant lieu si m_μ est fini. La première convergence est monotone car la fonction G_μ est croissante sur les réels positifs.

De plus, le produit de convolution, défini dans la proposition 1.12, de la mesure de probabilité μ avec une autre mesure de probabilité sur $\mathbb{N}$ vérifie la propriété suivante.

Lemme 4.4 (Fonction génératrice et produit de convolution). Soient μ et ν deux mesures de probabilité sur $\mathbb{N}$. Alors, leur produit de convolution $\mu * \nu$ est une mesure de probabilité sur $\mathbb{N}$ dont la fonction génératrice vérifie la relation

$$G_{\mu * \nu}(z) = G_\mu(z) G_\nu(z),$$

pour tout nombre complexe z tel que les séries définissant $G_\mu(z)$ et $G_\nu(z)$ comme dans (4.4) convergent absolument, ainsi que pour tout nombre réel $z \geq 0$ auquel cas les sommes peuvent éventuellement prendre la valeur ∞ .

Démonstration. Le fait que $\mu * \nu$ soit une mesure de probabilité sur $\mathbb{N}$ découle de la proposition 1.12. Considérons deux variables aléatoires X et Y à valeurs dans $\mathbb{N}$, indépendantes et distribuées respectivement selon les lois μ et ν . À nouveau d'après la proposition 1.12, la variable aléatoire $X + Y$ est distribuée selon la loi $\mu * \nu$. La proposition A.5 implique alors que

$$G_{\mu * \nu}(z) = G_{X+Y}(z) = G_X(z) G_Y(z) = G_\mu(z) G_\nu(z),$$

pour tout nombre z vérifiant les conditions de l'énoncé. $\square$

Pour décrire la loi de la taille de la population d'un processus de Galton-Watson $(Z_n)_{n \geq 0}$ à un instant $n \geq 0$ donné, on définit comme en (A.1) sa fonction génératrice

$$G_{Z_n}(z) = \mathbb{E}[z^{Z_n}] = \sum_{l=0}^{\infty} \mathbb{P}(Z_n = l) z^l.$$

Au vu de la remarque 4.1, on suppose toujours que $Z_0 = 1$ presque sûrement. On a la relation suivante entre les fonctions génératrices du processus et celle de sa loi de reproduction.

Proposition 4.5 (Fonctions génératrices d'un processus de Galton-Watson). Soit $(Z_n)_{n \geq 0}$ un processus de Galton-Watson vérifiant $Z_0 = 1$ presque sûrement, et de loi de reproduction μ . Les fonctions génératrices G_{Z_n} vérifient, pour tous $n \in \mathbb{N}$ et $z \geq 0$, la relation

$$G_{Z_{n+1}}(z) = G_{Z_n}(G_\mu(z)),$$

où G_μ est la fonction génératrice de la loi de reproduction μ . Lorsque la fonction intérieure vaut ∞ , on pose $G_Y(\infty) = \lim_{x \rightarrow \infty} G_Y(x) \in [0, \infty]$; dans les sommes de termes positifs, on adopte les conventions $\infty^0 = 1$ et $0 \cdot \infty = 0$. Ainsi, pour tous $n \in \mathbb{N}^*$ et $z \geq 0$, on a

$$G_{Z_n}(z) = \underbrace{G_\mu \circ \dots \circ G_\mu}_{n \text{ fois}}(z).$$

Démonstration. En distinguant suivant la valeur de la variable aléatoire Z_n , on peut écrire que

$$\begin{aligned} G_{Z_{n+1}}(z) &= \sum_{l=0}^{\infty} \sum_{k=0}^{\infty} \mathbb{P}(Z_{n+1} = l \text{ et } Z_n = k) z^l = \sum_{l=0}^{\infty} \sum_{k=0}^{\infty} \mathbb{P}(Z_n = k) \mu_l^{*k} z^l = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \mathbb{P}(Z_n = k) \mu_l^{*k} z^l \\ &= \sum_{k=0}^{\infty} \mathbb{P}(Z_n = k) \sum_{l=0}^{\infty} \mu_l^{*k} z^l = \sum_{k=0}^{\infty} \mathbb{P}(Z_n = k) G_{\mu^{*k}}(z) = \sum_{k=0}^{\infty} \mathbb{P}(Z_n = k) G_\mu(z)^k = G_{Z_n}(G_\mu(z)). \end{aligned}$$

On a utilisé la proposition 4.2 et le théorème B.3, ainsi que le lemme 4.4 qui permet de prouver par une récurrence élémentaire que la fonction génératrice de la puissance k -ième de convolution μ^{*k} est celle de μ à la puissance k . On obtient donc la relation de récurrence voulue, et on conclut aisément. $\square$

En dérivant la relation de récurrence de la proposition 4.5 sur les fonctions génératrices du processus de Galton-Watson, on peut en déduire son espérance et sa variance à chaque instant.

Corollaire 4.6 (Espérances et variances d'un processus de Galton-Watson). Soit $(Z_n)_{n \geq 0}$ un processus de Galton-Watson vérifiant $Z_0 = 1$ presque sûrement, et de loi de reproduction μ , dont la moyenne m_μ et la variance σ_μ^2 sont définies par (4.5). Si $m_\mu < \infty$, alors on a pour tout $n \in \mathbb{N}$ les propriétés suivantes :

- (i). L'espérance de la variable aléatoire Z_n est donnée par $\mathbb{E}[Z_n] = m_\mu^n$;
- (ii). Si de plus $\sigma_\mu^2 < \infty$, la variance de la variable aléatoire Z_n est donnée par

$$\text{Var}(Z_n) = \begin{cases} \sigma_\mu^2 m_\mu^{n-1} \frac{m_\mu^n - 1}{m_\mu - 1} & \text{si } m_\mu > 0, m_\mu \neq 1 \\ \sigma_\mu^2 n & \text{si } m_\mu = 1 \\ 0 & \text{si } m_\mu = 0. \end{cases}$$

Démonstration. Si $m_\mu = 0$, alors $\mu_0 = 1$: $Z_0 = 1$ et $Z_n = 0$ presque sûrement pour $n \geq 1$. Les formules sont immédiates, avec la convention $0^0 = 1$. Supposons désormais $m_\mu > 0$. Pour $0 < z < 1$, on a alors $0 < G_\mu(z) < 1$, ce qui permet de dériver la composition à l'intérieur des intervalles de convergence. En dérivant la relation de récurrence de la proposition 4.5, on a pour $n \in \mathbb{N}$ et $0 < z < 1$,

$$G'_{Z_{n+1}}(z) = G'_\mu(z) G'_{Z_n}(G_\mu(z)).$$

En faisant tendre z vers 1 par la gauche, à l'aide de la proposition A.3 et des limites (4.6), on obtient

$$\mathbb{E}[Z_{n+1}] = m_\mu \mathbb{E}[Z_n].$$

En partant de $\mathbb{E}[Z_0] = 1$, cette relation établit par récurrence la finitude des espérances et la formule $\mathbb{E}[Z_n] = m_\mu^n$. De plus, en dérivant une seconde fois, on a pour $n \in \mathbb{N}$ et $0 < z < 1$,

$$G''_{Z_{n+1}}(z) = G''_\mu(z) G'_{Z_n}(G_\mu(z)) + G'_\mu(z)^2 G''_{Z_n}(G_\mu(z)).$$

En supposant $\sigma_\mu^2 < \infty$, et en faisant à nouveau tendre z vers 1 par la gauche à l'aide de la proposition A.3 et des limites (4.6), on obtient maintenant

$$\mathbb{E}[Z_{n+1}(Z_{n+1} - 1)] = (\sigma_\mu^2 + m_\mu^2 - m_\mu) \mathbb{E}[Z_n] + m_\mu^2 \mathbb{E}[Z_n(Z_n - 1)].$$

Cette relation établit aussi par récurrence la finitude des moments d'ordre deux, puisque $Z_0 = 1$. Après simplifications et en utilisant l'expression trouvée précédemment pour l'espérance, cela donne

$$\text{Var}(Z_{n+1}) = m_\mu^2 \text{Var}(Z_n) + \sigma_\mu^2 m_\mu^n.$$

En divisant par $m_\mu^{2(n+1)}$, on peut former une identité qui fait intervenir une somme télescopique :

$$m_\mu^{-2n} \text{Var}(Z_n) = \sum_{j=0}^{n-1} \left(m_\mu^{-2(j+1)} \text{Var}(Z_{j+1}) - m_\mu^{-2j} \text{Var}(Z_j) \right) = \sigma_\mu^2 \sum_{j=0}^{n-1} m_\mu^{-j-2}.$$

On conclut avec la formule usuelle sur la somme des termes initiaux d'une suite géométrique. $\square$

Le corollaire 4.6 montre que la moyenne de la loi de reproduction du processus de Galton-Watson est un paramètre crucial pour déterminer l'évolution de la taille de la population. On définit plusieurs régimes pour le processus, selon la position de cette moyenne par rapport à la valeur 1.

Définition 4.2 (Régime d'un processus de Galton-Watson). Un processus de Galton-Watson de loi de reproduction μ dont la moyenne m_μ est définie par (4.5) est qualifié de *sous-critique* si $m_\mu < 1$, *critique* si $m_\mu = 1$, et *sur-critique* si $m_\mu > 1$.

Soit $(Z_n)_{n \geq 0}$ un processus de Galton-Watson tel que $Z_0 = 1$ presque sûrement, et dont la loi de reproduction μ a une moyenne m_μ et une variance σ_μ^2 finies. On déduit du corollaire 4.6 que, dans le cas sous-critique, les espérances et les variances convergent vers 0, c'est-à-dire que

$$\mathbb{E}[Z_n] \xrightarrow{n \rightarrow \infty} 0 \quad \text{et} \quad \text{Var}(Z_n) \xrightarrow{n \rightarrow \infty} 0.$$

La première convergence est de plus monotone. Si le processus est critique, il a en revanche une espérance constante égale à 1, et une variance qui croît linéairement comme $\sigma_\mu^2 n$. Enfin, si le processus est sur-critique et si $\sigma_\mu^2 > 0$, espérances et variances croissent et divergent vers ∞ , ce qui signifie que

$$\mathbb{E}[Z_n] \xrightarrow{n \rightarrow \infty} \infty \quad \text{et} \quad \text{Var}(Z_n) \xrightarrow{n \rightarrow \infty} \infty.$$

Si $\sigma_\mu^2 = 0$, la divergence de l'espérance persiste, mais la variance reste constante, donc le processus est déterministe. Ce cas, obtenu quand $\mu_l = 1$ pour un entier $l \in \mathbb{N}^*$, est traité dans la proposition 4.3.

Algorithme 4.1 (Simulation d'un processus de Galton-Watson). Afin de percevoir les différences de comportement d'un processus de Galton-Watson selon son régime, on simule r réalisations d'un tel processus, issu d'un seul individu et de loi de reproduction donnée par un vecteur de probabilité, puis on trace le nombre d'individus Z_n en fonction de n . On utilise pour cela le script C.12. On observe effectivement des comportements différents selon que le régime du processus est sous-critique, critique, ou sur-critique. On obtient les graphiques de la figure 4.4.

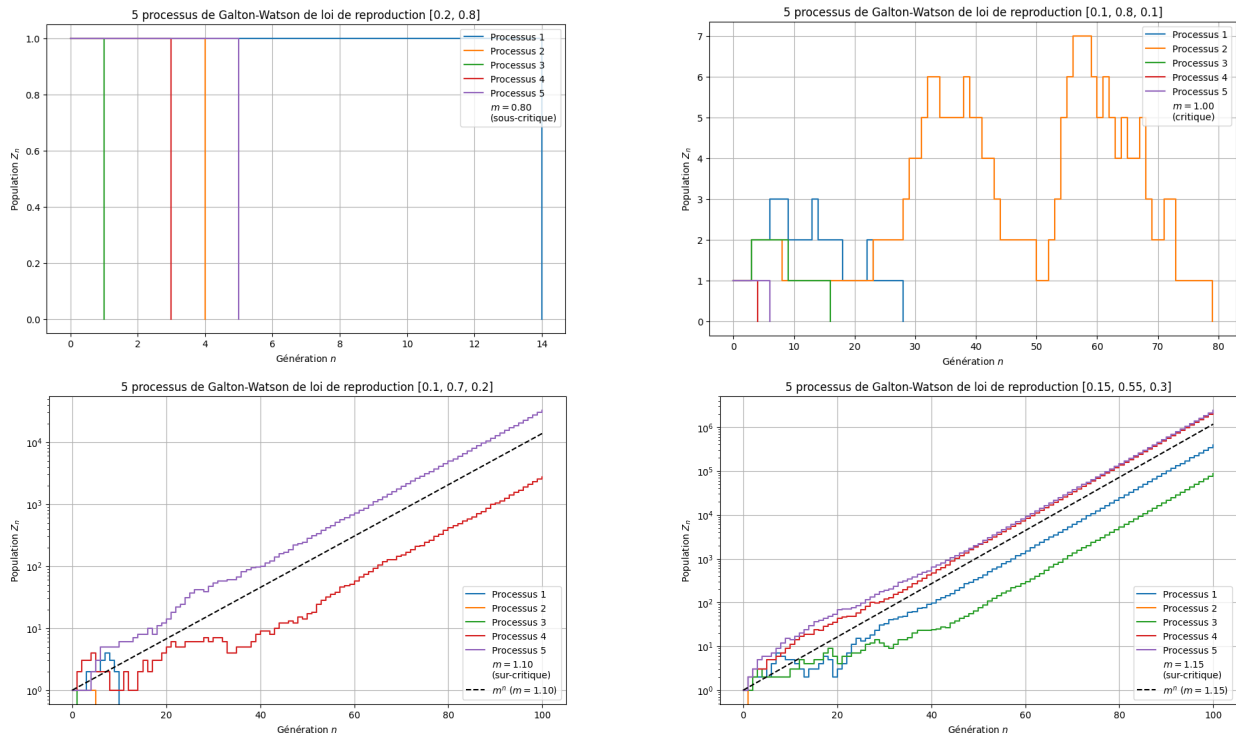


FIGURE 4.4 : Réalisations d'un processus de Galton-Watson issu d'un seul individu

4.4 Comportement à l'infini

On suppose dorénavant que les processus de Galton-Watson $(Z_n)_{n \geq 0}$ considérés ont une taille de population initiale et une loi de reproduction qui vérifient les conditions

$$Z_0 = 1 \quad \text{et} \quad 0 < \mu_0 \leq \mu_0 + \mu_1 < 1. \quad (4.7)$$

Cela implique en particulier que la taille de la population Z_1 à l'instant 1 suit la loi μ , et que cette mesure de probabilité vérifie $\mu_l < 1$ pour tout entier $l \in \mathbb{N}$. On peut faire l'hypothèse (4.7) au vu de la remarque 4.1 et de la proposition 4.3, qui évacue tous les cas où elle n'est pas vérifiée. Dans ce cas, un processus de Galton-Watson ne peut avoir que les deux comportements limites suivants.

Théorème 4.7 (Extinction ou explosion d'un processus de Galton-Watson). Soit $(Z_n)_{n \geq 0}$ un processus de Galton-Watson de loi de reproduction μ et vérifiant (4.7). On définit par (4.1) le temps d'extinction T du processus. Presque sûrement, on a l'un des deux cas suivants :

- (i). Le processus s'éteint : $T < \infty$ et $Z_n \xrightarrow{n \rightarrow \infty} 0$;
- (ii). Le processus explose : $T = \infty$ et $Z_n \xrightarrow{n \rightarrow \infty} \infty$.

Démonstration. Soit $(Y_n)_{n \geq 0}$ un processus de Galton-Watson de même loi de reproduction μ que le processus $(Z_n)_{n \geq 0}$, et tel que $\mathbb{P}(Y_0 = k) > 0$ pour tout $k \in \mathbb{N}$. Ces deux processus sont des chaînes de Markov homogènes à espace d'états $\mathbb{N}$ et de même matrice de transition, dont les coefficients sont donnés par la proposition 4.2. Définissons par (B.3) les instants de visite $T_1^{(k)}$ d'un état $k \in \mathbb{N}^*$ par la chaîne $(Y_n)_{n \geq 0}$. Pour tout $n \in \mathbb{N}$, on a $\{Y_n = 0\} \subseteq \{Y_{n+1} = 0\}$, donc $\{Y_1 = 0\} = \bigcap_{n=1}^{\infty} \{Y_n = 0\}$. Ainsi,

$$\begin{aligned} \mathbb{P}(T_1^{(k)} = \infty \text{ et } Y_0 = k) &\geq \mathbb{P}\left(\{Y_0 = k\} \cap \bigcap_{n=1}^{\infty} \{Y_n = 0\}\right) = \mathbb{P}(Y_1 = 0 \text{ et } Y_0 = k) \\ &= \mathbb{P}(Y_0 = k) \mu_0^{*k} = \mathbb{P}(Y_0 = k) \mu_0^k. \end{aligned}$$

En divisant par $\mathbb{P}(Y_0 = k) > 0$ et en utilisant le lemme B.11, le coefficient ρ_k défini par (B.7) vérifie

$$\rho_k = \mathbb{P}(T_1^{(k)} < \infty \mid Y_0 = k) \leq 1 - \mu_0^k.$$

La condition (4.7) garantit que $\mu_0 > 0$, de sorte que le majorant ci-dessus est strictement inférieur à 1. La proposition B.19 implique donc que l'état k est transient pour la chaîne de Markov $(Y_n)_{n \geq 0}$, donc le coefficient $p_{k,k}^*$ de la matrice potentielle associée à la matrice de transition de cette chaîne est fini. Le corollaire B.18 implique que le coefficient $p_{1,k}^*$ est fini aussi. En définissant par (B.4) le nombre de visites N_k de l'état k par la chaîne $(Z_n)_{n \geq 0}$, on déduit de la proposition B.15 que $\mathbb{E}[N_k] = p_{1,k}^* < \infty$. Ainsi, presque sûrement, pour tout $k \in \mathbb{N}^*$, le processus $(Z_n)_{n \geq 0}$ prend la valeur k un nombre fini de fois. Par conséquent, presque sûrement, pour tout entier $k \in \mathbb{N}^*$, on a

$$\#\{n \in \mathbb{N} \mid 1 \leq Z_n \leq k\} < \infty.$$

Finalement, si $T < \infty$, on a $Z_n = 0$ pour $n \geq T$, de sorte que $Z_n \xrightarrow{n \rightarrow \infty} 0$. Sinon, si $T = \infty$, alors pour tout entier $k \in \mathbb{N}^*$, on a $Z_n \leq k$ seulement pour un nombre fini d'indices n , donc $Z_n \xrightarrow{n \rightarrow \infty} \infty$. $\square$

On s'intéresse maintenant à la probabilité avec laquelle on est dans le cas du théorème 4.7 où le processus de Galton-Watson s'éteint. Il s'agit de la probabilité d'extinction ϕ définie par (4.3).

Théorème 4.8 (Probabilité d'extinction d'un processus de Galton-Watson). Soit $(Z_n)_{n \geq 0}$ un processus de Galton-Watson de loi de reproduction μ et vérifiant (4.7). On note ϕ la probabilité d'extinction du processus, définie par (4.3), et on note G_μ la fonction génératrice de la mesure de probabilité μ , définie par (4.4). On a les propriétés suivantes :

- (i). La probabilité d'extinction ϕ est strictement positive et c'est le plus petit point fixe de la fonction génératrice G_μ dans l'intervalle $[0, 1]$, c'est-à-dire que

$$\phi = \min\{z \in [0, 1] \mid G_\mu(z) = z\} > 0.$$

- (ii). Si le processus est sous-critique ou critique, alors G_μ admet 1 pour unique point fixe dans l'intervalle $[0, 1]$, de sorte que $\phi = 1$;
- (iii). Si le processus est sur-critique, alors G_μ admet deux points fixes distincts dans l'intervalle $[0, 1]$, qui sont 1 et $\phi \in]0, 1[$.

Démonstration. Les égalités (4.3) et la proposition 4.5 permettent d'écrire que

$$\phi = \mathbb{P}(T < \infty) = \lim_{n \rightarrow \infty} \mathbb{P}(Z_n = 0) = \lim_{n \rightarrow \infty} \uparrow G_{Z_n}(0) = \lim_{n \rightarrow \infty} \uparrow G_\mu(G_{Z_{n-1}}(0)) = G_\mu\left(\lim_{n \rightarrow \infty} \uparrow G_{Z_{n-1}}(0)\right) = G_\mu(\phi).$$

On a aussi utilisé la continuité et la croissance de la fonction génératrice G_μ sur l'intervalle $[0, 1]$. Ainsi, la probabilité d'extinction ϕ est bien un point fixe de G_μ dans l'intervalle $[0, 1]$. De plus, si $z \in [0, 1]$ est un point fixe de la fonction G_μ , une récurrence élémentaire montre que pour tout entier $n \in \mathbb{N}^*$,

$$\mathbb{P}(Z_n = 0) = G_{Z_n}(0) = \underbrace{G_\mu \circ \dots \circ G_\mu}_{n \text{ fois}}(0) \leq z.$$

En faisant tendre n vers l'infini, on obtient $\phi \leq z$. Ainsi, ϕ est le plus petit point fixe de G_μ dans l'intervalle $[0, 1]$. Comme $G_\mu(0) = \mu_0 > 0$ d'après (4.7), alors 0 n'est pas un point fixe, donc $\phi > 0$.

Le reste de la preuve se fait avec des arguments de convexité illustrés par la figure 4.5, et dont la rédaction nécessite quelques détails. Comme la fonction génératrice G_μ est une série entière de rayon de convergence au moins égal à 1, elle est indéfiniment dérivable sur l'intervalle $]-1, 1[$, et ses dérivées successives s'obtiennent en dérivant terme à terme l'expression (4.4). Ainsi, sa dérivée seconde est

$$G''_\mu(z) = \sum_{l=2}^{\infty} l(l-1)\mu_l z^{l-2} = \sum_{l=0}^{\infty} (l+2)(l+1)\mu_{l+2} z^l.$$

Les coefficients μ_{l+2} sont tous positifs, et au moins l'un d'eux, disons μ_{l_0+2} , est strictement positif d'après la condition (4.7). La fonction G''_μ est donc positive sur l'intervalle $[0, 1]$. Plus précisément, pour $0 \leq z < z_0 < 1$, on peut écrire que

$$(l_0 + 2)(l_0 + 1)\mu_{l_0+2} \int_z^{z_0} \int_t^{z_0} u^{l_0} du dt \leq \int_z^{z_0} \int_t^{z_0} G''_\mu(u) du dt = G'_\mu(z_0)(z_0 - z) - G_\mu(z_0) + G_\mu(z).$$

En faisant tendre z_0 vers 1 par la gauche et en utilisant les limites (4.6), on obtient

$$(l_0 + 2)(l_0 + 1)\mu_{l_0+2} \int_z^1 \int_t^1 u^{l_0} du dt \leq m_\mu(1 - z) - 1 + G_\mu(z),$$

où m_μ est la moyenne de la loi de reproduction μ , définie par (4.5). Si le processus est sous-critique ou critique, on a $m_\mu \leq 1$, donc pour tout $z \in [0, 1]$, on a

$$G_\mu(z) - z \geq (l_0 + 2)(l_0 + 1)\mu_{l_0+2} \int_z^1 \int_t^1 u^{l_0} du dt > 0.$$

Par conséquent, 1 est l'unique point fixe de G_μ dans l'intervalle $[0, 1]$, et donc $\phi = 1$.

Enfin, si le processus est sur-critique, alors la limite de $G'_\mu(z)$ quand z tend vers 1 par la gauche, qui vaut m_μ d'après (4.6), est strictement plus grande que 1. On en déduit qu'il existe $\delta > 0$ tel que $G'_\mu(z) > 1$ pour tout $z \in]1 - \delta, 1[$. Ainsi, pour $1 - \delta < z < z_0 < 1$, on a

$$0 \leq \int_z^{z_0} \int_z^t G''_\mu(u) du dt = G_\mu(z_0) - G_\mu(z) - (z_0 - z)G'_\mu(z).$$

En faisant tendre z_0 vers 1 par la gauche et en utilisant à nouveau (4.6), on obtient pour $1 - \delta < z < 1$,

$$0 \leq 1 - G_\mu(z) - (1 - z)G'_\mu(z) < 1 - G_\mu(z) - (1 - z) = z - G_\mu(z).$$

Ainsi, $G_\mu(z) - z < 0$ pour $1 - \delta < z < 1$, tandis que $G_\mu(0) = \mu_0 > 0$. Le théorème des valeurs intermédiaires montre que $z \mapsto G_\mu(z) - z$ s'annule donc dans $]0, 1[$. Ainsi, $\zeta = \inf\{z \in]0, 1[\mid G_\mu(z) = z\} < 1$. Par continuité de G_μ , on a $G_\mu(\zeta) = \zeta$, donc ζ est un point fixe de G_μ . Finalement, supposons qu'il existe un autre point fixe $z \in]\zeta, 1[$. Pour $z < z_0 < 1$, on peut écrire que

$$\begin{aligned} & (z - \zeta)(G_\mu(z_0) - G_\mu(z)) - (z_0 - z)(G_\mu(z) - G_\mu(\zeta)) \\ &= (z - \zeta) \int_z^{z_0} \left(G'_\mu(z) + \int_z^t G''_\mu(u) du \right) dt - (z_0 - z) \int_\zeta^z \left(G'_\mu(z) - \int_t^z G''_\mu(u) du \right) dt \\ &= (z - \zeta) \int_z^{z_0} \int_z^t G''_\mu(u) du dt + (z_0 - z) \int_\zeta^z \int_t^z G''_\mu(u) du dt \end{aligned}$$

$$\geq (l_0 + 2)(l_0 + 1)\mu_{l_0+2} \left((z - \zeta) \int_z^{z_0} \int_z^t u^{l_0} du dt + (z_0 - z) \int_\zeta^z \int_t^z u^{l_0} du dt \right).$$

En faisant tendre z_0 vers 1 par la gauche, et comme ζ , z et 1 sont des points fixes, on observe que le premier membre de ces relations tend vers 0, tandis que le dernier tend vers une valeur strictement positive, ce qui est contradictoire. Ainsi, ζ est le seul point fixe de G_μ dans $]0, 1[$, et donc $\phi = \zeta$. $\square$

Exemple 4.1 (Loi de reproduction géométrique). Suivant la définition 1.5, on note $\mu = (\mu_l)_{l \in \mathbb{N}}$ la mesure de probabilité correspondant à la loi géométrique $\text{Géo}_{\mathbb{N}}(p)$ de paramètre $p \in]0, 1[$, telle que $\mu_l = (1 - p)^l p$ pour tout entier $l \in \mathbb{N}$. Sa fonction génératrice est donnée par (1.11), et vérifie $G_\mu(z) = p/(1 - (1 - p)z)$ pour $(1 - p)|z| < 1$. En reprenant (1.12), on obtient $m_\mu = (1/p) - 1$ et $\sigma_\mu^2 = (1 - p)/p^2$ respectivement pour la moyenne et la variance de la mesure μ . Un processus de Galton-Watson de loi de reproduction μ est donc sous-critique, critique, ou sur-critique, selon respectivement que $p > 1/2$, $p = 1/2$, ou $p < 1/2$. Si le processus est issu d'un unique individu, la condition (4.7) est vérifiée. En appliquant le théorème 4.8, on montre que si $p \geq 1/2$, la probabilité d'extinction (4.3) du processus vaut $\phi = 1$, et si $p < 1/2$, il s'agit de la plus petite solution dans l'intervalle $[0, 1]$ de

$$\begin{aligned} \frac{p}{1 - (1 - p)\phi} = \phi &\iff (1 - p)\phi^2 - \phi + p = 0 \\ &\iff ((1 - p)\phi - p)(\phi - 1) = 0 \iff \phi \in \left\{ \frac{p}{1 - p}, 1 \right\}, \end{aligned}$$

de sorte que la probabilité d'extinction vaut $\phi = p/(1 - p)$. On trace sur la figure 4.5 la fonction génératrice de μ pour quelques valeurs de p . On utilise le script C.14.

Exemple 4.2 (Loi de reproduction poissonnienne). Notons $\mu = (\mu_l)_{l \in \mathbb{N}}$ la probabilité correspondant à la loi de Poisson $\text{Poi}(\lambda)$, donnée pour $\lambda \in]0, \infty[$ dans la définition 1.7. Sa fonction génératrice est $G_\mu(z) = e^{\lambda(z-1)}$ pour $z \in \mathbb{C}$, d'après (1.25). La moyenne et la variance de μ vérifient $m_\mu = \sigma_\mu^2 = \lambda$, d'après (1.27). Un processus de Galton-Watson de loi de reproduction μ est donc sous-critique, critique, ou sur-critique, selon respectivement que $\lambda < 1$, $\lambda = 1$, ou $\lambda > 1$. Si le processus est issu d'un unique individu, la condition (4.7) est satisfaite. On peut donc appliquer le théorème 4.8 pour déterminer que la probabilité d'extinction (4.3) vaut $\phi = 1$ si $\lambda \leq 1$, et est l'unique solution de l'équation $e^{\lambda(\phi-1)} = \phi$ dans $]0, 1[$ si $\lambda > 1$. On trace sur la figure 4.5 la fonction génératrice de μ , en utilisant le script C.13.

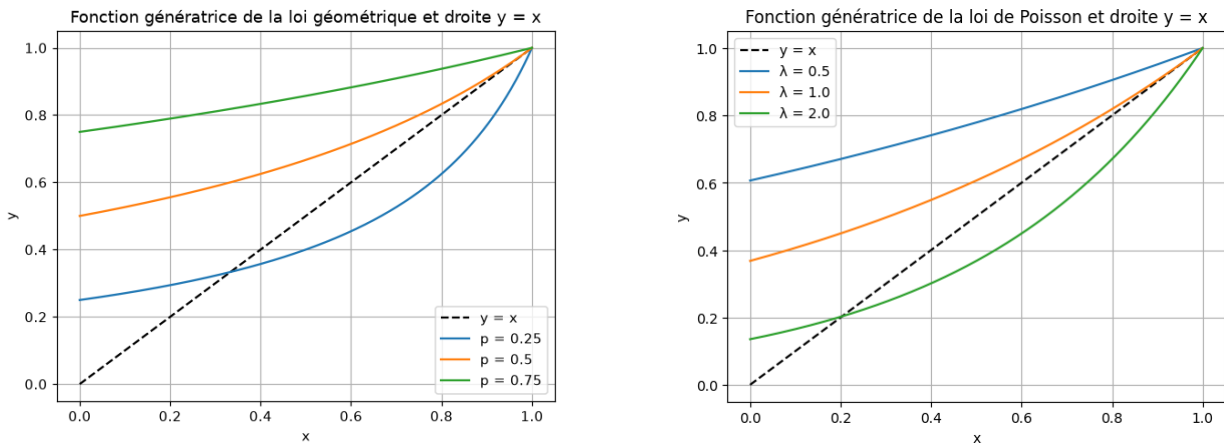


FIGURE 4.5 : Fonctions génératrices des lois géométrique et de Poisson

4.5 Taille totale de la population

On s'intéresse à la taille totale de la population d'un processus de Galton-Watson $(Z_n)_{n \geq 0}$ vérifiant $Z_0 = 1$ presque sûrement, et de loi de reproduction μ dont la moyenne m_μ définie par (4.5) est finie. Cette taille totale N est donnée par (4.2) comme la somme de toutes les tailles de population Z_n . On peut restreindre cette somme aux

indices $n < T$, où T est le temps d'extinction défini par (4.1). De plus, pour $n < T$, on a $Z_n \geq 1$, donc $N \geq T$. Comme les variables aléatoires $Z_0 + \dots + Z_n$ sont à valeurs dans $\mathbb{N}$ et croissent avec n vers la variable N , on peut alors appliquer le théorème B.16 et le corollaire 4.6 pour écrire que

$$\mathbb{E}[T] \leq \mathbb{E}[N] = \mathbb{E}\left[\sup_{n \in \mathbb{N}} Z_0 + \dots + Z_n\right] = \sup_{n \in \mathbb{N}} \mathbb{E}[Z_0 + \dots + Z_n] = \sum_{n=0}^{\infty} m_{\mu}^n = \begin{cases} \frac{1}{1 - m_{\mu}} & \text{si } m_{\mu} < 1 \\ \infty & \text{si } m_{\mu} \geq 1. \end{cases}$$

D'après le théorème 4.8, si le processus est sous-critique ou critique, alors presque sûrement, le processus s'éteint, c'est-à-dire que $T < \infty$, de sorte que la taille totale N est finie. Dans le cas critique, on observe que N est finie presque sûrement, mais que son espérance $\mathbb{E}[N]$ est infinie.

Le résultat suivant exploite la propriété de branchement pour caractériser la fonction génératrice de la population totale d'un processus de Galton-Watson.

Proposition 4.9 (Fonction génératrice de la taille totale de la population). *Soit $(Z_n)_{n \geq 0}$ un processus de Galton-Watson vérifiant $Z_0 = 1$ presque sûrement. La taille totale N , définie par (4.2), peut être infinie. On définit sa fonction génératrice par*

$$G_N(z) = \sum_{k=1}^{\infty} \mathbb{P}(N = k) z^k, \quad 0 \leq z \leq 1,$$

avec la convention $z^{\infty} = 0$, y compris pour $z = 1$. Ainsi, $G_N(1) = \mathbb{P}(N < \infty)$. Cette fonction vérifie

$$G_N(z) = z G_{\mu}(G_N(z)),$$

pour $z \in [0, 1]$, où G_{μ} désigne la fonction génératrice de la loi de reproduction μ du processus.

Démonstration. On observe que $N - 1 = N - Z_0$ est égal à la somme des tailles de population Z_n pour $n \geq 1$. Cependant, la propriété de branchement de la remarque 4.1 montre que, conditionnellement à l'événement $\{Z_1 = k\}$, le processus $(Z_n)_{n \geq 1}$ est équivalent à k processus de Galton-Watson indépendants, chacun issu d'un seul individu et de loi de reproduction μ . On en déduit que, conditionnellement à $\{Z_1 = k\}$, la variable $N - 1$ est donnée par la somme de k variables aléatoires indépendantes $N^{(1)}, \dots, N^{(k)}$, de même loi que N . En convenant que $N^{(1)} + \dots + N^{(k)} = 0$ si $k = 0$, on a pour tout $z \in [0, 1]$,

$$\begin{aligned} \mathbb{E}[z^{N-1}] &= \sum_{l=0}^{\infty} \mathbb{P}(N - 1 = l) z^l = \sum_{l=0}^{\infty} \sum_{k=0}^{\infty} \mathbb{P}(N - 1 = l \text{ et } Z_1 = k) z^l = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \mathbb{P}(N - 1 = l \text{ et } Z_1 = k) z^l \\ &= \sum_{\substack{k \in \mathbb{N} \\ \mathbb{P}(Z_1=k) > 0}} \mathbb{P}(Z_1 = k) \sum_{l=0}^{\infty} \mathbb{P}(N - 1 = l \mid Z_1 = k) z^l = \sum_{k=0}^{\infty} \mathbb{P}(Z_1 = k) \sum_{l=0}^{\infty} \mathbb{P}(N^{(1)} + \dots + N^{(k)} = l) z^l \\ &= \sum_{k=0}^{\infty} \mathbb{P}(Z_1 = k) G_{N^{(1)} + \dots + N^{(k)}}(z) = \sum_{k=0}^{\infty} \mathbb{P}(Z_1 = k) G_N(z)^k = G_{\mu}(G_N(z)). \end{aligned}$$

On a utilisé le théorème B.3. Le produit des fonctions génératrices reste valable pour les variables $N^{(j)}$ éventuellement infinies : avec la convention $z^{\infty} = 0$, on a $z^{N^{(1)} + \dots + N^{(k)}} = \prod_{j=1}^k z^{N^{(j)}}$ pour $0 \leq z \leq 1$, et les facteurs sont indépendants et bornés. L'espérance du produit est donc le produit des espérances, y compris en $z = 1$. Pour $k = 0$, le produit vide vaut 1. On conclut en observant que $G_N(z) = z \mathbb{E}[z^{N-1}]$. $\square$

Exemple 4.3 (Lois de reproduction poissonnienne et géométrique). On reprend les notations des exemples 4.1 et 4.2. La fonction génératrice de la taille totale N de la population d'un processus de Galton-Watson de loi de reproduction la loi $\text{Poi}(\lambda)$ vérifie

$$G_N(z) = z e^{\lambda(G_N(z)-1)}$$

pour $z \in [0, 1]$. De même, la fonction génératrice de la taille totale N de la population d'un processus de Galton-Watson de loi de reproduction la loi $\text{Géo}_{\mathbb{N}}(p)$ vérifie

$$G_N(z) = \frac{pz}{1 - (1-p)G_N(z)} \iff (1-p)G_N(z)^2 - G_N(z) + pz = 0$$

$$\Leftrightarrow G_N(z) = \frac{1 \pm \sqrt{1 - 4p(1-p)z}}{2(1-p)}.$$

Comme $N \geq 1$, on a $G_N(0) = 0$. La continuité de G_N sur $[0, 1]$ impose donc de retenir la racine avec le signe $-$; les deux racines sont distinctes pour $0 \leq z < 1$.

Génétique

Ce chapitre est consacré à l'étude des modèles d'évolution des génotypes au fil des générations dans une population de taille constante. Nous commençons par la loi de Hardy-Weinberg, qui est liée à la théorie de Mendel pour une population infinie. Ensuite, nous présentons quelques modèles classiques pour une population finie : le modèle de Moran et les modèles de Wright-Fisher.

5.1 Loi de Hardy-Weinberg

La génétique des populations repose sur l'étude de la transmission des gènes au sein des populations et de la manière dont leurs fréquences évoluent au fil des générations. Les travaux de Gregor Mendel ont posé les bases de l'hérédité en démontrant que les caractères sont transmis selon des lois précises, établissant ainsi la notion d'allèles dominants et récessifs. Toutefois, la théorie de Mendel ne permet pas, à elle seule, d'expliquer la stabilité des fréquences alléliques à l'échelle des populations. Un résultat fondamental dans ce domaine est la loi de Hardy-Weinberg, qui découle directement des principes de l'hérédité mendélienne et des mécanismes de reproduction sexuée. Cette loi établit que, sous certaines conditions idéalisées, les fréquences alléliques et génotypiques restent constantes d'une génération à l'autre. Ainsi, l'hérédité mendélienne préserve la variabilité des génotypes au sein d'une population qui satisfait des hypothèses précises.



FIGURE 5.1 : Gregor Mendel (1822-1884), Godfrey Harold Hardy (1877-1947) et Wilhelm Weinberg (1862-1937).

On considère une population dite *hermaphrodite* et *diploïde*, où chaque individu possède deux copies de chaque chromosome, héritées de ses parents. On s'intéresse dans cette population à un gène en particulier. Celui-ci est présent en deux formes, appelées *allèles*, qui peuvent être identiques ou différents et sont notées A et B . Ces deux copies forment un *génotype*, qui peut être AA , AB ou BB . On fait les hypothèses suivantes :

- (i). La population est de taille infinie ;
- (ii). Les générations sont *séparées et non chevauchantes*, c'est-à-dire que les reproductions n'ont lieu qu'entre individus de la même génération ;
- (iii). La population est *panmictique* et *pangamique*, c'est-à-dire que les accouplements se forment au hasard entre les individus et que la fécondation des gamètes est également aléatoire, sans sélection sexuelle, barrière géographique ou structure sociale influençant les croisements ;
- (iv). Il n'y a ni mutation, ni migration, ni sélection naturelle. Chaque individu transmet l'une de ses deux copies du gène avec probabilité $1/2$.

Théorème 5.1 (Loi de Hardy-Weinberg pour la théorie de Mendel). *Sous les conditions exprimées ci-dessus, on a les propriétés suivantes :*

- (i). *Les fréquences des allèles A et B au sein de la population restent constantes d'une génération à l'autre, et égales à leurs fréquences p et q = 1 - p dans la population initiale ;*
- (ii). *Si r, s et t désignent respectivement les fréquences 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 constantes et valent respectivement*

$$\left(r + \frac{s}{2}\right)^2, \quad 2\left(r + \frac{s}{2}\right)\left(t + \frac{s}{2}\right), \quad \left(t + \frac{s}{2}\right)^2;$$

- (iii). *On a la relation p = r + s/2 pour la contribution des génotypes AA et AB à la fréquence de l'allèle A. De même, on a q = t + s/2 pour les génotypes BB et AB et l'allèle B.*

Démonstration. Un individu de la première génération après la génération initiale dont le génotype est AA peut provenir d'un couple formé de deux individus AA, d'un couple formé d'un individu AA et d'un individu AB, ou d'un couple formé de deux individus AB. Ces unions se produisent avec des probabilités respectives r^2 , $2rs$ et s^2 . Lors de ces unions, la probabilité d'obtenir un individu AA est 1, 1/2 et 1/4, respectivement. On en déduit que la fréquence de AA parmi la première génération est

$$r^2 + \frac{1}{2}2rs + \frac{1}{4}s^2 = \left(r + \frac{s}{2}\right)^2,$$

voir la figure 5.2. Par symétrie, la fréquence de BB au sein de la première génération est $(t + s/2)^2$. Comme toutes les fréquences doivent sommer à 1, on en déduit que la fréquence de AB est

$$1 - \left(r + \frac{s}{2}\right)^2 - \left(t + \frac{s}{2}\right)^2 = 2\left(r + \frac{s}{2}\right)\left(t + \frac{s}{2}\right).$$

Autrement dit, en passant de la génération initiale à la première génération, les fréquences de AA, AB et BB subissent la transformation

$$\varphi : (r, s, t) \mapsto \left(\left(r + \frac{s}{2}\right)^2, 2\left(r + \frac{s}{2}\right)\left(t + \frac{s}{2}\right), \left(t + \frac{s}{2}\right)^2\right).$$

Les fréquences de AA, AB et BB pour toutes les générations suivantes s'obtiennent donc en itérant cette transformation. Cependant, on voit que les fréquences de première génération sont un point fixe de φ , puisque

$$\begin{cases} \left(\left(r + \frac{s}{2}\right)^2 + \frac{1}{2}2\left(r + \frac{s}{2}\right)\left(t + \frac{s}{2}\right)\right)^2 = \left(r + \frac{s}{2}\right)^2 \\ 2\left(\left(r + \frac{s}{2}\right)^2 + \frac{1}{2}2\left(r + \frac{s}{2}\right)\left(t + \frac{s}{2}\right)\right)\left(\left(t + \frac{s}{2}\right)^2 + \frac{1}{2}2\left(r + \frac{s}{2}\right)\left(t + \frac{s}{2}\right)\right) = 2\left(r + \frac{s}{2}\right)\left(t + \frac{s}{2}\right) \\ \left(\left(t + \frac{s}{2}\right)^2 + \frac{1}{2}2\left(r + \frac{s}{2}\right)\left(t + \frac{s}{2}\right)\right)^2 = \left(t + \frac{s}{2}\right)^2. \end{cases}$$

Les fréquences de AA, AB et BB restent donc constantes pour toutes les générations suivantes.

Enfin, comme chaque individu possède deux allèles, les fréquences des allèles A et B au sein de la génération initiale valent respectivement

$$p = \frac{2r + s}{2} = r + \frac{s}{2} \quad \text{et} \quad q = \frac{2t + s}{2} = t + \frac{s}{2}.$$

Pour les générations suivantes, elles valent respectivement

$$\left(r + \frac{s}{2}\right)^2 + \frac{1}{2}2\left(r + \frac{s}{2}\right)\left(t + \frac{s}{2}\right) = r + \frac{s}{2} = p \quad \text{et} \quad \left(t + \frac{s}{2}\right)^2 + \frac{1}{2}2\left(r + \frac{s}{2}\right)\left(t + \frac{s}{2}\right) = t + \frac{s}{2} = q.$$

Ces fréquences restent donc constantes, ce qui achève la démonstration. $\square$

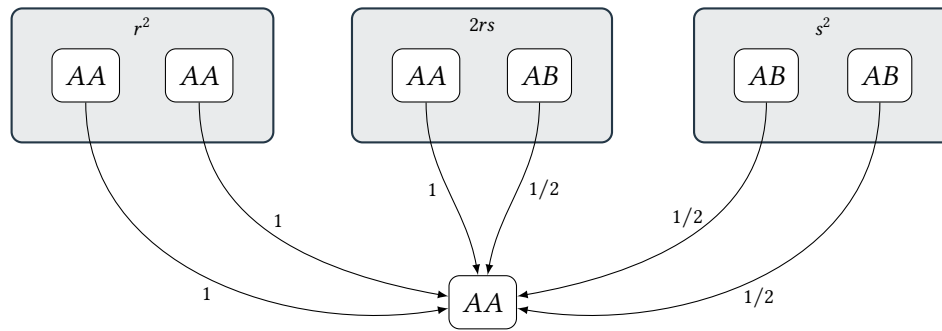


FIGURE 5.2 : Probabilités de produire le génotype AA à la première génération

Si les fréquences observées dans une population naturelle diffèrent significativement de celles attendues selon la loi de Hardy-Weinberg, cela peut indiquer l'action de forces évolutives telles que la sélection naturelle, la mutation ou la dérive génétique. La loi de Hardy-Weinberg, qui s'applique plus généralement à un locus (une position précise sur un chromosome, où se trouve un gène ou une séquence génétique) avec plusieurs allèles $A_1, \dots, A_k$, est un outil fondamental pour l'étude de l'évolution des populations. Elle permet d'identifier les déviations par rapport à l'équilibre génétique et d'inférer les pressions évolutives agissant sur une population. Les relations de dominance entre allèles n'affectent pas l'évolution des fréquences alléliques, et le phénomène de Hardy-Weinberg est fréquemment observé dans des populations réelles, même lorsque les hypothèses classiques sont assouplies. L'équilibre des fréquences génotypiques et la constance des fréquences alléliques sont deux propriétés distinctes : des accouplements non aléatoires peuvent modifier les premières sans modifier les secondes, tandis que la dérive génétique peut faire évoluer les fréquences alléliques dans une population finie. Cela motive l'introduction des modèles étudiés ci-après, qui décrivent l'évolution d'une population de taille finie.

5.2 Modèle de Moran

Le modèle de Moran est un modèle probabiliste décrivant l'évolution en temps discret de la fréquence des allèles dans une population finie de taille constante $N \geq 2$. On considère une population haploïde, où chaque individu porte une seule copie du gène étudié et où coexistent deux allèles, notés A et B , d'un même gène. On suppose qu'il n'y a ni migration, ni mutation, ni sélection naturelle.

À chaque instant n , on tire au hasard avec remise deux individus dans la population. Le premier individu choisi engendre un descendant, qui hérite exactement de son allèle. Le second individu choisi est remplacé par ce descendant. Tous les autres individus de la population se maintiennent, de sorte que la taille de la population reste constante égale à N . Les deux tirages étant effectués avec remise, il est possible que les deux individus coïncident. Dans ce cas, le descendant remplace son propre parent et la composition allélique de la population demeure inchangée à ce pas de temps.

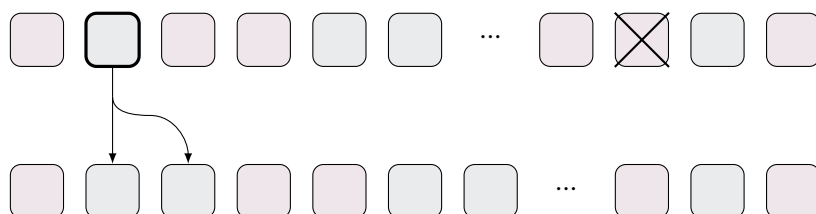


FIGURE 5.3 : Patrick Alfred Pierce Moran (1917–1988) et mécanisme de transition dans son modèle

En notant X_n le nombre d'individus portant l'allèle A dans la population à l'instant n , on a alors :

- $X_{n+1} = X_n + 1$ si l'individu choisi pour se reproduire porte l'allèle A et l'autre choisi porte B ;

- $X_{n+1} = X_n - 1$ si l'individu choisi pour se reproduire porte l'allèle B et l'autre choisi porte A ;
- $X_{n+1} = X_n$ si les deux individus choisis portent le même allèle.

Ces règles se traduisent par le résultat suivant.

Proposition 5.2 (Aspect markovien du modèle de Moran). *Sous les conditions exprimées ci-dessus, la suite $(X_n)_{n \geq 0}$ des nombres d'individus portant l'allèle A dans une population de taille $N \geq 2$ est une chaîne de Markov homogène à espace d'états $\{0, \dots, N\}$, dont la matrice de transition P a pour coefficients*

$$p_{k,l} = \begin{cases} \frac{k(N-k)}{N^2} & \text{si } |l - k| = 1 \\ \frac{k^2 + (N-k)^2}{N^2} & \text{si } l = k \\ 0 & \text{sinon.} \end{cases}$$

D'après la proposition 5.2, le graphe associé à la suite $(X_n)_{n \geq 0}$ des nombres d'individus portant l'allèle A , vue comme une chaîne de Markov homogène, admet l'espace d'états $\{0, \dots, N\}$ pour ensemble de sommets, et comporte une boucle sur chaque sommet $k \in \{0, \dots, N\}$, des arcs de chaque sommet $k \in \{1, \dots, N-1\}$ vers ses voisins $k-1$ et $k+1$, des arcs des sommets 1 et $N-1$ vers leurs voisins 0 et N , respectivement. On représente ce graphe sur la figure 5.4. L'ensemble des états $\{1, \dots, N-1\}$ est irréductible, tandis que les états 0 et N sont absorbants. Partant de l'état 1, si on passe directement à l'état 0, on ne peut jamais retourner à l'état 1. Ainsi, la probabilité ρ_1 de retour à l'état 1, qui est donnée par (B.7), vérifie $1 - \rho_1 \geq (N-1)/N^2 > 0$. D'après les propositions B.19 et B.20, l'état 1 est donc transient, de même que tous les états $k \in \{1, \dots, N-1\}$, avec qui il communique.

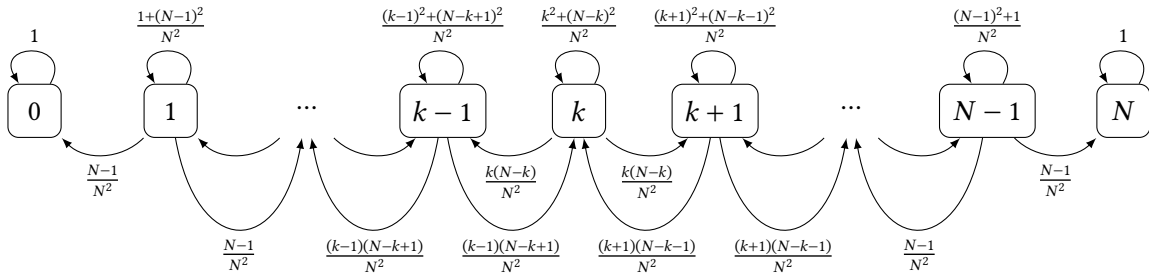


FIGURE 5.4 : Graphe de la chaîne de Markov associée au modèle de Moran

Lorsque la chaîne de Markov $(X_n)_{n \geq 0}$ atteint l'un des états absorbants 0 ou N , il n'y a plus qu'un seul allèle dans la population, et la dynamique s'arrête. Le premier instant où cela se produit est appelé le *temps de fixation* de l'allèle A ou B . Cet instant est la variable aléatoire définie par

$$T = \inf\{n \geq 0 \mid X_n \in \{0, N\}\}. \quad (5.1)$$

Considérons un entier $k \in \{0, \dots, N\}$ tel que $\mathbb{P}(X_0 = k) > 0$. Comme les états 0 et N sont absorbants, on peut procéder comme en (2.4) et déduire de (1.7) que

$$\begin{aligned} \mathbb{E}[T \mid X_0 = k] &= \sum_{n=1}^{\infty} \mathbb{P}(T \geq n \mid X_0 = k) = \sum_{n=0}^{\infty} \mathbb{P}(T \geq n+1 \mid X_0 = k) = \sum_{n=0}^{\infty} \mathbb{P}(X_n \notin \{0, N\} \mid X_0 = k) \\ &= \sum_{n=0}^{\infty} \sum_{l=1}^{N-1} \mathbb{P}(X_n = l \mid X_0 = k) = \sum_{l=1}^{N-1} \sum_{n=0}^{\infty} \mathbb{P}(X_n = l \mid X_0 = k) = \sum_{l=1}^{N-1} p_{k,l}^*. \end{aligned} \quad (5.2)$$

La dernière égalité découle de la proposition 5.2 et du théorème B.7. La dernière somme fait intervenir les coefficients $p_{k,l}^*$ de la matrice potentielle P^* associée à la matrice de transition P de la chaîne de Markov. Comme les états $l \in \{1, \dots, N-1\}$ sont tous transitoires, le corollaire B.18 et la proposition B.19 assurent que tous les coefficients $p_{k,l}^*$ dans la dernière somme sont finis. On en déduit que le temps de fixation de l'allèle A ou B vérifie

$$\mathbb{P}(T < \infty \mid X_0 = k) = 1 \quad \text{et} \quad \mathbb{E}[T \mid X_0 = k] < \infty.$$

À partir du temps de fixation, le nombre d'individus portant l'allèle A reste constant et égal à 0 ou N . On a alors presque sûrement,

$$X_n \xrightarrow{n \rightarrow \infty} X_T \in \{0, N\}. \quad (5.3)$$

La variable aléatoire X_T donne alors l'allèle fixé dans la population. En effet, elle prend la valeur N si l'allèle A est fixé et la valeur 0 si l'allèle B est fixé. Le résultat suivant donne la loi de la variable aléatoire X_T , ainsi qu'un équivalent de l'espérance du temps de fixation T .

Théorème 5.3 (Lieu et temps de fixation dans le modèle de Moran). *Sous les conditions exprimées ci-dessus, notons $(X_n)_{n \geq 0}$ la suite des nombres d'individus portant l'allèle A dans une population de taille $N \geq 2$, et notons T le temps de fixation de l'allèle A ou B . Pour tout entier $k \in \{0, \dots, N\}$, on a conditionnellement à l'événement $\{X_0 = k\}$, supposé de probabilité non nulle, les propriétés suivantes :*

(i). *La loi de la variable aléatoire X_T est donnée par*

$$\mathbb{P}(X_T = N \mid X_0 = k) = \frac{k}{N} \quad \text{et} \quad \mathbb{P}(X_T = 0 \mid X_0 = k) = \frac{N-k}{N};$$

(ii). *L'espérance du temps de fixation T est nulle pour $k = 0$ ou $k = N$. Pour $1 \leq k \leq N-1$, elle vérifie*

$$\mathbb{E}[T \mid X_0 = k] = N \left(\sum_{l=1}^k \frac{N-k}{N-l} + \sum_{l=k+1}^{N-1} \frac{k}{l} \right).$$

Démonstration. Pour tous entiers $n \geq 0$ et $k \in \{0, \dots, N\}$, considérons la variable aléatoire à valeurs entières positives $Y_n = (2N - |X_n - X_T|) \mathbf{1}_{\{X_0=k\}}$. D'après (5.3), on a presque sûrement $Y_n \xrightarrow{n \rightarrow \infty} 2N \mathbf{1}_{\{X_0=k\}}$, ce qui équivaut à l'égalité

$$2N \mathbf{1}_{\{X_0=k\}} = \sup_{n_0 \in \mathbb{N}} \inf_{n \geq n_0} Y_n.$$

En utilisant le théorème B.16, on peut alors écrire que

$$\mathbb{E}[2N \mathbf{1}_{\{X_0=k\}}] = \mathbb{E} \left[\sup_{n_0 \in \mathbb{N}} \inf_{n \geq n_0} Y_n \right] = \sup_{n_0 \in \mathbb{N}} \mathbb{E} \left[\inf_{n \geq n_0} Y_n \right] \leq \sup_{n_0 \in \mathbb{N}} \inf_{n \geq n_0} \mathbb{E}[Y_n].$$

On en déduit que

$$0 \leq \inf_{n_0 \in \mathbb{N}} \sup_{n \geq n_0} \mathbb{E}[|X_n - X_T| \mathbf{1}_{\{X_0=k\}}] = \inf_{n_0 \in \mathbb{N}} \sup_{n \geq n_0} \mathbb{E}[2N \mathbf{1}_{\{X_0=k\}} - Y_n] = \mathbb{E}[2N \mathbf{1}_{\{X_0=k\}}] - \sup_{n_0 \in \mathbb{N}} \inf_{n \geq n_0} \mathbb{E}[Y_n] \leq 0,$$

ce qui implique directement que $\mathbb{E}[|X_n - X_T| \mathbf{1}_{\{X_0=k\}}]$ tend vers 0 lorsque n tend vers l'infini. Ainsi,

$$N \mathbb{P}(X_T = N \text{ et } X_0 = k) = \mathbb{E}[X_T \mathbf{1}_{\{X_0=k\}}] = \lim_{n \rightarrow \infty} \mathbb{E}[X_n \mathbf{1}_{\{X_0=k\}}]. \quad (5.4)$$

Cependant, la proposition 5.2 implique que pour tous entiers $n \geq 1$ et $k_0, \dots, k_{n-1} \in \{0, \dots, N\}$, on a

$$\begin{aligned} \mathbb{E}[X_n \mathbf{1}_{\{X_0=k_0, \dots, X_{n-1}=k_{n-1}\}}] &= \sum_{l=1}^N l \mathbb{P}(X_0 = k_0, \dots, X_{n-1} = k_{n-1}, X_n = l) \\ &= \mathbb{P}(X_0 = k_0, \dots, X_{n-1} = k_{n-1}) \sum_{l=1}^N l p_{k_{n-1}, l}. \end{aligned}$$

On calcule alors, pour tout entier $k \in \{0, \dots, N\}$,

$$\sum_{l=1}^N l p_{k, l} = k \frac{k^2 + (N-k)^2}{N^2} + \mathbf{1}_{\{k \geq 2\}} (k-1) \frac{k(N-k)}{N^2} + \mathbf{1}_{\{k \leq N-1\}} (k+1) \frac{k(N-k)}{N^2} = k. \quad (5.5)$$

Il en ressort que pour tous entiers $n \geq 1$ et $k_0, \dots, k_{n-1} \in \{0, \dots, N\}$, on a

$$\mathbb{E}[X_n \mathbf{1}_{\{X_0=k_0, \dots, X_{n-2}=k_{n-2}\}} \mathbf{1}_{\{X_{n-1}=k_{n-1}\}}] = \mathbb{E}[X_{n-1} \mathbf{1}_{\{X_0=k_0, \dots, X_{n-2}=k_{n-2}\}} \mathbf{1}_{\{X_{n-1}=k_{n-1}\}}].$$

En sommant sur les entiers $k_{n-1} \in \{0, \dots, N\}$, puis en procédant comme ci-dessus, on obtient

$$\begin{aligned} \mathbb{E}[X_n \mathbf{1}_{\{X_0=k_0, \dots, X_{n-3}=k_{n-3}\}} \mathbf{1}_{\{X_{n-2}=k_{n-2}\}}] &= \mathbb{E}[X_{n-1} \mathbf{1}_{\{X_0=k_0, \dots, X_{n-3}=k_{n-3}\}} \mathbf{1}_{\{X_{n-2}=k_{n-2}\}}] \\ &= \mathbb{E}[X_{n-2} \mathbf{1}_{\{X_0=k_0, \dots, X_{n-3}=k_{n-3}\}} \mathbf{1}_{\{X_{n-2}=k_{n-2}\}}]. \end{aligned}$$

En itérant ce raisonnement, on obtient finalement que pour tous entiers $n \geq 1$ et $k \in \{0, \dots, N\}$,

$$\mathbb{E}[X_n \mathbf{1}_{\{X_0=k\}}] = \mathbb{E}[X_0 \mathbf{1}_{\{X_0=k\}}] = k \mathbb{P}(X_0 = k).$$

On en déduit la première partie du théorème en combinant cela avec (5.4).

Pour tout $k \in \{0, \dots, N\}$, désignons par θ_k la dernière somme dans (5.2) et procédons comme dans (2.5). La définition B.4 montre que $P^* = I + PP^*$, où I est la matrice identité. On obtient donc

$$\theta_k = \sum_{l=1}^{N-1} p_{k,l}^* = \sum_{l=1}^{N-1} \left(\mathbf{1}_{\{k=l\}} + \sum_{m=0}^N p_{k,m} p_{m,l}^* \right) = \mathbf{1}_{\{0 < k < N\}} + \sum_{m=0}^N p_{k,m} \sum_{l=1}^{N-1} p_{m,l}^* = \mathbf{1}_{\{0 < k < N\}} + \sum_{m=0}^N p_{k,m} \theta_m.$$

La proposition 5.2 montre alors que pour tout $k \in \{1, \dots, N-1\}$, on a

$$\theta_k = 1 + \frac{k^2 + (N-k)^2}{N^2} \theta_k + \frac{k(N-k)}{N^2} (\theta_{k-1} + \theta_{k+1}), \quad (5.6)$$

ce qui équivaut à

$$(\theta_{k+1} - \theta_k) - (\theta_k - \theta_{k-1}) = -\frac{N^2}{k(N-k)}.$$

Pour tout $l \in \{2, \dots, N\}$, on peut donc écrire que

$$(\theta_l - \theta_{l-1}) - (\theta_1 - \theta_0) = \sum_{k=1}^{l-1} ((\theta_{k+1} - \theta_k) - (\theta_k - \theta_{k-1})) = -\sum_{k=1}^{l-1} \frac{N^2}{k(N-k)}.$$

Ensuite, pour tout $m \in \{2, \dots, N\}$, on a

$$\begin{aligned} \theta_m &= \theta_1 + \sum_{l=2}^m (\theta_l - \theta_{l-1}) = \theta_1 + (\theta_1 - \theta_0)(m-1) - \sum_{l=2}^m \sum_{k=1}^{l-1} \frac{N^2}{k(N-k)} \\ &= \theta_1 + (\theta_1 - \theta_0)(m-1) - \sum_{k=1}^{m-1} \sum_{l=k+1}^m \frac{N^2}{k(N-k)} = m\theta_1 - \sum_{k=1}^{m-1} \frac{N^2(m-k)}{k(N-k)}. \end{aligned}$$

La dernière égalité vient de l'observation que $\theta_0 = 0$. Comme on a de même $\theta_N = 0$, on peut en prenant $m = N$ ci-dessus en déduire la valeur de θ_1 et la réinjecter ci-dessus. On obtient alors

$$\theta_m = m \sum_{k=1}^{N-1} \frac{N}{k} - \sum_{k=1}^{m-1} \frac{N^2(m-k)}{k(N-k)} = Nm \sum_{k=1}^{N-1} \frac{1}{k} - Nm \sum_{k=1}^{m-1} \left(\frac{1}{k} + \frac{1}{N-k} \right) + N^2 \sum_{k=1}^{m-1} \frac{1}{N-k}.$$

On en déduit le résultat annoncé, puisque $\theta_m = \mathbb{E}[T \mid X_0 = m]$ d'après (5.2), si l'événement par rapport auquel on conditionne est de probabilité non nulle. $\square$

On peut ensuite donner un équivalent de l'espérance du temps de fixation T lorsque le nombre initial d'individus portant l'allèle A est proportionnel à la taille N de la population, qui elle-même tend vers l'infini. Rappelons que $\lfloor \cdot \rfloor$ désigne la partie entière d'un nombre réel.

Corollaire 5.4 (Asymptotique de l'espérance du temps de fixation dans le modèle de Moran). *Sous les conditions exprimées ci-dessus, notons $(X_n)_{n \geq 0}$ la suite des nombres d'individus portant l'allèle A dans une population de taille $N \geq 2$, et notons T le temps de fixation de l'allèle A ou B . Pour tout nombre réel $p \in [0, 1]$, on appelle entropie de la loi de Bernoulli $\text{Ber}(p)$ la quantité*

$$H(p) = \begin{cases} -p \ln p - (1-p) \ln(1-p) & \text{si } 0 < p < 1 \\ 0 & \text{si } p \in \{0, 1\}. \end{cases}$$

On a alors le développement asymptotique

$$\mathbb{E}[T \mathbf{1}_{\{X_0 = \lfloor pN \rfloor\}}] = N^2 \left(H(p) + o_{N \rightarrow \infty}(1) \right) \mathbb{P}(X_0 = \lfloor pN \rfloor).$$

Démonstration. On peut supposer $0 < p < 1$, car le temps de fixation T est nul partant de $X_0 = 0$ ou $X_0 = N$. Il suffit d'étudier la limite lorsque N tend vers l'infini de la quantité $\theta_{[pN]}$ donnée dans la preuve du théorème 5.3, divisée par N^2 . On a

$$\frac{\theta_{[pN]}}{N^2} = \left(1 - \frac{[pN]}{N}\right) \sum_{l=1}^{[pN]} \frac{1}{N-l} + \frac{[pN]}{N} \sum_{l=[pN]+1}^{N-1} \frac{1}{l} = \left(1 - \frac{[pN]}{N}\right) \sum_{l=N-[pN]}^{N-1} \frac{1}{l} + \frac{[pN]}{N} \sum_{l=[pN]+1}^{N-1} \frac{1}{l}.$$

Cependant, d'après la définition 2.1 de la constante d'Euler-Mascheroni γ , on a

$$\sum_{l=N-[pN]}^{N-1} \frac{1}{l} = \ln(N-1) + \gamma - \ln(N - [pN] - 1) - \gamma + o_{N \rightarrow \infty}(1) = -\ln(1-p) + o_{N \rightarrow \infty}(1)$$

et de même

$$\sum_{l=[pN]+1}^{N-1} \frac{1}{l} = \ln(N-1) + \gamma - \ln([pN]) - \gamma + o_{N \rightarrow \infty}(1) = -\ln p + o_{N \rightarrow \infty}(1).$$

On en déduit directement le résultat annoncé. $\square$

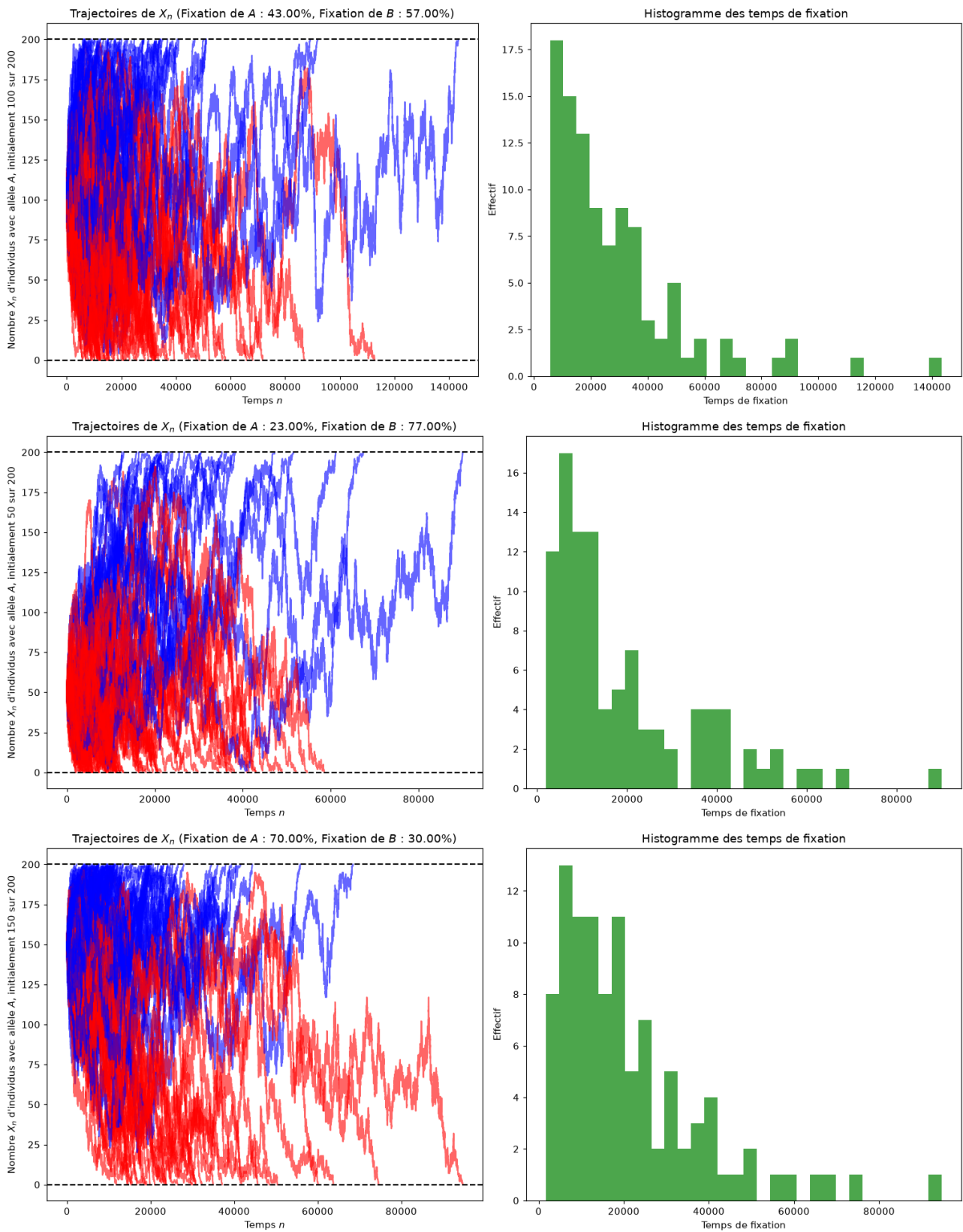
L'entropie $H(p)$ de la loi de Bernoulli de paramètre p , définie dans le corollaire 5.4, est une fonction continue et strictement concave par rapport à p . Elle est symétrique par rapport à $p = 1/2$, où elle atteint son unique maximum, égal à $\ln 2$. Elle est minimale et nulle en 0 et en 1. Ainsi, l'espérance du temps de fixation T dans le modèle de Moran est maximale lorsque le nombre initial d'individus portant l'allèle A est égal à $N/2$ environ, et minimale lorsque ce nombre est égal à 0 ou N . Le temps de fixation est donc d'autant plus long que la population est initialement équilibrée entre les deux allèles. Dans ce cas, le théorème 5.3 et le corollaire 5.4 montrent que la probabilité d'une fixation de l'allèle A est proche de $1/2$, et l'espérance du temps de fixation est proche de $N^2 \ln 2$. En revanche, si la population est initialement composée quasi uniquement d'individus portant l'allèle A , l'espérance du temps de fixation est petite devant N^2 lorsque la proportion de l'allèle minoritaire tend vers 0, et la probabilité de fixation de l'allèle A est proche de 1. Le même phénomène s'observe si la population initiale est composée quasi uniquement d'individus portant l'allèle B .

Le modèle de Moran est souvent utilisé en génétique des populations pour étudier la *dérive génétique*, c'est-à-dire les fluctuations aléatoires de la fréquence des allèles d'une population au fil des générations. Contrairement à la sélection naturelle, qui favorise la transmission des allèles conférant un avantage adaptatif, la dérive génétique agit de manière purement aléatoire, indépendamment de la valeur sélective des allèles. Elle est particulièrement marquée dans les populations de petite taille, où ces fluctuations aléatoires peuvent conduire à la fixation d'un allèle ou à sa disparition complète en un temps fini. Ainsi, un allèle potentiellement bénéfique peut être perdu simplement sous l'effet du hasard, sans qu'aucune pression sélective ne soit en jeu. L'étude du modèle de Moran nous a notamment permis de quantifier ces phénomènes en déterminant la probabilité de fixation d'un allèle ainsi que l'espérance du temps nécessaire pour atteindre cette fixation.

Algorithme 5.1 (Simulation du modèle de Moran). On simule m évolutions indépendantes du nombre d'individus porteurs de l'allèle A dans une population de taille N , sachant que ce nombre est initialement égal à k . On trace ces évolutions jusqu'au temps de fixation, en bleu quand il y a fixation de l'allèle A et en rouge quand il y a fixation de l'allèle B . On trace aussi l'histogramme du temps de fixation. On utilise le script C.15. Les graphiques correspondants, pour diverses valeurs des paramètres, sont sur la figure 5.5.

5.3 Modèle de Wright-Fisher

Tout comme le modèle de Moran, le modèle de Wright-Fisher est un modèle probabiliste décrivant l'évolution en temps discret de la fréquence des allèles dans une population finie. La population est haploïde, de taille constante $N \geq 1$, avec des générations non chevauchantes. Chaque individu porte une seule copie du gène étudié. De plus, il n'y a ni migration, ni mutation, ni sélection naturelle. On considère un gène pouvant exister sous deux formes alléliques, notées A et B . On définit la variable aléatoire X_n comme le nombre d'allèles A dans la population à la génération n . La population à la génération $n + 1$ s'obtient en tirant avec remise N individus de la génération

FIGURE 5.5 : Cent simulations indépendantes du modèle de Moran, avec $N = 200$

n , de sorte que chaque individu tiré porte l'allèle A avec probabilité X_n/N . Conditionnellement à l'événement $\{X_0 = k_0, \dots, X_n = k_n\}$, la variable aléatoire X_{n+1} suit alors une loi binomiale $\text{Bin}(N, k_n/N)$. Cela se traduit par le résultat suivant.

Proposition 5.5 (Aspect markovien du modèle de Wright-Fisher). *Sous les conditions exprimées ci-dessus, la suite $(X_n)_{n \geq 0}$ des nombres d'individus portant l'allèle A dans une population de taille $N \geq 1$ est une chaîne de Markov homogène à espace d'états $\{0, \dots, N\}$, dont la matrice de transition P a pour coefficients*

$$p_{k,l} = \binom{N}{l} \left(\frac{k}{N}\right)^l \left(1 - \frac{k}{N}\right)^{N-l}.$$

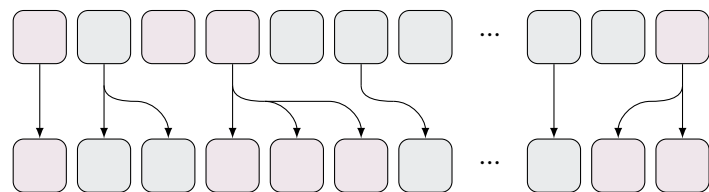


FIGURE 5.6 : Sewall Wright (1889–1988), Ronald Fisher (1890–1962) et le mécanisme de transition dans leur modèle

D'après la proposition 5.5, le graphe associé à la suite $(X_n)_{n \geq 0}$ des nombres d'individus portant l'allèle A , vue comme une chaîne de Markov homogène, admet l'espace d'états $\{0, \dots, N\}$ pour ensemble de sommets, et comporte un arc de chaque sommet $k \in \{1, \dots, N-1\}$ vers chaque sommet $l \in \{0, \dots, N\}$, et une boucle sur chacun des sommets 0 et N . Ce graphe est donc plus complexe que celui du modèle de Moran, puisqu'on peut accéder à partir d'un état $k \in \{1, \dots, N-1\}$ à tous les autres états et pas seulement les trois états $k-1, k, k+1$. Cependant, tout comme pour le modèle de Moran, les états 0 et N sont absorbants. Lorsque $N \geq 2$, les états $1, \dots, N-1$ sont transients et forment un ensemble irréductible. Si $N = 1$, la population est déjà fixée à l'instant initial et $T = 0$. Dans tous les cas, le temps de fixation T , défini à nouveau par (5.1), est d'espérance finie, donc presque sûrement fini. De plus, on a encore (5.3). La preuve de la première partie du théorème 5.3 reste valable ici, à condition de remplacer (5.5) par le calcul de l'espérance de la loi binomiale $\text{Bin}(N, k/N)$, c'est-à-dire

$$\sum_{l=1}^N l p_{k,l} = \sum_{l=1}^N l \binom{N}{l} \left(\frac{k}{N}\right)^l \left(1 - \frac{k}{N}\right)^{N-l} = N \frac{k}{N} = k.$$

On peut alors en déduire le résultat suivant.

Théorème 5.6 (Lieu de fixation dans le modèle de Wright-Fisher). *Sous les conditions exprimées ci-dessus, notons $(X_n)_{n \geq 0}$ la suite des nombres d'individus portant l'allèle A dans une population de taille $N \geq 1$, et notons T le temps de fixation de l'allèle A ou B . Pour tout entier $k \in \{0, \dots, N\}$ tel que $\mathbb{P}(X_0 = k) > 0$, on a*

$$\mathbb{P}(X_T = N \mid X_0 = k) = \frac{k}{N} \quad \text{et} \quad \mathbb{P}(X_T = 0 \mid X_0 = k) = \frac{N-k}{N}.$$

On peut aussi reprendre la preuve de la deuxième partie du théorème 5.3, mais ce faisant on obtient un système linéaire beaucoup plus compliqué que (5.6) pour lequel il n'y a pas de solution explicite simple. On peut cependant obtenir un développement asymptotique similaire à celui du corollaire 5.4. Plus précisément, on peut montrer que

$$\mathbb{E}[T \mathbf{1}_{\{X_0 = \lfloor pN \rfloor\}}] = 2N \left(H(p) + o_{N \rightarrow \infty}(1) \right) \mathbb{P}(X_0 = \lfloor pN \rfloor).$$

On voit encore apparaître ici l'entropie de la loi de Bernoulli de paramètre p . Les commentaires qui suivent le corollaire 5.4 s'appliquent donc également au modèle de Wright-Fisher. Cependant, l'asymptotique est en N plutôt qu'en N^2 . Comme le modèle de Wright-Fisher modifie davantage la proportion d'allèles à chaque étape que le modèle de Moran, la fixation est atteinte plus tôt.

Algorithme 5.2 (Simulation du modèle de Wright-Fisher). On simule m évolutions indépendantes du nombre d'individus porteurs de l'allèle A dans une population de taille N , sachant que ce nombre est initialement égal à k . On trace ces évolutions jusqu'au temps de fixation, en bleu quand il y a fixation de l'allèle A et en rouge quand il y a fixation de l'allèle B . On trace aussi l'histogramme du temps de fixation. On utilise le script C.16. Les graphiques correspondants, pour diverses valeurs des paramètres, sont sur la figure 5.7.

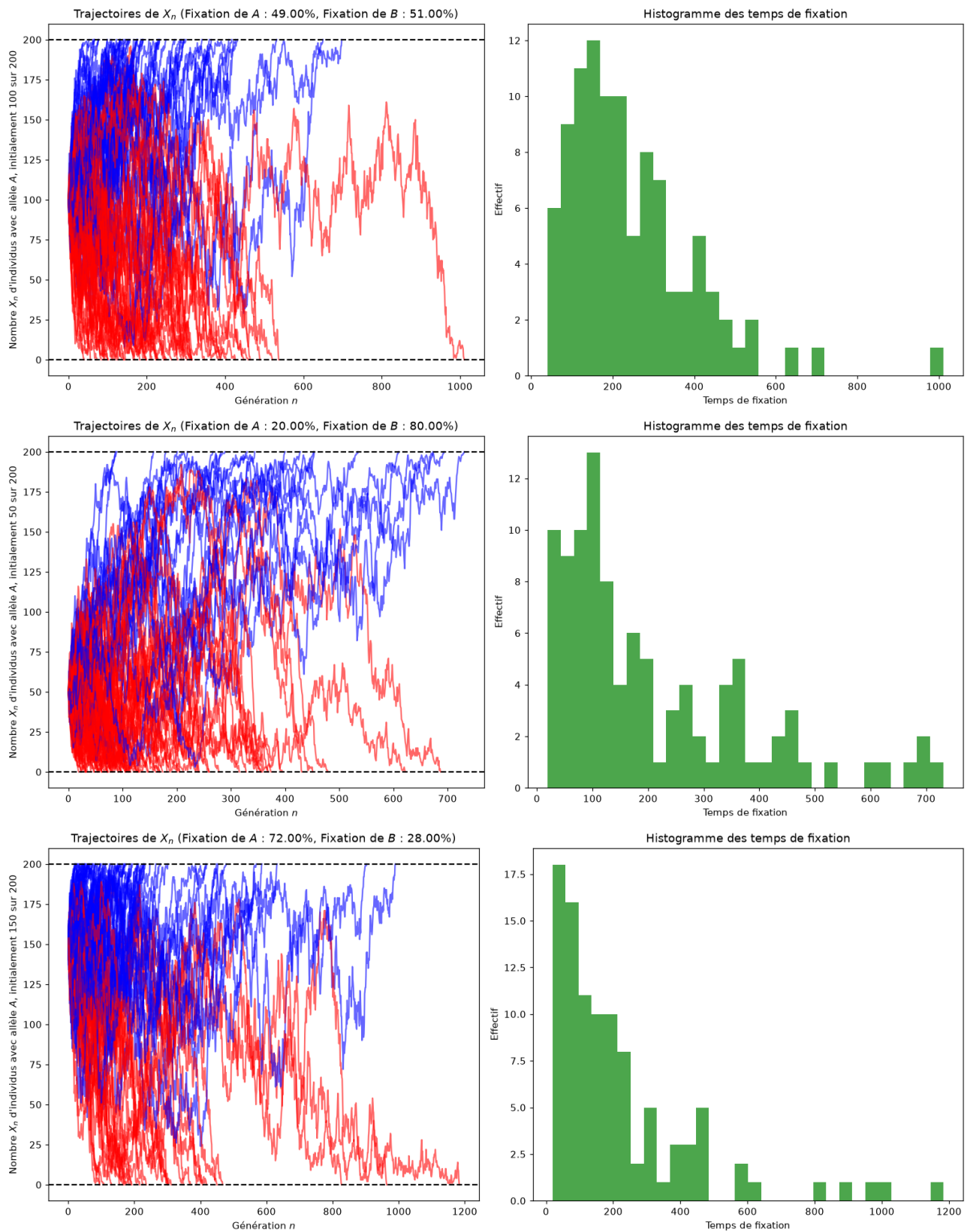


FIGURE 5.7 : Cent simulations indépendantes du modèle de Wright-Fisher, avec $N = 200$

Fonctions génératrices

Les fonctions génératrices fournissent un outil intéressant pour caractériser la loi d'une variable aléatoire à valeurs dans $\mathbb{N}$, et procéder à des calculs de moments (espérance, variance...). Elles sont de plus bien compatibles avec la notion d'indépendance. Elles sont définies de la manière suivante.

Définition A.1 (Fonction génératrice). Soit Y une variable aléatoire à valeurs dans l'ensemble $\mathbb{N}$. On appelle *fonction génératrice* de Y la série entière de la variable complexe définie par

$$G_Y(z) = \mathbb{E}[z^Y] = \sum_{k=0}^{\infty} \mathbb{P}(Y = k) z^k. \quad (\text{A.1})$$

Comme la série $\sum \mathbb{P}(Y = k)$ converge, la série entière qui définit $G_Y(z)$ converge absolument pour $|z| \leq 1$. Son rayon de convergence est donc au moins égal à 1. De plus, la fonction G_Y est continue à gauche au point 1, au sens que pour tout $\theta_0 \in [0, \pi/2[$,

$$G_Y(z) \xrightarrow[\substack{z \rightarrow 1 \\ z \in \Delta(\theta_0)}}{ } G_Y(1) = 1, \quad (\text{A.2})$$

où $\Delta(\theta_0)$ désigne le cône issu du point 1 et de demi-angle d'ouverture θ_0 , intersecté avec le disque unité ouvert, c'est-à-dire plus précisément le domaine

$$\Delta(\theta_0) = \{z \in \mathbb{C} \mid |z| < 1 \text{ et } z = 1 - \rho e^{i\theta}, \rho > 0, \theta \in [-\theta_0, \theta_0]\}, \quad (\text{A.3})$$

représenté sur la figure A.1. Ce résultat de continuité provient du théorème d'Abel ci-dessous.

Théorème A.1 (Théorème d'Abel). Soit $\sum a_k z^k$ une série entière de rayon de convergence au moins égal à 1 et de somme notée $f(z)$. Si la série $\sum a_k$ converge, alors pour tout $\theta_0 \in [0, \pi/2[$, on a

$$f(z) \xrightarrow[\substack{z \rightarrow 1 \\ z \in \Delta(\theta_0)}}{ } f(1).$$

Démonstration. On suppose que la série $\sum a_k$ converge. Pour tout $n \in \mathbb{N}$, on note respectivement S_n et R_n sa n -ième somme partielle et son n -ième reste, de sorte que $S_n + R_n = f(1)$. En effectuant une transformation d'Abel, on observe que pour tout $z \in \mathbb{C}$,

$$\begin{aligned} S_n - \sum_{k=0}^n a_k z^k &= \sum_{k=1}^n (R_{k-1} - R_k)(1 - z^k) = \sum_{k=0}^{n-1} R_k(1 - z^{k+1}) - \sum_{k=1}^n R_k(1 - z^k) \\ &= -R_n(1 - z^n) + (1 - z) \sum_{k=0}^{n-1} R_k z^k \end{aligned}$$

En supposant que $|z| < 1$ et en faisant tendre n vers l'infini, on obtient la relation

$$f(1) - f(z) = (1 - z) \sum_{k=0}^{\infty} R_k z^k.$$

Pour tout $n \in \mathbb{N}$, on en déduit à l'aide de l'inégalité triangulaire la majoration

$$|f(1) - f(z)| \leq |1 - z| \sum_{k=0}^n |R_k| + |1 - z| \left(\sup_{k \geq n+1} |R_k| \right) \sum_{k=n+1}^{\infty} |z|^k \leq |1 - z| \sum_{k=0}^n |R_k| + \frac{|1 - z|}{1 - |z|} \sup_{k \geq n+1} |R_k|.$$

De plus, si $z \in \Delta(\theta_0)$ et $|1 - z| \leq \cos \theta_0$, alors $z = 1 - \rho e^{i\theta}$ avec $\rho > 0$ et $\theta \in [-\theta_0, \theta_0]$, de sorte que

$$\frac{|1 - z|}{1 - |z|} = \frac{1 + |z|}{1 - |z|^2} |1 - z| \leq \frac{2\rho}{2\rho \cos \theta - \rho^2} \leq \frac{2}{2 \cos \theta_0 - |1 - z|}.$$

Ainsi, pour tout $\varepsilon > 0$, en choisissant $n \in \mathbb{N}$ assez grand pour que $\sup_{k \geq n+1} |R_k| \leq \varepsilon$, puis $\delta \in]0, \cos \theta_0]$ tel que $\delta \sum_{k=0}^n |R_k| \leq \varepsilon$, on obtient pour tout $z \in \Delta(\theta_0)$ vérifiant $|z - 1| \leq \delta$, la majoration

$$|f(z) - f(1)| \leq |z - 1| \sum_{k=0}^n |R_k| + \frac{2}{2 \cos \theta_0 - |z - 1|} \sup_{k \geq n+1} |R_k| \leq \varepsilon + \frac{2\varepsilon}{\cos \theta_0}.$$

On en conclut que $f(z)$ tend vers $f(1)$ quand z tend vers 1 en restant dans le domaine $\Delta(\theta_0)$. $\square$

On observe que toute fonction génératrice G_Y restreinte à l'intervalle réel $[0, \infty[$ est croissante et prend ses valeurs dans l'intervalle $[0, \infty]$. Pour $\theta_0 = 0$, on déduit de (A.2) que cette restriction est continue à gauche en 1. Cette propriété découle aussi du cas particulier, énoncé ci-dessous, du théorème de convergence monotone de Beppo Levi pour les séries.

Théorème A.2 (Théorème de Beppo Levi pour les séries). Soit $(a_{k,l})_{k,l \in \mathbb{N}}$ une famille à deux indices de nombres réels telle que $0 \leq a_{k,l} \leq a_{k,l+1}$ pour tous entiers $k, l \geq 0$. On a

$$\sum_{k=0}^{\infty} \sup_{l \in \mathbb{N}} a_{k,l} = \sup_{l \in \mathbb{N}} \sum_{k=0}^{\infty} a_{k,l} \in [0, \infty].$$

Démonstration. Il est clair que le membre de gauche de l'identité à démontrer est toujours supérieur au membre de droite. Réciproquement, comme par monotonie les suprema sur $l \in \mathbb{N}$ sont des limites quand $l \rightarrow \infty$, on a pour tout entier $n \in \mathbb{N}$,

$$\sum_{k=0}^n \sup_{l \in \mathbb{N}} a_{k,l} = \sum_{k=0}^n \lim_{l \rightarrow \infty} a_{k,l} = \lim_{l \rightarrow \infty} \sum_{k=0}^n a_{k,l} = \sup_{l \in \mathbb{N}} \sum_{k=0}^n a_{k,l} \leq \sup_{l \in \mathbb{N}} \sum_{k=0}^{\infty} a_{k,l},$$

ce qui conduit à l'identité annoncée, les deux membres pouvant éventuellement être égaux à ∞ . $\square$

Pour toute suite croissante de nombres réels positifs z_l de limite égale à 1, le théorème A.2 appliqué à la famille de réels $P(Y = k) z_l^k$ montre que la suite des réels $G_Y(z_l)$ converge vers $G_Y(1)$. On retrouve ainsi la continuité à gauche au point 1 de la restriction de G_Y à la variable réelle.

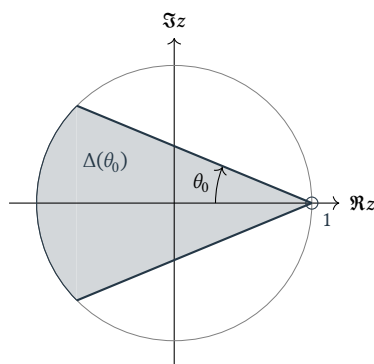


FIGURE A.1 : Niels Henrik Abel (1802–1829), le domaine défini par (A.3) et Beppo Levi (1875–1961)

Quels que soient $z \in \mathbb{C}$ et $r \in \mathbb{N}$, le *symbole de Pochhammer* $(z)_r$ est défini par la formule

$$(z)_r = \begin{cases} z(z-1) \cdots (z-r+1) & \text{si } r \geq 1 \\ 1 & \text{si } r = 0. \end{cases}$$

La fonction génératrice G_Y d'une variable aléatoire Y à valeurs dans $\mathbb{N}$ permet de calculer tous ses *moments factoriels* $E[(Y)_r]$, donc son espérance $E[Y]$ et sa variance $\text{Var}(Y)$.

Proposition A.3 (Dérivées de fonctions génératrices). Soit $r \in \mathbb{N}$ et soit Y une variable aléatoire à valeurs dans $\mathbb{N}$. On note $R_Y \geq 1$ le rayon de convergence de la série entière qui définit la fonction génératrice G_Y . La fonction G_Y , restreinte à la variable réelle, est r fois dérivable sur l'intervalle $]-R_Y, R_Y[$, et sa dérivée d'ordre r vérifie

$$G_Y^{(r)}(z) = \mathbb{E}[(Y)_r z^{Y-r}] \xrightarrow[z < 1]{z \rightarrow 1} \mathbb{E}[(Y)_r].$$

Dans ces expressions, le produit $(Y)_r z^{Y-r}$ est défini comme nul lorsque $Y < r$, et on adopte $0^0 = 1$ lorsque $Y = r$. De plus, si le moment factoriel $\mathbb{E}[(Y)_r]$ est fini, on a pour tout $\theta_0 \in [0, \pi/2[$,

$$\mathbb{E}[(Y)_r z^{Y-r}] \xrightarrow[z \in \Delta(\theta_0)]{z \rightarrow 1} \mathbb{E}[(Y)_r].$$

Démonstration. Une série entière d'une variable réelle est indéfiniment dérivable sur son intervalle ouvert de convergence, et les dérivées successives s'obtiennent en dérivant terme à terme. Ainsi,

$$G_Y^{(r)}(z) = \sum_{k=r}^{\infty} \mathbb{P}(Y = k) (k)_r z^{k-r} = \mathbb{E}[(Y)_r z^{Y-r}]$$

pour tout $z \in]-R_Y, R_Y[$. De plus, pour toute suite croissante de réels $z_l \in [0, 1[$ de limite égale à 1, le théorème A.2 appliqué à la famille $\mathbb{P}(Y = k) (k)_r z_l^{k-r}$ donne alors

$$\mathbb{E}[(Y)_r] = \sum_{k=r}^{\infty} \mathbb{P}(Y = k) (k)_r = \sum_{k=r}^{\infty} \sup_{l \in \mathbb{N}} \mathbb{P}(Y = k) (k)_r z_l^{k-r} = \sup_{l \in \mathbb{N}} G_Y^{(r)}(z_l) = \lim_{\substack{z \rightarrow 1 \\ z < 1}} G_Y^{(r)}(z).$$

La dernière égalité vient de la croissance de la fonction $G_Y^{(r)}$ sur l'intervalle $[0, 1[$. Le reste de l'énoncé découle du théorème A.1, appliqué à la série entière $\sum \mathbb{P}(Y = k) (k)_r z^{k-r}$ quand $\mathbb{E}[(Y)_r] < \infty$. $\square$

De plus, les fonctions génératrices caractérisent la loi d'une variable aléatoire à valeurs dans $\mathbb{N}$: deux variables aléatoires ayant la même fonction génératrice autour de 0 suivent la même loi.

Proposition A.4 (Caractérisation de la loi). Soient Y_1 et Y_2 deux variables aléatoires à valeurs dans $\mathbb{N}$. S'il existe $\delta > 0$ tel que $G_{Y_1}(z) = G_{Y_2}(z)$ pour tout $z \in]-\delta, \delta[$, alors Y_1 et Y_2 suivent la même loi, c'est-à-dire que

$$\forall k \in \mathbb{N} \quad \mathbb{P}(Y_1 = k) = \mathbb{P}(Y_2 = k).$$

Démonstration. Les fonctions génératrices de Y_1 et Y_2 coïncident au voisinage de 0, donc admettent la même valeur pour la dérivée d'ordre quelconque $r \in \mathbb{N}$ en ce point. D'après la proposition A.3,

$$G_{Y_1}^{(r)}(0) = \mathbb{E}[(Y_1)_r \mathbf{1}_{\{Y_1=r\}}] = r! \mathbb{P}(Y_1 = r),$$

et de même pour Y_2 . On en conclut que les probabilités $\mathbb{P}(Y_1 = r)$ et $\mathbb{P}(Y_2 = r)$ sont les mêmes. $\square$

Enfin, comme le montre le résultat suivant, les fonctions génératrices sont bien adaptées à l'étude des sommes de variables aléatoires indépendantes.

Proposition A.5 (Somme de variables indépendantes). Soit $n \in \mathbb{N}^*$, soient $Y_1, \dots, Y_n$ des variables aléatoires indépendantes à valeurs dans $\mathbb{N}$ et soit $z \in \mathbb{C}$. Si les séries définissant $G_{Y_1}(z), \dots, G_{Y_n}(z)$ comme dans (A.1) convergent absolument, alors la série définissant $G_{Y_1+\dots+Y_n}(z)$ converge absolument et on a

$$G_{Y_1+\dots+Y_n}(z) = G_{Y_1}(z) \cdots G_{Y_n}(z).$$

Ce résultat est aussi vrai si z est un nombre réel positif, auquel cas les fonctions génératrices peuvent éventuellement prendre la valeur ∞ .

Démonstration. Comme les variables aléatoires $Y_1, \dots, Y_n$ sont indépendantes, les variables aléatoires $Y_1 + \dots + Y_{n-1}$ et Y_n le sont. Le résultat se déduit alors par récurrence du cas où $n = 2$. Si les séries $\sum \mathbb{P}(Y_1 = k) z^k$ et $\sum \mathbb{P}(Y_2 = k) z^k$ convergent absolument, le produit de leurs sommes $G_{Y_1}(z)$ et $G_{Y_2}(z)$ est le produit de Cauchy des deux séries, qui est la série absolument convergente de terme général

$$\sum_{l=0}^k \mathbb{P}(Y_1 = l) z^l \mathbb{P}(Y_2 = k - l) z^{k-l} = z^k \sum_{l=0}^k \mathbb{P}(Y_1 = l \text{ et } Y_2 = k - l) = \mathbb{P}(Y_1 + Y_2 = k) z^k.$$

En sommant ces identités sur $k \in \mathbb{N}$, on en conclut que $G_{Y_1}(z) G_{Y_2}(z)$ est égal à $G_{Y_1+Y_2}(z)$. Si z est un nombre réel positif, on peut aussi écrire que

$$\begin{aligned} G_{Y_1+Y_2}(z) &= \sum_{k=0}^{\infty} \mathbb{P}(Y_1 + Y_2 = k) z^k = \sum_{k=0}^{\infty} \sum_{l=0}^k \mathbb{P}(Y_1 = l) z^l \mathbb{P}(Y_2 = k - l) z^{k-l} \\ &= \sum_{k=0}^{\infty} \sum_{\substack{(l_1, l_2) \in \mathbb{N}^2 \\ l_1 + l_2 = k}} \mathbb{P}(Y_1 = l_1) z^{l_1} \mathbb{P}(Y_2 = l_2) z^{l_2} = \sum_{(l_1, l_2) \in \mathbb{N}^2} \mathbb{P}(Y_1 = l_1) z^{l_1} \mathbb{P}(Y_2 = l_2) z^{l_2} \\ &= \sum_{l_1=0}^{\infty} \sum_{l_2=0}^{\infty} \mathbb{P}(Y_1 = l_1) z^{l_1} \mathbb{P}(Y_2 = l_2) z^{l_2} = \sum_{l_1=0}^{\infty} \mathbb{P}(Y_1 = l_1) z^{l_1} G_{Y_2}(z) = G_{Y_1}(z) G_{Y_2}(z), \end{aligned}$$

en raison de la proposition B.2 et du théorème B.3. □

Chaînes de Markov

Les chaînes de Markov sont des suites de variables aléatoires qui satisfont la propriété intuitive suivante : à tout instant donné, l'avenir ne dépend du passé qu'à travers l'état présent. Le but de cette annexe est de formaliser cette définition et d'examiner les propriétés qui en découlent.

B.1 Vecteurs et matrices indexés par un ensemble dénombrable

L'étude des chaînes de Markov nécessite de considérer des vecteurs et des matrices à coefficients dans l'intervalle $[0, \infty]$ et indexés par un ensemble dénombrable $\mathcal{S}$, fini ou infini. Pour les produits de coefficients, on adopte la convention $0 \cdot \infty = \infty \cdot 0 = 0$. On manipulera plus particulièrement des sommes de leurs coefficients. Considérons une famille $(u_s)_{s \in \mathcal{S}}$ d'éléments de l'intervalle $[0, \infty]$. Si l'ensemble $\mathcal{S}$ est fini, la définition de la somme $\sum_{s \in \mathcal{S}} u_s$ ne pose pas de problème puisqu'il n'y a qu'un nombre fini de termes. Si l'ensemble $\mathcal{S}$ est infini, on note

$$\sum_{s \in \mathcal{S}} u_s = \sup_{\substack{\mathcal{S}' \subseteq \mathcal{S} \\ \mathcal{S}' \text{ fini}}} \sum_{s \in \mathcal{S}'} u_s \in [0, \infty].$$

Comme l'indique le lemme suivant, on peut retrouver cette somme infinie à partir des énumérations de l'ensemble $\mathcal{S}$, c'est-à-dire des bijections de $\mathbb{N}$ dans $\mathcal{S}$. Cela permet de faire le lien avec la théorie des séries numériques à termes positifs.

Lemme B.1 (Somme sur un ensemble infini dénombrable). Soient $\mathcal{S}$ un ensemble dénombrable infini, et $(u_s)_{s \in \mathcal{S}}$ une famille d'éléments de $[0, \infty]$ indexée par $\mathcal{S}$. Pour toute énumération $(s_n)_{n \in \mathbb{N}}$ de $\mathcal{S}$, on a

$$\sum_{n=0}^{\infty} u_{s_n} = \sum_{s \in \mathcal{S}} u_s.$$

Démonstration. Soit $(s_n)_{n \in \mathbb{N}}$ une énumération de $\mathcal{S}$. On a d'abord

$$\sum_{s \in \mathcal{S}} u_s \geq \sum_{s \in \{s_n, 0 \leq n \leq N\}} u_s = \sum_{n=0}^N u_{s_n} \xrightarrow{N \rightarrow \infty} \sum_{n=0}^{\infty} u_{s_n}.$$

Réciproquement, pour tout ensemble fini $\mathcal{S}' \subseteq \mathcal{S}$, il existe $N \in \mathbb{N}$ tel que $\mathcal{S}' \subseteq \{s_n, 0 \leq n \leq N\}$, donc

$$\sum_{s \in \mathcal{S}'} u_s \leq \sum_{s \in \{s_n, 0 \leq n \leq N\}} u_s = \sum_{n=0}^N u_{s_n} \leq \sum_{n=0}^{\infty} u_{s_n}.$$

On conclut en prenant le supremum sur tous les ensembles finis $\mathcal{S}' \subseteq \mathcal{S}$. Dans tout ce qui précède, les sommes considérées peuvent éventuellement être égales à ∞ . $\square$

Remarque B.1 (Probabilité d'une union). Si $(B_n)_{n \in \mathbb{N}}$ est une suite d'événements disjoints, rappelons que $\mathbb{P}\left(\bigsqcup_{n=0}^{\infty} B_n\right) = \sum_{n=0}^{\infty} \mathbb{P}(B_n)$. Le symbole $\bigsqcup$ signifie que les événements constituant l'union sont disjoints. Plus généralement, pour toute famille $(A_s)_{s \in \mathcal{S}}$ d'événements disjoints indexés par un ensemble dénombrable $\mathcal{S}$, on a

$$\mathbb{P}\left(\bigsqcup_{s \in \mathcal{S}} A_s\right) = \sum_{s \in \mathcal{S}} \mathbb{P}(A_s). \quad (\text{B.1})$$

On applique en effet l'identité précédente à la suite des événements disjoints $B_n = A_{s_n}$, pour une énumération $(s_n)_{n \in \mathbb{N}}$ de $\mathcal{S}$, et on utilise le lemme B.1 pour conclure. Si les événements ne sont pas disjoints, rappelons que $\mathbb{P}\left(\bigcup_{n=0}^{\infty} B_n\right) \leq \sum_{n=0}^{\infty} \mathbb{P}(B_n)$, ce qui donne $\mathbb{P}\left(\bigcup_{s \in \mathcal{S}} A_s\right) \leq \sum_{s \in \mathcal{S}} \mathbb{P}(A_s)$.

On peut aussi regrouper par paquets disjoints les termes des sommes infinies.

Proposition B.2 (Regroupement par paquets). *Soit $(\mathcal{S}_t)_{t \in \mathcal{T}}$ une partition d'un ensemble dénombrable non vide $\mathcal{S}$, ce qui signifie que les ensembles $\mathcal{S}_t$ sont non vides, deux à deux disjoints et ont pour union l'ensemble $\mathcal{S}$. L'ensemble $\mathcal{T}$ est nécessairement dénombrable. De plus, pour toute famille $(u_s)_{s \in \mathcal{S}}$ d'éléments de l'intervalle $[0, \infty]$, on a*

$$\sum_{s \in \mathcal{S}} u_s = \sum_{t \in \mathcal{T}} \sum_{s \in \mathcal{S}_t} u_s.$$

Démonstration. Pour tout $t \in \mathcal{T}$, l'ensemble $\mathcal{S}_t$ contient un élément s_t . L'application $t \mapsto s_t$ est alors une injection de $\mathcal{T}$ dans $\mathcal{S}$. Comme $\mathcal{S}$ est dénombrable, $\mathcal{T}$ l'est aussi. Soit un ensemble fini $\mathcal{S}' \subseteq \mathcal{S}$. L'ensemble $\mathcal{T}'$ des indices $t \in \mathcal{T}$ tels que $\mathcal{S}' \cap \mathcal{S}_t \neq \emptyset$ est fini. Chaque ensemble $\mathcal{S}' \cap \mathcal{S}_t$, pour $t \in \mathcal{T}'$, est un sous-ensemble fini de $\mathcal{S}_t$. De plus, ces ensembles $\mathcal{S}' \cap \mathcal{S}_t$ forment une partition de $\mathcal{S}'$. Ainsi,

$$\sum_{s \in \mathcal{S}'} u_s = \sum_{t \in \mathcal{T}'} \sum_{s \in \mathcal{S}' \cap \mathcal{S}_t} u_s \leq \sum_{t \in \mathcal{T}'} \sum_{s \in \mathcal{S}_t} u_s \leq \sum_{t \in \mathcal{T}} \sum_{s \in \mathcal{S}_t} u_s.$$

On obtient une première inégalité en passant au supremum sur tous les ensembles finis $\mathcal{S}' \subseteq \mathcal{S}$.

Pour l'inégalité réciproque, on peut supposer que la somme $\sum_{s \in \mathcal{S}} u_s$ est finie. Considérons un ensemble fini $\mathcal{T}' \subseteq \mathcal{T}$ et un réel $\varepsilon > 0$. Pour $t \in \mathcal{T}'$, la somme $\sum_{s \in \mathcal{S}_t} u_s$ est finie, et il existe un ensemble fini $\mathcal{S}'_t \subseteq \mathcal{S}_t$ tel que $\sum_{s \in \mathcal{S}'_t} u_s \geq \sum_{s \in \mathcal{S}_t} u_s - \varepsilon$. L'union des $\mathcal{S}'_t$ sur $t \in \mathcal{T}'$ étant disjointe et finie,

$$\sum_{s \in \mathcal{S}} u_s \geq \sum_{s \in \bigcup_{t \in \mathcal{T}'} \mathcal{S}'_t} u_s = \sum_{t \in \mathcal{T}'} \sum_{s \in \mathcal{S}'_t} u_s \geq -\varepsilon \# \mathcal{T}' + \sum_{t \in \mathcal{T}'} \sum_{s \in \mathcal{S}_t} u_s.$$

On conclut en faisant tendre ε vers 0, puis en prenant le supremum sur les ensembles finis $\mathcal{T}' \subseteq \mathcal{T}$. $\square$

La proposition B.2 admet pour corollaire le résultat suivant, qui permet d'échanger l'ordre de sommation pour les séries à termes positifs à deux indices.

Théorème B.3 (Théorème de Fubini-Tonelli pour les séries à termes positifs). *Soit $(u_{k,l})_{k,l \in \mathbb{N}}$ une famille à deux indices d'éléments de l'intervalle $[0, \infty]$. On a*

$$\sum_{(k,l) \in \mathbb{N} \times \mathbb{N}} u_{k,l} = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} u_{k,l} = \sum_{l=0}^{\infty} \sum_{k=0}^{\infty} u_{k,l}.$$

Démonstration. On applique la proposition B.2 aux partitions $\mathcal{S} = \bigsqcup_{k=0}^{\infty} \{k\} \times \mathbb{N}$ et $\mathcal{S} = \bigsqcup_{l=0}^{\infty} \mathbb{N} \times \{l\}$ de l'ensemble dénombrable $\mathcal{S} = \mathbb{N} \times \mathbb{N}$, ainsi que le lemme B.1 aux énumérations $l \mapsto (k, l)$ et $k \mapsto (k, l)$ des ensembles $\{k\} \times \mathbb{N}$ et $\mathbb{N} \times \{l\}$, respectivement. On obtient

$$\sum_{(k,l) \in \mathbb{N} \times \mathbb{N}} u_{k,l} = \sum_{k=0}^{\infty} \sum_{(k,l) \in \{k\} \times \mathbb{N}} u_{k,l} = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} u_{k,l} \quad \text{et} \quad \sum_{(k,l) \in \mathbb{N} \times \mathbb{N}} u_{k,l} = \sum_{l=0}^{\infty} \sum_{(k,l) \in \mathbb{N} \times \{l\}} u_{k,l} = \sum_{l=0}^{\infty} \sum_{k=0}^{\infty} u_{k,l},$$

ce qui donne le résultat. $\square$

Remarque B.2 (Le théorème de Fubini-Tonelli implique celui de Beppo Levi pour les séries). Le théorème B.3 permet de retrouver le théorème A.2. En effet, considérons une famille à deux indices de nombres réels $(a_{k,l})_{k,l \in \mathbb{N}}$ telle que $0 \leq a_{k,l} \leq a_{k,l+1}$ pour tous entiers $k, l \geq 0$. Considérons $u_{k,0} = a_{k,0} \geq 0$ et $u_{k,l+1} = a_{k,l+1} - a_{k,l} \geq 0$ pour tous $k, l \geq 0$. Le théorème B.3 implique que

$$\sum_{k=0}^{\infty} \sup_{l \in \mathbb{N}} a_{k,l} = \sum_{k=0}^{\infty} \sup_{l \in \mathbb{N}} (u_{k,0} + \dots + u_{k,l}) = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} u_{k,l} = \sum_{l=0}^{\infty} \sum_{k=0}^{\infty} u_{k,l}$$



FIGURE B.1 : Guido Fubini (1879–1943) et Leonida Tonelli (1885–1946)

$$= \sup_{l \in \mathbb{N}} \left(\sum_{k=0}^{\infty} u_{k,0} + \cdots + \sum_{k=0}^{\infty} u_{k,l} \right) = \sup_{l \in \mathbb{N}} \sum_{k=0}^{\infty} (u_{k,0} + \cdots + u_{k,l}) = \sup_{l \in \mathbb{N}} \sum_{k=0}^{\infty} a_{k,l}.$$

On verra souvent comme des vecteurs lignes les familles d'éléments de l'intervalle $[0, \infty]$ indexés par un ensemble dénombrable $\mathcal{S}$. Considérons maintenant les matrices à coefficients dans $[0, \infty]$ et indexées par $\mathcal{S}$. Étant données deux telles matrices $P = (p_{s,t})_{s,t \in \mathcal{S}}$ et $Q = (q_{s,t})_{s,t \in \mathcal{S}}$, on définit le produit PQ comme étant la matrice dont les coefficients sont donnés par $\sum_{u \in \mathcal{S}} p_{s,u} q_{u,t}$, pour tous $s, t \in \mathcal{S}$. Cela permet de définir, pour toute matrice $P = (p_{s,t})_{s,t \in \mathcal{S}}$ et tout entier $m \in \mathbb{N}$, la puissance m -ième P^m , de coefficients notés $p_{s,t}^{(m)}$. On a alors le lemme suivant.

Lemme B.4 (Puissances itérées d'une matrice). Soit $P = (p_{s,t})_{s,t \in \mathcal{S}}$ une matrice à coefficients dans $[0, \infty]$ et indexée par un ensemble dénombrable $\mathcal{S}$. Pour tout $m \in \mathbb{N}$, le coefficient indexé par $s, t \in \mathcal{S}$ de la matrice P^m vérifie

$$p_{s,t}^{(m)} = \sum_{\substack{s_0, s_1, \dots, s_m \in \mathcal{S} \\ s_0 = s \text{ et } s_m = t}} \prod_{j=0}^{m-1} p_{s_j, s_{j+1}}.$$

Démonstration. La formule est vraie au rang $m = 0$. En effet, avec les conventions usuelles sur les sommes et produits indexés par l'ensemble vide, le membre de droite vaut 1 si $s = t$, et vaut 0 si $s \neq t$, tout comme le coefficient $p_{s,t}^{(0)}$. Supposons que la formule est vraie pour un certain entier $m \in \mathbb{N}$ et pour tous les indices $s, t \in \mathcal{S}$. Grâce à la proposition B.2, on a alors

$$p_{s,t}^{(m+1)} = \sum_{u \in \mathcal{S}} p_{s,u}^{(m)} p_{u,t} = \sum_{u \in \mathcal{S}} p_{u,t} \sum_{\substack{s_0, s_1, \dots, s_m \in \mathcal{S} \\ s_0 = s \text{ et } s_m = u}} \prod_{j=0}^{m-1} p_{s_j, s_{j+1}} = \sum_{\substack{s_0, s_1, \dots, s_{m+1} \in \mathcal{S} \\ s_0 = s \text{ et } s_{m+1} = t}} \prod_{j=0}^m p_{s_j, s_{j+1}},$$

ce qui montre que la formule est vraie au rang $m + 1$. On conclut grâce au principe de récurrence. $\square$

On peut maintenant fixer quelques définitions, utiles dans tout le chapitre, concernant les vecteurs et matrices à coefficients dans l'intervalle $[0, \infty]$ et indexés par un ensemble dénombrable. Dans toute la suite, les ensembles d'indices des matrices stochastiques sont supposés non vides.

Définition B.1 (Mesure, probabilité, matrice stochastique). Soit $\mathcal{S}$ un ensemble dénombrable. On définit les notions suivantes :

- (i). On appelle *mesure* sur $\mathcal{S}$ tout vecteur ligne $\mu = (\mu_s)_{s \in \mathcal{S}}$ à coefficients dans $[0, \infty]$;
- (ii). Une *mesure de probabilité*, ou plus simplement une *probabilité*, sur $\mathcal{S}$ est une mesure $\mu = (\mu_s)_{s \in \mathcal{S}}$ sur $\mathcal{S}$ dont la somme des coefficients est égale à 1, c'est-à-dire telle que

$$\sum_{s \in \mathcal{S}} \mu_s = 1.$$

- (iii). On appelle *matrice stochastique* sur $\mathcal{S}$ toute matrice $P = (p_{s,t})_{s,t \in \mathcal{S}}$ à coefficients réels positifs dont les sommes le long des lignes sont égales à 1, c'est-à-dire telle que

$$\forall s \in \mathcal{S} \quad \sum_{t \in \mathcal{S}} p_{s,t} = 1.$$

Les produits, donc les puissances, de matrices stochastiques sont des matrices stochastiques. Le produit d'une mesure de probabilité par une matrice stochastique est une mesure de probabilité.

B.2 Définitions et premières propriétés des chaînes de Markov

On définit la notion de chaîne de Markov de la manière suivante.

Définition B.2 (Chaîne de Markov, loi initiale). Soit $\mathcal{S}$ un ensemble dénombrable. On appelle *chaîne de Markov* à espace d'états $\mathcal{S}$ toute suite $(Y_n)_{n \geq 0}$ de variables aléatoires à valeurs dans $\mathcal{S}$ qui vérifie la *propriété de Markov* ci-après : pour tous $n \in \mathbb{N}$ et $y_0, \dots, y_n, t \in \mathcal{S}$ tels que $P(Y_0 = y_0, \dots, Y_n = y_n) > 0$, on a

$$P(Y_{n+1} = t \mid Y_0 = y_0, \dots, Y_n = y_n) = P(Y_{n+1} = t \mid Y_n = y_n). \quad (\text{B.2})$$

On appelle *loi initiale* d'une chaîne de Markov $(Y_n)_{n \geq 0}$ à espace d'états $\mathcal{S}$ la mesure de probabilité $\mu = (\mu_s)_{s \in \mathcal{S}}$ sur $\mathcal{S}$ telle que $\mu_s = P(Y_0 = s)$ pour tout $s \in \mathcal{S}$.

Le fait que la loi initiale soit une mesure de probabilité résulte de la formule (B.1). La propriété de Markov (B.2) stipule que, conditionnellement au passé, l'évolution future de la chaîne ne dépend que de l'état présent. Elle équivaut à la formulation plus générale suivante.

Lemme B.5 (Formulation équivalente de la propriété de Markov). Une suite $(Y_n)_{n \geq 0}$ de variables aléatoires à valeurs dans un ensemble dénombrable $\mathcal{S}$ est une chaîne de Markov si et seulement si pour tous les entiers $0 \leq n_0 < \dots < n_l$, avec $l \in \mathbb{N}$, et $m \in \mathbb{N}^*$, et pour tous les états $y_{n_0}, \dots, y_{n_l}, t \in \mathcal{S}$ tels que $P(Y_{n_0} = y_{n_0}, \dots, Y_{n_l} = y_{n_l}) > 0$, on a

$$P(Y_{n_l+m} = t \mid Y_{n_0} = y_{n_0}, \dots, Y_{n_l} = y_{n_l}) = P(Y_{n_l+m} = t \mid Y_{n_l} = y_{n_l}).$$

Démonstration. Il est clair que la propriété de Markov (B.2) est un cas particulier de la condition du lemme, obtenue en prenant $n_j = j$, pour $j \in \{0, \dots, n\}$, et $m = 1$. Réciproquement, supposons (B.2) et considérons des entiers $0 \leq n_0 < \dots < n_l$, avec $l \in \mathbb{N}$, et $m \in \mathbb{N}^*$, et des états $z_0, \dots, z_{n_l+m} \in \mathcal{S}$ tels que $P(Y_{n_l} = z_{n_l}, \dots, Y_{n_l+m-1} = z_{n_l+m-1}) > 0$. On observe que

$$\begin{aligned} P(Y_0 = z_0, \dots, Y_{n_l+m} = z_{n_l+m}) &= P(Y_0 = z_0, \dots, Y_{n_l+m-1} = z_{n_l+m-1}) \\ &\quad \cdot P(Y_{n_l+m} = z_{n_l+m} \mid Y_{n_l+m-1} = z_{n_l+m-1}). \end{aligned}$$

Cette identité résulte de la condition (B.2) utilisée à l'instant $n_l + m - 1$ si la probabilité en premier facteur à droite est non nulle, et est triviale sinon. Notons que la probabilité conditionnelle à droite est toujours correctement définie. En procédant de même aux instants $n_l + m - 2, \dots, n_l$, on obtient

$$P(Y_0 = z_0, \dots, Y_{n_l+m} = z_{n_l+m}) = P(Y_0 = z_0, \dots, Y_{n_l} = z_{n_l}) \prod_{j=n_l}^{n_l+m-1} P(Y_{j+1} = z_{j+1} \mid Y_j = z_j).$$

Considérons $y_{n_0}, \dots, y_{n_l}, t \in \mathcal{S}$ tels que $\mathbb{P}(Y_{n_0} = y_{n_0}, \dots, Y_{n_l} = y_{n_l}) > 0$. On choisit $z_{n_0} = y_{n_0}, \dots, z_{n_l} = y_{n_l}$ et $z_{n_l+m} = t$, puis on somme l'identité précédente pour toutes les valeurs possibles des autres états $z_j \in \mathcal{S}$ telles que $\mathbb{P}(Y_{n_l} = z_{n_l}, \dots, Y_{n_l+m-1} = z_{n_l+m-1}) > 0$. Grâce à la formule (B.1), on obtient

$$\mathbb{P}(Y_{n_l+m} = t \mid Y_{n_0} = y_{n_0}, \dots, Y_{n_l} = y_{n_l}) = \sum_{\substack{z_{n_l} = y_{n_l}, z_{n_l+1}, \dots, z_{n_l+m-1} \in \mathcal{S}, z_{n_l+m} = t \\ \mathbb{P}(Y_{n_l} = z_{n_l}, \dots, Y_{n_l+m-1} = z_{n_l+m-1}) > 0}} \prod_{j=n_l}^{n_l+m-1} \mathbb{P}(Y_{j+1} = z_{j+1} \mid Y_j = z_j).$$

En choisissant $z_{n_l} = y_{n_l}$ et $z_{n_l+m} = t$, puis en sommant de même l'identité précédente pour toutes les valeurs possibles des autres états $z_j \in \mathcal{S}$ telles que $\mathbb{P}(Y_{n_l} = z_{n_l}, \dots, Y_{n_l+m-1} = z_{n_l+m-1}) > 0$, on a aussi

$$\mathbb{P}(Y_{n_l+m} = t \mid Y_{n_l} = y_{n_l}) = \sum_{\substack{z_{n_l} = y_{n_l}, z_{n_l+1}, \dots, z_{n_l+m-1} \in \mathcal{S}, z_{n_l+m} = t \\ \mathbb{P}(Y_{n_l} = z_{n_l}, \dots, Y_{n_l+m-1} = z_{n_l+m-1}) > 0}} \prod_{j=n_l}^{n_l+m-1} \mathbb{P}(Y_{j+1} = z_{j+1} \mid Y_j = z_j).$$

Les deux probabilités conditionnelles de l'énoncé du lemme sont donc égales. $\square$

On s'intéresse désormais uniquement aux chaînes de Markov dont les lois probabilistes d'évolution ne varient pas au cours du temps. Pour être plus précis, on adopte la définition suivante.

Définition B.3 (Chaîne de Markov homogène, matrice de transition). Une chaîne de Markov $(Y_n)_{n \geq 0}$ à espace d'états dénombrable $\mathcal{S}$ est dite *homogène* s'il existe une matrice stochastique $P = (p_{s,t})_{s,t \in \mathcal{S}}$, appelée *matrice de transition* de la chaîne de Markov, telle que pour tous $n \in \mathbb{N}$ et $s, t \in \mathcal{S}$ vérifiant $\mathbb{P}(Y_n = s) > 0$, on a

$$\mathbb{P}(Y_{n+1} = t \mid Y_n = s) = p_{s,t}.$$

Les lois en dimension finie d'une chaîne de Markov homogène sont déterminées de manière unique par sa matrice de transition et sa loi initiale.

Proposition B.6 (Lois en dimension finie). Soit $(Y_n)_{n \geq 0}$ une chaîne de Markov homogène à espace d'états dénombrable $\mathcal{S}$, de matrice de transition $P = (p_{s,t})_{s,t \in \mathcal{S}}$ et de loi initiale $\mu = (\mu_s)_{s \in \mathcal{S}}$. Alors, pour tous $n \in \mathbb{N}$ et $s_0, \dots, s_n \in \mathcal{S}$,

$$\mathbb{P}(Y_0 = s_0, \dots, Y_n = s_n) = \mu_{s_0} \prod_{j=0}^{n-1} p_{s_j, s_{j+1}}.$$

Démonstration. La preuve se fait par récurrence sur n . Le résultat est clair au rang $n = 0$, par définition de la loi initiale d'une chaîne de Markov, puisque la formule devient simplement $\mathbb{P}(Y_0 = s_0) = \mu_{s_0}$ pour tout $s_0 \in \mathcal{S}$. Supposons la formule vraie au rang n , et choisissons des états $s_0, \dots, s_{n+1} \in \mathcal{S}$. Si on a $\mathbb{P}(Y_0 = s_0, \dots, Y_n = s_n) > 0$, la propriété de Markov (B.2) et de l'hypothèse de récurrence donnent

$$\begin{aligned} \mathbb{P}(Y_0 = s_0, \dots, Y_{n+1} = s_{n+1}) &= \mathbb{P}(Y_{n+1} = s_{n+1} \mid Y_n = s_n) \mathbb{P}(Y_0 = s_0, \dots, Y_n = s_n) \\ &= p_{s_n, s_{n+1}} \mu_{s_0} \prod_{j=0}^{n-1} p_{s_j, s_{j+1}} = \mu_{s_0} \prod_{j=0}^n p_{s_j, s_{j+1}}. \end{aligned}$$

Sinon, le membre de gauche ci-dessus est nul, et son membre de droite l'est aussi, puisque le produit $\mu_{s_0} \prod_{j=0}^{n-1} p_{s_j, s_{j+1}}$ est nul, par hypothèse de récurrence. Le résultat est donc vrai au rang $n + 1$. $\square$

D'après le résultat ci-dessous, les puissances de la matrice de transition d'une chaîne de Markov homogène donnent les probabilités de transition entre états après un certain nombre d'étapes. Elles permettent aussi de relier la loi de la chaîne à un instant donné à celles des instants ultérieurs.

Théorème B.7 (Équations de Chapman-Kolmogorov). Soit $(Y_n)_{n \geq 0}$ une chaîne de Markov homogène à espace d'états dénombrable $\mathcal{S}$, de matrice de transition $P = (p_{s,t})_{s,t \in \mathcal{S}}$. Pour tout $m \in \mathbb{N}$, on note $p_{s,t}^{(m)}$ les coefficients de la matrice P^m . Alors, pour tous $n, m \in \mathbb{N}$ et $s, t \in \mathcal{S}$ tels que $\mathbb{P}(Y_n = s) > 0$, on a

$$\mathbb{P}(Y_{n+m} = t \mid Y_n = s) = p_{s,t}^{(m)}.$$

De plus, si $\mu^{(n)}$ est la mesure de probabilité sur $\mathcal{S}$ dont les composantes sont $\mu_s^{(n)} = \mathbb{P}(Y_n = s)$, alors on a $\mu^{(n+m)} = \mu^{(n)} P^m$. On a en particulier $\mu^{(m)} = \mu P^m$, où μ est la loi initiale de la chaîne.

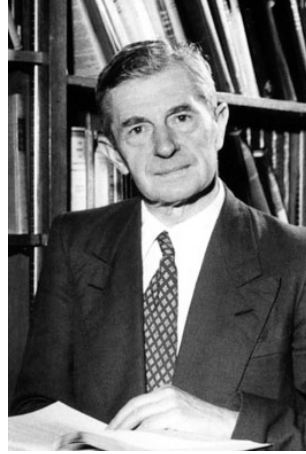


FIGURE B.2 : Andreï Andreïevitch Markov (1856–1922), Sydney Chapman (1888–1970) et Andreï Nikolaïevitch Kolmogorov (1903–1987)

Démonstration. On prouve le résultat portant sur les probabilités conditionnelles par récurrence sur m , l'hypothèse de récurrence étant que le résultat est vrai quels que soient s, t et n . Le résultat est clair si $m = 0$. Supposons qu'il est vrai pour un certain entier $m \in \mathbb{N}$. Pour $n \in \mathbb{N}$ et $s, t \in \mathcal{S}$, on a

$$\begin{aligned} \mathbb{P}(Y_{n+m+1} = t \text{ et } Y_n = s) &= \sum_{u \in \mathcal{S}} \mathbb{P}(Y_{n+m+1} = t, Y_{n+m} = u, Y_n = s) \\ &= \sum_{\substack{u \in \mathcal{S} \\ \mathbb{P}(Y_{n+m}=u \text{ et } Y_n=s) > 0}} \mathbb{P}(Y_{n+m+1} = t \mid Y_{n+m} = u \text{ et } Y_n = s) \mathbb{P}(Y_{n+m} = u \text{ et } Y_n = s) \\ &= \sum_{\substack{u \in \mathcal{S} \\ \mathbb{P}(Y_{n+m}=u \text{ et } Y_n=s) > 0}} \mathbb{P}(Y_{n+m+1} = t \mid Y_{n+m} = u) \mathbb{P}(Y_{n+m} = u \text{ et } Y_n = s) \\ &= \sum_{\substack{u \in \mathcal{S} \\ \mathbb{P}(Y_n=s) > 0}} p_{u,t} \mathbb{P}(Y_{n+m} = u \mid Y_n = s) \mathbb{P}(Y_n = s) = \mathbb{P}(Y_n = s) \sum_{u \in \mathcal{S}} p_{s,u}^{(m)} p_{u,t}. \end{aligned}$$

La première égalité résulte de la formule (B.1). La dernière somme ci-dessus est égale à $p_{s,t}^{(m+1)}$, donc le résultat est vrai au rang $m + 1$. Le principe de récurrence permet de conclure que le résultat est vrai pour tout $m \in \mathbb{N}$. Enfin, la formule (B.1) implique aussi que pour tous $n, m \in \mathbb{N}$ et $t \in \mathcal{S}$, on a

$$\mu_t^{(n+m)} = \mathbb{P}(Y_{n+m} = t) = \sum_{s \in \mathcal{S}} \mathbb{P}(Y_{n+m} = t \text{ et } Y_n = s) = \sum_{s \in \mathcal{S}} \mu_s^{(n)} p_{s,t}^{(m)}.$$

On a donc, en écriture matricielle, $\mu^{(n+m)} = \mu^{(n)} P^m$. Pour $n = 0$, on a en particulier $\mu^{(m)} = \mu^{(0)} P^m$, où le vecteur $\mu^{(0)}$ s'identifie à la loi initiale de la chaîne de Markov. $\square$

B.3 Visites successives d'un état

Soit $(Y_n)_{n \geq 0}$ une chaîne de Markov à espace d'états dénombrable $\mathcal{S}$. Pour tout état $s \in \mathcal{S}$ et tout entier $l \in \mathbb{N}$, considérons les variables aléatoires

$$T_0^{(s)} = \inf\{n \geq 0 \mid Y_n = s\} \quad \text{et} \quad T_{l+1}^{(s)} = \inf\{n \geq T_l^{(s)} + 1 \mid Y_n = s\}. \quad (\text{B.3})$$

On adopte la convention usuelle que l'infimum vaut ∞ si l'ensemble est vide. En particulier, si une variable $T_l^{(s)}$ est infinie, alors les variables ultérieures $T_{l+1}^{(s)}, T_{l+2}^{(s)}, \dots$ le sont aussi. Les variables $T_l^{(s)}$ qui sont finies donnent les instants de visite de l'état s par la chaîne de Markov. Le nombre total N_s de visites de l'état s est une variable aléatoire à valeurs dans $\overline{\mathbb{N}}$ telle que

$$N_s = \sum_{n=0}^{\infty} \mathbf{1}_{\{Y_n=s\}} = \sum_{l=0}^{\infty} \mathbf{1}_{\{T_l^{(s)} < \infty\}}. \quad (\text{B.4})$$

Le but de cette section est d'étudier la probabilité que les instants de visite soient finis, de décrire la loi du nombre total de visites, et d'exprimer son espérance.

B.3.1 Accessibilité d'un état

Pour commencer, intéressons-nous à la probabilité que la chaîne de Markov, partant d'un état s donné, accède à un état t donné. Cela revient au fait que le premier instant de visite $T_0^{(t)}$ de l'état t soit fini, et que le nombre de visites N_t de l'état t soit au moins 1. Étant donnée une matrice stochastique $P = (p_{s,t})_{s,t \in \mathcal{S}}$ indexée par un ensemble dénombrable $\mathcal{S}$, posons pour tous $s, t \in \mathcal{S}$,

$$\alpha_{s,t} = \begin{cases} p_{s,t}^{\rightarrow} & \text{si } s \neq t \\ 1 & \text{si } s = t \end{cases} \quad \text{avec} \quad p_{s,t}^{\rightarrow} = \sum_{n=1}^{\infty} \sum_{s_0, \dots, s_n \in \{s\} \times (\mathcal{S} \setminus \{t\})^{n-1} \times \{t\}} \prod_{j=0}^{n-1} p_{s_j, s_{j+1}}. \quad (\text{B.5})$$

Comme l'indique le résultat ci-dessous, le coefficient $\alpha_{s,t}$ défini par (B.5) donne la probabilité qu'une chaîne de Markov homogène de matrice de transition P qui part de l'état s accède à l'état t .

Lemme B.8 (Probabilité d'accès à un état). Soit $(Y_n)_{n \geq 0}$ une chaîne de Markov homogène à espace d'états dénombrable $\mathcal{S}$ et de matrice de transition $P = (p_{s,t})_{s,t \in \mathcal{S}}$. Les instants de visite et le nombre de visites d'un état $t \in \mathcal{S}$, définis par (B.3) et (B.4), sont notés $T_l^{(t)}$ et N_t . Pour tous $s, t \in \mathcal{S}$ tels que $\mathbb{P}(Y_0 = s) > 0$, on a

$$\mathbb{P}(N_t \geq 1 \mid Y_0 = s) = \mathbb{P}(T_0^{(t)} < \infty \mid Y_0 = s) = \alpha_{s,t},$$

où les coefficients $\alpha_{s,t}$ sont ceux définis dans (B.5).

Démonstration. D'après (B.4), les événements $\{N_t \geq 1\}$ et $\{T_0^{(t)} < \infty\}$ sont égaux. La première égalité est donc vraie. Si $s = t$, la seconde égalité est triviale, puisque $T_0^{(s)} = 0$ quand $Y_0 = s$. Si $s \neq t$, on a

$$\{Y_0 = s \text{ et } T_0^{(t)} < \infty\} = \bigsqcup_{n=1}^{\infty} \{Y_0 = s \text{ et } T_0^{(t)} = n\} = \bigsqcup_{n=1}^{\infty} \bigsqcup_{s_0, \dots, s_n \in \{s\} \times (\mathcal{S} \setminus \{t\})^{n-1} \times \{t\}} \{Y_0 = s_0, \dots, Y_n = s_n\}.$$

La proposition B.6 et la formule (B.1) donnent

$$\mathbb{P}(Y_0 = s \text{ et } T_0^{(t)} < \infty) = \sum_{n=1}^{\infty} \sum_{s_0, \dots, s_n \in \{s\} \times (\mathcal{S} \setminus \{t\})^{n-1} \times \{t\}} \mathbb{P}(Y_0 = s_0) \prod_{j=0}^{n-1} p_{s_j, s_{j+1}} = \mathbb{P}(Y_0 = s) \alpha_{s,t}.$$

On conclut en divisant par la probabilité de l'événement $\{Y_0 = s\}$. □

On peut comparer les coefficients donnés par (B.5) avec ceux des puissances de la matrice stochastique qui sert à les définir, et ceux de la matrice potentielle associée, définie comme suit.

Définition B.4 (Matrice potentielle). Soit $P = (p_{s,t})_{s,t \in \mathcal{S}}$ une matrice stochastique indexée par un ensemble dénombrable $\mathcal{S}$. Pour tout entier $m \in \mathbb{N}$, notons $p_{s,t}^{(m)}$ les coefficients de la matrice P^m . La *matrice potentielle* associée à P , notée P^* , est la matrice indexée par $\mathcal{S}$ dont les coefficients sont

$$p_{s,t}^* = \sum_{m=0}^{\infty} p_{s,t}^{(m)} \in [0, \infty].$$

De plus, on appelle matrice potentielle d'une chaîne de Markov homogène la matrice potentielle associée à la matrice de transition de cette chaîne.

On a alors le résultat de comparaison suivant.

Proposition B.9 (Caractérisations matricielles de l'accessibilité d'un état). Soit $P = (p_{s,t})_{s,t \in \mathcal{S}}$ une matrice stochastique indexée par un ensemble dénombrable $\mathcal{S}$. Pour tout $m \in \mathbb{N}$, notons $p_{s,t}^{(m)}$ les coefficients de la matrice P^m . Notons $p_{s,t}^*$ les coefficients de la matrice potentielle P^* , et $\alpha_{s,t}$ les coefficients définis par (B.5). Pour tous $s, t \in \mathcal{S}$, on a les inégalités

$$\sup_{m \in \mathbb{N}} p_{s,t}^{(m)} \leq \alpha_{s,t} \leq p_{s,t}^*,$$

et donc les équivalences

$$\begin{aligned} \alpha_{s,t} > 0 &\iff p_{s,t}^* > 0 \iff [\exists m \in \mathbb{N} : p_{s,t}^{(m)} > 0] \\ &\iff \left[\exists m \in \mathbb{N} \exists s_0, s_1, \dots, s_m \in \mathcal{S} : \begin{cases} s_0 = s \text{ et } s_m = t \\ \forall j \in \{0, \dots, m-1\}, p_{s_j, s_{j+1}} > 0 \end{cases} \right]. \end{aligned}$$

Démonstration. Considérons une chaîne de Markov homogène $(Y_n)_{n \geq 0}$ à espace d'états $\mathcal{S}$, de matrice de transition P , et telle que $\mathbb{P}(Y_0 = s) > 0$. En définissant $T_0^{(t)}$ par (B.3), on a pour tout $m_0 \in \mathbb{N}$,

$$\mathbb{P}(Y_{m_0} = t \mid Y_0 = s) \leq \mathbb{P}(T_0^{(t)} < \infty \mid Y_0 = s) = \mathbb{P}\left(\bigcup_{m=0}^{\infty} \{Y_m = t\} \mid Y_0 = s\right) \leq \sum_{m=0}^{\infty} \mathbb{P}(Y_m = t \mid Y_0 = s).$$

D'après le théorème B.7, le premier terme est égal à $p_{s,t}^{(m_0)}$ et le dernier terme est égal à $p_{s,t}^*$. D'après le lemme B.8, le deuxième terme est égal à $\alpha_{s,t}$. On en déduit les inégalités, puis les deux premières équivalences annoncées. La troisième équivalence découle du lemme B.4. $\square$

La notion d'accessibilité suivante exprime que toutes les assertions équivalentes mentionnées dans la proposition B.9 sont vérifiées.

Définition B.5 (Relations d'accessibilité et de communication). Soit $P = (p_{s,t})_{s,t \in \mathcal{S}}$ une matrice stochastique indexée par un ensemble dénombrable $\mathcal{S}$. Pour tous $s, t \in \mathcal{S}$, en définissant $\alpha_{s,t}$ par (B.5), on dit que :

- (i). t est *accessible* à partir de s via la matrice P , et on note $s \xrightarrow{P} t$, si $\alpha_{s,t} > 0$;
- (ii). s et t *communiquent* via la matrice P , et on note $s \leftrightarrow^P t$, si on a $s \xrightarrow{P} t$ et $t \xrightarrow{P} s$.

La relation $\leftrightarrow^P$ est une relation d'équivalence sur l'ensemble $\mathcal{S}$. En particulier, tout élément de $\mathcal{S}$ communique toujours avec lui-même. On étend naturellement la relation d'accessibilité aux classes d'équivalence en notant $C \xrightarrow{P} C'$ si et seulement si $s \xrightarrow{P} s'$ pour des états $s \in C$ et $s' \in C'$. On obtient ainsi une relation d'ordre entre les classes d'équivalence.

Les relations d'accessibilité et de communication entre éléments de $\mathcal{S}$ permettent ensuite de définir les propriétés ci-dessous sur les sous-ensembles de $\mathcal{S}$.

Définition B.6 (Fermeture et irréductibilité, éléments absorbants). Soient $\mathcal{S}$ un ensemble dénombrable et $P = (p_{s,t})_{s,t \in \mathcal{S}}$ une matrice stochastique indexée par $\mathcal{S}$. On introduit les qualificatifs suivants :

- (i). Un sous-ensemble C l'ensemble $\mathcal{S}$ est dit :
 - (a). *fermé* pour la matrice P si on a $p_{s,t} = 0$ quels que soient $s \in C$ et $t \in \mathcal{S} \setminus C$;
 - (b). *irréductible* pour la matrice P si on a $s \xleftrightarrow{P} t$ quels que soient $s, t \in C$;
- (ii). La matrice P est dite *irréductible* si l'ensemble $\mathcal{S}$ est irréductible pour P ;
- (iii). Un élément $s \in \mathcal{S}$ est dit *absorbant* si le singleton $\{s\}$ est un sous-ensemble fermé.

De plus, une chaîne de Markov homogène à espace d'états $\mathcal{S}$ est dite *irréductible* si sa matrice de transition est irréductible.

Les classes d'équivalence de l'ensemble $\mathcal{S}$ pour la relation $\xleftrightarrow{P}$ sont clairement des sous-ensembles irréductibles. De plus, la matrice P est irréductible si et seulement si $\mathcal{S}$ est la seule classe d'équivalence. Enfin, la proposition B.9 permet de montrer qu'un sous-ensemble C de l'ensemble $\mathcal{S}$ est fermé si et seulement si pour tous $s, t \in \mathcal{S}$,

$$[s \in C \text{ et } s \xrightarrow{P} t] \implies t \in C. \quad (\text{B.6})$$

Ainsi, une classe d'équivalence pour la relation $\xleftrightarrow{P}$ est un sous-ensemble fermé si et seulement si elle est maximale pour la relation d'ordre $\xrightarrow{P}$, c'est-à-dire qu'elle ne conduit à aucune autre classe. Dans ce cas, on dit que la classe est *finale*. Une classe non finale est constituée d'états transients. La réciproque est fautive en général : une classe finale peut être transiente lorsque l'espace d'états est infini.

B.3.2 Graphe d'accessibilité

Les propriétés d'accessibilité entre éléments via une matrice stochastique peuvent s'interpréter à l'aide d'un graphe orienté défini comme suit à partir de cette matrice.

Définition B.7 (Graphe d'accessibilité). Soit $P = (p_{s,t})_{s,t \in \mathcal{S}}$ une matrice stochastique indexée par un ensemble dénombrable $\mathcal{S}$. Le *graphe d'accessibilité* de P est le graphe orienté dont l'ensemble des sommets est $\mathcal{S}$, et l'ensemble des arêtes orientées est $\{(s, t) \in \mathcal{S}^2 \mid p_{s,t} > 0\}$.

L'existence de chemins dans ce graphe caractérise l'accessibilité entre états.

Proposition B.10 (Caractérisation par le graphe de l'accessibilité d'un état). Soit $P = (p_{s,t})_{s,t \in \mathcal{S}}$ une matrice stochastique indexée par un ensemble dénombrable $\mathcal{S}$. Pour tous $s, t \in \mathcal{S}$, on a la relation $s \xrightarrow{P} t$ si et seulement si $s = t$ ou il existe un chemin orienté allant de s à t dans le graphe d'accessibilité de P .

Démonstration. On rappelle que $s \xrightarrow{P} s$ pour tout $s \in \mathcal{S}$. De plus, pour $s \neq t$, la proposition B.9 montre que $s \xrightarrow{P} t$ si et seulement s'il existe $m \in \mathbb{N}^*$ et $s_0, s_1, \dots, s_m \in \mathcal{S}$ vérifiant $s_0 = s$ et $s_m = t$, ainsi que $p_{s_j, s_{j+1}} > 0$ pour tout entier $j \in \{0, \dots, m-1\}$. Cette dernière condition équivaut à l'existence d'un chemin orienté parcourant les sommets $s = s_0, s_1, \dots, s_m = t$ dans le graphe d'accessibilité de P . $\square$

D'après les propositions B.9 et B.10, les classes d'équivalence pour $\xleftrightarrow{P}$ donnent les composantes fortement connexes du graphe d'accessibilité de la matrice P , c'est-à-dire les sous-graphes maximaux dans lesquels il existe un chemin orienté d'un sommet quelconque à un autre.

B.3.3 Retour à un état

Intéressons-nous maintenant à la probabilité qu'une chaîne de Markov, partant d'un état s , revienne ultérieurement à cet état s . Cela revient au fait que $T_1^{(s)}$, le deuxième instant de visite de l'état s qui est donné par (B.3) soit fini. À

partir d'une matrice stochastique $P = (p_{s,t})_{s,t \in \mathcal{S}}$ et des formules (B.5), définissons pour tout $s \in \mathcal{S}$,

$$\rho_s = p_{s,s}^{\rightarrow} = \sum_{n=1}^{\infty} \sum_{s_0, \dots, s_n \in \{s\} \times (\mathcal{S} \setminus \{s\})^{n-1} \times \{s\}} \prod_{j=0}^{n-1} p_{s_j, s_{j+1}}. \quad (\text{B.7})$$

Comme l'indique le résultat ci-dessous, le coefficient ρ_s défini par (B.7) donne la probabilité qu'une chaîne de Markov homogène de matrice de transition P qui part de l'état s revienne à l'état s .

Lemme B.11 (Probabilité de retour à un état). Soit $(Y_n)_{n \geq 0}$ une chaîne de Markov homogène à espace d'états dénombrable $\mathcal{S}$ et de matrice de transition $P = (p_{s,t})_{s,t \in \mathcal{S}}$. Les instants de visite et le nombre de visites d'un état $s \in \mathcal{S}$, définis respectivement par (B.3) et (B.4), sont notés $T_1^{(s)}$ et N_s . Pour tout $s \in \mathcal{S}$ tels que $\mathbb{P}(Y_0 = s) > 0$, on a

$$\mathbb{P}(N_s \geq 2 \mid Y_0 = s) = \mathbb{P}(T_1^{(s)} < \infty \mid Y_0 = s) = \rho_s,$$

où les coefficients ρ_s sont ceux définis dans (B.7).

Démonstration. D'après (B.4), les événements $\{N_s \geq 2\}$ et $\{T_1^{(s)} < \infty\}$ sont égaux. La première égalité est donc vraie. Pour établir la seconde égalité, observons que

$$\{Y_0 = s \text{ et } T_1^{(s)} < \infty\} = \bigsqcup_{n=1}^{\infty} \{Y_0 = s \text{ et } T_1^{(s)} = n\} = \bigsqcup_{n=1}^{\infty} \bigsqcup_{s_0, \dots, s_n \in \{s\} \times (\mathcal{S} \setminus \{s\})^{n-1} \times \{s\}} \{Y_0 = s_0, \dots, Y_n = s_n\}.$$

En utilisant la proposition B.6 et la formule (B.1), comme dans la preuve du lemme B.8, on obtient

$$\mathbb{P}(Y_0 = s \text{ et } T_1^{(s)} < \infty) = \sum_{n=1}^{\infty} \sum_{s_0, \dots, s_n \in \{s\} \times (\mathcal{S} \setminus \{s\})^{n-1} \times \{s\}} \mathbb{P}(Y_0 = s_0) \prod_{j=0}^{n-1} p_{s_j, s_{j+1}} = \mathbb{P}(Y_0 = s) \rho_s,$$

et on conclut la preuve en divisant par la probabilité de l'événement $\{Y_0 = s\}$. $\square$

B.3.4 Nombre de visites d'un état

En combinant avec la propriété de Markov les propriétés d'accessibilité et de retour à un état étudiées précédemment, on peut déterminer la loi du nombre de visites d'un état par une chaîne de Markov homogène. C'est l'objet du résultat suivant.

Proposition B.12 (Loi du nombre de visites d'un état). Soit $(Y_n)_{n \geq 0}$ une chaîne de Markov homogène à espace d'états dénombrable $\mathcal{S}$ et de matrice de transition $P = (p_{s,t})_{s,t \in \mathcal{S}}$, à partir de laquelle on définit des coefficients $\alpha_{s,t}$ et ρ_t via (B.5) et (B.7). Les instants de visite et le nombre de visites d'un état $t \in \mathcal{S}$, définis par (B.3) et (B.4), sont notés $T_l^{(t)}$ et N_t . Pour tous les états $s, t \in \mathcal{S}$ tels que $\mathbb{P}(Y_0 = s) > 0$, et pour tout entier $l \in \mathbb{N}$, on a

$$\mathbb{P}(N_t \geq l + 1 \mid Y_0 = s) = \mathbb{P}(T_l^{(t)} < \infty \mid Y_0 = s) = \alpha_{s,t} \rho_t^l.$$

Par conséquent, on a pour tout $l \in \mathbb{N}^*$,

$$\mathbb{P}(N_t = l \mid Y_0 = s) = \alpha_{s,t} \rho_t^{l-1} (1 - \rho_t),$$

ainsi que

$$\mathbb{P}(N_t = 0 \mid Y_0 = s) = 1 - \alpha_{s,t},$$

et enfin

$$\mathbb{P}(N_t = \infty \mid Y_0 = s) = \begin{cases} \alpha_{s,t} & \text{si } \rho_t = 1 \\ 0 & \text{sinon.} \end{cases}$$

Démonstration. Soit $l \in \mathbb{N}$. La première égalité provient de (B.4), qui montre que les événements $\{N_t \geq l+1\}$ et $\{T_l^{(t)} < \infty\}$ sont égaux. De plus, on a $T_{l+1}^{(t)} < \infty$ si et seulement si il existe des entiers $0 \leq n_0 < \dots < n_{l+1}$ tels que pour tout $n \in \{0, \dots, n_{l+1}\}$, on a $Y_n = t$ si et seulement si $n \in \{n_0, \dots, n_{l+1}\}$. Pour $s_0, \dots, s_{n_{l+1}} \in \mathcal{S}$, la proposition B.6 donne

$$\begin{aligned} \mathbb{P}(Y_0 = s_0, \dots, Y_{n_{l+1}} = s_{n_{l+1}}) &= \mathbb{P}(Y_0 = s_0) \prod_{j=0}^{n_{l+1}-1} p_{s_j, s_{j+1}} \\ &= \left(\mathbb{P}(Y_0 = s_0) \prod_{j=0}^{n_l-1} p_{s_j, s_{j+1}} \right) \cdot \frac{1}{\mathbb{P}(\tilde{Y}_0 = s_{n_l})} \cdot \left(\mathbb{P}(\tilde{Y}_0 = s_{n_l}) \prod_{j=n_l}^{n_{l+1}-1} p_{s_j, s_{j+1}} \right) \\ &= \mathbb{P}(Y_0 = s_0, \dots, Y_{n_l} = s_{n_l}) \frac{1}{\mathbb{P}(\tilde{Y}_0 = s_{n_l})} \mathbb{P}(\tilde{Y}_0 = s_{n_l}, \dots, \tilde{Y}_{n_{l+1}-n_l} = s_{n_{l+1}}), \end{aligned}$$

pour toute chaîne de Markov homogène $(\tilde{Y}_n)_{n \geq 0}$ à espace d'états $\mathcal{S}$, de matrice de transition P , et vérifiant $\mathbb{P}(\tilde{Y}_0 = s_{n_l}) > 0$. Sommons ces identités sur tous les choix possibles d'états $s_0, \dots, s_{n_{l+1}} \in \mathcal{S}$ tels que $s_0 = s$, et $s_n = t$ si et seulement si $n \in \{n_0, \dots, n_{l+1}\}$. À l'aide de la formule (B.1), on obtient

$$\mathbb{P}(Y_0 = s, T_0^{(t)} = n_0, \dots, T_{l+1}^{(t)} = n_{l+1}) = \mathbb{P}(Y_0 = s, T_0^{(t)} = n_0, \dots, T_l^{(t)} = n_l) \frac{\mathbb{P}(\tilde{Y}_0 = t \text{ et } \tilde{T}_1^{(t)} = n_{l+1} - n_l)}{\mathbb{P}(\tilde{Y}_0 = t)},$$

si la chaîne auxiliaire $(\tilde{Y}_n)_{n \geq 0}$ est choisie de sorte que $\mathbb{P}(\tilde{Y}_0 = t) > 0$, et les instants de visite de l'état t par cette chaîne, définis comme en (B.3), sont notés $\tilde{T}_l^{(t)}$. En sommant l'identité ci-dessus sur tous les entiers $0 \leq n_0 < \dots < n_{l+1}$, on en déduit encore à l'aide de la formule (B.1) et de la proposition B.2 que

$$\begin{aligned} \mathbb{P}(Y_0 = s \text{ et } T_{l+1}^{(t)} < \infty) &= \sum_{0 \leq n_0 < \dots < n_{l+1}} \mathbb{P}(Y_0 = s, T_0^{(t)} = n_0, \dots, T_{l+1}^{(t)} = n_{l+1}) \sum_{n_{l+1}=n_l+1}^{\infty} \mathbb{P}(\tilde{T}_1^{(t)} = n_{l+1} - n_l \mid \tilde{Y}_0 = t) \\ &= \sum_{0 \leq n_0 < \dots < n_l} \mathbb{P}(Y_0 = s, T_0^{(t)} = n_0, \dots, T_l^{(t)} = n_l) \sum_{n=1}^{\infty} \mathbb{P}(\tilde{T}_1^{(t)} = n \mid \tilde{Y}_0 = t) \\ &= \mathbb{P}(Y_0 = s \text{ et } T_l^{(t)} < \infty) \mathbb{P}(\tilde{T}_1^{(t)} < \infty \mid \tilde{Y}_0 = t). \end{aligned}$$

En divisant par la probabilité, strictement positive, de l'événement $\{Y_0 = s\}$, et en appliquant le lemme B.11 à la chaîne $(\tilde{Y}_n)_{n \geq 0}$ qui a aussi P pour matrice de transition, on obtient

$$\mathbb{P}(T_{l+1}^{(t)} < \infty \mid Y_0 = s) = \mathbb{P}(T_l^{(t)} < \infty \mid Y_0 = s) \rho_t.$$

On peut alors facilement établir la deuxième égalité annoncée, par récurrence sur l , en utilisant la relation ci-dessus et la véracité de cette égalité au rang $l = 0$, en conséquence du lemme B.8.

Pour en déduire l'expression de la loi de la variable N_t conditionnellement à $Y_0 = s$, on observe simplement que l'événement $\{N_t = 0\}$ est le complémentaire de l'événement $\{N_t \geq 1\}$, et que $\{N_t \geq l\}$ est l'union disjointe de $\{N_t = l\}$ et $\{N_t \geq l+1\}$, pour tout $l \in \mathbb{N}^*$. Enfin, l'événement $\{N_t = \infty\}$ est le complémentaire de l'union disjointe des événements $\{N_t = l\}$, pour $l \in \mathbb{N}$. Cela permet de se ramener à des expressions de probabilités conditionnelles obtenues précédemment. $\square$

Si l'état t dont on étudie le nombre de visites est le même que l'état s de départ, le premier instant de visite est forcément nul. La proposition B.12 conduit donc au résultat suivant.

Corollaire B.13 (Loi du nombre de visites de l'état de départ). Soit $(Y_n)_{n \geq 0}$ une chaîne de Markov homogène à espace d'états dénombrable $\mathcal{S}$ et de matrice de transition $P = (p_{s,t})_{s,t \in \mathcal{S}}$, à partir de laquelle on définit des coefficients ρ_s via (B.7). Le nombre de visites N_s , défini par (B.4), d'un état $s \in \mathcal{S}$ tel que $\mathbb{P}(Y_0 = s) > 0$ vérifie :

- (i). Si $\rho_s = 1$, alors $\mathbb{P}(N_s = \infty \mid Y_0 = s) = 1$;
- (ii). Si $\rho_s < 1$, alors conditionnellement à $Y_0 = s$, la variable N_s suit la loi géométrique $\text{Géo}_{\mathbb{N}^*}(1 - \rho_s)$, c'est-à-dire que pour tout entier $l \in \mathbb{N}^*$, on a

$$\mathbb{P}(N_s = l \mid Y_0 = s) = \rho_s^{l-1}(1 - \rho_s).$$

Par conséquent, conditionnellement à $Y_0 = s$, le nombre $N_s - 1$ de retours à l'état s est presque sûrement infini si $\rho_s = 1$, et suit la loi géométrique $\text{Géo}_{\mathbb{N}}(1 - \rho_s)$ sinon.

Démonstration. C'est une conséquence directe de la proposition B.12, et du fait que $\alpha_{s,s} = 1$. □

Déduisons maintenant de la proposition B.12 une expression de l'espérance du nombre de visites d'un état par une chaîne de Markov.

Proposition B.14 (Première formule pour l'espérance du nombre de visites d'un état). Soit $(Y_n)_{n \geq 0}$ une chaîne de Markov homogène à espace d'états dénombrable $\mathcal{S}$ et de matrice de transition $P = (p_{s,t})_{s,t \in \mathcal{S}}$, à partir de laquelle on définit des coefficients $\alpha_{s,t}$ et ρ_t via (B.5) et (B.7). Les instants de visite et le nombre de visites d'un état $t \in \mathcal{S}$, définis par (B.3) et (B.4), sont notés $T_l^{(t)}$ et N_t . Pour tous états $s, t \in \mathcal{S}$ tels que $\mathbb{P}(Y_0 = s) > 0$, on a

$$\mathbb{E}[N_t \mid Y_0 = s] = \begin{cases} 0 & \text{si } \rho_t = 1 \text{ et } \alpha_{s,t} = 0 \\ \infty & \text{si } \rho_t = 1 \text{ et } \alpha_{s,t} > 0 \\ \frac{\alpha_{s,t}}{1 - \rho_t} & \text{sinon.} \end{cases}$$

Démonstration. On combine la formule (1.7) avec la proposition B.12. Ainsi,

$$\mathbb{E}[N_t \mid Y_0 = s] = \sum_{k=1}^{\infty} \mathbb{P}(N_t \geq k \mid Y_0 = s) = \sum_{l=0}^{\infty} \mathbb{P}(N_t \geq l+1 \mid Y_0 = s) = \sum_{l=0}^{\infty} \alpha_{s,t} \rho_t^l,$$

ce qui conduit directement à la formule annoncée. □

En exploitant la première égalité de (B.4), on peut aussi exprimer l'espérance du nombre de visites d'un état en fonction de la matrice potentielle de la chaîne, comme l'indique le résultat suivant.

Proposition B.15 (Deuxième formule pour l'espérance du nombre de visites d'un état). Soit $(Y_n)_{n \geq 0}$ une chaîne de Markov homogène à espace d'états dénombrable $\mathcal{S}$ et de matrice de transition $P = (p_{s,t})_{s,t \in \mathcal{S}}$. Notons $p_{s,t}^*$ les coefficients de la matrice potentielle P^* . Pour tous les états $s, t \in \mathcal{S}$, si $\mathbb{P}(Y_0 = s) > 0$, le nombre total de visites de t , défini par (B.4), vérifie la relation

$$\mathbb{E}[N_t \mid Y_0 = s] = p_{s,t}^*.$$

Démonstration. D'après le théorème B.7, quand la chaîne part de l'état s , l'espérance du nombre de visites de l'état t entre l'instant 0 et un instant $n \in \mathbb{N}$ est donnée par

$$\mathbb{E}\left[\mathbf{1}_{\{Y_0=s\}} \sum_{m=0}^n \mathbf{1}_{\{Y_m=t\}}\right] = \sum_{m=0}^n \mathbb{E}[\mathbf{1}_{\{Y_0=s\}} \mathbf{1}_{\{Y_m=t\}}] = \sum_{m=0}^n \mathbb{P}(Y_m = t \text{ et } Y_0 = s) = \mathbb{P}(Y_0 = s) \sum_{m=0}^n p_{s,t}^{(m)}.$$

Dans le membre de gauche, la variable aléatoire dont on prend l'espérance croît avec n et tend vers $\mathbf{1}_{\{Y_0=s\}} N_t$. Le théorème B.16 ci-après implique que

$$\mathbb{E}[\mathbf{1}_{\{Y_0=s\}} N_t] = \mathbb{E}\left[\sup_{n \in \mathbb{N}} \mathbf{1}_{\{Y_0=s\}} \sum_{m=0}^n \mathbf{1}_{\{Y_m=t\}}\right] = \sup_{n \in \mathbb{N}} \mathbb{E}\left[\mathbf{1}_{\{Y_0=s\}} \sum_{m=0}^n \mathbf{1}_{\{Y_m=t\}}\right]$$

$$= \sup_{n \in \mathbb{N}} \mathbb{P}(Y_0 = s) \sum_{m=0}^n p_{s,t}^{(m)} = \mathbb{P}(Y_0 = s) p_{s,t}^*.$$

On conclut en divisant par la probabilité, strictement positive, de l'événement $\{Y_0 = s\}$. $\square$

On a utilisé dans la preuve ci-dessus l'analogie suivant du théorème A.2 pour les espérances.

Théorème B.16 (Théorème de Beppo Levi pour les espérances). Soit $(Z_j)_{j \geq 0}$ une suite de variables aléatoires à valeurs dans $\overline{\mathbb{N}}$ telle que $0 \leq Z_j \leq Z_{j+1}$ avec probabilité 1, pour tout entier $j \in \mathbb{N}$. On a

$$\mathbb{E}\left[\sup_{j \in \mathbb{N}} Z_j\right] = \sup_{j \in \mathbb{N}} \mathbb{E}[Z_j] \in [0, \infty].$$

Démonstration. La variable aléatoire $Z_\infty = \sup_{j \in \mathbb{N}} Z_j$ est à valeurs dans $\overline{\mathbb{N}}$, et vérifie pour tout $k \in \mathbb{N}^*$,

$$\mathbb{P}(Z_\infty \geq k) = \mathbb{P}\left(\bigcup_{j=0}^{\infty} \uparrow \{Z_j \geq k\}\right) = \sup_{j \in \mathbb{N}} \mathbb{P}(Z_j \geq k),$$

où $\bigcup \uparrow$ signifie que les événements constituant l'union sont croissants. La dernière égalité découle du lemme B.17 ci-dessous. En combinant ces identités avec le lemme 1.1 et le théorème A.2, on a

$$\mathbb{E}[Z_\infty] = \sum_{k=1}^{\infty} \mathbb{P}(Z_\infty \geq k) = \sum_{k=1}^{\infty} \sup_{j \in \mathbb{N}} \mathbb{P}(Z_j \geq k) = \sup_{j \in \mathbb{N}} \sum_{k=1}^{\infty} \mathbb{P}(Z_j \geq k) = \sup_{j \in \mathbb{N}} \mathbb{E}[Z_j],$$

ce qui donne l'identité annoncée, avec éventuellement les deux membres égaux à ∞ . $\square$

La preuve précédente utilise le lemme ci-dessous, qui montre que la probabilité d'une union croissante d'événements est égale au supremum des probabilités de ces événements.

Lemme B.17 (Probabilité d'une union croissante d'événements). Soit $(A_j)_{j \geq 0}$ une suite d'événements telle que $A_j \subseteq A_{j+1}$ pour tout $j \in \mathbb{N}$. Alors,

$$\mathbb{P}\left(\bigcup_{j=0}^{\infty} \uparrow A_j\right) = \lim_{j \rightarrow \infty} \uparrow \mathbb{P}(A_j) = \sup_{j \in \mathbb{N}} \mathbb{P}(A_j),$$

où $\lim \uparrow$ signifie que les termes de la suite dont on considère la limite sont croissants.

Démonstration. Posons $B_0 = A_0$ ainsi que $B_j = A_j \setminus A_{j-1}$ pour $j \in \mathbb{N}^*$. On a $A_j = B_0 \sqcup \dots \sqcup B_j$. Ainsi,

$$\mathbb{P}\left(\bigcup_{j=0}^{\infty} \uparrow A_j\right) = \mathbb{P}\left(\bigsqcup_{l=0}^{\infty} B_l\right) = \sum_{l=0}^{\infty} \mathbb{P}(B_l) = \sup_{j \in \mathbb{N}} \sum_{l=0}^j \mathbb{P}(B_l) = \sup_{j \in \mathbb{N}} \mathbb{P}\left(\bigsqcup_{l=0}^j B_l\right) = \sup_{j \in \mathbb{N}} \mathbb{P}(A_j).$$

Le supremum est aussi une limite, puisque les probabilités $\mathbb{P}(A_j)$ forment une suite croissante. $\square$

Les formules données par les propositions B.14 et B.15 pour l'espérance du nombre de visites d'un état révèlent deux propriétés des matrices potentielles : chacun des coefficients diagonaux domine les autres coefficients de sa colonne ; la somme des coefficients sur chaque ligne est infinie.

Corollaire B.18 (Propriétés remarquables des coefficients de la matrice potentielle). Soient $\mathcal{S}$ un ensemble dénombrable et $P = (p_{s,t})_{s,t \in \mathcal{S}}$ une matrice stochastique indexée par $\mathcal{S}$. Définissons à partir de P des coefficients $\alpha_{s,t}$ via (B.5), et notons $p_{s,t}^*$ les coefficients de la matrice potentielle P^* . Pour tous $s, t \in \mathcal{S}$, avec la convention $0 \cdot \infty = 0$, on a

$$p_{s,t}^* = \alpha_{s,t} p_{t,t}^* \leq p_{t,t}^* \quad \text{et} \quad \sum_{u \in \mathcal{S}} p_{s,u}^* = \infty.$$

Démonstration. Soit $(Y_n)_{n \geq 0}$ une chaîne de Markov homogène à espace d'états dénombrable $\mathcal{S}$, de matrice de transition P , et de loi initiale telle que $\mathbb{P}(Y_0 = s) > 0$ et $\mathbb{P}(Y_0 = t) > 0$. Notons N_t le nombre de visites de t par cette chaîne, et ρ_t le coefficient défini par (B.7) à partir de P . Les propositions B.14 et B.15 donnent

$$p_{s,t}^* = \mathbb{E}[N_t | Y_0 = s] = \begin{cases} 0 & \text{si } \rho_t = 1 \text{ et } \alpha_{s,t} = 0 \\ \infty & \text{si } \rho_t = 1 \text{ et } \alpha_{s,t} > 0 \\ \frac{\alpha_{s,t}}{1 - \rho_t} & \text{sinon} \end{cases} \quad \text{et} \quad p_{t,t}^* = \mathbb{E}[N_t | Y_0 = t] = \begin{cases} \infty & \text{si } \rho_t = 1 \\ \frac{1}{1 - \rho_t} & \text{sinon.} \end{cases}$$

On en déduit la première partie du résultat en observant que $\alpha_{s,t} \leq 1$ à l'aide du lemme B.8. De plus,

$$\sum_{u \in \mathcal{S}} p_{s,u}^* = \sum_{u \in \mathcal{S}} \mathbb{E}[N_u | Y_0 = s] = \mathbb{E} \left[\sum_{u \in \mathcal{S}} \sum_{m=0}^{\infty} \mathbf{1}_{\{Y_m = u\}} \mid Y_0 = s \right] = \mathbb{E} \left[\sum_{m=0}^{\infty} \mathbf{1}_{\{Y_m \in \mathcal{S}\}} \mid Y_0 = s \right] = \infty.$$

La deuxième égalité provient du lemme B.1 et du théorème B.16. La troisième provient du théorème B.3 et du fait que les événements $\{Y_m = u\}$, pour $u \in \mathcal{S}$, forment une partition de $\{Y_m \in \mathcal{S}\}$. $\square$

B.4 Récurrence et transience

On peut distinguer les états d'une chaîne de Markov selon qu'en espérance, la chaîne y retourne un nombre fini ou infini de fois quand elle en part. D'après la proposition B.15, cela revient à distinguer selon la finitude des coefficients diagonaux de la matrice potentielle.

Définition B.8 (Récurrence et transience). Soient $\mathcal{S}$ un ensemble dénombrable et $P = (p_{s,t})_{s,t \in \mathcal{S}}$ une matrice stochastique indexée par $\mathcal{S}$. Notons $p_{s,t}^*$ les coefficients de la matrice potentielle P^* . On définit les notions suivantes :

- (i). Un élément $s \in \mathcal{S}$ est qualifié de *récurrent* pour P si $p_{s,s}^* = \infty$, et de *transient* si $p_{s,s}^* < \infty$;
- (ii). La matrice P est dite *récurrente* si tous les éléments de $\mathcal{S}$ sont récurrents pour cette matrice P , et *transiente* si tous les éléments sont transients.

Soit $(Y_n)_{n \geq 0}$ une chaîne de Markov homogène à espace d'états dénombrable $\mathcal{S}$ et de matrice de transition P . On définit les notions suivantes :

- (i). Un état $s \in \mathcal{S}$ est dit *récurrent* ou *transient* pour la chaîne selon qu'il est récurrent ou transient pour la matrice P , respectivement;
- (ii). La chaîne est dite *récurrente* si tous les états de $\mathcal{S}$ sont récurrents pour cette chaîne, et *transiente* s'ils sont tous transients.

La proposition B.15 indique que l'espérance du nombre de visites d'un état dont on est parti est finie si l'état est transient, et qu'elle est infinie si l'état est récurrent. Le résultat suivant montre que les nombres de visites sont en fait eux-mêmes finis ou infinis avec probabilité 1, selon que l'état est transient ou récurrent, respectivement. Il donne aussi une caractérisation de la transience et de la récurrence en fonction du premier instant de retour à un état dont on est parti.

Proposition B.19 (Caractérisations de la récurrence et la transience). Soit $(Y_n)_{n \geq 0}$ une chaîne de Markov homogène à espace d'états dénombrable $\mathcal{S}$, et soit $s \in \mathcal{S}$ un état tel que $\mathbb{P}(Y_0 = s) > 0$. Notons respectivement $T_1^{(s)}$ et N_s les instants de visite et le nombre de visites de s par la chaîne, définis par (B.3) et (B.4). Notons aussi $p_{s,t}^*$ les coefficients de la matrice potentielle de la chaîne, et ρ_s les coefficients définis par (B.7) à partir de sa matrice de transition. On a alors les caractérisations suivantes :

(i). L'état s est récurrent si et seulement si on a chacune des propriétés équivalentes suivantes :

$$p_{s,s}^* = \mathbb{E}[N_s | Y_0 = s] = \infty, \quad \mathbb{P}(N_s = \infty | Y_0 = s) = 1, \quad \rho_s = \mathbb{P}(T_1^{(s)} < \infty | Y_0 = s) = 1;$$

(ii). L'état s est transient si et seulement si on a chacune des propriétés équivalentes suivantes :

$$p_{s,s}^* = \mathbb{E}[N_s | Y_0 = s] < \infty, \quad \mathbb{P}(N_s < \infty | Y_0 = s) = 1, \quad \rho_s = \mathbb{P}(T_1^{(s)} < \infty | Y_0 = s) < 1.$$

Démonstration. On sait que $\rho_s = \mathbb{P}(T_1^{(s)} < \infty | Y_0 = s)$ et $p_{s,s}^* = \mathbb{E}[N_s | Y_0 = s]$, d'après le lemme B.11 et la proposition B.15, respectivement. En outre, le corollaire B.13 donne les implications suivantes :

$$\rho_s = \mathbb{P}(T_1^{(s)} < \infty | Y_0 = s) = 1 \implies \mathbb{P}(N_s = \infty | Y_0 = s) = 1 \implies p_{s,s}^* = \mathbb{E}[N_s | Y_0 = s] = \infty.$$

Si au contraire $\rho_s < 1$, la proposition B.14 appliquée avec $t = s$ donne

$$\infty \cdot \mathbb{P}(N_s = \infty | Y_0 = s) \leq \mathbb{E}[N_s | Y_0 = s] = \frac{1}{1 - \rho_s} < \infty,$$

ce qui conduit directement aux implications suivantes :

$$\rho_s = \mathbb{P}(T_1^{(s)} < \infty | Y_0 = s) < 1 \implies p_{s,s}^* = \mathbb{E}[N_s | Y_0 = s] < \infty \implies \mathbb{P}(N_s < \infty | Y_0 = s) = 1.$$

On en déduit aisément les équivalences annoncées. $\square$

Le corollaire B.18 sur les coefficients des matrices potentielles permet de montrer que la propriété de récurrence se perpétue via la relation d'accessibilité. On a en effet le résultat suivant.

Proposition B.20 (Récurrence et accessibilité). Soit $P = (p_{s,t})_{s,t \in \mathcal{S}}$ une matrice stochastique indexée par un ensemble dénombrable $\mathcal{S}$. Notons $p_{s,t}^*$ les coefficients de la matrice potentielle P^* , et $\alpha_{s,t}$ les coefficients définis par (B.5). Pour tous $s, t \in \mathcal{S}$, on a les équivalences

$$\begin{cases} p_{s,s}^* = \infty \\ \alpha_{s,t} = 1 \end{cases} \iff \begin{cases} s \text{ est récurrent} \\ s \xrightarrow{P} t \end{cases} \iff \begin{cases} t \text{ est récurrent} \\ t \xrightarrow{P} s \end{cases} \iff \begin{cases} p_{t,t}^* = \infty \\ \alpha_{t,s} = 1. \end{cases}$$

Ainsi, les éléments d'une même classe d'équivalence pour la relation de communication $\overset{P}{\longleftrightarrow}$ sur l'ensemble $\mathcal{S}$ sont soit tous récurrents, soit tous transients. De plus, les classes formées d'éléments récurrents sont des sous-ensembles fermés et irréductibles de $\mathcal{S}$.

Démonstration. Notons $p_{s,t}^{(m)}$ les coefficients de la matrice P^m . Si $s \xrightarrow{P} t$, la proposition B.9 donne l'existence d'un entier $m_0 \in \mathbb{N}$ tel que $p_{s,t}^{(m_0)} > 0$, et le corollaire B.18 implique que

$$p_{t,t}^* \geq p_{s,t}^* = \sum_{m=0}^{\infty} p_{s,t}^{(m)} \geq \sum_{m=m_0}^{\infty} p_{s,t}^{(m)} = \sum_{m=0}^{\infty} p_{s,t}^{(m+m_0)} = \sum_{m=0}^{\infty} \sum_{u \in \mathcal{S}} p_{s,u}^{(m)} p_{u,t}^{(m_0)} \geq \sum_{m=0}^{\infty} p_{s,s}^{(m)} p_{s,t}^{(m_0)} = p_{s,s}^* p_{s,t}^{(m_0)}.$$

La récurrence de s , caractérisée par $p_{s,s}^* = \infty$, implique donc celle de t , caractérisée par $p_{t,t}^* = \infty$.

Le cas $s = t$ étant immédiat, supposons désormais $s \neq t$. Considérons une chaîne de Markov homogène $(Y_n)_{n \geq 0}$ à espace d'états $\mathcal{S}$, de matrice de transition P , et vérifiant $\mathbb{P}(Y_0 = s) > 0$. Notons $T_0^{(t)}$ le premier instant de visite de t , défini par (B.3), puis le premier instant $U^{(s)}$ de visite de s après $T_0^{(t)}$, donné par

$$U^{(s)} = \inf\{n \geq T_0^{(t)} + 1 \mid Y_n = s\}.$$

On a $Y_0 = s$, $T_0^{(t)} < \infty$ et $U^{(s)} < \infty$ si et seulement s'il existe des entiers $1 \leq n_0 < n_1$ et des états $s_0, \dots, s_{n_1} \in \mathcal{S}$ tels que : $s_0 = s$; n_0 est l'unique indice $j \in \{1, \dots, n_0\}$ tel que $s_j = t$; n_1 est l'unique indice $j \in \{n_0 + 1, \dots, n_1\}$ tel que $s_j = s$. On procède maintenant comme dans la preuve de la proposition B.12. Pour tous $s_0, \dots, s_{n_1} \in \mathcal{S}$, la proposition B.6 donne

$$\mathbb{P}(Y_0 = s_0, \dots, Y_{n_1} = s_{n_1}) = \mathbb{P}(Y_0 = s_0, \dots, Y_{n_0} = s_{n_0}) \frac{1}{\mathbb{P}(\tilde{Y}_0 = s_{n_0})} \mathbb{P}(\tilde{Y}_0 = s_{n_0}, \dots, \tilde{Y}_{n_1-n_0} = s_{n_1}),$$

pour toute chaîne de Markov homogène $(\tilde{Y}_n)_{n \geq 0}$ à espace d'états $\mathcal{S}$, de matrice de transition P , et vérifiant $\mathbb{P}(\tilde{Y}_0 = s_{n_0}) > 0$. En sommant sur tous les choix convenables d'états $s_0, \dots, s_{n_1} \in \mathcal{S}$ et en utilisant la formule (B.1) et la proposition B.2, on obtient

$$\mathbb{P}(Y_0 = s, T_0^{(t)} = n_0, U^{(s)} = n_1) = \mathbb{P}(Y_0 = s \text{ et } T_0^{(t)} = n_0) \frac{\mathbb{P}(\tilde{Y}_0 = t \text{ et } \tilde{T}_0^{(s)} = n_1 - n_0)}{\mathbb{P}(\tilde{Y}_0 = t)},$$

si la chaîne auxiliaire $(\tilde{Y}_n)_{n \geq 0}$ est choisie de sorte que $\mathbb{P}(\tilde{Y}_0 = t) > 0$, et le premier instant de visite s par cette chaîne, défini comme en (B.3), est noté $\tilde{T}_0^{(s)}$. En sommant sur les entiers $1 \leq n_0 < n_1$ et en utilisant encore la formule (B.1) et la proposition B.2, on obtient

$$\begin{aligned} \mathbb{P}(Y_0 = s, T_0^{(t)} < \infty, U^{(s)} < \infty) &= \sum_{n_0=1}^{\infty} \mathbb{P}(Y_0 = s \text{ et } T_0^{(t)} = n_0) \sum_{n_1=n_0+1}^{\infty} \mathbb{P}(\tilde{T}_0^{(s)} = n_1 - n_0 \mid \tilde{Y}_0 = t) \\ &= \mathbb{P}(Y_0 = s \text{ et } T_0^{(t)} < \infty) \mathbb{P}(\tilde{T}_0^{(s)} < \infty \mid \tilde{Y}_0 = t) \\ &= \mathbb{P}(Y_0 = s) \alpha_{s,t} \alpha_{t,s}, \end{aligned}$$

où les coefficients $\alpha_{s,t}$ sont définis par (B.5). La dernière égalité découle du lemme B.8. Si l'état s est récurrent, la proposition B.19 montre qu'il est visité une infinité de fois par la chaîne $(Y_n)_{n \geq 0}$, conditionnellement à $\{Y_0 = s\}$. Dans ce cas, $U^{(s)}$ est fini si $T_0^{(t)}$ l'est. Le membre de gauche ci-dessus est donc égal à $\mathbb{P}(Y_0 = s) \alpha_{s,t}$ d'après le lemme B.8. On en déduit que $\alpha_{s,t} = \alpha_{s,t} \alpha_{t,s}$. Si on suppose de plus que $s \xrightarrow{P} t$, alors $\alpha_{s,t} > 0$, ce qui donne finalement $\alpha_{t,s} = 1 > 0$, donc en particulier $t \xrightarrow{P} s$. Comme t est récurrent, le même argument en échangeant s et t donne $\alpha_{s,t} = 1$. Cela achève la preuve des équivalences.

D'après les équivalences prouvées ci-dessus, si une classe d'équivalence pour la relation $\overset{P}{\leftrightarrow}$ contient un élément récurrent, tous les éléments de cette classe sont récurrents aussi. Enfin, une classe d'équivalence est toujours irréductible, et la caractérisation (B.6) montre directement qu'une classe formée d'éléments récurrents est un sous-ensemble fermé. $\square$

Le corollaire B.18 implique aussi l'existence d'au moins un élément récurrent dans les ensembles d'indices de matrices stochastiques qui sont finis. Combinée à la proposition B.20, cette propriété implique la récurrence de ces matrices quand elles sont de plus irréductibles.

Proposition B.21 (Récurrence des matrices indexées par un ensemble fini). Soient $\mathcal{S}$ un ensemble fini non vide et P une matrice stochastique indexée par $\mathcal{S}$. Alors :

- (i). L'ensemble $\mathcal{S}$ contient un élément récurrent pour la matrice P ;
- (ii). Si la matrice P est irréductible, alors elle est récurrente.

Démonstration. Notons $p_{s,t}^*$ les coefficients de la matrice potentielle P^* . Pour tout élément $s \in \mathcal{S}$, le corollaire B.18 montre que $\sum_{t \in \mathcal{S}} p_{s,t}^* = \infty$, donc que $p_{s,t}^* = \infty$ pour un élément $t \in \mathcal{S}$. Le corollaire B.18 donne $p_{t,t}^* \geq p_{s,t}^* = \infty$, ce qui signifie que l'élément t est récurrent.

Si la matrice P est irréductible, alors tous les éléments $s \in \mathcal{S}$ sont accessibles à partir de l'élément récurrent t via la matrice P , et sont donc récurrents aussi, d'après la proposition B.20. $\square$

Enfin, en mettant l'accent sur les chaînes de Markov plutôt que sur les matrices stochastiques, la proposition B.20 peut s'interpréter de la manière suivante.

Corollaire B.22 (Décomposition de l'espace d'états). Soit $(Y_n)_{n \geq 0}$ une chaîne de Markov homogène à espace d'états dénombrable $\mathcal{S}$. Notons N_t le nombre de visites de l'état t , défini par (B.4). L'espace $\mathcal{S}$ se décompose, à l'ordre des classes près, de manière unique en

$$\mathcal{S} = T \sqcup \bigsqcup_{i \in I} R_i,$$

où T est l'ensemble des états transients, I est un ensemble au plus dénombrable et les R_i sont les classes de communication récurrentes, non vides, fermées et irréductibles. Elles vérifient :

- (i). Pour tout $i \in I$ et tous $s, t \in R_i$ tels que $\mathbb{P}(Y_0 = s) > 0$, on a $\mathbb{P}(N_t = \infty \mid Y_0 = s) = 1$;
- (ii). Pour tous $i, j \in I$, $s \in R_i$ et $t \in R_j$ tels que $i \neq j$ et $\mathbb{P}(Y_0 = s) > 0$, on a $\mathbb{P}(N_t = 0 \mid Y_0 = s) = 1$.

L'ensemble T peut être vide. S'il n'y a aucun état récurrent, on prend $I = \emptyset$.

Démonstration. Notons d'une part T l'ensemble, éventuellement vide, des états transients pour la chaîne. D'autre part, considérons les classes d'équivalence pour la relation de communication $\overset{P}{\leftrightarrow}$ sur l'ensemble $\mathcal{S}$, où P désigne la matrice de transition de la chaîne. D'après la proposition B.20, ces classes sont formées d'éléments qui sont soit tous transients, soit tous récurrents. Dans le second cas, elles sont fermées et irréductibles. Notons $(R_i)_{i \in I}$ les classes constituées d'éléments récurrents. Comme $\mathcal{S}$ est dénombrable, l'ensemble d'indices I est au plus dénombrable. Ces classes sont déterminées de manière unique par la relation de communication, à leur indexation près.

Considérons deux états $s \in R_i$ et $t \in R_j$, avec $\mathbb{P}(Y_0 = s) > 0$. Si $i = j$, les états s et t sont récurrents et communiquent. D'après les propositions B.19 et B.20, les coefficients $\alpha_{s,t}$ et ρ_t , définis respectivement par (B.5) et (B.7) à partir de la matrice de transition de la chaîne valent tous les deux 1. Si au contraire $i \neq j$, les états s et t ne communiquent pas, donc on ne peut pas avoir simultanément $s \overset{P}{\rightarrow} t$ et $t \overset{P}{\rightarrow} s$. Cependant, la proposition B.20 montre que ces deux conditions d'accessibilité sont équivalentes pour des états récurrents. Ainsi, $\alpha_{s,t} = \alpha_{t,s} = 0$. On conclut à l'aide de la proposition B.12. $\square$

De même, la proposition B.21 peut s'interpréter comme suit en considérant des chaînes de Markov plutôt que des matrices stochastiques.

Corollaire B.23 (Existence d'un état récurrent dans un espace d'états fini). (i). Toute chaîne de Markov homogène à espace d'états fini admet un état récurrent.
(ii). Toute chaîne de Markov homogène et irréductible à espace d'états fini est récurrente.

Démonstration. Il suffit d'appliquer la proposition B.21 à la matrice stochastique de la chaîne. $\square$

B.5 Probabilité invariante

On s'intéresse dans le reste de ce chapitre au comportement asymptotique des chaînes de Markov. Si $(Y_n)_{n \geq 0}$ désigne une chaîne de Markov homogène à espace d'états dénombrable $\mathcal{S}$ et de matrice de transition P , et si $\mu^{(n)}$ désigne la mesure de probabilité sur $\mathcal{S}$ telle que $\mu_s^{(n)} = \mathbb{P}(Y_n = s)$, le théorème B.7 donne la relation de récurrence $\mu^{(n+1)} = \mu^{(n)}P$. Supposons pouvoir passer à la limite quand $n \rightarrow \infty$ dans cette relation. On obtient ainsi $\mu = \mu P$, où la mesure $\mu = (\mu_s)_{s \in \mathcal{S}}$ est la limite de $\mu^{(n)}$, composante par composante. Il s'agit d'une mesure à valeurs dans l'intervalle $[0, 1]$, qui n'est pas nécessairement une mesure de probabilité. Si μ est non nulle, elle peut être vue comme un vecteur propre à gauche de la matrice P , associé à la valeur propre 1. La mesure nulle est également invariante, mais ce n'est pas un vecteur propre. Une telle mesure μ relève de la définition ci-dessous.

Définition B.9 (Mesure invariante, probabilité invariante). Soit $\mathcal{S}$ un ensemble dénombrable. On définit les notions suivantes :

- (i). Une mesure μ sur $\mathcal{S}$ est dite *invariante* pour une matrice stochastique $P = (p_{s,t})_{s,t \in \mathcal{S}}$ indexée par $\mathcal{S}$ si on a la relation $\mu = \mu P$, c'est-à-dire que pour tout $t \in \mathcal{S}$,

$$\mu_t = \sum_{s \in \mathcal{S}} \mu_s p_{s,t};$$

- (ii). Une mesure μ sur $\mathcal{S}$ est dite *invariante* pour une chaîne de Markov homogène à espace d'états $\mathcal{S}$ si μ est mesure invariante pour la matrice de transition de la chaîne.

Une mesure invariante qui est une mesure de probabilité est appelée *probabilité invariante*.

La mesure nulle est invariante pour toute matrice stochastique. La mesure dont toutes les composantes valent ∞ est invariante si et seulement si aucune colonne de la matrice n'est identiquement nulle, avec la convention $0 \cdot \infty = 0$. De plus, l'ensemble des mesures invariantes pour une matrice stochastique est un *cône convexe*, ce qui signifie que toute combinaison linéaire à coefficients positifs de mesures invariantes est une mesure invariante.

Remarque B.3 (Invariance des lois unidimensionnelles). Soit P une matrice stochastique indexée par un ensemble dénombrable $\mathcal{S}$. Supposons que P admette une mesure invariante $\mu = (\mu_s)_{s \in \mathcal{S}}$ telle que $0 < \sum_{s \in \mathcal{S}} \mu_s < \infty$. En posant $\pi_s = \frac{\mu_s}{\sum_{t \in \mathcal{S}} \mu_t}$, on obtient une probabilité invariante $\pi = (\pi_s)_{s \in \mathcal{S}}$ pour la matrice P . Le théorème B.7 implique que pour toute chaîne de Markov $(Y_n)_{n \geq 0}$ à espace d'états $\mathcal{S}$, de matrice de transition P et de loi initiale π , on a $\mathbb{P}(Y_n = s) = \pi_s^{(n)} = (\pi P^n)_s = \pi_s$ pour tous $n \in \mathbb{N}$ et $s \in \mathcal{S}$.

Le résultat suivant implique qu'une chaîne de Markov homogène, irréductible et récurrente admet nécessairement une mesure invariante.

Proposition B.24 (Mesures invariantes dans le cas irréductible et récurrent). Soit P une matrice stochastique indexée par un ensemble dénombrable $\mathcal{S}$. Soient $t \in \mathcal{S}$ et $(Y_n)_{n \geq 0}$ la chaîne de Markov homogène à espace d'états $\mathcal{S}$, de matrice de transition P , et de loi initiale telle que $\mathbb{P}(Y_0 = t) = 1$. Notons $T_1^{(t)}$ les instants de visite définis par (B.3). Posons alors

$$\tau_t = \mathbb{E}[T_1^{(t)}] \quad \text{et} \quad \mu_s^{(t)} = \mathbb{E} \left[\sum_{n=1}^{T_1^{(t)}} \mathbf{1}_{\{Y_n = s\}} \right], \quad (\text{B.8})$$

pour tout $s \in \mathcal{S}$. Supposons que t est récurrent pour P . On a les propriétés suivantes :

- (i). La mesure $\mu^{(t)} = (\mu_s^{(t)})_{s \in \mathcal{S}}$ est invariante pour P , et elle vérifie $\mu_t^{(t)} = 1$ et $\sum_{s \in \mathcal{S}} \mu_s^{(t)} = \tau_t \geq 1$;

- (ii). Pour tout $s \in \mathcal{S}$, on a les implications
$$\begin{cases} s \xrightarrow{P} t \implies \mu_s^{(t)} < \infty \\ t \xrightarrow{P} s \implies \mu_s^{(t)} > 0 \\ \mu_s^{(t)} > 0 \implies s \xleftrightarrow{P} t; \end{cases}$$

- (iii). Si P est irréductible, alors $\mu^{(t)}$ est l'unique mesure invariante pour P telle que $\mu_t^{(t)} = 1$.

Démonstration. Prouvons (i). Comme t est récurrent et $\mathbb{P}(Y_0 = t) = 1$, la proposition B.19 implique que $\mathbb{P}(T_1^{(t)} < \infty) = 1$. De plus, pour $n \in \{1, \dots, T_1^{(t)}\}$, on a $Y_n = t$ si et seulement si $n = T_1^{(t)}$. Ainsi,

$$\mu_t^{(t)} = \mathbb{E} \left[\sum_{n=1}^{T_1^{(t)}} \mathbf{1}_{\{Y_n = t\}} \right] = \mathbb{E}[\mathbf{1}_{\{T_1^{(t)} < \infty\}}] = \mathbb{P}(T_1^{(t)} < \infty) = 1.$$

En utilisant le lemme B.1 et les théorèmes B.3 et B.16, on calcule ensuite

$$\sum_{s \in \mathcal{S}} \mu_s^{(t)} = \sum_{s \in \mathcal{S}} \mathbb{E} \left[\sum_{n=1}^{T_1^{(t)}} \mathbf{1}_{\{Y_n=s\}} \right] = \mathbb{E} \left[\sum_{s \in \mathcal{S}} \sum_{n=1}^{T_1^{(t)}} \mathbf{1}_{\{Y_n=s\}} \right] = \mathbb{E} \left[\sum_{n=1}^{T_1^{(t)}} \sum_{s \in \mathcal{S}} \mathbf{1}_{\{Y_n=s\}} \right] = \mathbb{E}[T_1^{(t)}] = \tau_t.$$

On a clairement $\tau_t \geq 1$ puisque $T_1^{(t)} \geq 1$ presque sûrement. Pour prouver l'invariance, posons $T = T_1^{(t)}$. Comme $Y_0 = Y_T = t$ presque sûrement, on a pour tout état u

$$\sum_{m=0}^{T-1} \mathbf{1}_{\{Y_m=u\}} = \sum_{m=1}^T \mathbf{1}_{\{Y_m=u\}},$$

donc les deux sommes ont pour espérance $\mu_u^{(t)}$. Pour tout $m \geq 0$, l'événement $\{T > m\}$ est déterminé par $Y_0, \dots, Y_m$. En décomposant cet événement selon les trajectoires jusqu'au temps m , la propriété de Markov donne, pour tous états u, s ,

$$\mathbb{P}(Y_m = u, Y_{m+1} = s, T > m) = p_{u,s} \mathbb{P}(Y_m = u, T > m).$$

Cette identité vaut aussi pour $m = 0$: sa contribution est $p_{t,s}$, puisque $Y_0 = t$ et $T \geq 1$. Par le théorème B.16 et le théorème B.3, on obtient

$$\begin{aligned} \mu_s^{(t)} &= \sum_{m=0}^{\infty} \mathbb{P}(Y_{m+1} = s, T > m) \\ &= \sum_{m=0}^{\infty} \sum_{u \in \mathcal{S}} p_{u,s} \mathbb{P}(Y_m = u, T > m) \\ &= \sum_{u \in \mathcal{S}} p_{u,s} \mathbb{E} \left[\sum_{m=0}^{T-1} \mathbf{1}_{\{Y_m=u\}} \right] = \sum_{u \in \mathcal{S}} \mu_u^{(t)} p_{u,s}. \end{aligned}$$

La mesure $\mu^{(t)}$ est donc invariante pour P .

Prouvons (ii). Considérons un entier $n \geq 0$ et notons $p_{s,t}^{(n)}$ les coefficients de P^n . Comme $\mu^{(t)}$ est une mesure invariante pour P , on a pour tout $u \in \mathcal{S}$,

$$\mu_u^{(t)} = (\mu^{(t)} P^n)_u = \sum_{v \in \mathcal{S}} \mu_v^{(t)} p_{v,u}^{(n)}. \quad (\text{B.9})$$

On en déduit que pour tout $s \in \mathcal{S}$,

$$1 = \mu_t^{(t)} \geq \mu_s^{(t)} p_{s,t}^{(n)} \quad \text{et} \quad \mu_s^{(t)} \geq \mu_t^{(t)} p_{t,s}^{(n)} = p_{t,s}^{(n)}.$$

Si $s \xrightarrow{P} t$, la proposition B.9 montre qu'il existe un entier $n \in \mathbb{N}$ tel que $p_{s,t}^{(n)} > 0$, donc $\mu_s^{(t)} < \infty$. Si $t \xrightarrow{P} s$, il existe de même un entier $n \in \mathbb{N}$ tel que $p_{t,s}^{(n)} > 0$, donc $\mu_s^{(t)} > 0$. Enfin, observons que

$$\sum_{n=1}^{T_1^{(t)}} \mathbf{1}_{\{Y_n=s\}} \leq \sum_{n=0}^{\infty} \mathbf{1}_{\{Y_n=s\}} = N_s,$$

où N_s désigne le nombre total de visites de s par la chaîne, défini par (B.4). Cette inégalité se transmet aux espérances, par exemple grâce au lemme 1.1. On en déduit que $\mu_s^{(t)} \leq p_{t,s}^*$, où $p_{t,s}^*$ est le coefficient de la matrice potentielle P^* . La valeur de l'espérance de N_s découle de la proposition B.15 et du fait que la chaîne a pour état initial t . Par conséquent, si $\mu_s^{(t)} > 0$, la proposition B.9 implique que $t \xrightarrow{P} s$. Dans ce cas, comme t est récurrent, la proposition B.20 montre que $s \xrightarrow{P} t$, donc finalement $s \leftrightarrow t$.

Pour prouver (iii), supposons que la matrice P est irréductible et notons ν une mesure invariante pour P , telle que $\nu_t = 1$. Alors, pour tout état $s \in \mathcal{S}$, on a

$$\nu_s = \nu_t p_{t,s} + \sum_{\substack{u \in \mathcal{S} \\ u \neq t}} \nu_u p_{u,s} = p_{t,s} + \sum_{\substack{u \in \mathcal{S} \\ u \neq t}} \nu_u p_{u,s} \geq p_{t,s}.$$

Cependant, la preuve de l'invariance de $\mu^{(t)}$ montre que pour tout entier $j \geq 1$ et tout état $s \in \mathcal{S}$,

$$\sum_{n=1}^j \mathbb{P}(Y_n = s \text{ et } n \leq T_1^{(t)}) = p_{t,s} + \sum_{\substack{u \in \mathcal{S} \\ u \neq t}} p_{u,s} \sum_{n=1}^{j-1} \mathbb{P}(Y_n = u \text{ et } n \leq T_1^{(t)})$$

et que le membre de gauche de cette égalité tend vers $\mu_s^{(t)}$ quand $j \rightarrow \infty$. On en déduit, par une récurrence élémentaire sur l'entier $j \geq 1$ que pour tout état $s \in \mathcal{S}$,

$$v_s \geq \sum_{n=1}^j \mathbb{P}(Y_n = s \text{ et } n \leq T_1^{(t)}) \xrightarrow{j \rightarrow \infty} \mu_s^{(t)}.$$

L'irréductibilité et l'invariance de v donnent, pour chaque s , un entier n tel que $p_{s,t}^{(n)} > 0$ et $1 = v_t \geq v_s p_{s,t}^{(n)}$. Ainsi, $v_s < \infty$, et $\mu_s^{(t)} \leq v_s$ est finie aussi. La différence $\lambda = v - \mu^{(t)}$ est donc bien définie coefficient par coefficient et est une mesure invariante pour la matrice P , telle que $\lambda_t = v_t - \mu_t^{(t)} = 0$. De même qu'en (B.9), on a $\lambda_t = \sum_{s \in \mathcal{S}} \lambda_s p_{s,t}^{(n)}$ pour tout entier $n \geq 0$. Pour tout état $s \in \mathcal{S}$, on a finalement $\lambda_s = 0$ dès qu'il existe un entier $n \geq 0$ tel que $p_{s,t}^{(n)} > 0$, ce qui est toujours le cas d'après la proposition B.9, puisque la matrice P est irréductible. Ainsi, $\lambda = 0$ et $v = \mu^{(t)}$. $\square$

L'existence d'une mesure invariante est aussi garantie dans le cadre de la définition suivante.

Définition B.10 (Réversibilité). Soit $\mathcal{S}$ un ensemble dénombrable. Une matrice stochastique $P = (p_{s,t})_{s,t \in \mathcal{S}}$ indexée par $\mathcal{S}$ est *réversible*, s'il existe une mesure $\mu = (\mu_s)_{s \in \mathcal{S}}$ non nulle, à coefficients finis $\mu_s \in [0, \infty]$, dite *réversible*, telle que pour tous $s, t \in \mathcal{S}$,

$$\mu_s p_{s,t} = \mu_t p_{t,s}.$$

De plus, une chaîne de Markov homogène à espace d'états $\mathcal{S}$ et de matrice de transition P est dite *réversible* si sa matrice de transition P et sa loi initiale μ vérifient la relation ci-dessus.

On peut renverser dans le temps une chaîne de Markov homogène et réversible $(Y_n)_{n \geq 0}$ à espace d'états dénombrable $\mathcal{S}$. En notant $P = (p_{s,t})_{s,t \in \mathcal{S}}$ la matrice de transition et $\mu = (\mu_s)_{s \in \mathcal{S}}$ la loi initiale de la chaîne, la relation de réversibilité et la proposition B.6 donnent pour tous $n \in \mathbb{N}$ et $s_0, \dots, s_n \in \mathcal{S}$,

$$\begin{aligned} \mathbb{P}(Y_0 = s_0, \dots, Y_n = s_n) &= \mu_{s_0} p_{s_0,s_1} p_{s_1,s_2} \cdots p_{s_{n-2},s_{n-1}} p_{s_{n-1},s_n} = p_{s_1,s_0} \mu_{s_1} p_{s_1,s_2} \cdots p_{s_{n-2},s_{n-1}} p_{s_{n-1},s_n} \\ &= \cdots = p_{s_1,s_0} p_{s_2,s_1} \cdots p_{s_{n-1},s_{n-2}} \mu_{s_n} p_{s_n,s_{n-1}} = \mu_{s_n} p_{s_n,s_{n-1}} p_{s_{n-1},s_{n-2}} \cdots p_{s_2,s_1} p_{s_1,s_0} \\ &= \mathbb{P}(Y_0 = s_n, \dots, Y_n = s_0). \end{aligned}$$

Le résultat suivant indique en outre qu'une telle chaîne admet une mesure invariante.

Proposition B.25 (Mesures invariantes dans le cas réversible). Soit P une matrice stochastique réversible indexée par un ensemble dénombrable. Alors, toute mesure réversible associée à P est invariante pour P .

Démonstration. Pour tout $t \in \mathcal{S}$, on calcule $\sum_{s \in \mathcal{S}} \mu_s p_{s,t} = \sum_{s \in \mathcal{S}} \mu_t p_{t,s} = \mu_t \sum_{s \in \mathcal{S}} p_{t,s} = \mu_t$, grâce à la relation de réversibilité et au fait que la matrice P est stochastique. $\square$

B.6 Récurrence positive

D'après la proposition B.19, un état récurrent s d'une chaîne de Markov homogène est caractérisé par un temps de retour presque sûrement fini à s quand la chaîne est conditionnée à partir de s . On peut alors distinguer les états récurrents selon que ce temps de retour a une espérance finie ou non.

Définition B.11 (Récurrence positive et récurrence nulle). Soit P une matrice stochastique indexée par un ensemble dénombrable $\mathcal{S}$. On définit ces notions :

- (i). Un élément $s \in \mathcal{S}$ récurrent pour la matrice P est dit *récurrent positif* pour la matrice P si $\tau_s < \infty$, où τ_s est défini par (B.8). Sinon, s est dit *récurrent nul* pour P ;
- (ii). La matrice P est dite *récurrente positive* si tous les éléments de $\mathcal{S}$ sont récurrents positifs pour P , et *récurrente nulle* s'ils sont tous récurrents nuls.

Soit $(Y_n)_{n \geq 0}$ une chaîne de Markov homogène à espace d'états dénombrable $\mathcal{S}$ et de matrice de transition P . On définit les notions suivantes :

- (i). Un état $s \in \mathcal{S}$ est dit *récurrent positif* ou *récurrent nul* pour la chaîne selon qu'il est récurrent positif ou récurrent nul pour la matrice P , respectivement;
- (ii). La chaîne est dite *récurrente positive* si tous les états de $\mathcal{S}$ sont récurrents positifs pour cette chaîne, et *récurrente nulle* s'ils sont tous récurrents nuls.

Tout comme les propriétés de récurrence et de transience, les propriétés de récurrence positive ou nulle dépendent d'une chaîne de Markov uniquement par sa matrice de transition. On peut cependant mettre l'accent sur la chaîne par l'intermédiaire du résultat suivant.

Proposition B.26 (Formulation équivalente de la récurrence positive). Soit $(Y_n)_{n \geq 0}$ une chaîne de Markov homogène à espace d'états dénombrable $\mathcal{S}$. Notons $T_1^{(s)}$ les instants de visite par la chaîne, définis par (B.3), d'un état $s \in \mathcal{S}$ tel que $\mathbb{P}(Y_0 = s) > 0$. On a

$$\mathbb{E}[T_1^{(s)} | Y_0 = s] = \mathbb{E}[\tilde{T}_1^{(s)}],$$

où $\tilde{T}_1^{(s)}$ désigne le premier instant de retour à l'état s pour la chaîne de Markov homogène $(\tilde{Y}_n)_{n \geq 0}$ à espace d'états $\mathcal{S}$, de même matrice de transition que $(Y_n)_{n \geq 0}$, et de loi initiale telle que $\mathbb{P}(\tilde{Y}_0 = s) = 1$. Par conséquent, si s est un état récurrent pour la chaîne $(Y_n)_{n \geq 0}$, il est récurrent positif pour cette chaîne si et seulement si $\mathbb{E}[T_1^{(s)} | Y_0 = s] < \infty$.

Démonstration. On observe grâce à la proposition B.6 que pour tout $n \in \mathbb{N}^*$ et tous $s_1, \dots, s_n \in \mathcal{S}$,

$$\mathbb{P}(Y_1 = s_1, \dots, Y_n = s_n | Y_0 = s) = \mathbb{P}(\tilde{Y}_1 = s_1, \dots, \tilde{Y}_n = s_n).$$

En sommant sur tous les états $s_1, \dots, s_n \in \mathcal{S} \setminus \{s\}$, on obtient directement

$$\mathbb{P}(T_1^{(s)} > n | Y_0 = s) = \mathbb{P}(\tilde{T}_1^{(s)} > n),$$

pour tout entier $n \geq 1$. L'égalité vaut aussi pour $n = 0$, puisque le temps de retour est toujours au moins 1. En remplaçant n par $n - 1$, on obtient donc l'égalité des probabilités $\mathbb{P}(T_1^{(s)} \geq n | Y_0 = s)$ et $\mathbb{P}(\tilde{T}_1^{(s)} \geq n)$ pour tout $n \geq 1$. On conclut grâce au lemme 1.1 et à (1.7). $\square$

Le résultat suivant met en évidence le lien entre récurrence positive et probabilité invariante.

Théorème B.27 (Récurrence positive et probabilité invariante). Pour toute matrice stochastique irréductible P indexée par un ensemble dénombrable non vide $\mathcal{S}$, les propriétés suivantes sont équivalentes :

- (i). La matrice P admet une probabilité invariante;
- (ii). Il existe dans $\mathcal{S}$ un élément récurrent positif pour la matrice P ;
- (iii). La matrice P est récurrente positive.

De plus, si la matrice P vérifie l'une des propriétés ci-dessus, alors sa probabilité invariante est unique, et il s'agit de la mesure $\pi = (\pi_t)_{t \in \mathcal{S}}$ telle que $\pi_t = 1/\tau_t$ pour tout $t \in \mathcal{S}$, avec τ_t défini par (B.8) et la convention $1/\infty = 0$.

Démonstration. L'implication (iii) $\Rightarrow$ (ii) est évidente. Pour prouver l'implication (ii) $\Rightarrow$ (i), supposons qu'il existe un élément $t \in \mathcal{S}$ récurrent positif pour la matrice P . La proposition B.24 donne une mesure invariante $\mu^{(t)}$ pour la matrice P , telle que $\sum_{s \in \mathcal{S}} \mu_s^{(t)} = \tau_t < \infty$. Il en ressort que la mesure $\pi_t \mu^{(t)}$ est une probabilité invariante pour la matrice P .

Pour établir l'implication (i) $\Rightarrow$ (iii), supposons que P admet une probabilité invariante $\nu = (\nu_t)_{t \in \mathcal{S}}$. De même qu'en (B.9), en notant $p_{s,t}^{(n)}$ les coefficients de P^n , on a pour tout $t \in \mathcal{S}$ et $n \in \mathbb{N}$,

$$\nu_t = \sum_{s \in \mathcal{S}} \nu_s p_{s,t}^{(n)}. \quad (\text{B.10})$$

D'après le corollaire B.18, les coefficients de la matrice potentielle P^* vérifient $p_{s,t}^* \leq p_{t,t}^* < \infty$ pour tous $s, t \in \mathcal{S}$ avec t transient pour P . La série définissant $p_{s,t}^*$ est alors convergente, donc son terme général vérifie $p_{s,t}^{(n)} \xrightarrow{n \rightarrow \infty} 0$. Si l'ensemble $\mathcal{S}$ est fini, on trouve directement $\nu_t = 0$ en passant à la limite dans (B.10). Sinon, pour toute énumération $(s_k)_{k \in \mathbb{N}}$ de $\mathcal{S}$ et tout $K \in \mathbb{N}$, le lemme B.1 donne

$$0 \leq \nu_t = \sum_{k=0}^{\infty} \nu_{s_k} p_{s_k,t}^{(n)} = \sum_{k=0}^K \nu_{s_k} p_{s_k,t}^{(n)} + \sum_{k=K+1}^{\infty} \nu_{s_k} p_{s_k,t}^{(n)} \leq \sum_{k=0}^K \nu_{s_k} p_{s_k,t}^{(n)} + \sum_{k=K+1}^{\infty} \nu_{s_k} \xrightarrow{n \rightarrow \infty} \sum_{k=K+1}^{\infty} \nu_{s_k}, \quad (\text{B.11})$$

et cette dernière limite tend vers 0 quand $K \rightarrow \infty$. On en déduit que $\nu_t = 0$ pour tout $t \in \mathcal{S}$ transient pour P . Ainsi, tout élément $t \in \mathcal{S}$ tel que $\nu_t > 0$ est récurrent pour la matrice P . De plus, dans ce cas, la mesure μ définie par $\mu_s = \nu_s / \nu_t$ est invariante pour P et vérifie $\mu_t = 1$. Comme P est irréductible, l'unicité dans la proposition B.24 implique que μ est la mesure $\mu^{(t)}$ donnée par (B.8). On a alors

$$\tau_t = \mathbb{E}[T_1^{(t)}] = \sum_{s \in \mathcal{S}} \mu_s^{(t)} = \sum_{s \in \mathcal{S}} \mu_s = \sum_{s \in \mathcal{S}} \frac{\nu_s}{\nu_t} = \frac{1}{\nu_t} < \infty. \quad (\text{B.12})$$

Tout élément $t \in \mathcal{S}$ tel que $\nu_t > 0$ est donc récurrent positif pour P . Il reste à montrer que tous les éléments $t \in \mathcal{S}$ vérifient $\nu_t > 0$. Pour cela, on observe à partir de (B.10) que $\nu_s p_{s,t}^{(n)} \leq \nu_t$ pour tous $s, t \in \mathcal{S}$ et $n \in \mathbb{N}$. Il existe nécessairement $s \in \mathcal{S}$ tel que $\nu_s > 0$. Pour tout $t \in \mathcal{S}$, comme P est irréductible, on a $p_{s,t}^{(n)} > 0$ pour un entier $n \in \mathbb{N}$, d'après la proposition B.9. Ainsi, $\nu_t > 0$.

Finalement, le calcul (B.12) montre que la probabilité invariante ν vérifie $\nu_t = \pi_t$ pour tout $t \in \mathcal{S}$. Il y a donc unicité de la probabilité invariante, qui est la mesure π . $\square$

La proposition B.20 indique que les éléments d'une même classe d'équivalence pour la relation de communication sont tous récurrents ou tous transients. On peut maintenant compléter ce résultat en distinguant les classes d'équivalence selon qu'elles sont récurrentes positives ou nulles.

Proposition B.28 (Récurrence positive et communication). Soit P une matrice stochastique indexée par un ensemble dénombrable $\mathcal{S}$. Les éléments d'une même classe d'équivalence pour la relation de communication $\overset{P}{\leftrightarrow}$ sur l'ensemble $\mathcal{S}$ sont soit tous récurrents positifs, soit tous récurrents nuls, soit tous transients.

Démonstration. Soit C une classe d'équivalence pour la relation de communication $\overset{P}{\leftrightarrow}$ sur l'ensemble $\mathcal{S}$. D'après la proposition B.20, les éléments de C sont tous récurrents ou tous transients. Supposons que C contient un élément récurrent positif t , et considérons un élément $s \in C$. Comme s et t sont récurrents, la proposition B.24 donne des mesures invariantes $\mu^{(s)}$ et $\mu^{(t)}$ pour la matrice P . Comme $s \overset{P}{\leftrightarrow} t$, on a $\mu_s^{(t)} > 0$, ce qui permet de considérer la mesure $\mu^{(t)} / \mu_s^{(t)}$. Il s'agit d'une mesure invariante pour la matrice P , qui vaut 1 en s . En inspectant la preuve de (iii) dans la proposition B.24, on voit que $\mu_u^{(t)} / \mu_s^{(t)} = \mu_u^{(s)}$ si $u \overset{P}{\rightarrow} s$, donc si $u \in C$. De plus, $\mu_u^{(s)} = \mu_u^{(t)} = 0$ si $u \notin C$. Par conséquent, on a

$$\tau_s = \sum_{u \in \mathcal{S}} \mu_u^{(s)} = \sum_{u \in C} \mu_u^{(s)} = \frac{1}{\mu_s^{(t)}} \sum_{u \in C} \mu_u^{(t)} = \frac{1}{\mu_s^{(t)}} \sum_{u \in \mathcal{S}} \mu_u^{(t)} = \frac{\tau_t}{\mu_s^{(t)}} < \infty.$$

Ainsi, tous les éléments de la classe C sont récurrents positifs, dès lors qu'un parmi eux l'est. $\square$

B.7 Apériodicité et convergence

Le théorème B.27 montre que toute matrice stochastique irréductible et récurrente positive admet une probabilité invariante donnée par les inverses des espérances des temps de retour. On peut alors s'intéresser à la convergence des puissances de la matrice vers la probabilité invariante. Cela revient à étudier la convergence des lois de la chaîne de Markov associée. Une condition suffisante de convergence est définie comme suit.

Définition B.12 (Période et apériodicité). Soit P une matrice stochastique indexée par un ensemble dénombrable $\mathcal{S}$. Notons $p_{s,t}^{(n)}$ les coefficients de la matrice P^n . La *période* d'un élément $s \in \mathcal{S}$ pour la matrice P est l'entier

$$d_s = \text{pgcd}\{n \in \mathbb{N} \mid p_{s,s}^{(n)} > 0\}.$$

Un élément $s \in \mathcal{S}$ est dit *apériodique* pour la matrice P si $d_s = 1$. La matrice P est dite *apériodique* si tous les éléments de $\mathcal{S}$ sont apériodiques pour P . Enfin, une chaîne de Markov à espace d'états $\mathcal{S}$ est dite *apériodique* si sa matrice de transition est apériodique.

Le résultat suivant montre que les coefficients des puissances d'une matrice stochastique irréductible et apériodique sont strictement positifs à partir d'un certain rang.

Lemme B.29 (Positivité des puissances d'une matrice irréductible et apériodique). Soit P une matrice stochastique indexée par un ensemble dénombrable $\mathcal{S}$. Notons $p_{s,t}^{(n)}$ les coefficients de la matrice P^n . Si P est irréductible et apériodique, alors

$$\forall s, t \in \mathcal{S} \quad \exists n_{s,t} \in \mathbb{N} \quad \forall n \geq n_{s,t} \quad p_{s,t}^{(n)} > 0.$$

Démonstration. Soit $t \in \mathcal{S}$. Comme P est apériodique, les entiers $m \in \mathbb{N}$ tels que $p_{t,t}^{(m)} > 0$ sont premiers entre eux. Il existe donc un nombre fini de ces entiers, notés $m_1, \dots, m_k \geq 1$, qui sont premiers entre eux. Si $k = 1$, on a nécessairement $m_1 = 1$ et tout entier naturel est un multiple de m_1 . Si $k \geq 2$, on applique le théorème B.30 ci-dessous. Dans les deux cas, il existe $n_t \in \mathbb{N}$ tel que pour tout entier $n \geq n_t$,

$$\exists a_1, \dots, a_k \in \mathbb{N} \quad n = a_1 m_1 + \dots + a_k m_k.$$

En écrivant la matrice P^n sous la forme $(P^{m_1})^{a_1} \dots (P^{m_k})^{a_k}$, on obtient alors

$$p_{t,t}^{(n)} \geq (p_{t,t}^{(m_1)})^{a_1} \dots (p_{t,t}^{(m_k)})^{a_k} > 0.$$

De plus, comme P est irréductible, pour tout $s \in \mathcal{S}$, il existe un entier $m_0 \in \mathbb{N}$ tel que $p_{s,t}^{(m_0)} > 0$. Ainsi,

$$p_{s,t}^{(n)} \geq p_{s,t}^{(m_0)} p_{t,t}^{(n-m_0)} > 0$$

pour tout entier $n \geq m_0 + n_t$. On conclut en posant $n_{s,t} = m_0 + n_t$. □

On a utilisé dans la preuve du lemme ci-dessus le théorème de Schur suivant sur le problème des pièces de monnaie de Frobenius.

Théorème B.30 (Théorème de Schur sur le problème des pièces de monnaie de Frobenius). Pour tout entier $k \geq 2$ et pour tous entiers $m_1, \dots, m_k \geq 1$ premiers entre eux, on a

$$\#\{(a_1, \dots, a_k) \in \mathbb{N}^k \mid n = a_1 m_1 + \dots + a_k m_k\} = \frac{n^{k-1}}{(k-1)! m_1 \dots m_k} + O_{n \rightarrow \infty}(n^{k-2}).$$

Par conséquent, on peut payer n'importe quelle somme d'argent suffisamment grande en utilisant des pièces de valeurs $m_1, \dots, m_k$.



FIGURE B.3 : Ferdinand Georg Frobenius (1849–1917), Issai Schur (1875–1941) et Wolfgang Döblin (1915–1940)

Démonstration. Notons b_n le cardinal dans le membre de gauche. La série entière associée vérifie

$$\sum_{n=0}^{\infty} b_n z^n = \sum_{a_1, \dots, a_k \in \mathbb{N}} z^{a_1 m_1 + \dots + a_k m_k} = \left(\sum_{a_1=0}^{\infty} z^{a_1 m_1} \right) \dots \left(\sum_{a_k=0}^{\infty} z^{a_k m_k} \right) = \frac{1}{1 - z^{m_1}} \dots \frac{1}{1 - z^{m_k}}.$$

Ce calcul est valable au moins pour $0 \leq z < 1$, par le théorème B.3. Le membre de droite est une fraction rationnelle qui admet 1 pour pôle d'ordre k , et les racines m_i -ièmes de l'unité pour pôles d'ordre strictement inférieur à k puisque les entiers m_i sont premiers entre eux. La décomposition en éléments simples de cette fraction rationnelle est de la forme

$$\sum_{\omega} \sum_{p=1}^{r_{\omega}} \frac{c_{\omega,p}}{(1 - z/\omega)^p} = \sum_{\omega} \sum_{p=1}^{r_{\omega}} \sum_{n=0}^{\infty} \binom{n+p-1}{p-1} c_{\omega,p} \omega^{-n} z^n = \sum_{n=0}^{\infty} \left(\sum_{\omega} \sum_{p=1}^{r_{\omega}} \binom{n+p-1}{p-1} c_{\omega,p} \omega^{-n} \right) z^n,$$

où la somme porte sur tous les pôles ω , dont les ordres respectifs sont notés r_{ω} . On a utilisé ici le développement en série entière (1.18). Par identification, on obtient

$$b_n = \sum_{\omega} \sum_{p=1}^{r_{\omega}} \binom{n+p-1}{p-1} c_{\omega,p} \omega^{-n} \quad \text{avec} \quad \binom{n+p-1}{p-1} \underset{n \rightarrow \infty}{\sim} \frac{n^{p-1}}{(p-1)!}.$$

Le terme dominant dans b_n est donc celui pour lequel p est maximal, c'est-à-dire $p = k$, obtenu uniquement pour le pôle $\omega = 1$. On conclut en observant que $c_{1,k} = 1/(m_1 \dots m_k)$. $\square$

Sous la condition d'apériodicité, on a alors le résultat de convergence suivant.

Théorème B.31 (Convergence dans le cas irréductible et apériodique). Soit P une matrice stochastique indexée par un ensemble dénombrable $\mathcal{S}$. Notons $p_{s,t}^{(n)}$ les coefficients de la matrice P^n , et τ_t l'espérance du premier temps de retour définie par (B.8). Si P est irréductible et apériodique, alors pour tous $s, t \in \mathcal{S}$, avec la convention $1/\infty = 0$, on a

$$p_{s,t}^{(n)} \xrightarrow{n \rightarrow \infty} \frac{1}{\tau_t}.$$

Démonstration. Supposons P irréductible. D'après la proposition B.28, la matrice P est transiente, récurrente positive, ou récurrente nulle. Si elle est transiente, le corollaire B.18 implique que tous les coefficients de la matrice potentielle P^* vérifient $p_{s,t}^* \leq p_{t,t}^* < \infty$. Comme la série qui définit $p_{s,t}^*$ est convergente, son terme général vérifie $p_{s,t}^{(n)} \xrightarrow{n \rightarrow \infty} 0$. De plus, la proposition B.19 montre que le temps de retour en t est infini avec une probabilité non nulle, ce qui donne $\tau_t = \infty$, et donc $1/\tau_t = 0$.

Supposons de plus P apériodique et récurrente positive. La preuve utilise un argument de *couplage* dû à Wolfgang Döblin. Considérons deux chaînes de Markov homogènes indépendantes $(Y_n)_{n \geq 0}$ et $(Y'_n)_{n \geq 0}$ à espace d'états $\mathcal{S}$ et

de matrice de transition P . Alors, $Y_n^\times = (Y_n, Y'_n)$ définit une chaîne de Markov homogène à espace d'états $\mathcal{S} \times \mathcal{S}$ dont la matrice de transition $P^\times$ a pour coefficients

$$p_{(s,s'),(t,t')}^\times = p_{s,t} p_{s',t'}.$$

Comme la matrice P est irréductible et apériodique, le lemme B.29 montre que pour tous $s, t, s', t' \in \mathcal{S}$ fixés, il existe des entiers $n_{s,t}, n_{s',t'} \in \mathbb{N}$ tels que pour tous entiers $n \geq n_{s,t}$ et $n' \geq n_{s',t'}$, on a $p_{s,t}^{(n)} > 0$ et $p_{s',t'}^{(n')} > 0$. On en déduit que pour tout $n \geq \max\{n_{s,t}, n_{s',t'}\}$, le coefficient de $(P^\times)^n$ vérifie

$$p_{(s,s'),(t,t')}^{\times(n)} = p_{s,t}^{(n)} p_{s',t'}^{(n)} > 0,$$

de sorte que la matrice $P^\times$ est irréductible et apériodique. De plus, d'après le théorème B.27, la matrice P admet pour probabilité invariante la mesure $\pi = (\pi_t)_{t \in \mathcal{S}}$ telle que $\pi_t = 1/\tau_t$ pour tout $t \in \mathcal{S}$. On en déduit par un simple calcul que $P^\times$ admet pour probabilité invariante la mesure donnée par

$$\pi_{(t,t')}^\times = \pi_t \pi_{t'}.$$

À nouveau d'après le théorème B.27, la matrice $P^\times$ est donc récurrente positive.

Pour tout $s \in \mathcal{S}$, notons $T_l^{(s,s)}$ les instants de visite par la chaîne $(Y_n^\times)_{n \geq 0}$ de l'état (s, s) , définis par (B.3). Considérons $s_0, t_0 \in \mathcal{S}$ tels que $(s_0, t_0) \neq (s, s)$ et $\mathbb{P}(Y_0^\times = (s_0, t_0)) > 0$. Comme la chaîne de Markov $(Y_n^\times)_{n \geq 0}$ est irréductible et récurrente, le lemme B.8 et la proposition B.20 montrent que

$$\mathbb{P}(T_0^{(s,s)} < \infty \mid Y_0^\times = (s_0, t_0)) = 1.$$

Ainsi, presque sûrement, conditionnellement à l'événement $\{Y_0^\times = (s_0, t_0)\}$, il existe un instant $n \geq 1$ tel que $Y_n^\times = (s, s)$. Cela reste clairement vrai si $s_0 = t_0 = s$. On en déduit que presque sûrement, le premier instant de passage T^Δ de la chaîne de Markov $(Y_n^\times)_{n \geq 0}$ sur la diagonale $\Delta = \{(s, s), s \in \mathcal{S}\}$ est fini, c'est-à-dire que presque sûrement,

$$T^\Delta = \inf\{n \geq 0 \mid Y_n = Y'_n\} < \infty.$$

Considérons des entiers tels que $0 \leq m < n$ et un état $t \in \mathcal{S}$. D'après la proposition B.6, comme les deux chaînes de Markov $(Y_n)_{n \geq 0}$ et $(Y'_n)_{n \geq 0}$ ont la même matrice de transition, on a pour tous états $s_0, s'_0, \dots, s_m, s'_m, s_{m+1}, \dots, s_{n-1} \in \mathcal{S}$ tels que $s_m = s'_m$,

$$\begin{aligned} & \mathbb{P}(Y_n = t, Y_{n-1} = s_{n-1}, \dots, Y_{m+1} = s_{m+1}, Y_m^\times = (s_m, s'_m), \dots, Y_0^\times = (s_0, s'_0)) \\ &= \mathbb{P}(Y'_n = t, Y'_{n-1} = s_{n-1}, \dots, Y'_{m+1} = s_{m+1}, Y_m^\times = (s_m, s'_m), \dots, Y_0^\times = (s_0, s'_0)). \end{aligned}$$

En sommant sur $(s_0, s'_0), \dots, (s_{m-1}, s'_{m-1}) \notin \Delta, (s_m, s'_m) \in \Delta, s_{m+1}, \dots, s_{n-1} \in \mathcal{S}$, on obtient

$$\begin{aligned} \mathbb{P}(Y_n = t \text{ et } T^\Delta = m) &= \mathbb{P}(Y_n = t, Y_{n-1} \in \mathcal{S}, \dots, Y_{m+1} \in \mathcal{S}, Y_m^\times \in \Delta, Y_{m-1}^\times \notin \Delta, \dots, Y_0^\times \notin \Delta) \\ &= \mathbb{P}(Y'_n = t, Y'_{n-1} \in \mathcal{S}, \dots, Y'_{m+1} \in \mathcal{S}, Y_m^\times \in \Delta, Y_{m-1}^\times \notin \Delta, \dots, Y_0^\times \notin \Delta) \\ &= \mathbb{P}(Y'_n = t \text{ et } T^\Delta = m). \end{aligned}$$

Cette égalité vaut aussi pour $m = n$, puisque $Y_n = Y'_n$ sur l'événement $\{T^\Delta = n\}$. En sommant sur tous les entiers $m \in \{0, \dots, n\}$, on en déduit que

$$\mathbb{P}(Y_n = t \text{ et } T^\Delta \leq n) = \mathbb{P}(Y'_n = t \text{ et } T^\Delta \leq n).$$

Supposons que $\mathbb{P}(Y_0 = s) = 1$ pour un état $s \in \mathcal{S}$ quelconque, et que la chaîne de Markov $(Y'_n)_{n \geq 0}$ admet π pour loi initiale. Comme π est invariante pour la chaîne de Markov $(Y'_n)_{n \geq 0}$, la remarque B.3 montre que $\mathbb{P}(Y'_n = t) = \pi_t$ pour tout $n \in \mathbb{N}$. D'après le théorème B.7 et l'égalité ci-dessus, on a alors

$$p_{s,t}^{(n)} - \pi_t = \mathbb{P}(Y_n = t) - \mathbb{P}(Y'_n = t) = \mathbb{P}(Y_n = t \text{ et } T^\Delta > n) - \mathbb{P}(Y'_n = t \text{ et } T^\Delta > n).$$

On en conclut avec l'inégalité triangulaire que

$$|p_{s,t}^{(n)} - \pi_t| \leq \mathbb{P}(Y_n = t \text{ et } T^\Delta > n) + \mathbb{P}(Y'_n = t \text{ et } T^\Delta > n) \leq 2 \mathbb{P}(T^\Delta > n) \xrightarrow{n \rightarrow \infty} 0.$$

La dernière convergence découle du lemme B.17 et du fait que les événements $\{T^\Delta \leq n\}$ sont croissants, et que leur union est l'événement $\{T^\Delta < \infty\}$ qui est de probabilité 1.

Finalement, supposons que la matrice P est récurrente nulle. Si la chaîne $(Y_n^\times)_{n \geq 0}$ est transiente, le même argument qu'au début de la preuve montre que $(p_{s,t}^{(n)})^2 = p_{(s,s),(t,t)}^{(n)} \xrightarrow{n \rightarrow \infty} 0$, et donc $p_{s,t}^{(n)} \xrightarrow{n \rightarrow \infty} 0$. Par récurrence nulle, on a $\tau_t = \infty$, donc $1/\tau_t = 0$, ce qui permet de conclure. Supposons que la chaîne $(Y_n^\times)_{n \geq 0}$ est récurrente. Elle est alors forcément récurrente nulle. En effet, son temps de retour à un état quelconque (s, s) est supérieur ou égal au temps de retour à l'état s pour la chaîne de Markov $(Y_n)_{n \geq 0}$, qui a une espérance infinie. Supposons par l'absurde ne pas avoir $p_{s,t}^{(n)} \xrightarrow{n \rightarrow \infty} 0$. Alors, par extraction diagonale, il existe une sous-suite d'entiers $0 \leq n_0 < n_1 < \dots$ et une mesure $\alpha = (\alpha_t)_{t \in \mathcal{S}}$ non nulle telle que pour tous $s, t \in \mathcal{S}$,

$$p_{s,t}^{(n_k)} \xrightarrow{k \rightarrow \infty} \alpha_t.$$

Les valeurs de ces limites sont indépendantes de s , car en faisant partir les chaînes $(Y_n)_{n \geq 0}$ et $(Y'_n)_{n \geq 0}$ d'états s et s' quelconques dans $\mathcal{S}$ respectivement, le raisonnement du cas précédent donne

$$p_{s,t}^{(n)} - p_{s',t}^{(n)} \xrightarrow{n \rightarrow \infty} 0.$$

En notant σ la somme de toutes les limites α_t , on observe alors que

$$0 < \sigma = \sum_{t \in \mathcal{S}} \alpha_t = \sup_{\substack{\mathcal{S}' \subseteq \mathcal{S} \\ \mathcal{S}' \text{ fini}}} \sum_{t \in \mathcal{S}'} \alpha_t = \sup_{\substack{\mathcal{S}' \subseteq \mathcal{S} \\ \mathcal{S}' \text{ fini}}} \lim_{k \rightarrow \infty} \sum_{t \in \mathcal{S}'} p_{s,t}^{(n_k)} \leq 1.$$

De plus, pour tout sous-ensemble fini $\mathcal{S}' \subseteq \mathcal{S}$, on a

$$\sum_{u \in \mathcal{S}'} p_{s,u}^{(n_k)} p_{u,t} \leq p_{s,t}^{(n_k+1)} = \sum_{u \in \mathcal{S}} p_{s,u} p_{u,t}^{(n_k)}.$$

Par passage à la limite quand $k \rightarrow \infty$ comme dans (B.11) puis au supremum sur $\mathcal{S}'$, on obtient

$$\sum_{u \in \mathcal{S}} \alpha_u p_{u,t} = \sup_{\substack{\mathcal{S}' \subseteq \mathcal{S} \\ \mathcal{S}' \text{ fini}}} \sum_{u \in \mathcal{S}'} \alpha_u p_{u,t} \leq \sum_{u \in \mathcal{S}} p_{s,u} \alpha_t = \alpha_t.$$

L'inégalité ci-dessus est toujours une égalité. En effet, si on a une inégalité stricte pour un état $t_0 \in \mathcal{S}$, alors le théorème B.3 donne une contradiction :

$$\sigma = \sum_{u \in \mathcal{S}} \alpha_u = \sum_{u \in \mathcal{S}} \alpha_u \sum_{t \in \mathcal{S}} p_{u,t} = \sum_{t \in \mathcal{S}} \sum_{u \in \mathcal{S}} \alpha_u p_{u,t} < \sum_{t \in \mathcal{S}} \alpha_t = \sigma.$$

On en déduit que la mesure $(\alpha_t/\sigma)_{t \in \mathcal{S}}$ est une probabilité invariante pour la matrice P , ce qui est impossible d'après le théorème B.27 puisque cette matrice est récurrente nulle. $\square$

L'énoncé du théorème B.31 ci-dessus peut se reformuler dans le vocabulaire de la convergence des lois des chaînes de Markov associées, plutôt que celui des matrices stochastiques.

Corollaire B.32 (Limites des probabilités de transition et convergence en loi). Soit $(Y_n)_{n \geq 0}$ une chaîne de Markov homogène, irréductible et apériodique à espace d'états $\mathcal{S}$. Pour tout état $t \in \mathcal{S}$, notons τ_t l'espérance du premier temps de retour définie par (B.8) pour la matrice de transition de la chaîne. Avec la convention $1/\infty = 0$, on a

$$\mathbb{P}(Y_n = t) \xrightarrow{n \rightarrow \infty} \frac{1}{\tau_t}.$$

Dans le cas récurrent positif, ces limites définissent la probabilité invariante et il y a convergence en loi vers celle-ci. Dans les cas transient et récurrent nul, toutes ces limites sont nulles : elles ne définissent pas une loi de probabilité sur $\mathcal{S}$.

Démonstration. Notons $p_{s,t}^{(n)}$ les coefficients de la matrice de transition de la chaîne. La formule des probabilités totales et le théorème B.7 impliquent que pour tout $t \in \mathcal{S}$,

$$\mathbb{P}(Y_n = t) = \sum_{s \in \mathcal{S}} p_{s,t}^{(n)} \mathbb{P}(Y_0 = s) \xrightarrow{n \rightarrow \infty} \sum_{s \in \mathcal{S}} \frac{1}{\tau_t} \mathbb{P}(Y_0 = s) = \frac{1}{\tau_t}.$$

Cette convergence se prouve de la même manière que (B.11). $\square$

On rappelle que, d'après la proposition B.21, toute matrice stochastique irréductible indexée par un ensemble fini est récurrente. De plus, le corollaire B.23 réinterprète ce résultat en indiquant que toutes les chaînes de Markov homogènes et irréductibles à espace d'états fini sont récurrentes. Le résultat suivant montre qu'il y a en fait récurrence positive.

Corollaire B.33 (Récurrence positive pour un espace d'états fini). *Toute matrice stochastique irréductible indexée par un ensemble fini est récurrente positive. Ainsi, toute chaîne de Markov homogène et irréductible à espace d'états fini est récurrente positive.*

Démonstration. Soit P une matrice stochastique irréductible indexée par un ensemble fini $\mathcal{S}$, et soit ν une probabilité sur $\mathcal{S}$. Pour $N \geq 1$, posons

$$\pi^{(N)} = \frac{1}{N} \sum_{n=0}^{N-1} \nu P^n.$$

Ces vecteurs appartiennent au simplexe compact des probabilités sur $\mathcal{S}$. Une sous-suite converge donc vers une probabilité π . Or

$$\pi^{(N)} P - \pi^{(N)} = \frac{\nu P^N - \nu}{N} \xrightarrow{N \rightarrow \infty} 0.$$

Le passage à la limite le long de cette sous-suite donne $\pi P = \pi$. Le théorème B.27 implique alors que P est récurrente positive. Cette preuve ne nécessite aucune hypothèse d'apériodicité. $\square$

Scripts Python

Ce chapitre regroupe les scripts Python utilisés dans les différents chapitres de ce cours. Ces scripts sont disponibles sous la forme de bloc-notes Jupyter sur la [page eCampus du cours](#). Vous pouvez les exécuter sur votre ordinateur pour mieux comprendre les concepts abordés.

C.1 Scripts utilisés dans le chapitre 1

SCRIPT C.1 : Simulation du processus de Bernoulli

```

1 # Simule un processus de Bernoulli et affiche le nombre de succès et la fréquence des succès en fonction du nombre d'épreuves,
  compare avec le nombre et la fréquence espérés de succès.
2 import numpy as np
3 import matplotlib.pyplot as plt
4
5 # Demande à l'utilisateur la probabilité de succès p et le nombre d'épreuves n
6 p = float(input("Entrez la probabilité de succès (p): "))
7 n = int(input("Entrez le nombre d'épreuves (n): "))
8
9 # Valide les entrées
10 if not (0 <= p <= 1):
11     raise ValueError("La probabilité p doit être entre 0 et 1.")
12 if n <= 0:
13     raise ValueError("La taille de l'échantillon n doit être un entier positif.")
14
15 # Simule les n épreuves de Bernoulli indépendantes de paramètre p
16 trials = np.random.rand(n) < p
17
18 # Calcule le processus de Bernoulli (nombre de succès en fonction du temps)
19 cumulative_successes = np.cumsum(trials)
20
21 # Crée une figure avec 2 sous-graphes
22 fig, ax = plt.subplots(1, 2, figsize=(14, 6))
23
24 # Tracé du premier graphique (gauche) : Nombre de succès vs nombre d'épreuves
25 ax[0].step(range(1, n+1), cumulative_successes, where='post', label='Nombre de succès')
26 ax[0].plot(range(1, n+1), np.arange(1, n+1) * p, label=f'Probabilité de succès (p={p})', linestyle='--', color='red')
27
28 # Ajoute les étiquettes et le titre du premier graphique
29 ax[0].set_xlabel('Nombre d\'épreuves')
30 ax[0].set_ylabel('Nombre de succès')
31 ax[0].set_title('Processus de Bernoulli')
32 ax[0].legend()
33
34 # Tracé du second graphique (droite) : Fréquence des succès vs nombre d'épreuves
35 ax[1].plot(range(1, n+1), cumulative_successes / np.arange(1, n+1), label='Fréquence des succès')
36 ax[1].axhline(y=p, color='red', linestyle='--', label=f'Limite théorique (p={p})')
37
38 # Ajoute les étiquettes et le titre du second graphique
39 ax[1].set_xlabel('Nombre d\'épreuves')
40 ax[1].set_ylabel('Fréquence des succès')
41 ax[1].set_title('Convergence de la fréquence des succès')
42 ax[1].legend()
43
44 # Affiche les tracés
45 plt.tight_layout()
46 plt.show()

```

SCRIPT C.2 : Représentation des lois binomiale et normale

```

1 # Calcule et trace la loi binomiale et la densité normale approchante
2 import numpy as np
3 import matplotlib.pyplot as plt
4 from scipy.stats import binom, norm

```

```

5
6 # Demande les paramètres à l'utilisateur
7 n = int(input("Entrez la taille de l'échantillon (n) : "))
8 p = float(input("Entrez la probabilité de succès (p, entre 0 et 1) : "))
9
10 # Valide les entrées
11 if not (0 < p < 1):
12     raise ValueError("La densité normale exige 0 < p < 1 (variance non nulle).")
13 if n <= 0:
14     raise ValueError("La taille de l'échantillon n doit être un entier positif.")
15
16 q = 1 - p
17
18 # Calcule les paramètres de la loi normale correspondante
19 mean = n * p
20 stddev = np.sqrt(n * p * q)
21
22 # Domaine d'affichage : autour de l'espérance plus ou moins 4 fois l'écart-type
23 low = max(0, int(mean - 4 * stddev))
24 high = min(n, int(mean + 4 * stddev))
25 x = np.arange(low, high + 1)
26
27 # Probabilités de la loi binomiale
28 binom_probs = binom.pmf(x, n, p)
29
30 # Densité de la loi normale
31 x_continuous = np.linspace(low, high, 500)
32 normal_density = norm.pdf(x_continuous, mean, stddev)
33
34 # Tracé
35 plt.figure(figsize=(12, 8))
36
37 # Histogramme pour la loi binomiale
38 plt.bar(x, binom_probs, width=0.8, color='blue', alpha=0.5, label=f'Loi binomiale (n={n}, p={p:.2f})')
39
40 # Courbe pour la loi normale
41 plt.plot(x_continuous, normal_density, color='red', lw=2, label=f'Densité normale ( $\mu$ ={mean:.2f},  $\sigma$ ={stddev:.2f})')
42
43 # Personnalisation
44 plt.xlabel('$k$ (succès)', fontsize=14)
45 plt.ylabel('Probabilité / densité', fontsize=14)
46 plt.title('Loi binomiale et densité normale approchante', fontsize=16)
47 plt.grid(True, which='both', linestyle='--', linewidth=0.5)
48 plt.legend(fontsize=12, loc='upper right')
49
50 # Affichage
51 plt.tight_layout()
52 plt.show()

```

SCRIPT C.3 : Simulation d'une planche de Galton

```

1 # Planche de Galton : simule m lancers d'une jeton sur une planche de Galton avec n bifurcations et une probabilité p d'aller à
   droite, puis affiche l'histogramme des positions finales des jetons ; cela revient à simuler m réalisations d'une variable
   aléatoire Bin(n,p) et à afficher l'histogramme des résultats.
2 import numpy as np
3 import matplotlib.pyplot as plt
4 from scipy.stats import norm
5
6 def galton_board(p, n, m):
7     # Initialisation des positions finales des jetons
8     final_positions = np.zeros(m, dtype=int)
9     # Simulation des trajectoires des jetons
10    for i in range(m):
11        # À chaque bifurcation, on décide d'aller à gauche (0) ou à droite (+1)
12        steps = np.random.choice([0, 1], size=n, p=[1 - p, p])
13        final_positions[i] = np.sum(steps) # Position finale après toutes les bifurcations
14    # Renormalisation des positions finales
15    mean = n * p
16    std_dev = np.sqrt(n * p * (1 - p))
17    Z = (final_positions - mean) / std_dev
18    # Tracé des histogrammes
19    fig, axes = plt.subplots(1, 2, figsize=(14, 6))
20    # Histogramme des positions finales
21    axes[0].hist(final_positions, bins=np.arange(n + 2) - 0.5, rwidth=0.8, color='skyblue', edgecolor='black')
22    axes[0].set_title(f"Positions finales des m = {m} jetons (p = {p}, niveaux = {n})")

```

```

23 axes[0].set_xlabel("Position finale $S_n$ (cases)")
24 axes[0].set_ylabel("Nombre de jetons dans chaque case par les $m$")
25 axes[0].grid(True)
26 # Histogramme renormalisé
27 x = np.linspace(-4, 4, 500) # Plage pour la densité normale
28 axes[1].hist(Z, bins=int(np.ceil(np.log2(m) + 1)), density=True, color='lightcoral', alpha=0.7, edgecolor='black',
29             label="Histogramme renormalisé")
30 axes[1].plot(x, norm.pdf(x), 'b-', lw=2, label="Densité normale centrée réduite")
31 axes[1].set_title(f"Histogramme renormalisé vs loi normale (p = {p}, niveaux = n = {n})")
32 axes[1].set_xlabel(r"Position finale renormalisée  $S_n^{\ast} = \frac{S_n - np}{\sqrt{np(1-p)}}$ ")
33 axes[1].set_ylabel("Densité")
34 axes[1].legend()
35 axes[1].grid(True)
36
37 plt.tight_layout()
38 plt.show()
39
40 # Demander les paramètres à l'utilisateur
41 try:
42     p = float(input("Entrez la probabilité d'aller à droite (p entre 0 et 1) : "))
43     m = int(input("Entrez le nombre de jetons à faire descendre : "))
44     n = int(input("Entrez le nombre de bifurcations (niveaux) : "))
45     if not (0 < p < 1):
46         raise ValueError("La renormalisation exige 0 < p < 1 (variance non nulle).")
47     if m <= 0 or n <= 0:
48         raise ValueError("Le nombre de jetons et de niveaux doit être strictement positif.")
49     # Appeler la fonction pour simuler la planche de Galton
50     galton_board(p, n, m)
51 except ValueError as e:
52     print(f"Erreur : {e}")

```

SCRIPT C.4 : Représentation des lois binomiale et de Poisson

```

1 # Calcule et trace la loi binomiale et la loi de Poisson approchante
2 import numpy as np
3 import matplotlib.pyplot as plt
4 from scipy.stats import binom, poisson
5
6 # Demande les paramètres à l'utilisateur
7 n = int(input("Entrez la taille de l'échantillon (n) : "))
8 p = float(input("Entrez la probabilité de succès (p, entre 0 et 1) : "))
9
10 # Valide les entrées
11 if not (0 <= p <= 1):
12     raise ValueError("La probabilité p doit être entre 0 et 1.")
13 if n <= 0:
14     raise ValueError("La taille de l'échantillon n doit être un entier positif.")
15
16 # Calcul du paramètre de la loi de Poisson
17 lambda_param = n * p
18
19 # Domaine d'affichage : autour de l'espérance plus ou moins 4 fois la racine carrée de lambda
20 low = max(0, int(lambda_param - 4 * np.sqrt(lambda_param)))
21 high = min(n, int(lambda_param + 4 * np.sqrt(lambda_param)))
22 x = np.arange(low, high + 1)
23
24 # Probabilités de la loi binomiale
25 binom_probs = binom.pmf(x, n, p)
26
27 # Probabilités de la loi de Poisson
28 poisson_probs = poisson.pmf(x, lambda_param)
29
30 # Tracé
31 plt.figure(figsize=(12, 8))
32
33 # Histogramme pour la loi binomiale
34 plt.bar(x, binom_probs, width=0.8, color='blue', alpha=0.5, label=f'Loi binomiale (n={n}, p={p:.2f})')
35
36 # Points pour la loi de Poisson
37 plt.scatter(x, poisson_probs, color='orange', label=f'Loi de Poisson  $\lambda$ ={lambda_param:.2f}', zorder=3)
38
39 # Personnalisation
40 plt.xlabel('$k$ (succès)', fontsize=14)
41 plt.ylabel('Probabilité', fontsize=14)

```

```

42 plt.title('Loi binomiale et loi de Poisson approchante', fontsize=16)
43 plt.grid(True, which='both', linestyle='--', linewidth=0.5)
44 plt.legend(fontsize=12, loc='upper right')
45
46 # Affichage
47 plt.tight_layout()
48 plt.show()

```

C.2 Scripts utilisés dans le chapitre 2

SCRIPT C.5 : Simulation d'une collecte de coupons

```

1 # Simule une expérience de collection de coupons et trace le nombre de coupons collectés en fonction du nombre de tirages
2 import matplotlib.pyplot as plt
3 import numpy as np
4 import math
5 import random
6
7 # Demander à l'utilisateur d'entrer le nombre de coupons
8 r = int(input("Entrez le nombre de coupons à collecter (r) : "))
9 if r < 2:
10     raise ValueError("Le quotient T(r)/(r*ln(r)) exige r >= 2.")
11
12 # Initialisation des variables
13 coupons = set(range(1, r + 1)) # Ensemble des coupons disponibles
14 collected_coupons = set() # Ensemble des coupons collectés
15 n = 0 # Compteur de tirages
16 collected_counts = [] # Liste pour stocker le nombre de coupons collectés à chaque tirage
17
18 # Simulation des tirages
19 while len(collected_coupons) < r:
20     n += 1
21     new_coupon = random.randint(1, r) # Tirer un coupon au hasard parmi tous les coupons possibles
22     if new_coupon not in collected_coupons: # Vérifier si le coupon est déjà collecté
23         collected_coupons.add(new_coupon) # Ajouter le coupon à l'ensemble des coupons collectés
24         collected_counts.append(len(collected_coupons))
25
26 # Calcul de r*ln(r) et du quotient n/(r*ln(r))
27 r_ln_r = r * math.log(r)
28 ratio = n / r_ln_r
29
30 # Tracer la courbe
31 plt.figure(figsize=(10, 6))
32 plt.plot(range(1, n + 1), collected_counts, label="Coupons collectés", color="blue")
33 plt.xlabel("Nombre de tirages (n)")
34 plt.ylabel("Nombre de coupons collectés")
35 plt.title("Problème du collectionneur de coupons")
36 plt.axhline(y=r, color='green', linestyle='--', label=f"Nombre total de coupons r = {r}")
37 plt.axvline(x=n, color='green', linestyle='--', label=f"Temps de recouvrement T(r) = {n} donc T(r)/(r*ln(r)) = {ratio:.2f}")
38 plt.legend()
39 plt.grid()
40 plt.show()

```

SCRIPT C.6 : Tracé de la densité de la loi de Gumbel

```

1 # Trace la densité de la loi de Gumbel
2 import numpy as np
3 import matplotlib.pyplot as plt
4
5 def gumbel_density(x):
6     # Calcule la densité de la loi de Gumbel : exp(-x - exp(-x))
7     return np.exp(-x - np.exp(-x))
8
9 # Définir un intervalle de valeurs pour x
10 x = np.linspace(-5, 5, 500)
11
12 # Calculer la densité pour chaque valeur de x
13 y = gumbel_density(x)
14
15 # Tracer la courbe de densité
16 plt.figure(figsize=(8, 6))

```

```

17 plt.plot(x, y, label="Densité de la loi de Gumbel", color="blue")
18 plt.title("Densité de la loi de Gumbel")
19 plt.xlabel("x")
20 plt.ylabel("Densité")
21 plt.axhline(0, color="black", linewidth=0.8, linestyle="--") # Ligne horizontale à y=0
22 plt.grid()
23 plt.legend()
24 plt.show()

```

SCRIPT C.7 : Comparaison de la fonction de répartition empirique des fluctuations renormalisées et celle de la loi de Gumbel

```

1 # Répète plusieurs fois l'expérience de collecte de coupons et compare la fonction de répartition empirique des fluctuations
  renormalisées des temps de recouvrement avec celle de la loi de Gumbel
2 import math
3 import random
4 import numpy as np
5 import matplotlib.pyplot as plt
6
7 def simulate_coupon_collector(r):
8     # Initialisation des variables
9     collected_coupons = set() # Ensemble des coupons collectés
10    n = 0 # Compteur de tirages
11
12    # Simulation des tirages
13    while len(collected_coupons) < r:
14        n += 1
15        new_coupon = random.randint(1, r) # Tirer un coupon au hasard parmi tous les coupons possibles
16        if new_coupon not in collected_coupons: # Vérifier si le coupon est déjà collecté
17            collected_coupons.add(new_coupon) # Ajouter le coupon à l'ensemble des coupons collectés
18
19    return n
20
21 # Demander à l'utilisateur d'entrer le nombre de coupons et le nombre d'expériences
22 r = int(input("Entrez le nombre de coupons à collecter (r) : "))
23 m = int(input("Entrez le nombre d'expériences à répéter (m) : "))
24 if r < 1 or m < 1:
25     raise ValueError("Les nombres de coupons et d'expériences doivent être positifs.")
26
27 # Répéter l'expérience m fois
28 results = [simulate_coupon_collector(r) for _ in range(m)]
29
30 # Fluctuations T(r)/r - ln(r), également définies pour r = 1
31 adjustment_vector = [n / r - math.log(r) for n in results]
32
33 # Fonction de répartition empirique, avec regroupement des valeurs répétées
34 values, counts = np.unique(adjustment_vector, return_counts=True)
35 t_min, t_max = values[0] - 1, values[-1] + 1
36 cdf_x = np.concatenate([t_min, values, [t_max]])
37 cdf_y = np.concatenate([0, np.cumsum(counts) / m, [1]])
38
39 # Fonction de répartition de Gumbel sur une grille indépendante
40 x = np.linspace(t_min, t_max, 500)
41 gumbel_cdf = np.exp(-np.exp(-x))
42
43 # Tracer les deux fonctions, même pour une seule observation
44 plt.figure(figsize=(8, 6))
45 plt.step(cdf_x, cdf_y, where="post", color="b", label="Fonction de répartition empirique")
46 plt.plot(x, gumbel_cdf, color="r", label="Fonction de répartition de la loi de Gumbel")
47 plt.title("Comparaison entre les fonctions de répartition empirique et de Gumbel")
48 plt.xlabel("Fluctuations renormalisées (t)")
49 plt.ylabel("Probabilité cumulée")
50 plt.legend(title=f"Nombre de coupons = {r}, nombre d'expériences = {m}")
51 plt.grid()
52 plt.show()

```

C.3 Scripts utilisés dans le chapitre 3

SCRIPT C.8 : Simulation d'une marche aléatoire unidimensionnelle

```

1 # Simule m réalisations d'une marche aléatoire simple sur Z et trace la position de la particule en fonction du temps
2 import numpy as np

```

```

3 import matplotlib.pyplot as plt
4
5 def simulate_random_walks(m, n, p):
6     # Initialisation de la figure
7     plt.figure(figsize=(10, 6))
8     # Génération des marches aléatoires
9     for i in range(m):
10        # Génération de n sauts pour une réalisation
11        steps = np.random.choice([-1, 1], size=n, p=[1 - p, p])
12        walk = np.cumsum(steps) # Position cumulée
13        walk = np.insert(walk, 0, 0) # Ajouter 0 au début pour partir de la position initiale
14        plt.plot(range(n + 1), walk, label=f'Marche {i+1}')
15    # Ajout des titres et légendes
16    plt.title(f"{m} réalisations d'une marche aléatoire simple (p = {p})")
17    plt.xlabel(f"Instant (maximal = {n})")
18    plt.ylabel("Position")
19    # plt.legend(loc="upper left", fontsize=6)
20    plt.grid(True)
21    plt.show()
22
23 # Demander les paramètres à l'utilisateur
24 try:
25     m = int(input("Entrez le nombre de réalisations (m) : "))
26     n = int(input("Entrez la durée de chaque marche (n) : "))
27     p = float(input("Entrez la probabilité de faire un saut +1 (p entre 0 et 1) : "))
28     if m <= 0 or n <= 0 or not (0 <= p <= 1):
29         raise ValueError("Les paramètres doivent respecter : m > 0, n > 0, et 0 <= p <= 1.")
30     # Appeler la fonction pour simuler les marches
31     simulate_random_walks(m, n, p)
32 except ValueError as e:
33     print(f"Erreur dans les paramètres : {e}")

```

SCRIPT C.9 : Simulation du problème de la ruine du joueur

```

1 # Simule m marches aléatoires pour le problème de la ruine du joueur. La marche part de k et s'arrête dès qu'elle atteint a ou b.
  La probabilité de faire un pas +1 est p.
2 import numpy as np
3 import matplotlib.pyplot as plt
4
5 def simulate_ruin(p, a, k, b, m):
6     hitting_times = [] # Liste des instants où chaque marche s'arrête
7     reaches_a = 0      # Compteur du nombre de marches atteignant a en premier
8     marches = []       # Stocke les trajectoires pour les tracer
9     colors = []        # Stocke les couleurs (bleu ou rouge) des marches
10
11    for _ in range(m):
12        position = k
13        trajectory = [position] # Trajectoire complète de cette marche
14        time = 0
15        while position != a and position != b:
16            step = 1 if np.random.rand() < p else -1
17            position += step
18            trajectory.append(position)
19            time += 1
20
21        hitting_times.append(time)
22        marches.append(trajectory)
23        if position == a:
24            reaches_a += 1
25            colors.append('blue') # Atteint la borne a
26        else:
27            colors.append('red')  # Atteint la borne b
28
29    # Calcul des proportions
30    proportion_a = reaches_a / m
31    proportion_b = 1 - proportion_a
32
33    # Tracé des marches et histogramme des instants d'arrêt
34    fig, axes = plt.subplots(1, 2, figsize=(14, 6))
35
36    # Graphique des marches simulées
37    for march, color in zip(marches, colors):
38        axes[0].plot(range(len(march)), march, color=color, alpha=0.7)
39        axes[0].axhline(a, color='blue', linestyle='--', label=f"Borne a = {a}")
40        axes[0].axhline(b, color='red', linestyle='--', label=f"Borne b = {b}")

```

```

41 axes[0].set_title(f"m = {m} parties, proba de gagner un tour p = {p}, capital initial k = {k}")
42 axes[0].set_xlabel("Tours")
43 axes[0].set_ylabel("Capital")
44 axes[0].legend()
45 axes[0].grid(True)
46
47 # Histogramme des instants d'arrêt
48 axes[1].hist(hitting_times, bins=int(np.sqrt(m)), color='lightblue', edgecolor='black', alpha=0.7)
49 axes[1].set_title(f"Histogramme des instants d'arrêt")
50 axes[1].set_xlabel("Premier instant d'arrêt")
51 axes[1].set_ylabel("Effectif")
52
53 # Ajouter les proportions directement dans la légende
54 axes[1].text(
55     0.95, 0.95,
56     f"Proportion atteignant a = {a}: {proportion_a:.2f}\n"
57     f"Proportion atteignant b = {b}: {proportion_b:.2f}",
58     ha='right', va='top', transform=axes[1].transAxes,
59     bbox=dict(boxstyle="round,pad=0.3", edgecolor='none', facecolor='white')
60 )
61 axes[1].grid(True)
62
63 plt.tight_layout()
64 plt.show()
65
66 # Demande des paramètres à l'utilisateur
67 try:
68     p = float(input("Entrez la probabilité de gagner un tour (p entre 0 et 1) : "))
69     a = int(input("Entrez le capital minimal a : "))
70     b = int(input("Entrez le capital maximal b : "))
71     k = int(input("Entrez le capital initial k (k doit être entre a et b) : "))
72     m = int(input("Entrez le nombre de parties : "))
73
74     if not (0 <= p <= 1):
75         raise ValueError("La probabilité p doit être entre 0 et 1.")
76     if not (a <= k <= b):
77         raise ValueError("Le capital initial k doit être compris entre a et b.")
78     if m <= 0:
79         raise ValueError("Le nombre de parties m doit être strictement positif.")
80
81     # Simulation des marches
82     simulate_ruin(p, a, k, b, m)
83
84 except ValueError as e:
85     print(f"Erreur : {e}")

```

SCRIPT C.10 : Simulation d'une marche aléatoire dans le plan

```

1 # Simule et trace les n premiers pas d'une marche aléatoire simple symétrique dans le plan, en partant de l'origine (0, 0).
  L'évolution temporelle est représentée par un dégradé de couleurs. Les directions possibles sont haut, bas, gauche, droite.
  Elles sont équiprobables et représentées par des flèches.
2 import numpy as np
3 import matplotlib.pyplot as plt
4 # Demander la longueur de la marche avec validation
5 while True:
6     try:
7         n = int(input("Entrez la longueur n de la marche aléatoire : "))
8         if n > 0:
9             break
10        else:
11            print("Veuillez entrer un entier positif.")
12    except ValueError:
13        print("Veuillez entrer un nombre entier valide.")
14 # Génération des directions des pas (haut, bas, gauche, droite)
15 directions = np.random.choice(['haut', 'bas', 'gauche', 'droite'], size=n)
16 # Initialisation des coordonnées
17 x = [0] # Coordonnées x (origine)
18 y = [0] # Coordonnées y (origine)
19 # Calcul des positions successives
20 for direction in directions:
21     if direction == 'haut':
22         x.append(x[-1])
23         y.append(y[-1] + 1)
24     elif direction == 'bas':
25         x.append(x[-1])

```

```

26     y.append(y[-1] - 1)
27     elif direction == 'gauche':
28         x.append(x[-1] - 1)
29         y.append(y[-1])
30     elif direction == 'droite':
31         x.append(x[-1] + 1)
32         y.append(y[-1])
33 # Génération des couleurs pour le dégradé
34 colors = plt.cm.Blues(np.linspace(0.3, 1, len(x))) # Dégradé de bleu
35 # Création du graphique
36 plt.figure(figsize=(8, 8))
37 # Tracer les segments avec des flèches pour chaque pas
38 for i in range(len(x) - 1):
39     plt.plot(x[i:i+2], y[i:i+2], color=colors[i], linewidth=2) # Segment
40     # Ajout de flèches au milieu du segment
41     dx = x[i+1] - x[i]
42     dy = y[i+1] - y[i]
43     plt.arrow(x[i] + dx / 2, y[i] + dy / 2, dx * 0.2, dy * 0.2,
44              head_width=0.15, head_length=0.15, fc=colors[i], ec=colors[i])
45 # Marquer l'origine et la fin
46 plt.scatter(0, 0, color=colors[0], label='Origine', zorder=5) # Origine
47 plt.scatter(x[-1], y[-1], color=colors[-1], label=f'Position finale ({x[-1]}, {y[-1]})', zorder=5) # Position finale
48 # Ajustements du graphique
49 plt.title(f"Trajectoire d'une marche aléatoire dans le plan sur n = {n} pas")
50 plt.xlabel("Position x")
51 plt.ylabel("Position y")
52 plt.axhline(0, color='gray', linewidth=0.5, linestyle='--') # Ligne y=0
53 plt.axvline(0, color='gray', linewidth=0.5, linestyle='--') # Ligne x=0
54 plt.grid(color='lightgray', linestyle='--', linewidth=0.5)
55 plt.xticks(np.arange(min(x)-1, max(x)+1, 1)) # Espacement horizontal de 1
56 plt.yticks(np.arange(min(y)-1, max(y)+1, 1)) # Espacement vertical de 1
57 plt.legend()
58 plt.axis('equal') # Échelle égale pour x et y
59 plt.show()

```

SCRIPT C.11 : Simulation de marches aléatoires dans l'espace

```

1 # Simule et trace m trajectoires de marches aléatoires en 3D, chacune de n pas. Les directions possibles sont haut, bas, gauche,
  droite, avant, arrière. Elles sont équiprobables et représentées par des flèches.
2 import numpy as np
3 import matplotlib.pyplot as plt
4 from mpl_toolkits.mplot3d import Axes3D
5
6 # Demander à l'utilisateur le nombre de pas et le nombre de trajectoires
7 n = int(input("Entrez le nombre de pas pour chaque trajectoire : "))
8 m = int(input("Entrez le nombre de trajectoires à tracer : "))
9
10 # Directions possibles (haut, bas, gauche, droite, avant, arrière)
11 directions = np.array([
12     [1, 0, 0], # Droite
13     [-1, 0, 0], # Gauche
14     [0, 1, 0], # Haut
15     [0, -1, 0], # Bas
16     [0, 0, 1], # Avant
17     [0, 0, -1] # Arrière
18 ])
19
20 # Initialisation de la figure 3D
21 fig = plt.figure(figsize=(10, 8))
22 ax = fig.add_subplot(111, projection='3d')
23
24 # Générer et tracer chaque trajectoire
25 for trajectory in range(m):
26     # Initialisation de la trajectoire
27     positions = np.zeros((n + 1, 3)) # Positions (x, y, z)
28
29     # Génération de la marche aléatoire
30     for i in range(1, n + 1):
31         # Choisir une direction aléatoire
32         direction = directions[np.random.choice(len(directions))]
33         # Mettre à jour la position
34         positions[i] = positions[i - 1] + direction
35
36     # Tracer la trajectoire avec une couleur aléatoire
37     ax.plot(positions[:, 0], positions[:, 1], positions[:, 2], lw=1, alpha=0.7, label=f'Trajectoire {trajectory + 1}')

```

```

38
39 # Marquer l'origine
40 ax.scatter(0, 0, 0, color='red', s=100, label='Origine')
41
42 # Ajouter des labels et une légende
43 ax.set_xlabel('Axe X')
44 ax.set_ylabel('Axe Y')
45 ax.set_zlabel('Axe Z')
46 ax.set_title(f"Marches aléatoires en 3D ({n} pas, {m} trajectoires)")
47 ax.legend()
48
49 # Afficher le graphe
50 plt.show()

```

C.4 Scripts utilisés dans le chapitre 4

SCRIPT C.12 : Simulation d'un processus de Galton-Watson

```

1 # Simule r réalisations d'un processus de Galton-Watson en fonction des paramètres fournis par l'utilisateur ; trace l'évolution
  de la population Z_n en fonction de la génération n pour chaque processus, ainsi que la valeur m de l'espérance de la loi de
  reproduction
2 import numpy as np
3 import matplotlib.pyplot as plt
4
5 def input_reproduction_law():
6     while True:
7         law_input = input("Entrez les coefficients de la loi de reproduction (séparés par des points-virgules) : ")
8         try:
9             coefficients = list(map(float, law_input.split(';')))
10            if not all(0 <= p <= 1 for p in coefficients):
11                print("Les coefficients doivent être entre 0 et 1.")
12                continue
13            if not np.isclose(sum(coefficients), 1.0, atol=1e-5):
14                print("La somme des coefficients doit être égale à 1.")
15                continue
16            # Normaliser les arrondis acceptés avant le tirage.
17            return np.asarray(coefficients) / sum(coefficients)
18        except ValueError:
19            print("Veuillez entrer des nombres valides séparés par des points-virgules.")
20
21 def simulate_galton_watson(r, coefficients, max_generations):
22     results = []
23     for _ in range(r):
24         population = [1] # Commence avec 1 individu (Z_0 = 1)
25         for n in range(1, max_generations + 1):
26             next_population = sum(np.random.choice(len(coefficients), size=population[-1], p=coefficients))
27             population.append(next_population)
28             if next_population == 0:
29                 break # Extinction
30         results.append(population)
31     return results
32
33 def plot_results(results, coefficients, max_generations):
34     plt.figure(figsize=(10, 6))
35     # Calcul de l'espérance m
36     expectation = sum(k * p for k, p in enumerate(coefficients))
37
38     for i, population in enumerate(results):
39         # Préparer les données pour un tracé par morceaux
40         n_values = np.arange(len(population)) # Générations
41         Z_values = population # Populations
42
43         # Tracer en escalier
44         plt.step(n_values, Z_values, where='post', label=f'Processus {i+1}')
45
46     # Déterminer le régime (sous-critique, critique, sur-critique)
47     if expectation < 1:
48         regime = "sous-critique"
49     elif expectation == 1:
50         regime = "critique"
51     else:
52         regime = "sur-critique"

```

```

53
54 # Ajouter la valeur de m et le régime dans la légende
55 plt.plot([], [], ' ', label=f'$m = {expectation:.2f}$\n({regime})') # Ligne vide pour ajouter m et le régime
56
57 # Tracer m^n dans le cas sur-critique
58 if expectation > 1:
59     generations = np.arange(0, max_generations + 1)
60     m_n = expectation ** generations
61     plt.plot(generations, m_n, 'k--', label=f'$m^n$ ($m = {expectation:.2f}$)')
62
63 plt.xlabel('Génération $n$')
64 plt.ylabel('Population $Z_n$')
65 plt.title(f'{len(results)} processus de Galton-Watson de loi de reproduction {coefficients}')
66
67 # Positionner la légende en fonction de m
68 if expectation > 1:
69     plt.legend(loc='lower right') # Légende en bas à droite si m > 1
70 else:
71     plt.legend(loc='upper right') # Légende en haut à droite si m <= 1
72
73 plt.grid(True)
74
75 # Choisir l'échelle en fonction de m
76 if expectation > 1:
77     plt.yscale('log') # Échelle logarithmique si m > 1
78 else:
79     plt.yscale('linear') # Échelle linéaire si m <= 1
80
81 plt.show()
82
83 if __name__ == "__main__":
84     # Entrée des paramètres par l'utilisateur
85     r = int(input("Entrez le nombre de processus à simuler : "))
86     coefficients = input_reproduction_law()
87     max_generations = int(input("Entrez la génération maximale à simuler : "))
88
89     # Simulation et tracé des résultats
90     results = simulate_galton_watson(r, coefficients, max_generations)
91     plot_results(results, coefficients, max_generations)

```

SCRIPT C.13 : Tracé de la fonction génératrice d'une loi de Poisson

```

1 # Trace la fonction génératrice de la loi de Poisson et la compare avec la première bissectrice
2 import numpy as np
3 import matplotlib.pyplot as plt
4
5 # Demander à l'utilisateur la ou les valeurs de lambda
6 lambda_input = input("Entrez la ou les valeurs de λ séparées par ';': ")
7 lambda_values = [float(val.strip()) for val in lambda_input.split(';')]
8 if not all(np.isfinite(lam) and lam >= 0 for lam in lambda_values):
9     raise ValueError("Les paramètres de Poisson doivent être finis et positifs ou nuls.")
10
11 # Définir l'intervalle [0, 1] pour x
12 x = np.linspace(0, 1, 500)
13
14 # Tracer la droite y = x
15 plt.plot(x, x, label='y = x', color='black', linestyle='--')
16
17 # Tracer la fonction génératrice pour chaque valeur de lambda
18 for lam in lambda_values:
19     y = np.exp(lam * (x - 1)) # Fonction génératrice de la loi de Poisson
20     plt.plot(x, y, label=f'λ = {lam}')
21
22 # Ajouter des labels et une légende
23 plt.xlabel('x')
24 plt.ylabel('y')
25 plt.title('Fonction génératrice de la loi de Poisson et droite y = x')
26 plt.legend()
27 plt.grid(True)
28
29 # Afficher le graphe
30 plt.show()

```

SCRIPT C.14 : Tracé de la fonction génératrice d'une loi géométrique

```

1 # Trace la fonction génératrice de la loi géométrique et la compare avec la première bissectrice
2 import numpy as np
3 import matplotlib.pyplot as plt
4
5 # Demander à l'utilisateur la ou les valeurs de p
6 p_input = input("Entrez la ou les valeurs de p séparées par ';' : ")
7 p_values = [float(val.strip()) for val in p_input.split(';')]
8 if not all(0 < p <= 1 for p in p_values):
9     raise ValueError("Une loi géométrique exige 0 < p <= 1.")
10
11 # La fonction génératrice est définie sur tout [0, 1] pour p > 0
12 x = np.linspace(0, 1, 500)
13
14 # Tracer la droite y = x
15 plt.plot(x, x, label='y = x', color='black', linestyle='--')
16
17 # Tracer la fonction génératrice pour chaque valeur de p
18 for p in p_values:
19     y = p / ((1 - x) + p * x) # Forme stable, y compris en x = 1
20     plt.plot(x, y, label=f'p = {p}')
21
22 # Ajouter des labels et une légende
23 plt.xlabel('x')
24 plt.ylabel('y')
25 plt.title('Fonction génératrice de la loi géométrique et droite y = x')
26 plt.legend()
27 plt.grid(True)
28
29 # Afficher le graphe
30 plt.show()

```

C.5 Scripts utilisés dans le chapitre 5

SCRIPT C.15 : Simulation du modèle de Moran

```

1 # Modèle de Moran. Simule l'évolution du nombre de porteurs de l'allèle A dans une population de taille N. On note k le nombre
   d'individus portant l'allèle A au départ, et m le nombre d'évolutions à simuler. Tracer ces évolutions jusqu'au temps de
   fixation, en rouge quand fixation à 0 (extinction) et en bleu quand fixation à N. Trace l'historique du temps de fixation.
2 import numpy as np
3 import matplotlib.pyplot as plt
4
5 def moran_simulation(N, k, m):
6     if N < 1 or not (0 <= k <= N) or m < 1:
7         raise ValueError("Il faut N >= 1, 0 <= k <= N et m >= 1.")
8     # Initialisation des trajectoires
9     trajectories = []
10    fixation_times = []
11    fixation_types = [] # 0 pour extinction, 1 pour fixation
12
13    for _ in range(m):
14        X = [k] # X_0 = k
15        while True:
16            current_X = X[-1]
17            if current_X == 0 or current_X == N:
18                fixation_times.append(len(X) - 1)
19                fixation_types.append(0 if current_X == 0 else 1)
20                break
21            # Tirage de deux individus
22            chosen = np.random.choice(N, 2, replace=True) + 1 # Numérotation de 1 à N
23            # Conversion en allèles
24            allele1 = 1 if chosen[0] <= current_X else 0
25            allele2 = 1 if chosen[1] <= current_X else 0
26            # Mise à jour de X_n
27            if allele1 == 1 and allele2 == 0:
28                new_X = current_X + 1
29            elif allele1 == 0 and allele2 == 1:
30                new_X = current_X - 1
31            else:
32                new_X = current_X
33            X.append(new_X)
34    trajectories.append(X)

```

```

35
36 # Calcul des proportions
37 prop_extinct = fixation_types.count(0) / m
38 prop_fix = fixation_types.count(1) / m
39
40 # Tracé des trajectoires
41 plt.figure(figsize=(14, 6))
42 plt.subplot(1, 2, 1)
43 for i, traj in enumerate(trajectories):
44     color = 'red' if fixation_types[i] == 0 else 'blue'
45     plt.plot(traj, color=color, alpha=0.6)
46 plt.title(f'Trajectoires de $X_n$ (Fixation de $A$ : {prop_fix:.2%}, Fixation de $B$ : {prop_extinct:.2%})')
47 plt.xlabel('Temps $n$')
48 plt.ylabel(f'Nombre $X_n$ d\'individus avec allèle $A$, initialement {k} sur {N}')
49 plt.axhline(y=0, color='black', linestyle='--')
50 plt.axhline(y=N, color='black', linestyle='--')
51
52 # Histogramme des temps de fixation
53 plt.subplot(1, 2, 2)
54 plt.hist(fixation_times, bins=30, color='green', alpha=0.7)
55 plt.title('Histogramme des temps de fixation')
56 plt.xlabel('Temps de fixation')
57 plt.ylabel('Effectif')
58
59 plt.tight_layout()
60 plt.show()
61
62 # Demande des paramètres à l'utilisateur
63 N = int(input("Entrez la taille de la population N: "))
64 k = int(input("Entrez le nombre initial d'individus portant l'allèle A (k): "))
65 m = int(input("Entrez le nombre de simulations à effectuer (m): "))
66
67 # Exécution de la simulation
68 moran_simulation(N, k, m)

```

SCRIPT C.16 : Simulation du modèle de Wright-Fisher

```

1 # Modèle de Wright-Fisher. Simule l'évolution du nombre de porteurs de l'allèle A dans une population de taille N. On note k le
  nombre d'individus portant l'allèle A au départ, et m le nombre d'évolutions à simuler. Tracer ces évolutions jusqu'au temps
  de fixation, en rouge quand fixation à 0 (extinction) et en bleu quand fixation à N. Trace l'historique du temps de fixation.
2 import numpy as np
3 import matplotlib.pyplot as plt
4
5 def wright_fisher_simulation(N, k, m):
6     if N < 1 or not (0 <= k <= N) or m < 1:
7         raise ValueError("Il faut N >= 1, 0 <= k <= N et m >= 1.")
8     # Initialisation des trajectoires
9     trajectories = []
10    fixation_times = []
11    fixation_types = [] # 0 pour extinction, 1 pour fixation
12
13    for _ in range(m):
14        X = [k] # X_0 = k
15        while True:
16            current_X = X[-1]
17            if current_X == 0 or current_X == N:
18                fixation_times.append(len(X) - 1)
19                fixation_types.append(0 if current_X == 0 else 1)
20                break
21            # Tirage de la nouvelle génération selon une loi binomiale
22            new_X = np.random.binomial(N, current_X / N)
23            X.append(new_X)
24        trajectories.append(X)
25
26    # Calcul des proportions
27    prop_extinct = fixation_types.count(0) / m
28    prop_fix = fixation_types.count(1) / m
29
30    # Tracé des trajectoires
31    plt.figure(figsize=(14, 6))
32    plt.subplot(1, 2, 1)
33    for i, traj in enumerate(trajectories):
34        color = 'red' if fixation_types[i] == 0 else 'blue'
35        plt.plot(traj, color=color, alpha=0.6)
36    plt.title(f'Trajectoires de $X_n$ (Fixation de $A$ : {prop_fix:.2%}, Fixation de $B$ : {prop_extinct:.2%})')

```

```
37 plt.xlabel('Génération $n$')
38 plt.ylabel(f'Nombre $X_n$ d\'individus avec allèle $A$, initialement {k} sur {N}')
39 plt.axhline(y=0, color='black', linestyle='--')
40 plt.axhline(y=N, color='black', linestyle='--')
41
42 # Histogramme des temps de fixation
43 plt.subplot(1, 2, 2)
44 plt.hist(fixation_times, bins=30, color='green', alpha=0.7)
45 plt.title('Histogramme des temps de fixation')
46 plt.xlabel('Temps de fixation')
47 plt.ylabel('Effectif')
48
49 plt.tight_layout()
50 plt.show()
51
52 # Demande des paramètres à l'utilisateur
53 N = int(input("Entrez la taille de la population N: "))
54 k = int(input("Entrez le nombre initial d'individus portant l'allèle A (k): "))
55 m = int(input("Entrez le nombre de simulations à effectuer (m): "))
56
57 # Exécution de la simulation
58 wright_fisher_simulation(N, k, m)
```

Bibliographie

- [1] J. Angst, Modèles aléatoires en biologie, Notes de cours, 2012.
- [2] N. Berglund, Processus aléatoires et applications – algorithmes MCMC et vitesse de convergence, Notes de cours, 2024.
- [3] D. Chafaï et F. Malrieu, Modèles stochastiques, Notes de cours, 2014.
- [4] D. Chafaï et F. Malrieu, *Recueil de modèles aléatoires*, Number 78 in Mathématiques et applications, Springer, 2016.
- [5] D. Chafaï et P.-A. Zitt, *Probabilités Préparation à l'agrégation interne*, AFNIL, 2016.
- [6] F. Malrieu, Quelques modèles aléatoires célèbres pour les sciences de la vie et de la terre, Notes de cours, 2008.
- [7] S. Méléard, *Modèles aléatoires en écologie et évolution*, Springer, 2016.

Crédits des illustrations. Les liens donnent accès aux sources et aux licences des illustrations. L'abréviation DP désigne le domaine public selon la notice citée, avec les limites territoriales indiquées. Les images sont redimensionnées pour la mise en page ; les recadrages sont signalés dans les crédits. Pour certaines illustrations, les conditions de reproduction ne sont pas établies : leur présence dans ce document ne vaut pas autorisation de réutilisation.

Couverture. Logo de couverture : Université Paris-Saclay, Faculté des sciences d'Orsay (identité institutionnelle).

Fig. 1.1. Jacques Bernoulli : Niklaus Bernoulli (DP); *Ars conjectandi* (1713) : Jakob Bernoulli (DP).

Fig. 1.2. Blaise Pascal : anonyme, d'après François II Quesnel; château de Versailles (DP).

Fig. 1.4. Émile Borel : Agence Mondial Photo-Presse / BnF; restauration PBriex (DP); *Francesco Paolo Cantelli* : photographe inconnu (DP en Italie selon la notice); *Chimpanzé à la machine à écrire* (vers 1906) : New York Zoological Society / Library of Congress (DP aux États-Unis; aucune restriction connue selon la Library of Congress).

Fig. 1.5. Abraham de Moivre : auteur non identifié; University of York (DP); *The Doctrine of Chances* (1718) : Abraham de Moivre / MacTutor (DP); *Pierre-Simon de Laplace* : James Posselwhite (DP); *Théorie analytique des probabilités* (1812) : Pierre-Simon Laplace, Paris, Courcier (source de la numérisation non identifiée).

Fig. 1.6. *Methodus differentialis* (1764) : James Stirling / BEIC (DP); *John Wallis* : anonyme, d'après Godfrey Kneller / National Portrait Gallery, NPG 578 (DP); revendication sur la reproduction signalée par Commons).

Fig. 1.8. Francis Galton (années 1890) : Eveleen Myers (née Tennant) / National Portrait Gallery (DP); *Planche de Galton* : Antoine Tavenaux (CC BY-SA 3.0); *The Wall : Face au mur* : photographe et source non identifiés; conditions de reproduction non établies.

Fig. 1.10. Lucien Le Cam, Berkeley (1987) : George M. Bergman (CC BY-SA 4.0; numérisation 2018, réduite); *An approximation theorem for the Poisson binomial distribution*, p. 1181 (1960) : Lucien Le Cam; © 1960 Pacific Journal of Mathematics (droits réservés; autorisation de reproduction non établie).

Fig. 2.1. *Paninipedia Dinos* : Panini / National Geographic Kids (produit représenté); auteur et source du visuel non identifiés; conditions de reproduction non établies; *Carte des Départ'aimants Le Gaulois* : Le Gaulois (produit représenté); photographe et source non identifiés; conditions de reproduction non établies.

Fig. 2.2. Lorenzo Mascheroni : graveur inconnu / Università di Genova (DP); *Leonhard Euler* : Jakob Emanuel Handmann / Deutsches Museum (DP).

Fig. 2.4. Emil Julius Gumbel : photographe anonyme; E. Schultz (dir.), *Das Gesicht der Demokratie*, 1931, p. 149 (DP).

Fig. 3.4. Johann Andreas von Segner : Fuger (peinture), C. G. Rasp (gravure) / Dibner Library (DP); *Walther von Dyck* : photographe inconnu / MacTutor (DP).

Fig. 3.5. Eugène Charles Catalan (1884) : Émile Delpérée / Université de Liège (DP); *Désiré André* : photographe inconnu; H. Carnoy, *Dictionnaire biographique international des écrivains*, 1902, p. 121 (DP).

Fig. 3.6. Joseph Bertrand : photographe inconnu / Wolfram ScienceWorld (DP).

Fig. 3.9. George Pólya : photographe inconnu / ETH-Bibliothek Zürich, Bildarchiv, Portr_13887-19-AL (DP).

Fig. 4.2. Irénée-Jules Bienaimé : photographe inconnu (DP); *Henry William Watson* : photographe inconnu / MacTutor (DP); *Francis Galton (milieu du XIXe siècle)* : photographe non identifié; K. Pearson, *The Life, Letters, and Labours of Francis Galton* (DP).

Fig. 5.1. Gregor Mendel : photographe inconnu (DP aux États-Unis selon la notice); *Godfrey Harold Hardy* : photographe inconnu; *A Mathematician's Apology* (DP); *Wilhelm Weinberg* : photographe inconnu; © Genetics Society of America (autorisation de reproduction non établie).

Fig. 5.3. Patrick Alfred Pierce Moran : photographe non identifié; source : Freebase, via Wikipédia (image non libre; autorisation de reproduction non établie).

Fig. 5.6. Sewall Wright (1965) : photographe inconnu; courtesy of the University of Wisconsin-Madison / Encyclopædia Britannica (image non libre; autorisation de reproduction non établie); *Ronald Fisher* (1913) : photographe inconnu / University of Adelaide (DP).

Fig. A.1. Niels Henrik Abel : Johan Gørbitz / MacTutor (DP); *Beppo Levi* : photographe inconnu / Università di Torino (DP).

Fig. B.1. Guido Fubini : photographe inconnu (DP en Italie selon la notice); *Leonida Tonelli* : photographe anonyme / Università Bocconi (DP).

Fig. B.2. *Andrei Markov* : photographe inconnu; Histoire de l'Académie des sciences de l'URSS, t. 2, 1964, p. 475 (DP); *Sydney Chapman* : NOAA (DP aux États-Unis, œuvre fédérale américaine); *Andrei Kolmogorov* : Konrad Jacobs / Mathematisches Forschungsinstitut Oberwolfach (MFO) (CC BY-SA 2.0 DE; recadrage).

Fig. B.3. *Georg Frobenius* : photographe non identifié; reproduction Furfur / MFO, album de la Mathematische Gesellschaft Hamburg (CC BY-SA 3.0; version recadrée); *Issai Schur* : photographe non identifié / MFO, photo 12209 (CC BY-SA 2.0 DE; recadrage); *Wolfgang Döblin (vers 1938)* : photographie familiale, communiquée par Ulrich Krengel / MFO, photo 9417 (CC BY-SA 2.0 DE).

Schémas et simulations. Les schémas et les graphiques de simulation ont été réalisés pour ce cours. Des programmes permettant de reproduire les simulations sont présentés en annexe C.