Guy Melançon

Processus stochastiques - MIAGe / e-MIAGe

10 janvier 2012

Département informatique
UFR Mathématiques Informatique
Université Bordeaux I
Année académique 2010 - 2011

Table des matières

1 Introduction 5

Bibliographie 7

Partie I Rappels

2 Calcul et dénombrement 11
2.1 Calcul, algèbre et combinatoire 11
2.2 Dénombrement 12

3 Probabilités : propriétés élémentaires 17
3.1 Propriétés élémentaires 20

4 Indépendance et probabilités conditionnelles 29
4.1 Formule de Bayes 31

5 Variables aléatoires, espérance et variance 37
5.1 Distribution de probabilité 39
5.2 Espérance 40
5.3 Variance 47
5.4 Somme et produit de variables aléatoires 48

6 Lois des grands nombres 51
6.1 Simulation et méthodes Monte Carlo 54

7 Probabilités et simulation, génération aléatoire 55
7.1 Génération aléatoire 56
7.2 Méthodes à rejet 60

8 Variable aléatoire réelle, densité de probabilité 61
8.1 Densité de probabilité et fonction de répartition 61
8.2 Lois continues classiques 62

8.2.1 Loi de distribution uniforme 62
8.2.2 Loi de distribution non uniforme et fonction de répartition inverse 64
8.2.3 Méthodes à rejet 67
8.2.4 Loi exponentielle 68
8.3 Loi de distribution de Gauss (loi normale) 73

Partie II Processus stochastiques

9 Introduction 77

10 Calcul matriciel, graphes et chemins 79

11 Chaînes de Markov 83
11.1 Chaîne de Markov à deux états 83
11.1.1 Condition de Markov 85
11.2 Fonction de transition et matrices 87
11.3 Distribution stationnaire 96
11.3.1 Etats transitoires, récurrents, absorbants 100
11.4 Chaînes de Markov et optimisation 106
11.5 Méthodes Monte Carlo et chaînes de Markov 110
11.5.1 Chaînes de Markov et marche aléatoire uniforme 111
11.5.2 L'algorithme de Metropolis 112

12 Chaînes de Markov à temps continu 113
12.1 Condition de Markov et loi exponentielle 114
12.1.1 Backward et forward équations. 115
12.1.2 Système à deux états 115
12.2 Description matricielle 117
12.2.1 Le cas à deux états 118
12.3 Une chaîne continue à trois états 119

Chapitre 1
Introduction

The capricious gods that were previously invoked to explain the lack of predictability in the world have been replaced by mathematical, statistical and computer-based models that allow us to understand and manipulate uncertain events.

D. Hand, H. Mannila, P. Smyth (2001). Principles of Data Mining, MIT Press.

L'un des aspects les plus présents dans ce cours devrait être la simulation : les processus stochastiques prennent tout leur sens en informatique dès lors que l'on peut leur donner corp et les "faire vivre" en machine. La simulation apparaît comme une alternative, parfois obligée, à une analyse mathématique (analytique) en proposant plutôt une imitation d'un système. Elle met en oeuvre un modèle (agençant mathématique et informatique), est souvent conceptuellement plus simple et permet d'obtenir des informations quantitatives là où l'analyse mathématique échoue.

Cette vision s'accorde aux objectifs de la formation MIAGe :

A la croisée de plusieurs domaines – sciences et technologie de l'information et de la communication, sciences de gestion, économie, communication – l'objectif du master MIAGe est de préparer l'étudiant à la maîtrise, voire l'expertise, des méthodes d'ingénierie et à leurs applications dans la construction et les adaptations des systèmes d'information.

Corine Cauvet, Daniel Marquié, Jean-Pierre Peyrin. La formation MIAGE, Bulletin SPECIF (62), Décembre 2009, pp. 7–11.

Les simulations sont aujourd'hui très populaires et sont pratiquées dans de nombreux domaines scientifiques, technologiques, industriels et économiques. Une simulation génère un volume important de données décrivant un système (à travers le modèle que l'on en a formulé). Contrairement à une étude analytique, les inférences liées à la simulation (les conclusions tirées des valeurs observées) sont de nature statistiques et peuvent être interprétées de façon équivoque. Les résultats obtenus sont intimement liés aux postulats de départ et une légère variation de ceux-ci peut modifier radicalement les résultats obtenus (selon la sensibilité du modèle). La mise au point de simulations requière la génération de valeurs possibles de certains paramètres dont on suppose souvent qu'ils suivent une loi de probabilité, ou de certains processus que l'on modélise à l'aide de processus stochastiques dont on maîtrise la portée dans le modèle. Le cours nous donnera l'occasion de mettre un pied sur ce terrain en soulignant la nécessité de bien maîtriser les aspects stochastiques de ces ingrédients d'une simulation.

Le cours s'appuie nécessairement sur la théorie des probabilités sur laquelle nous reviendrons au début. Cette partie du cours coincide avec le cours de probabilités de la licence informatique. Les ouvrages de Richard Isaac [Isa05] ou de Pierre Brémaud [Bré09] sont conseillés pour revoir ces notions de base. Le logiciel opensource R [1], ou le site web SMEL [2] sont conseillés pour mettre en pratique les notions probabilistes (et statistiques, le cas échéant). Nous serons amenés à utiliser ces outils lorsqu'il s'agira de manipuler/simuler les lois de probabilités ou les processus stochastiques dont nous aborderons les fondements. L'utilisation de tableurs pourra même être envisagée [BMPS98] [BPSD07].

Nous adopterons aussi le point de vue développé par Bercu et Chafaï [BC07] tourné vers la simulation des processus stochastiques. La partie du cours abordant les chaînes de Markov s'appuie abondamment sur les ouvrages de Pardoux [Par07] et Hoel *et al.* [HPS72]. Des portions de l'ouvrage de Foata et Fuchs [FF02] sont à la base des chapitres sur les processus stochastiques, en particulier sur les processus de Poisson.

1. The R Project for Statistical Computing, `http://www.r-project.org/`.
2. “Statistiques Médicale en Ligne”, `http://mistis.inrialpes.fr/software/SMEL/`.

Bibliographie

[BC07] Bernard Bercu and Djalil Chafaï. *Modélisation stochastique et simulation – Cours et applications*. Sciences Sup. Dunod, 2007.

[BMPS98] Anne Brygoo, Michelle Morcrette, Odile Paliès, and Michèle Soria. *Initiation à la programmation par Word et Excel, Principes et macros*. Vuibert, 1998.

[BPSD07] Anne Brygoo, Maryse Pelletier, Michèle Soria, and Séverine Dubuisson. *Programmation et Algorithmique en VBA pour Excel*. Dunod, 2007.

[Bré09] Pierrre Brémaud. *Initiation aux probabilités et aux chaînes de Markov*. Springer, 2009.

[FF02] Dominique Foata and Aimé Fuchs. *Processus stochastiques – Processus de Poisson, chaînes de Markov et martingales – Cours et exercices corrigés*. Dunod, 2002.

[HPS72] Paul G. Hoel, Sidnay C. Port, and Charles J. Stone. *Introduction to stochastic processes*. Waveland Press Inc., 1972.

[Isa05] Richard Isaac. *Une initiation aux probabilités*. Vuibert, 2005.

[Par07] Etienne Pardoux. *Processus de Markov et applications – Algorithmes, réseaux, génome et finance*. Mathématiques pour le Master / SMAI. Dunod, 2007.

Partie I
Rappels

Chapitre 2
Calcul et dénombrement

2.1 Calcul, algèbre et combinatoire

Cette section rappelle quelques identités remarquables que l'on retrouve fréquemment en calcul de probabilité. Il est primordial d'être à l'aise avec ces manipulations algébriques élémentaires afin de conduire correctement le calcul qui suit du raisonnement probabiliste.

Exercice 2.1 Identités remarquables *Montrez (par récurrence) :*

$$\sum_{k=0}^{n} 2k+1 = (n+1)^2$$

$$\sum_{k=0}^{n} k = \frac{n(n+1)}{2}$$

$$\sum_{k=0}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$$

Série géométrique. Soit une valeur q avec $0 \leq q < 1$ alors la puissance q^n tend vers 0 lorsque $n \to \infty$. Mais plus encore, on peut faire la somme de toutes ces valeurs q^n et on obtient :

$$\lim_{n\to\infty} 1 + q + \cdots + q^n = \sum_{k\geq 0} q^n = \frac{1}{1-q}$$

En effet, désignons par $S_n = 1 + q + \cdots + q^n = \sum_{k=0}^{n} q^k$. On a alors $S_{n+1} = S_n + q^{n+1}$ et $S_{n+1} = 1 + qS_n$. Les deux membres droit de ces égalités coincident, d'où l'on tire $S_n = \frac{1-q^{n+1}}{1-q}$. En prenant la limite, et puisque $q^{n+1} \to 0$ lorsque $n \to \infty$, on obtient bien l'identité annoncée.

Exercice 2.2 *Calculez S_n et la série géométrique pour $q = 1/2$, $q = 1/3$, $q = 1/4$.*

Série harmonique. Montrez que la série :

$$\sum_{n=1}^{\infty} \frac{1}{n}$$

diverge. Comparez-là à la série :

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{4} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \dots$$

qui ne converge pas (puisqu'en regroupant ses termes on trouve $1/2 + 1/2 + 1/2 + \dots$).

En revanche, la suite $\sum_{n=1}^{p} \frac{1}{n} - \log(n)$ converge. Sa limite, notée γ est appelée la constante d'Euler.

Exercice 2.3 *Ecrivez un court programme qui évalue la constante d'Euler pour différentes valeurs de n (et en précision croissante).*

Coefficients binomiaux. Les coefficients binomiaux sont au centre de l'étude de la loi de probabilité du même nom. Ils permettent d'énumérer les sous-ensembles d'un ensemble donné (voir la prochaine section). On peut les définir de manière algébrique, comme suit. On définit le coefficient $\binom{n}{k}$ pour tout $n \geq 0$ et tout $0 \leq k \leq n$ en posant $\binom{n}{0} = \binom{n}{n} = 1$ et $\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$. On peut écrire ces nombres dans un grand tableau triangulaire, chaque ligne débutant et se terminant par 1. La récurrence donne une façon facile de remplir les case de ce tableau : le nombre de la ligne n en position k s'obtient des nombres de la ligne précédente en position $k-1$ et k.

Exercice 2.4 Triangle de Pascal *Calculez les valeurs des 10 premières lignes du tableau de coefficient binomiaux.*

Exercice 2.5 Identité du binôme *L'identité du binôme généralise l'identité remarquable $(x+y)^2 = x^2 + 2xy + y^2$. Utilisez la récurrence définissant les coefficients binomiaux pour établir l'identité :*

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

Ecrivez l'identité particulière que l'on obtient lorsque $x = p$ et $y = 1 - p$.

2.2 Dénombrement

Le calcul des probabilités discrètes, dont il sera beaucoup question dans le cours, requière souvent que l'on puisse énumérer tous les objets d'un événement, sous-ensemble de l'espace de probabilités (l'ensemble de toutes les observations possibles). Ces opérations de dénombrement s'appuient le plus souvent sur l'utilisation de coefficients qui calculent des configurations typiques :

- le *produit cartésien* de deux ensembles E, F noté $E \times F$: il est formé de la liste des paires ordonnées (e, f) avec $e \in E, f \in F$; son cardinal est $|E| \cdot |F|$ (où $|E|$ désigne le cardinal de l'ensemble E) ; plus généralement, le produit cartésien $E_1 \times \cdots \times E_k$ est formé des k-uplets $(e_1, \ldots, e_k)$ avec $e_i \in E_i$ et contient $|E_1| \cdots |E_k|$ éléments.
- le nombre de suites ordonnées de n éléments distincts (on parle aussi de *permutations* de ces éléments) est $n! = n(n-1)(n-2)\cdots 2 \cdot 1$. On convient que $0! = 1$;
- le nombre d'*arrangements* de k éléments parmi n : ce sont des suites ordonnées de k éléments distincts choisis parmi n (à la différence d'une permutation, on ne prend pas tous les n éléments) ; le nombre d'arrangements k éléments parmi n est $\frac{n!}{(n-k)!}$;
- le *coefficient binomial* $\binom{n}{k}$ calculant le nombre de *sous-ensembles* à k éléments parmi n.

Dans chacun des exercices suivants, rapportez-vous à des situations modélisées par un produit cartésien, une permutation, un arrangement ou un sous-ensemble [1].

Exercice 2.6 Morpion *On peut choisir de mettre ou non une croix dans chacune des cases d'un carré* 3×3 *(morpion). Combien y a-t-il de façons distinctes de procéder ?*

Exercice 2.7 Lancers de dé *On lance trois fois de suite un dé numéroté de 1 à 6 et on note les triplets ainsi obtenus. Combien y a-t-il de tels triplets ? (en tenant compte de l'ordre).*

Exercice 2.8 Football *Pour constituer une équipe de football, on a le choix entre 20 postulants. En supposant que chaque joueur est polyvalent (il peut jouer à tous les postes), combien peut-on constituer d'équipes* [2] *différentes ?*

Si parmi les 20 postulants, 17 sont joueurs de champ et 3 sont gardiens. Combien d'équipes distinctes peut-on alors constituer ?

Solution Tous les postes sont différents, on peut donc les ordonner et les énumérer 1, 2, ..., 11. Il s'agit donc de choisir un joueur pour le premier poste – on a 20 candidats, puis un joueur pour le poste 2 – il reste 19 candidats, etc. Le nombre total d'équipe que l'on peut former est donc $20 \times 19 \times \cdots \times 10$.

Si on distingue le poste de gardien pour lequel on n'a que 3 candidats, on réduit ce nombre à $3\times$ (on choisit le gardien) $17 \times 16 \times \cdots \times 8$. $\square$

1. Les exercices proposés ici sont tirés du web (`http://pagesperso-orange.fr/gilles.costantini`)
2. Par équipe, il faut entendre "qui joue à quel poste", chaque poste étant distinct.

Exercice 2.9 Chiffres *Avec les nombres de 1 à 6, on veut constituer un nombre de 3 chiffres distincts. Combien de nombres distincts peut-on réaliser ?*

Sans répétitions, combien de nombres de 3 chiffres peut-on former à l'aide des six chiffres 2, 3 ,5, 6, 7, 9 ? Combien de ces nombres sont :

- *inférieurs à 500 ?*
- *impairs ?*
- *pairs ?*
- *multiples de 5 ?*

Solution On peut fabriquer 6^3 nombres de trois chiffres (on est autorisé à répéter le même chiffre), peu importe que l'on utilise les chiffres $\{1, \ldots, 6\}$ ou $\{2, 3, 5, 6, 7, 9\}$. Si un chiffre ne peut être utilisé qu'une fois, alors on a plus que $6 \times 5 \times 4$ possibilités.

Si on doit fabriquer un nombre inférieur à 500, alors on ne peut plus utiliser que les chiffres $2, 3$ en position des centaines, faisant tomber le nombre de possibilités à $2 \times 5 \times 4$. □

Exercice 2.10 Boîtes et craies *On dispose de trois boîtes et de cinq craies de couleur bleue, rouge, jaune, verte et orange. 1) De combien de façons distinctes peut-on ranger les cinq craies dans les trois boîtes ? 2) Même question en laissant l'une des boîtes vides. 3) Même question si la bleue et la rouge sont rangées ensembles. 4) Même question si la bleue et la rouge sont rangées ensembles, mais seules.*

Exercice 2.11 Mots et alphabets *Combien de mots de 6 lettres peut-on écrire en utilisant 3 lettres D et 3 lettres H ?*

Combien de mots de 4 lettres peut-on former avec les 26 lettres de l'alphabet : a) En admettant les répétitions de lettres ? b) Sans lettres répétées ? Quelle est la probabilité qu'un mot de 4 lettres n'aie pas de lettres répétées ?

- *Combien de mots de 5 lettres peut-on faire avec les 26 lettres de l'alphabet ?*
- *Combien de ces mots ne comportent que des lettres distinctes ?*
- *Combien de ces mots comportent exactement 4 lettres distinctes (et donc une lettre répétée) ?*

Exercice 2.12 Triangles *Quel est le nombre de triangles que l'on peut former avec 10 points distincts (on supposera que trois points distincts ne se trouvent jamais sur une même droite, pour éviter les triangles dégénérés . . .) ?*

Exercice 2.13 Chemins discrets *On considère les chemins sur le carré* 3×3 *vu comme un grille reliant les points (0, 0), (0, 1), (1, 0), . . . , (2, 2). Combien y a t-il de chemins qui vont de (0,0) à (2,2) ? Plus généralement, combien de chemins vont de* $(0, 0)$ *à* (m, n) *?*

Exercice 2.14 Trafic aérien *En hiver une compagnie aérienne dessert 6 villes. Quel est le nombre de lignes en service (elle propose tous les vols reliant deux de ces villes) ? En été, la compagnie a 45 lignes en service. Quel est le nombre de villes desservies ?*

Exercice 2.15 Loto *Au loto, quelle est le nombre de grilles qui ont k numéros gagnants ? (pour* $k = 3, 4, 5, 6$ *sur une grille de taille 49).*

Exercice 2.16 Jeu de cartes *Dans un jeu de 32 cartes, on choisit 5 cartes au hasard (ces 5 cartes s'appellent une "main").*
- *Quel est le nombre total de mains que l'on peut obtenir ?*
- *Combien de mains contiennent exactement 4 as ?*
- *Combien de mains contiennent exactement 3 as et 2 rois ?*
- *Combien de mains contiennent au moins 3 rois ?*
- *Combien de mains contiennent au moins un as ?*

Exercice 2.17 Boules et urnes *Une urne contient 49 boules numérotées de 1 à 49. On tire successivement 6 boules, sans remise. On appelle "tirage" cet ensemble de 6 numéros obtenus (sans tenir compte de l'ordre).*
- *Combien y a-t-il de tirages au total ?*
- *Combien y a-t-il de tirages qui contiennent 3 numéros pairs et 3 numéros impairs ?*
- *Combien y a-t-il de tirages qui contiennent au moins 5 numéros pairs ? (C'est-à-dire 5 numéros pairs ou 6 numéros pairs)*

Chapitre 3
Probabilités : propriétés élémentaires

Ce chapitre présente la théorie des probabilités, discrètes et continues. Il aborde d'abord les probabilités discrètes : celles qui décrivent des phénomènes dont les "possibilités de jeu" sont finies. Il est utile d'avoir une vision ensembliste des probabilités pour raisonner et calculer. Calculer une probabilités c'est d'abord déterminer l'espace que l'on observe, ou celui dans lequel se déroule une expérience. Il s'agit de décrire quelles sont toutes ces choses qui peuvent être observées (ou qui peuvent effectivement arriver) : tous les états possibles que peut prendre un système, ou toutes les configurations que l'on peut observer après que ce soit produit certains événements, etc. Pour raisonner, on se ramène souvent à des modèles très simples (états 0, 1 ; des configurations décrits sous forme de k-uplets de valeurs simples, etc.) qui seuls autorisent une solution numérique exacte ou approchée.

Ainsi, les exemples illustrant le calcul des probabilités s'appuie souvent sur les "jeux de hasard", les dés, les cartes, le tirage aléatoire de boules de couleurs dans des urnes, etc. Le travail consiste ensuite à savoir relever une situation concrète vers l'un de ces modèles simplistes. Par exemple, par souci de performance ou de robustesse des structures de données en informatique, on peut vouloir calculer la probabilité que deux clés de hachage (attribuée lors du stockage dans une table de hachage) entrent en collision. Du point de vue des probabilités, on peut de manière équivalente se poser la question : quelle est la probabilité que deux personnes dans une assemblée de N personnes aient la même date de naissance ? Du point de vue méthodologique, si l'on sait résoudre l'un des problèmes, on aura la clé pour résoudre l'autre.

De même, les exemples illustrant la théorie des processus stochastiques s'inspirent souvent des jeux de hasard. Du point de vue méthodologique, on peut penser à un hamster qui change de pièce dans un labyrinthe pour y manger des graines (un système changeant d'état), à des puces sautant d'un chien à un autre (une entreprise gagnant des parts de marché sur sa concurrente), ...

Développons maintenant la discussion autour d'un exemple. Imaginons que nous observions un système et qu'il s'agisse de comprendre comment il évolue. Un dispositif nous permet de savoir à tout moment s'il est actif ou passif (s'il émet ou non, une information, une onde, etc., par exemple). On peut donc imaginer effectuer une série de N observations du système. L'ensemble de ces observations, chacune étant codée par un 0 ou par un 1, nous donne une liste ordonnée de N valeurs. Les informaticiens parleront d'un vecteur de N bits ; les mathématiciens d'un élément du produit cartésien $\{0, 1\}^N$. A priori, nous ne connaissons rien du système observé et faisons l'hypothèse que toutes les possibilités d'observations (tous les vecteurs de bits ; tous les N-uplets du produit cartésien) sont possibles. Observez au passage que l'expérience consistant à jouer N fois à pile ou face peut être décrite exactement de la même manière et que l'analyse que nous nous apprêtons à faire s'applique tout aussi bien aux observations du système qu'au jeu de hasard.

Imaginez que vous observiez le N-uplet uniquement formé de 0. En d'autres mots, à chaque fois que vous avez noté l'état du système il était inactif. Bien que ce N-uplets soient aussi probables que toutes les autres séquences d'observations, vous êtes autorisés à vous questionner. C'est bien le caractère aléatoire de votre expérience qu'il s'agit de comprendre : votre ignorance du comportement du système vous autorise à modéliser la situation en ces termes, vous avez des raisons de croire que vos observations doivent vous livrer une suite au caractère erratique plutôt qu'une séquence trop ordonnée, régulière ou périodique.

Ainsi, comprendre le système en termes probabilistes, c'est pouvoir répondre à des questions comme : quelle est la probabilité d'observer une séquence comptant k valeurs à 1 ? quelle est la probabilité d'observer une séquence comptant au moins k valeurs à 1 ? des séquences comptant au moins k valeurs à 1 consécutives ? etc.

Chacune de ces questions portent sur un *évènement* : c'est un sous-ensemble de l'ensemble de toutes les *épreuves* ou *observations* possibles. Introduisons quelques notations. On désignera par Ω l'espace de probabilité, c'est-à-dire l'ensemble de toutes les épreuves $\omega \in \Omega$. C'est, dans l'exemple que nous décrivions à l'instant, l'ensemble de tous les éléments du produit cartésien. Un *évènement* A est un sous-ensemble de Ω (les N-uplets comptant k valeurs à 1, ou les N-uplets comptant au moins k valeurs à 1, par exemple). La *probabilité* que A se réalise sera notée $P(A)$. La théorie des probabilités apporte un formalisme rigoureux à cette formulation des choses. Dans certains cas, il faut porter attention aux évènements que l'on peut considérer : puisque ce sont des ensembles contenus dans Ω, qu'ils peuvent être inclus les uns dans les autres, s'intersecter, etc. On pourrait aussi avoir à considérer le cas où Ω est infini, où le nombre d'évènements $A, B, C \ldots$ est lui aussi infini, etc. [1]

1. La théorie introduit la notion de tribus et de boréliens, puisqu'il faut s'assurer de pouvoir considérer l'union et l'intersection des évènements que l'on observe, de la cal-

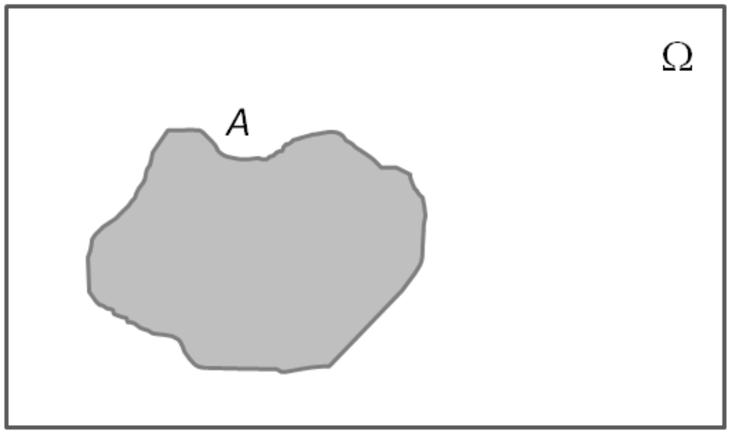


Fig. 3.1 Un évènement représenté classiquement à l'aide d'un diagramme de Venn. Intuitivement, sa probabilité est proportionnelle à son aire (relative au rectangle englobant représentant Ω.

Classiquement et en s'inspirant de la théorie des ensembles, on représente un évènement par un sous-ensemble A du référentiel Ω comme à la figure 3.4. Il faut lire cette illustration en pensant à la probabilité $P(A)$ comme étant égale à la part de l'aire de l'ensemble A dans le rectangle Ω. Cette façon de voir les choses peut être utile pour motiver ou justifier certains résultats ou définitions.

Concluons par un dernier exemple illustrant bien l'utilité des exercices parfois "ludiques" des cours de probabilités. *Q. Dans une assemblée de N personnes, quelle la probabilité que deux d'entre elles soient nées le même jour ?* A première vue, voilà un exercice de lycée sans lien évident avec les objectifs d'une licence en informatique. Mais pourtant, la question n'est étrangère à une question beaucoup plus informatique : étant donné une fonction de hachage $h : U \to [1, m]$ quelle est la probabilité d'une collision ? (deux éléments de U qui seraient associés à une même clé). On pense ici à une fonction définit sur un domaine assez grand – tous les mots possibles sur un alphabet, beaucoup plus grand que m. On exige que la fonction soit uniforme, c'est-à-dire que le nombre de clé associé à un entier $i \in [1, m]$ ne dépende pas de i et soit $|h^{-1}(i) = \frac{|U|}{m}|$.

Le coup des anniversaires dans une soirée est un cas particulier : la clé de hachage associée à une personne est la date du jour de sa naissance

culabilité des probabilité lorsque l'on fait des unions et/ou des intersections finies ou infinies. Ces calculs font aussi appel à la théorie de l'intégration (la théorie de la mesure), incontournable lorsqu'il s'agit d'utiliser les probabilités pour décrire le monde des particules en physique, par exemple.

(jour/mois), et on fait l'hypothèse que le nombre de personnes nés un jour donné est à peu près toujours le même peu importe le jour – ce qui semble raisonnable.

Exercice 3.1 Anniversaire et hachage *Résolvez les deux problèmes précédents. Précisez à chaque fois ce que doit être Ω et ce que sont les évènements à considérer.*

- *Dans une assemblée de N personnes, quelle la probabilité que deux d'entre elles soient nées le même jour ?*
- *Etant donné une fonction de hachage $h : U \to [1, m]$ quelle est la probabilité d'une collision ? (Deux éléments de U sont associés à une même clé.)*

3.1 Propriétés élémentaires

Plus formellement, un espace de probabilité est formé de trois ingrédients. Un ensemble Ω rassemblant les *épreuves* (les observations possibles), un ensemble d'*évènements* $\mathcal{U} = \{A, B, C, \ldots\}$, où $A, B, C, \ldots$ sont des sous-ensembles de Ω, et une *mesure de probabilité* P.

Les probabilités s'intéressent beaucoup à la théorie de la mesure (les conditions que doit satisfaire P, la façon de la calculer selon les propriétés de l'espace Ω et l'algèbre des évènements, entraînant une formalisation de l'algèbre des évènements $\mathcal{U}$). L'ensemble $\mathcal{U}$ est souvent infini, et il faut prendre des précautions lorsque l'on effectue une union, ou une intersection infinie d'éléments de $\mathcal{U}$. L'ensemble $\mathcal{U}$ est souvent appelé une *tribu*. Nous n'aborderons toutefois pas (directement) ces subtilités mathématiques.

Toujours en raisonnant de manière ensembliste (et à partir de la figure 3.4), si A et B sont deux évènements disjoints ($A \cap B = \emptyset$) alors on a $P(A \cup B) = P(A) + P(B)$, puisque l'aire totale de l'ensemble $A \cup B$ est la somme des aires de A et de B.

L'ensemble $\Omega \setminus A$ est le complément de A, qu'on note A^c. On a donc $P(A^c) = 1 - P(A)$.

Exercice 3.2 Décrire un espace de probabilités

Dans une fabrique de processeurs, on prélève toutes les heures les trois derniers processeurs produits. Ceux-ci sont classés dans deux catégories : fonctionnel, codé 1 et défectueux, codé 0.

1. *Décrivez l'espace associé à cette expérience aléatoire (par quoi peut-on représenter l'expérience (les états des trois processeurs p_1, p_2, p_3 sélectionnés)*
   , ?)
2. *Décrivez (en termes ensemblistes) les évènements suivants :*

– $A =$ *"le premier processeur est défectueux"*
– $B =$ *"le dernier est fonctionnel"*
– $C =$ *"deux processeurs sont défectueux"*
– $D =$ *"au moins deux processeurs sont fonctionnels".*

Exercice 3.3 Le problème du chevalier de Méré *Lequel de ces deux événements est le plus probable : "Obtenir au moins une fois un six en quatre lancers de dés", ou "Obtenir au moins un double six en vingt quatre lancers de deux dés".*

Solution On note par $\Omega = \{1, 2, \ldots, 6\}$ l'ensemble des valeurs possibles prises par un dé (lors d'un lancer unique), pour lequel on a $P(\omega) = 1/6$ $(\omega \in \Omega)$.

L'espace de probabilité qui décrit l'expérience "4 lancers de dés" est Ω^4. L'évènement $A =$ "avoir au moins un six" est complémentaire de l'évènement $A^c =$ "ne jamais avoir de six". Or, il y a 5^4 séquences $(\omega_1, \omega_2, \omega_3, \omega_4)$ dont aucune composante ne comporte un 6. On a donc $P(A) = 1 - P(A^c) = 1 - (5/6)^4$.

L'espace de probabilité qui décrit l'expérience "24 lancers de 2 dés" est $(\Omega \times \Omega)^{24}$. L'évènement $B =$ "avoir au moins un double six" est complémentaire de l'évènement $B^c =$ "ne jamais avoir de double six". Or, il y a 35 paires de dés qui ne correspondent pas à un double 6, et donc 35^{24} séquences $(\omega_{1,1}, \omega_{2,1}), (\omega_{2,1}, \omega_{2,2}), \ldots, (\omega_{24,1}, \omega_{24,2})$ dont aucune composante n'est un double six. On a donc $P(B) = 1 - P(B^c) = 1 - (35/36)^{24}$.

Reste à comparer ces deux valeurs, et on trouve $P(A) > P(B)$. Question subsidiaire : combien de lancers additionnels de double dés faut-il pour que les probabilités s'inversent ? $\square$

Exercice 3.4 Le bon choix *Vous jouez à un jeu télévisé pour gagner un PC de dernière génération. Le PC se trouve derrière une porte parmi trois A, B ou C ; les deux autres portes cachent des babioles. Après avoir fait votre choix, le présentateur télé ouvre une porte derrière laquelle se trouve un première babiole.*

Il vous propose soit de rester sur votre premier choix, soit d'échanger et d'opter plutôt pour la porte qu'il n'a pas ouverte ... Avez-vous avantage à rester sur votre premier choix ou à changer ?

Etudiez d'abord vos chances de gagner en faisant l'hypothèse que vous changez de porte quoiqu'il arrive. Puis étudiez le cas où vous ne changez pas de porte.

Exercice 3.5 Urnes et boules

Une urne contient 8 boules blanches et 6 boules noires, chaque boule ayant la même probabilité d'être tirée.

1. *On tire simultanément 5 boules. Quelle est la probabilité d'obtenir :*
   – *3 blanches et 2 noires ?*
   – *des boules de couleurs différentes ?*

2. *On tire successivement 5 boules avec remise de chaque boule tirée. Quelle est la probabilité d'avoir :*
   – *3 blanches puis 2 noires ?*
   – *3 blanches et 2 noires dans un ordre quelconque ?*

Exercice 3.6 Jeux de données

On considère des jeux de données permettant de tester la robustesse de l'implémentation d'un algorithme, dont 8 sont jugés d'une complexité "élevée" et 6 sont jugés de complexité "moyenne". On choisit au hasard des jeux de données qu'on soumet à l'algorithme, chaque jeu de données ayant la même probabilité d'être choisi.

1. *On tire simultanément 5 jeux de données. Quelle est la probabilité d'obtenir :*
   – *3 jeux de complexité élevée et 2 de complexité moyenne ?*
   – *des jeux de données de complexités différentes (pas tous de complexité élevée ou tous de complexité moyenne) ?*
2. *On tire successivement 5 jeu de données, en s'autorisant à choisir le même jeu de données plus d'une fois. Quelle est la probabilité d'avoir :*
   – *3 jeux de complexité élevée puis 2 de complexité moyenne (dans cet ordre) ?*
   – *3 jeux de complexité élevée puis 2 de complexité moyenne (peu importe l'ordre dans lequel ils sont apparus) ?*

Exercice 3.7 Tirage aléatoire en machine *Explorez la (les) librairie(s) mises à disposition dans divers langages de programmation (C, Java, python, etc.).*

A minima, le langage vous permet de tirer au hasard (et uniformément, c'est-à-dire que le tirage est équiprobable) un nombre réel de l'intervalle [0, 1).

Utilisez cette méthode de la librairie pour écrire un court programme qui permet de simuler un évènement de Bernouilli avec probabilité p (le lancer d'un pièce au jeu de pile ou face, où pile est obtenu avec probabilité p).

En répétant ce tirage aléatoire un grand nombre de fois, et en cumulant les fréquences des résultats, vous prendrez soin de vérifier empiriquement que votre tirage aléatoire suit bien la probabilité prescrite p.

Solution On peut simuler le tirage de Bernouilli de paramètre $0 < p < 1$ si on dispose d'un générateur de nombre réel : une fonction retournant un nombre réel de l'intervalle $[0, 1)$. C'est ce que fait la fonction python `random` (de la librairie `random`). Après avoir défini la fonction Bernouilli, on l'appelle m fois (avec m de plus en plus grand, pour constater empiriquement que la fréquence de succès s'approche de plus en plus de la probabilité théorique p).

□

Exercice 3.8 Rademacher *Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires indépendantes et de même loi de* Rademacher $R(p)$, *c'est-à-dire* $P(X_n =$

```python
from random import *

def Bernouilli(p):
    if random() < p:
        return 1
    else:
        return 0

def freqEmpirique(p, m):
    s = 0
    for i in range(m):
        s += Bernouilli(p)
    return s/m
```

Fig. 3.2 Code python réalisant un tirage de Bernouilli de paramètre p.

$1) = p$ *et* $P(X_n = -1) = 1 - p$ *avec* $0 < p < 1$. *Ecrivez un court programme qui permet de simuler la loi de Rademacher (que l'on peut comparer à un lancer à pile ou face mais à valeur 1 ou -1 avec probabilité* p*).*

Exercice 3.9 Tirage aléatoire sur l'intervalle $[a, b)$ *Utilisez le tirage aléatoire d'un réel de l'intervalle* $[0, 1)$ *pour construire une méthode pour tirer aléatoirement et uniformément un réel de l'intervalle* $[a, b)$ *où* $a < b$ *sont deux nombres réels ou entiers quelconques.*

Exercice 3.10 Histogramme *Un histogramme permet de rendre compte de la distribution de valeurs dans un échantillon numérique. Plus précisément, supposons que nous ayons une suite de nombres* $\mathcal{S} = x_1, x_2, \ldots$ *entiers ou réels. On peut alors fabriquer des couples* (x, n_x) *où* $x \in E$ *et* n_x *indique le nombre de valeurs* x_n *qui coincide avec* x. *Le nombre* n_x *est appelé la* fréquence *de* x *dans la suite* $\mathcal{S}$.

Pour visualiser les fréquences on peut placer les couples (x, n_x) *dans le plan. Cette visualisation peut être toutefois irrégulière et difficile à interpréter. On peut alors regrouper les valeurs* x_n *par intervalles disjoints et cumuler les fréquences des éléments appartenant à un même intervalle.*

Reprenez le tirage aléatoire de nombre réels de l'intervalle $[a, b)$ *et construisez un histogramme rendant compte des valeurs obtenues au cours d'un grand nombre de tirages. Vous trouverez un utilitaire permettant de visualiser l'histogramme*[2]*.*

Exercice 3.11 Simuler un lancer de dé *Selon les possibilités offertes par la (les) librairie(s) du langage utilisé, vous pouvez tirer au hasard un nombre réel de l'intervalle* $[0, 1)$ *ou un entier naturel (borné par la taille des mots mémoire et le format de stockage interne).*

2. On peut penser à `gnuplot`, par exemple.

A partir de l'un (ou des deux) de ces moyens de base, fabriquer une méthode permettant de simuler le lancer d'un dé équiprobable à six faces. Vous prendrez soin de vérifier empiriquement que votre tirage aléatoire est bien équiprobable.

Simulez l'expérience du chevalier de Méré et constater le résultat théorique de l'exercice 3.3.

Solution □

```python
# reproduit l'expérience de Méré
# 'simple' ou 'double'
def mere(type):
    if type == 'simple':
        for i in range(4):
            if lancerDe() == 6:
                return 1
        return 0
    else:
        for i in range(24):
            l1 = lancerDe()
            l2 = lancerDe()
            if l1 == 6 and l2 == 6:
                return 1
        return 0
```

Fig. 3.3 Code python simulant l'expérience de Méré. On peut, comme pour le tirage de Bernouilli (Fig. 3.2), calculer empiriquement la fréquence de succès de l'expérience.

Exercice 3.12 Générer des entiers *En machine, les entiers sont décrits par des vecteurs de bits de taille N (= 32 ou 64). On peut donc penser générer aléatoirement des entiers en jouant à pile ou face N fois avec $p = 1/2$ et en prenant le résultat du jeu comme résultat. Une suite générée ainsi forme t-elle une suite aléatoire d'entiers ?*

Ecrivez un court programme qui tire au hasard un nombre entiers de l'ensemble $\{0, 2^N - 1\}$ de cette façon ($N = 32$ ou 64 selon l'architecture de la machine utilisée).

Solution On suppose disposer d'une fonction `Bernouilli(p)` qui permet de simuler un évènement de Bernouilli de paramètre p. Il suffit donc de répéter N tirages de Bernouilli avec $p = 1/2$, et de faire la conversion de la base 2 à la base 10.

Si, par exemple, on utilise cette procédure pour générer un milliers d'entiers ($N = 10$), on doit pouvoir observer le caractère uniforme de la génération. On dresse un histogramme des nombres générés et on le trace :

□

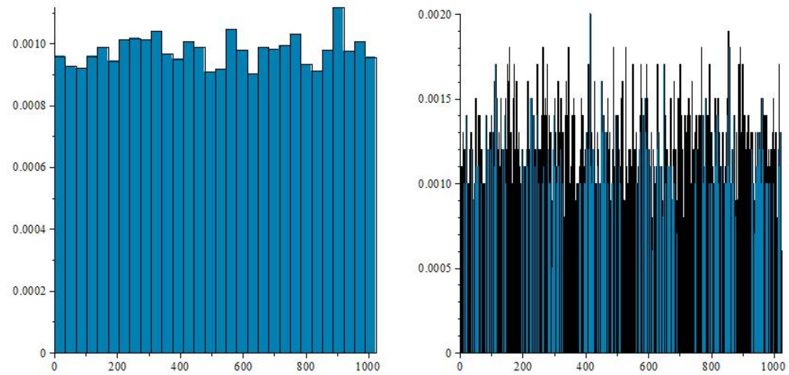


Fig. 3.4 Ces deux histogrammes illustrent la répartition des entiers générés par la procédure. Le caractère uniforme de la génération aléatoire est révélé par l'allure “plate” de l'histogramme. Remarquez que l'on perd ce caractère plat et permet de voir l'aspect chaotique du tirage aléatoire dès lors que l'histogramme contient trop de classes.

```
# retourne un entier choisi uniformément aléatoirement
# en fabriquant un vecteur de bits de taille N
# (donné en paramètre)
def randomVectInt(N):
    t = [0] * N
    for i in range(N):
        if random.random() < 0.5:
            t[i] = 0
        else:
            t[i] = 1
    s = 0
    b = 1
    for i in range(N):
        s += t[i] * b
        b *= 2
    return s
```

Fig. 3.5 Code python générant des entiers aléatoires en base 2, en simulant une loi binomiale de paramètre $p = 1/2$ (puis en retournant l'entier converti en base 10).

Exercice 3.13 Probabilités et modèles physiques *Une corde dans un cercle est un segment de droite le coupant en deux points de sa circonférence. On choisit “au hasard” une corde d'un cercle de rayon unitaire. Quelle est la probabilité qu'elle soit moins longue que le côté d'un triangle équilatéral inscrit dans le cercle*

Nous allons voir que la réponse que l'on peut donner à cette question dépend de ce que l'on entend par “choisit au hasard”.

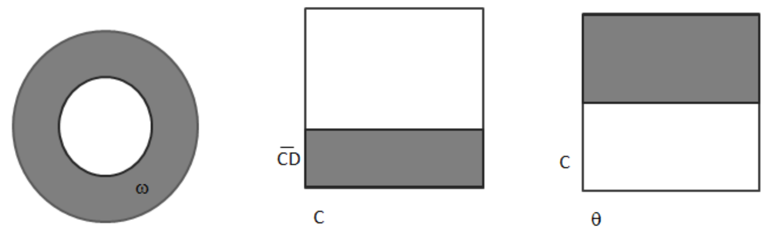


Fig. 3.6 On choisit une corde aléatoirement en choisissant au hasard un point du cercle. Les trois figures illustrent les trois espaces de probabilités considérés pour donner une solution. Chacune décrit ce que l'on entend par "choisir une corde au hasard" (un point ω dans le cercle ; un couple de points $(C, D) \in [0, 2\pi] \times [0, 2\pi]$ sur la circonférence du cercle ; un rayon déterminé par un angle $\theta \in [0, 2\pi]$ et un point $C \in [0, 1]$ sur ce rayon). La zone grisée correspond à chaque fois à l'évènement A.

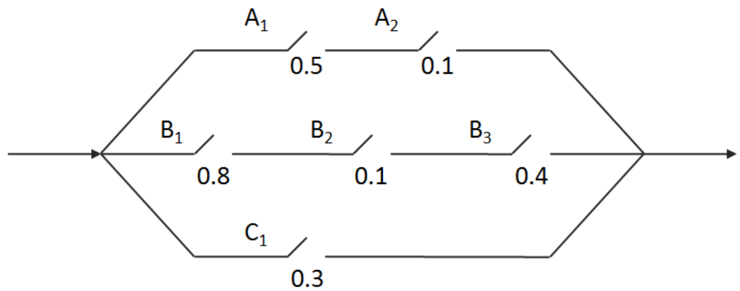
Solution 1. *Pour toute corde, il existe un unique rayon qui la croise à angle droit. Ainsi, à tout point ω d'un cercle $\mathcal{C}$ centré en O, on peut faire correspondre une unique corde perpendiculaire au rayon passant par ω. On peut donc interpréter "choisir une corde au hasard" comme étant "choisir un point ω du cercle et tracer l'unique corde passant par le rayon $O\omega$. Cette corde est de longueur supérieure au côté du triangle à condition que ce point se situe hors du cercle $\mathcal{C}'$ centré en O et de rayon $1/2$. Calculez la probabilité cherchée selon ce modèle.*

Solution 2. *Toute corde est déterminée par deux points C, D de la circonférence du cercle $\mathcal{C}$. Observez que le côté du triangle est lui-même une corde de longueur exactement $\sqrt{3}$ qui correspond à une portion de la circonférence de longueur $2\pi/3$. Etant donné un point C, les points D qui sont à une distance inférieure à $2\pi/3$ de C sur la circonférence permettent de construire une corde satisfaisant la condition. Choisir une corde au hasard correspond ici au choix d'un couple ordonné de points C, D sur la circonférence. Calculez la probabilité cherchée selon ce modèle.*

Solution 3. *Toute corde est déterminée par un point C se situant sur un rayon du cercle, lui-même déterminée par un angle $\theta \in [0, 2\pi)$. La corde correspond donc à un couple (θ, C) avec $C \in [0, 1)$, qui satisfait la condition dès lors que $C > 1/2$. Choisir une corde au hasard correspond ici au choix d'un couple ordonné de points C, θ. Calculez la probabilité cherchée selon ce modèle.*

Exercice 3.14 *On considère l'ensemble de relais de la figure 3.1. Les nombres indiquent les probabilités que le relais soient "ouverts" (ça ne circule pas). Les relais sont indépendants (formalisez cette notion). Calculez la probabilité pour qu'au moins une branche soit "fermée" (et que ça circule).*

Exercice 3.15 *Simulez par programmation le tirage aléatoire d'une corde selon les trois modèles de l'exercice précédent.*



Exercice 3.16 Guerre de processus *Imaginons deux threads s'exécutant simultanément, chacun réclamant au système une portion mémoire à différents moments. On suppose que les threads t_a et t_b demande à chaque fois une portion mémoire de taille fixe et qu'à la fin de leur exécution t_a aura demandé m portions mémoires, alors que t_b en aura demandé n, avec $m \geq n$. On s'intéresse à la probabilité que le thread t_a ait constamment plus de mémoire à sa disposition que le threads t_b.*

Nous allons modéliser ce problème de la façon suivante : les demandes successives des threads sont décrites par une séquence de a et de b (un mot). Une séquence qui maintient l'avantage à t_a (plus de mémoire accordée à t_a à tout moment) est un mot dont les préfixes comptent toujours plus de a que de b.

Soient $n, m \geq 0$, avec $m \geq n$. On note $C(m,n)$ l'ensemble des mots u de $\{a,b\}^*$ tels que $|u|_a = m$ et $|u|_b = n$, i.e. u est constitué de m lettres a et de n lettres b.

Soit $p : \{a,b\}^* \to Z$ l'application définie par $p(u) = |u|_a - |u|_b$. On propose de calculer la probabilité de l'évènement

$$D = \{u \in C(m,n) | \forall v \neq \epsilon \text{ prefixe de } u, p(v) > 0\}.$$

- *Soit $C_b(m,n)$ l'ensemble des mots de $C(m,n)$ commençant par b. Décrire l'évènement $E = C(m,n) \backslash D$ et montrer que $P(E) = 2P(C_b(m,n))$.*
- *En déduire la probabilité que le threads t_a dispose toujours de plus de mémoire que le threads t_b.*

Exercice 3.17 Quidam et casino *(Cet exercice est emprunté à et adapté de [Bré09, Chap.,3.1].) Au casino, un croupier choisit deux nombres $m, n \in [0, N]$, avec $m < n$ disons, qu'ils ne vous dévoilent pas. Il écrit ces deux nombres sur des bouts de papier qu'il place secrètement sous deux chapeaux. Il vous demande ensuite de choisir l'un des chapeaux et vous dévoile le nombre qui y était caché. Vous devez ensuite parier en répondant à la question : le nombre resté sous l'autre chapeau est-il plus grand que celui que l'on vient de vous dévoiler ?*

A priori, à choisir sans stratégie particulière mais seulement "au hasard" (en jouant à pile ou face, par exemple), vous n'avez qu'une chance sur deux de gagner. Existe t-il une stratégie de réponse qui vous garantit que vous avez plus de chances de donner la bonne réponse ? C'est ce que nous allons voir ...

Choisissez un nombre Z dans $[0, N]$ au hasard et selon une distribution de probabilité quelconque. Si Z est strictement plus grand que le nombre dévoilé, pariez que le nombre resté secret est plus grand que celui qui a été dévoilé. Montrez que cette stratégie est avantageuse.

Notons X le nombre qui vous a été dévoilé. Il s'agit de calculer la probabilité $P(X = m, Z > m) + P(X = n, Z < n)$ (ce sont les deux cas gagnants). Le croupier choisit l'un des deux chapeaux de manière équiprobable et les variables X et Z sont indépendantes, par conséquent :

$$
\begin{aligned}
& P(X = m, Z > m) + P(X = n, Z < n) \\
= \quad & P(X = m)P(Z > m) + P(X = n)P(Z \leq n) \\
= \quad & \tfrac{1}{2} \textstyle\sum_{k=m+1}^{N} P(Z = k) + \tfrac{1}{2} \textstyle\sum_{k=0}^{n} P(Z = k) \\
= & \; 1/2 + \textstyle\sum_{k=m+1}^{n} = 1/2 + P(m + 1 \leq Z \leq n) \geq 1/2
\end{aligned}
$$

Chapitre 4
Indépendance et probabilités conditionnelles

Les situations plus “réalistes” exigent d’étudier les *probabilités conditionnelles*. Des exemples illustreront bien ce propos. Imaginons que vous observiez le système de l’exemple en page 17 décrit à la section précédente et que vous constater que le système est actif environ deux fois sur trois. La question de connaître la probabilité d’observer le système en état actif quatre fois de suite ne se pose plus de la même manière. En termes probabilistes, vous cherchez à connaître la probabilité d’observer une séquence comptant quatre valeurs à 1 consécutives, sachant que le nombre total de valeurs à 1 est $2N/3$. Deux évènements sont parties prenantes ici. L’ensemble A des séquences comptant 4 valeurs à 1 consécutives, et l’ensemble B des séquences comptant $2N/3$ valeurs à 1. Et l’on cherche à calculer la probabilité de A *sachant* que l’on est dans la situation décrite par B.

Pensez à ces ensembles sous forme graphique. Puisque l’on suppose que l’évènement B s’est produit, on cherche à calculer combien d’éléments de A sont susceptibles de se produire. En termes ensemblistes, les évènements que l’on pourra observer sont ceux de $A \cap B$. La probabilité cherchée est ici $\frac{|A \cap B|}{|B|}$, ce qui est aussi égal à $\frac{P(A \cap B)}{P(B)}$.

C’est ce qu’on appelle la *probabilité conditionnelle de A sachant B*, et qu’on note $P(A|B)$:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{4.1}$$

Notez au passage qu’on a

$$P(A \cap B) = P(A|B)P(B).$$

Evènements indépendants L’évènement A ne dépend pas de l’évènement B si on a $P(A|B) = P(A)$: le fait qu’on suppose que B se soit produit n’influe pas sur la probabilité d’observer A. On en déduit $P(A \cap B) =$

$P(A)P(B)$ et que par conséquent, on a aussi $P(B|A) = P(B)$. On dit alors que ces évènements sont *mutuellement indépendants.*

Exercice 4.1 Poker *Vous jouez au poker, version Texas Hold'em*[1]*. Le croupier distribue les cartes et vous recevez vos deux cartes : un as et un valet de pique (!).*

1. *Quelle est la probabilité que vous obteniez une quinte flush royale ? (une série de cinq cartes dans la même couleur allant de l'as au 10).*
2. *Quelle est la probabilité que vous obteniez un full formé de trois as et de deux valets ?*

Exercice 4.2 Bits *Des bits sont envoyés à la suite sur un canal, et sont égaux soit à 0, soit à 1 avec même probabilité.*

1. *Vous enregistrés deux bits à la suite* b_0, b_1.
   - *Quelle est la probabilité que les deux bits aient des valeurs différentes (l'un vaut 0, l'autre vaut 1) ?*
   - *Quelle est la probabilité que le second bits soit égal à 1 si le premier est égal à 0 ?*
2. *On note maintenant la valeur de quatre bits envoyés à la suite.*
   - *Trouvez la probabilité d'avoir au moins deux bits égaux à 1 parmi les quatre.*
   - *Trouvez la probabilité d'avoir au moins deux bits égaux à 1 tout en étant certain (sachant que) que l'un d'entre eux l'est.*
   - *Trouvez la probabilité que tous les bits soient effectivement égaux à 1 sachant qu'au moins deux d'entre eux sont effectivement égaux à 1.*
   - *Trouvez la probabilité d'avoir au moins deux bits égaux à 1 tout en étant certain (sachant que) que l'un d'entre eux l'est et que le quatrième bit est à 0.*

Exercice 4.3 *D'après le tableau suivant, dans quel(s) cas les évènements* A *et* B *sont-ils indépendants ?*

	$P(A)$	$P(B)$	$P(A \cup B)$
cas I	*0.1*	*0.9*	*0.91*
cas II	*0.4*	*0.6*	*0.76*
cas III	*0.5*	*0.3*	*0.73*

Exercice 4.4 Simuler des probabilités conditionnelles *On considère un espace de probabilité* Ω *sur lequel on définit deux évènements* A *et* B. *On répète* N *fois une expérience* $\omega \in \Omega$ *et on note le nombre de fois où se produisent :*

1. Dans cette version du jeu, chaque joueur reçoit d'abord deux cartes. Puis cinq cartes sont découvertes au centre de la table. Le joueur qui "gagne" – on ignore ici l'effet des mises – est celui qui fabrique la meilleure main de cinq cartes en composant avec les siennes et les cartes communes à tous les joueurs.

– $n(A)$ = *nombre de réalisations de* A
– $n(B)$ = *nombre de réalisations de* B
– $n(A \cap B)$ = *nombre de réalisations simultanées de* A *et* B

puis on calcule les fréquences qui approxime les probabilités théoriques $P(A) \sim \frac{n(A)}{N}$, $P(B) \sim \frac{n(B)}{N}$, $P(A \cap B) \sim \frac{n(A\cap B)}{N}$ *et* $P(A|B) \sim \frac{n(A\cap B)}{n(A)}$.

(a) Simuler la probabilité conditionnelle d'obtenir, en lançant deux dés, une somme supérieure à 6 sachant qu'au moins un des deux dés est supérieur ou égal à 3.

(b) Simuler la probabilité conditionnelle d'obtenir, en jouant 100 *fois à pile ou face, la probabilité d'avoir pile 52 fois sachant que vous avez eu pile aux deux premiers lancers.*

Exercice 4.5 $P(A \cap B \cap C)$

Soient A, B, C *des évènements quelconques, tels que* A *et* $A \cap B$ *sont de probabilités non nulles. Utilisez la formule de probabilité conditionnelle pour déduire l'identité* $P(A)P(B|A)P(C|A \cap B) = P(A \cap B \cap C)$.

Exercice 4.6 Emission de bits *Sur une séquence* S *de* n *bits, on définit une procédure d'ajout d'un* $(n+1)$*ème bit* b. *La séquence* S *est reçue (sur un canal), puis on construit une séquence* S' *qu'on renvoie.*

Si $|S|_1$ *(le nombre de bits égaux à 1 dans* S*) est impair alors* $b = 1$ *sinon* $b = 0$. *On obtient ainsi une séquence de bit* S'. *L'émission des bits n'étant pas sécurisée, il se peut que la séquence* S' *ne soit pas cohérente. On suppose qu'un bit 1 de la séquence* S' *est émis avec la probabilité* p. *Dans quelle mesure l'ajout du* $(n+1)$*ème bit permet-t-il de contrôler que la séquence* S *n'a pas été altérée ?*

Exercice 4.7 *Montrer :*

$$P(A|B \cap C) = P(A|C) \cdot \frac{P(B|A \cap C)}{P(B|C)}$$

4.1 Formule de Bayes

Les probabilités conditionnelles ne sont pas symétriques, c'est-à-dire qu'il n'y a a priori aucune raison pour qu'on ait $P(A|B) = P(B|A)$.

Prenons encore une fois un exemple. Supposons que l'on mette au point un algorithme d'analyse du langage naturel qui détermine (de manière probabiliste) si une page web qu'il analyse est écrit en anglais "natif". Les concepteurs de l'algorithme sont assez confiant et estiment que l'algorithme dit juste 95% du temps. On estime par ailleurs que 55% des pages écrites en anglais et publiées sur le web sont en anglais natif. On peut donc chercher à évaluer la robustesse de l'algorithme en calculant la probabilité qu'il affirme qu'un page est en anglais natif alors que ce n'est pas le cas (*faux positif*), ou qu'une page soit écrite en anglais natif alors que l'algorithme affirme le contraire (*faux négatif*). De la même manière, on peut considérer les cas où l'algorithme dit juste (*vrai positif* : l'algorithme affirme que la page est en anglais natif et c'est bien le cas). ; ou le cas où la page n'est pas en anglais natif comme l'affirme l'algorithme (*vrai négatif*).

Soit Ω l'ensemble des pages web écrites en anglais, et désignons par A l'évènement l'algorithme affirme qu'une page est en anglais natif, et B l'évènement "une page est écrite en anglais natif". La probabilité d'observer un faux positif est donc $P(A|B^c)$. La probabilité d'observer un faux négatif est $P(B|A^c)$.

Dans l'exemple, il est vraisemblable d'évaluer les probabilités $P(A|B)$ et $P(A|B^c)$ en faisant tourner l'algorithme sur un échantillon de pages web dont on contrôle *a priori* la qualité de l'anglais. (Il nous faut aussi pouvoir évaluer la probabilité $P(B)$). Mais qu'en est-il de $P(B|A)$? Il se trouve que ces deux probabilités sont liés par la **formule de Bayes** :

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} \tag{4.2}$$

que l'on déduit simplement de la définition des probabilités $P(A|B)$ et $P(B|A)$.

Ainsi, on peut faire jouer la dépendance d'une variable sur l'autre à condition de pouvoir connaître au moins en partie (et dans l'un des deux sens) les probabilités conditionnelles.

Selon que l'on connaît la valeur $P(A)$ ou pas, on pourra utiliser la formule (4.2) ou une forme étendue :

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)} \tag{4.3}$$

Pour montrer la formule (4.3), il suffit d'utiliser l'identité (4.1). On trouve :

$$\begin{aligned}
P(B|A) &= \frac{P(B \cap A)}{P(A)} \\
&= \frac{P(A \cap B)P(B)}{P(A)P(B)} \\
&= \frac{P(A|B)P(B)}{P(A)} \\
&= \frac{P(A|B)P(B)}{P(A \cap B) + P(A \cap B^c)} \\
&= \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)}
\end{aligned}$$

Exercice 4.8 Langue naturelle . . .

On met au point une heuristique (un algorithme) qui détecte si une page web écrite en anglais est rédigée par un natif *(quelqu'un dont l'anglais est la langue maternelle). On évalue à 55% le pourcentage de pages sur le web qui sont écrites en anglais par des natifs. L'heuristique réussit à détecter correctement que la page est écrite par un natif dans 95% des cas lorsque la page est effectivement écrite par un natif. Elle affirme cependant incorrectement que la page est écrite par un natif alors que ce n'est pas le cas avec probabilité 1%.*

Quelle est la probabilité qu'une page soit écrite par un natif lorsque l'heuristique l'affirme ?

Exercice 4.9 Processeurs en panne *Un contrôle systématique est effectuée sur un ensemble de processeurs dont 15% présentent une panne non apparente. Ce dépistage est débuté par un test qui donne 95% de résultats positifs pour les processeurs défectueux, et 10% de résultats positifs pour les processeurs non défectueux. Quelle est la probabilité (conditionnelle) qu'un processeur pris au hasard présente la panne sachant que le test a donné un résultat positif ?*

Exercice 4.10 Bayes encore *Montrez une version généralisée de la formule de Bayes qui implique trois évènements,* A*,* B*,* C *disjoints et complémentaires, c'est-à-dire que* $A \cup B \cup C = \Omega$ *et* $A \cap B \cap C = \emptyset$*. Soit un autre évènement* D*. Exprimez* $P(A|D)$ *en fonction des probabilités conditionnelles* $P(B|D), P(C|D)$ *(et des probabilités* $P(A), P(B), P(C)$*, etc.).*

Exercice 4.11 Barrettes mémoire *Une usine fabrique des barrettes mémoire à l'aide de trois machines* A*,* B *et* C*. La machine* A *assure 20% de la production et 5% des barrettes fabriquées par A sont défectueuses. La machine B assure 30% de la production et 4% des barrettes fabriquées par B sont défectueuses. La machine C assure 50% de la production et 1% des barrettes fabriquées par C sont défectueuses.*

1. *On choisit au hasard une barrette. Calculer les probabilités :*
   - *pour que la barrette soit défectueuse et produite par* A*,*
   - *pour que la barrette soit défectueuse et produite par* B*,*
   - *pour que la barrette soit défectueuse et produite par* C*.*
2. *Calculer les probabilités pour qu'une barrette défectueuse :*
   - *provienne de* A*,*
   - *provienne de* B*,*
   - *provienne de* C*.*

Exercice 4.12 Règles des causes totales *Soient* A, B, C, D *des évènements tels que* B, C *et* D *sont mutuellement exclusifs et tels que* $B \cup C \cup D = \Omega$*. Montrez qu'on a :*

$$P(A) = P(A|B)P(B) + P(A|C)P(C) + P(A|D)P(D).$$

Proposez une généralisation de cette règle des causes totales.

Exercice 4.13 Paquets perdus *L'envoi d'un paquet du serveur S_1 au serveur S_2 sur internet passe par deux routeurs intermédiaires R_1 et R_2. La probabilité que le paquet se perde au niveau S_1, R_1 ou R_2 est p. On constate, au niveau du serveur S_2 la perte du paquet. Quelle est la probabilité qu'il ait été perdu au niveau de S_1 ? de R_1 ? de R_2 ?*

Solution On sait que la probabilité qu'un paquet se perde sur chaque segment du chemin $S_1 \to R_1 \to R_2 \to S_2$ est p. Notons par $P(S_1 \to R_1)$, $P(R_1 \to R_2)$, $P(R_2 \to S_2)$ les probabilités que la perte se produise sur un des segments du chemin. On peut calculer la probabilité que le paquet se perde (constaté au niveau du serveur S_2) : soit le paquet est perdu sur le premier segment, soit sur le deuxième, soit sur le troisième. Cette probabilité est donc égale à $P(S_1 \to R_1) + P(R_1 \to R_2) + P(R_2 \to S_2) = p + (1-p)p + (1-p)^2 p$.

Les probabilités conditionnelles cherchées sont donc $P(\text{perte}|S_1 \to R_1) = P(\text{perte})/P(S_1 \to R_1) = p/(p+(1-p)p+(1-p)^2 p)$. Les autres valeurs se calculent de manière similaire $P(\text{perte}|S_1 \to R_1) = P(\text{perte})/P(S_1 \to R_1) = p/(p+(1-p)p+(1-p)^2 p)$., $P(\text{perte}|R_1 \to R_2) = (1-p)p/(p+(1-p)p+(1-p)^2 p)$, $P(\text{perte})/P(R_2 \to S_2) = (1-p)^2 p/(p+(1-p)p+(1-p)^2 p)$.

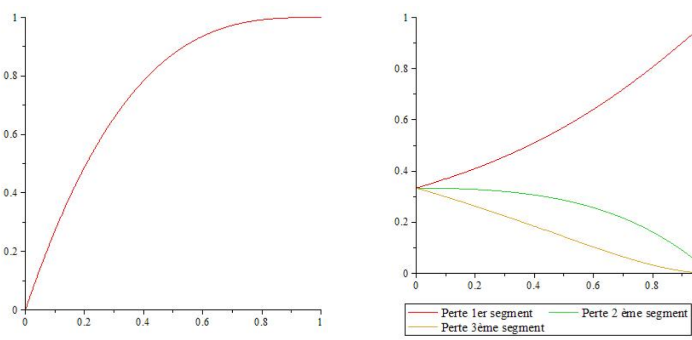


Fig. 4.1 La courbe de gauche indique comment évolue la probabilité de perdre un paquet en fonction de p (la probabilité de perdre un paquet sur l'un des segments. Les courbes de la figures de droite donne l'évolution des probabilités conditionnelles en fonction de p. Les probabilités conditionnelles concernant les deux derniers segments décroissent alors que la probabilité conditionnelle concernant le premier segment tend vers 1.

Il est intéressant de voir comment évolue ces probabilités conditionnelles en fonction de la valeur de p. A l'évidence, plus p augmente, plus il est raisonnable d'attribuer la perte d'un paquet au segment $S_1 \to R_1$. On observe alors naturellement une décroissance des probabilités conditionnelles impliquant les deux derniers segments du chemin [2]. □

Exercice 4.14 Paquets endommagés *Un routeur reçoit l'essentiel des paquets qu'il fait transiter depuis deux autres routeurs R_1 et R_2. On constate qu'environ un paquet sur 200 est endommagé* [3] *lorsque ceux-ci proviennent de R_1, alors que seulement un paquet sur 1000 est endommagé lorsqu'ils proviennent de R_2. On considère un lot de paquets reçus de l'un de ces deux routeurs R_1 ou R_2. Le premier paquet inspecté n'est pas endommagé. Quelle est la probabilité pour que le second paquet ne soit pas, lui non plus, endommagé ?*

Exercice 4.15 Moteurs de recherche *On met au point un algorithme qui lance un même requête sur trois moteurs de recherche M_1, M_2 et M_3. La précision de ces moteurs de recherche est évaluée en fonction de la pertinence des réponses qu'ils retournent, et de manière probabiliste : M_1 renvoie une réponse pertinente 3 fois sur 4, M_2 1 fois sur 4 et M_3 1 fois sur 2. Après avoir lancé ses requêtes, l'algorithme retient deux des réponses comme étant pertinentes et rejette la troisième. Quelle est la probabilité que la réponses rejetée proviennent de M_1 ? de M_2 ? de M_3 ?*

Exercice 4.16 *On tire au hasard deux points sur le segment $[0, 1)$ indépendamment l'un de l'autre. Le plus petit nombre est supérieur à 1/3. Quelle est la probabilité pour que le plus grand soit supérieur à 3/4 ?*

Exercice 4.17 Spam *Vous venez d'installer un module de détection de courriers indésirables (spam) dans votre client de courrier électronique. Le module réussit à identifier les courriers indésirables dans 99% des cas. Son taux de* faux positifs *est toutefois de 2%* [4]*. Les statistiques officielles indiquent que 10% du courriers électroniques reçus est indésirable. Quelle est la probabilité qu'un message soit effectivement indésirable lorsque le module indique que c'est le cas ?*

---

2. Les expressions pour les probabilités conditionnelles convergent vers 1/3 lorsque $p \to 0$. Elles sont donc non continues en ce point puisque ces probabilités sont évidemment nulles si $p = 0$.

3. On suppose disposer d'un procédé pour déterminer si les données d'un paquet ont subi un dommage.

4. Ce sont les cas où le module annonce qu'un message est indésirable alors qu'il ne l'est pas.

Chapitre 5
Variables aléatoires, espérance et variance

Connaître la probabilité d'un évènement A, c'est pouvoir calculer la "taille" de l'ensemble sous-jacent à A. Les choses sont parfois moins "directes", et décrites de manière composite : dans un casino, vous misez 2 € et vous lancez trois dés. Si vous avez un triple vous empochez 5 € (vous gagnez 3 €), si vous avez un double vous empochez 3 € (vous gagnez 1 €). Sinon, vous perdez votre mise (vous gagnez -2 €).

Naturellement, on s'intéresse à la probabilité de gain. Quelle est la probabilité de perdre sa mise, de gagner 1 € ou 3 € ? Les gains dépendent évidemment des évènements possibles lors du lancer des dés, mais les probabilités cherchées sont formulées en fonction du gain, et non pas en fonction des évènements eux-mêmes. En d'autres termes, les probabilités dépendent toujours des évènements sur $\Omega = \{1, \ldots, 6\}^3$ mais les probabilités sont formulées en fonction du gain possible du joueur.

Il faut voir le gain comme une fonction $X : \Omega \to \mathbb{N}$ dont le domaine est l'espace où sont lancés les dés et à valeurs dans l'ensemble des entiers. Cette fonction est une *variable aléatoire*. Ainsi, si le lancer de dés est $\omega = (3, 3, 3)$ (un triple de 3) alors le gain du joueur est de $X(\omega) = 3$.

On peut ainsi se poser la question de connaître la probabilité d'avoir un gain de 3 € qu'on écrit $P(X = 3)$. La réponse se fait en calculant l'évènement $A = X^{-1}(3)$ dans Ω, c'est l'ensemble de tous les lancers de dés qui donne un gain de 3 €. Dans cet exemple, on a $A = \{(1, 1, 1), \ldots, (6, 6, 6)\}$ et la probabilité est donc $P(X = 3) = |A|/|\Omega| = 1/36$.

Exercice 5.1 Casino *Reprenez l'exemple du joueur de casino qui lance les dés. Calculez les probabilités $P(X = 1)$ et $P(X = -2)$.*

Exercice 5.2 Système actif/passif *On effectue une séquence de N observations d'un système qui se trouve dans l'un de deux états 0 ou 1, avec probabilité p d'être actif (valeur observée = 1). On considère la variable aléatoire qui donne le nombre de fois où le système est actif (on observe la valeur 1).*

Précisez ce qu'est Ω, définissez la variable aléatoire X. Calculez la probabilité $P(X = k)$.

Exercice 5.3 Phénotypes et génotypes *Certains caractères héréditaires sont portés par paires de gènes, un gène pouvant prendre deux formes (allèles), dominantes ou récessives (couleur des yeux, caractère de la peau des petits pois, etc.). Ces gènes sont donc groupés par paires et se présentent dans l'organisme sous trois* génotypes *possibles : AA, Aa, aa (on ne distingue pas Aa et aA). Le phénotype suit de la dominance du caractère A sur a. Lorganisme porteur montrera le phénotype qui suit du gène dominant A mais pourra tout de même transmettre la gène récessif a. Ainsi, une personne aux yeux marron peut très bien avoir des enfants aux yeux bleus.*

Chaque parent transmet à un descendant l'un des deux gènes qu'il porte. On suppose que le gène transmis par le parent est choisi au hasard parmi les deux possibles.

On définit deux variables aléatoires X_1, X_2 qui prennent les valeurs des phénotypes AA, Aa ou aa des deux parents et Y_1, Y_2 qui prennent les valeurs du gène A ou a transmis par chacun des deux parents à leur descendant.

Sachant que le gène transmis par le parent est choisi au hasard parmi les gènes de son génotype, et sachant que le gène transmis par l'un des parents ne dépend pas de celui transmis par l'autre, donnez les probabilités conditionnelles des variables Y_i sachant X_j.

On suppose que les génotypes AA, Aa, aa sont répartis dans la population avec probabilité $p, 2r, q$ (avec $p + 2r + q = 1$). Exprimez p, q et r en fonction des probabilité sur les Y_i. (Vous utiliserez la règle des causes totales.)

Exercice 5.4 Guêpe et probabilité conditionnelle *Deux pièces A et B sont reliées entre elles par une porte ouverte. Seule la pièce B possède une issue vers l'extérieur. Une guêpe initialement dans la pièce A voudrait sortir à l'air libre. Son trajet obéit aux règles suivantes :*

- *Lorsqu'elle est en A au temps $t = n$, alors au temps $t = n + 1$, elle reste en A avec une probabilité égale à 1/3 ou elle passe en B avec une probabilité égale à 2/3,*
- *Lorsqu'elle est en B au temps $t = n$, alors au temps $t = n + 1$, elle retourne en A avec une probabilité égale à 1/4, ou elle reste en B avec une probabilité égale à 1/2, ou elle sort à l'air libre avec une probabilité égale à 1/4.*

Au temps $t = 0$, la guêpe est dans la pièce A. Lorsqu'elle est sortie, elle ne revient plus.

1. *Calculez explicitement les distributions de probabilités des variables X_0 et X_1,*
2. *Exprimez $P(X_{n+1} = A)$ et $P(X_{n+1} = B)$ en fonction de $P(X_n = A)$ et $P(X_n = B)$ (avec des notations évidentes),*
3. *Vérifiez que la suite $\frac{6}{10}P(X_n = A) - \frac{3}{10}P(X_n = B)$ est constante,*
4. *Vérifiez que la suite $\frac{4}{10}P(X_n = A) + \frac{3}{10}P(X_n = B)$ est géométrique de raison $\frac{5}{6}$,*
5. *En déduire l'expression de $P(X_n = A)$ et $P(X_n = B)$,*

6. *Montrer que pour* $n \geq 2$, $P(X_n = S) = \frac{1}{4}\sum_{k=1}^{n-1} P(X_{n-1} = B)$. *En déduire* $P(X_n = S)$.

Exercice 5.5 *Soit* $a \in (0, 1/2)$ *un nombre réel.*

Dans une bourse de valeurs, un titre donné peut monter, *rester* stable *ou* baisser.

Dans un modèle mathématique, on considère que :

- *le premier jour le titre est stable*
- *si un jour* n *le titre monte, le jour* $n+1$ *; il montera avec probabilité* $1-2a$, *restera stable avec probabilité* a *et baissera avec probabilité* a
- *si un jour* n *le titre est stable, le jour* $n+1$ *il montera avec probabilité* a, *restera stable avec probabilité* $1-2a$ *et baissera avec probabilité* a
- *si un jour* n *le titre baisse, le jour* $n+1$ *il montera avec probabilité* a, *restera stable avec probabilité* a *et baissera avec la probabilité* $1-2a$

On note M_n *(resp.* S_n*, resp.* B_n*) l'événement "le titre donné monte" (resp. reste stable, resp. baisse) le jour* n. *On pose* $p_n = P(M_n)$, $q_n = P(S_n)$ *et* $r_n = P(B_n)$.

a) *Que vaut* $p_n + q_n + r_n$ *? En déduire l'expression de* r_n *en fonction de* p_n *et* q_n,
b) *Expliciter* p_{n+1} *(resp.* q_{n+1}*) en fonction de* p_n, q_n, r_n,
c) *En déduire* p_n, q_n *puis* r_n,
d) *Donner la limite de ces trois suites et interpréter le résultat.*

5.1 Distribution de probabilité

Une variable aléatoire calcule donc une valeur associée à toute épreuve $\omega \in \Omega$ ou à tout évènement $A \subset \Omega$. Elle peut prendre ses valeurs dans tout ensemble, bien que nous considérerons le plus souvent des variables aléatoires à valeurs dans $\mathbb{N}$ ou $\mathbb{R}$.

La distribution de probabilité d'une variable aléatoire décrit la probabilité associée à chacune des valeurs qu'elle peut prendre. certaines distributions de probabilités ont un caractère générique.

5.1.0.1 Distribution binomiale

Un évènement aléatoire à valeur 0 ou 1, avec probabilité p de succès (valeur = 1) est appelé un *évènement de Bernouilli* de probabilité p. L'évènement de Bernouilli par excellence est le tirage d'une pièce à pile ou face. Observez l'état 0 ou 1 d'un système est aussi modélisé par un évènement de Bernouilli.

On considère une suite de variables aléatoires $X_1, \ldots, X_N$ où les variables X_i correspondent à des évènements de Bernouilli *indépendants* de probabilité p. On considère la variable aléatoire $S = X_1 + \cdots X_N$.

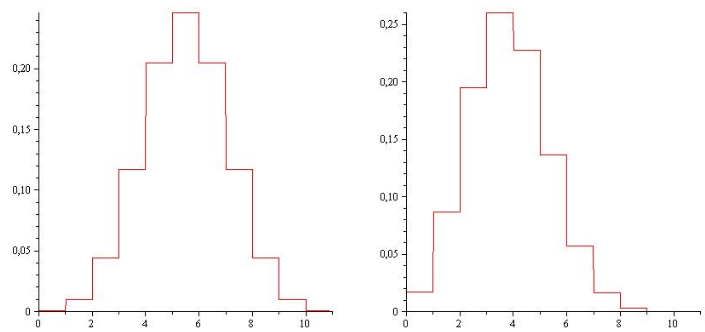


Fig. 5.1 La distribution binomiale peut être décrite à l'aide d'un histogramme ou d'une courbe brisée. Observez sa symétrie lorsque $p = 1/2$ (gauche), qui est perdue lorsque $p \neq 1/2$ (droite avec $p = 1/3$).

Exercice 5.6 *Donnez la distribution de probabilité de S. Cette distribution est la* distribution binomiale *de paramètres (N, p) que l'on a déjà vu à l'exercice 5.2.*

Une variable aléatoire est elle-même une fonction $X : \Omega \to \mathbb{N}$ (ou $\mathbb{R}$), que l'on peut composer avec une seconde fonction $f : \mathbb{R} \to \mathbb{R}$. On peut, par exemple, considérer la fonction $f : x \mapsto x^2$ et considérer une nouvelle variable aléatoire $f(X)$ qui calcule le carré de la valeur retournée par X. Il faut souvent s'assurer de conditions satisfaite par la fonction f, mais nous aurons le plus souvent affaire à des cas simples où la fonction est continue et monotone croissante.

5.2 Espérance

Dès lors que l'on considère une variable aléatoire, on s'intéresse aux valeurs qu'elle peut prendre et aux probabilités qu'elle atteigne effectivement ces valeurs. On s'intéresse naturellement à la valeur moyenne de la variable, qui donne une première indication de son "comportement". Dans l'exemple du joueur au casino, on peut effectivement se demander de combien sera le gain du joueur en moyenne. En réalité, ce sont les gérants du casino qui doivent prendre soin de calculer cette valeur, puisque le joueur par nature espère pouvoir échapper à la loi des grands nombres !

Chacun sait calculer la moyenne de la variable aléatoire qui donne le résultat d'un lancer de dé. (Cette variable est horriblement simple, puisqu'elle est définie sur l'ensemble $\Omega = \{1, \ldots, 6\}$ et retourne la valeur du dé lancer). Vous faites la somme des valeurs possibles divisée par le cardinal de l'ensemble, et vous trouvez 3,5. Ce cas parait facile car il est particulier : les valeurs prises par la variable sont équiprobables. Vous procédez de même si vous calculez la moyenne de votre semestre, *lorsque toutes les matières ont même coefficient.* Et si les coefficients varient ? Vous en tenez compte dans le calcul, vous effectuez alors une somme pondérée – ce raisonnement vaut aussi pour calculer la moyenne d'une variable aléatoire.

Prenons un autre exemple, celui où la variable aléatoire Y retourne la somme de deux dés lancés sur le tapis. L'espace de probabilités est $\Omega = \{(i,j)|i,j \in \{1,\ldots,6\}\}$. Elle prend donc ses valeurs dans $\{2,\ldots,12\}$. On a $P(Y=2) = 1/36$, $P(Y=3) = 2/36$, $P(Y=4) = 3/36, \ldots$ La moyenne de la variable doit tenir compte des valeurs prises par la variable, mais aussi des probabilités d'atteindre chacune de ces valeurs. On définit l'espérance d'un variable aléatoire X :

$$\mathbb{E}(X) = \sum_k k \cdot P(X = k) \tag{5.1}$$

où la somme parcourt l'ensemble des valeurs prises par la variable (nous n'avons considéré que des valeurs entières jusqu'à maintenant et ça nous suffira pendant un moment – les choses ne sont pas foncièrement différente si on passe aux variables à valeurs dans R).

Exercice 5.7 Casino – encore *Quel est l'espérance du gain du joueur de dés au casino ?*

Exercice 5.8 Une formule simple *Soit $X : \Omega \to \mathbb{N}$ une variable aléatoire, montrez que $\mathbb{E}(X) = \sum_{n\geq 0} P(X \geq n)$.*

Définition 1. Fonction indicatrice Sur tout espace de probabilité Ω, et pour tout évènement $A \subset \Omega$, on peut construire une variable aléatoire $1_A : \Omega \to 0, 1$ qui vaut $1_A(\omega) = 1 \iff \omega \in A$. En d'autres mots la variable indique si une épreuve ω appartient à A. Ainsi, on a $P(1_A = 1) = P(A)$.

Exercice 5.9 *Observez $\mathbb{E}(1_A) = P(A)$.*

Exercice 5.10 Loi binomiale *Calculer l'espérance d'une variable suivant la loi binomiale de paramètres (N, p).*

Exercice 5.11 Analyse d'algorithme *On se propose d'analyser la complexité (temps de calcul) en moyenne d'un algorithme. On considère l'algorithme ci-dessous pour trouver le plus grand élément dans une liste non vide L de n entiers :*

```
début
```

```
(1) M := L[1]
(2) pour j := 2 à n faire
(3)   si L[j] > M alors M := L[j]
fin
```

Lors du traitement d'une liste de taille n *:*

– *Quel est le nombre de comparaisons effectuées (ligne 3) ?*
– *Quel est le nombre d'affectations (lignes 1 et 3) dans le cas le plus favorable ? le plus défavorable ?*

Dans la suite, on suppose que les éléments de la liste sont deux-à-deux distincts et que les $n!$ *permutations possibles ont la même probabilité d'apparaître dans la liste. Soit* $p(n,k)$ *la probabilité pour que, lors du déroulement de l'algorithme sur une liste (aléatoire)* L *de taille* n*,* k *affectations soient nécessaires.*

– *Montrez la récurrence suivante :*

$$p(n,k) = \frac{1}{n}p(n-1,k-1) + \frac{n-1}{n}p(n-1,k)$$

pour $n \geq 2$ *et* $1 \leq k \leq n$*. Il faut distinguer deux cas de figures et raisonner selon que le* n^{me} *élément est maximal ou non.*

– *Soit* $\mathbb{E}_n$ *l'espérance mathématique (la moyenne) du nombre d'affectations effectuées lors du traitement d'une liste de taille* n*. Déduire de la récurrence ci-dessus la récurrence sur* $\mathbb{E}_n$

$$\mathbb{E}_n = \frac{1}{n} + \mathbb{E}_{n-1},$$

dont on peut déduire $\mathbb{E}_n = \sum_{k=1}^{n} \frac{1}{k}$*.*

– *Montrez que l'ordre de grandeur de* $\mathbb{E}_n$ *se compare à* $\log n$ *lorsque* $n \to \infty$*.*

Exercice 5.12 Simuler un loi binomiale *Soit la variable aléatoire* $S = X_1 + \cdots + X_n$ *où les* X_i *correspondent à des évènements de Bernouilli avec probabilité* p*. Donnez un procédé permettant de simuler cette variable aléatoire. En d'autres mots vous tirez au hasard des entiers variant de 0 à* N*, non pas de manière uniforme mais de façon à ce que la probabilité d'obtenir l'entier* k *soit égale à* $P(S = k)$*.*

5.2.0.2 Loi géométrique

Une variable aléatoire X suit une loi géométrique de paramètre p si $P(X = k) = (1-p)^k p$. Cette loi a une interprétation assez intuitive :

– Vous jouez à pile ou face, et vous lancez la pièce tant que vous n'avez pas obtenu pile (avec probabilité p d'obtenir pile). La variable qui indique le nombre de fois où vous avez lancé la pièce avant d'obtenir pile suit une loi géométrique.

– Vous observez un signal (0 ou 1), et vous poursuivez l'observation en notant à quel moment (combien de signaux reçus) vous avez vu avant d'observer la valeur 1.

Exercice 5.13 *Calculez l'espérance et la variance de la loi géométrique.*

Solution Ce calcul fait appel, sans surprise, à la manipulation de séries géométriques (voir section 2.1). Désignons par X la variable de loi géométrique de paramètre p. La série $s(q) = \sum_{i\geq 0} q^i = \frac{1}{1-q}$ peut être vue comme une fonction réelle de la variable $q \in [0,1)$, que l'on peut donc dériver $s'(q) = \sum_{i\geq 1} iq^{i-1} = \frac{-1}{(1-q)^2}$, $s''(q) = \frac{1}{(1-q)^3}$.

$$\mathbb{E}(X) = \sum_{k\geq 1} k(1-p)^{k-1}p = ps'(1-p) = \frac{1}{p}$$

Le même type de raisonnement (en prenant soin de remarquer que $k^2 = k(k-1)+k$ permet de trouver $\mathbb{V}(X) = \frac{1-p}{p^2}$. □

Exercice 5.14 Le collectionneur *Imaginez que vous collectionnez des objets offerts à chaque achat au supermarché. La collection compte N objets différents et vous vous mettez en tête de continuer à faire vos courses à ce supermarché pour faire vos courses, chaque semaine disons, jusqu'à ce que vous ayez la collection complète. Pendant combien de semaines aurez-vous à faire vos courses avant d'avoir obtenu la collection complète (en moyenne) ?*

Modélisez la situation ainsi. Vous considérez les variables $X_1, \ldots, X_N$ qui indique le temps d'attente (en nombre de semaines) avant d'obtenir une $i$ème carte. On a nécessairement $X_1 = 1$ puisque la première carte démarre la collection. Qu'en est-il de X_2 ?..., puis X_9 ?

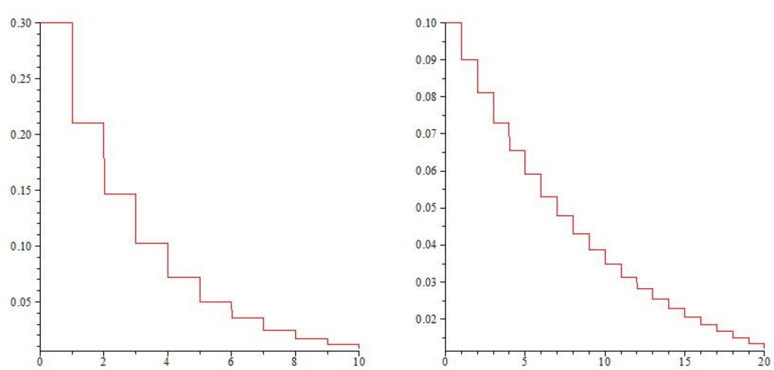


Fig. 5.2 La distribution de la loi géométrique suit une courbe qui décroit exponentiellement (gauche $p = 0.3$, droite $p = 0.1$).

Solution Le nombre d'achats effectués avant d'avoir obtenu enfin la $i$ème carte est donné par la variable X_i. On a donc $X_1 = 1$ avec probabilité 1 puisque c'est la carte qui démarre la collection. Supposons que la collection contiennent i cartes distinctes (attention, on peut avoir acheté bien plus de i carte pour avoir ces i cartes différentes). Lors d'un prochain achat, on aura une nouvelle carte pour la collection avec probabilité $p = \frac{N-i}{N}$ puisqu'il reste $N - i$ cartes à collectionner ; on aura une carte déjà présente dans la collection avec probabilité $1 - p = \frac{i}{N}$. La variable X_{i+1} qui donne le nombre d'achats effectués avant d'avoir une nouvelle carte suit donc une loi géométrique de paramètre p : *elle indique le nombre de fois où il faut jouer avant d'avoir un succès – obtenir enfin une nouvelle carte.* Son espérance $\mathbb{E}(X_{i+1}) = 1/p = N/(N - i)$ correspond au nombre moyen de fois où il faudra faire un achat pour avoir une nouvelle carte à ce stade de la collection. Le nombre total d'achats à effectuer en moyenne est $\mathbb{E}(X_1 + \cdots + X_N) = \sum_{i=1}^{N} \mathbb{E}(X_i) = N \sum_{i=1}^{N} \frac{1}{i}$. Or la somme $H_N = \sum_{i=1}^{N} \frac{1}{i}$ connue sous le nom "série harmonique" s'approche de la valeur $\log(N)$ à une constante près (la constante d'Euler). Le nombre moyen d'achats à effectuer pour compléter la collection est donc de l'ordre de $N \log(N)$. $\square$

Exercice 5.15 Simuler un loi géométrique *Simulez une loi géométrique. En d'autres mots, donnez un procédé qui permet de choisir un entier au hasard tel que l'entier k soit choisi avec probabilité $(1-p)^k p$. On peut procéder de deux manières différentes. Rappelons d'abord que $\sum_{k=0}^{\infty}(1-p)^k p = 1$.*

- *On choisit au hasard un nombre réel $r \in [0, 1)$. On calcule l'entier i tel que $i = \inf_i \sum_{k=0}^{i}(1-p)^k p > r$. C'est l'entier choisit au hasard par la loi géométrique.*
- *On choisit un nombre réel $r \in [0, 1)$. On calcule $i = \lfloor \frac{\log U}{\log 1-p} \rfloor + 1$. C'est l'entier retourné par la loi géométrique.*
- *Montrer que la variable $X = \lfloor \frac{\log U}{\log 1-p} \rfloor + 1$ suit bien une loi géométrique (de paramètre p). En effet, puisque $P(X = k) = P((1-p)^{k-1} \leq U < (1-p)^k)$.*

5.2.0.3 Distribution de Poisson

Imaginons des évènements de Bernouilli qui se produisent de manière indépendante et qui se répètent à des temps réguliers. On peut penser à la réception de paquets provenant de N canaux distincts, à la détection de noyaux d'hélium émis par des atomes d'uranium, etc. Cette façon de voir les choses modélise bien la réalité sur un temps très court. Un canal émettra, ou non, un seul paquet ; un atome d'uranium émettra, ou non, un seul noyau d'hélium.

Comme précédemment, on peut considérer la variable $S_N = X_1 + \cdots + X_N$ qui compte le nombre de succès observés sur un intervalle de temps très court, la difficulté étant qu'on ne connaît pas le nombre N d'évènements – puisqu'on observe que le nombre k de succès. En revanche, on suppose pouvoir évaluer le nombre moyen λ de succès observés sur un intervalle de temps.

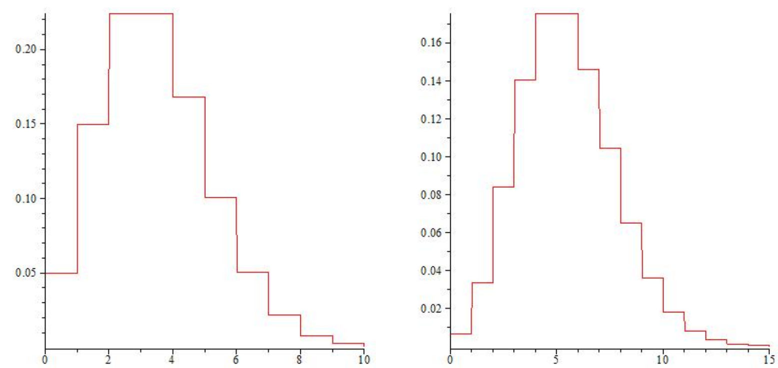


Fig. 5.3 La distribution de Poisson suit une courbe centrée en λ, qui décroît exponentiellement et tend asymptotiquement vers 0 lorsque $k \to +\infty$ (gauche $\lambda = 2$, droite $\lambda = 5$).

Or, on sait que la variable S_N suit une loi binomiale de paramètres (N, p_N), et par conséquent on a $\mathbb{E}(S_N) = Np_N$, et donc, $p_N = \lambda/N$.

Cela étant posé, la probabilité de n'observer aucun succès est égale à $P(S_N = 0) = (1 - p_N)^N = (1 - \lambda/N)^N$. On a donc :

$$\lim_{N \to \infty} P(S_N = 0) = e^{-\lambda}$$

et, pour $N > k$:

$$\begin{aligned}
\frac{P(S_N = k+1)}{P(S_N = k)} &= \frac{\binom{N}{k+1} p^{k+1}(1-p)^{N-k-1}}{\binom{N}{k} p^k (1-p)^{N-k}} \\
&= \frac{N-k}{k+1} \frac{p}{1-p} \\
&= \frac{\lambda}{k+1}
\end{aligned}$$

L'unique solution satisfaisant ces conditions nous donne :

$$\lim_{N \to \infty} P(S_N = k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

La distribution $P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$ où λ est un nombre réel positif est appelé la *distribution de Poisson* de paramètre λ.

Exercice 5.16 Dimensionner un tableau *On doit prédimensionner un tableau, qui est une ressource pour un ensemble de processus : un processus qui s'exécute a besoin d'une entrée dans le tableau. Si aucune entrée n'est disponible alors il est mis en attente dans une file. On sait qu'en moyenne λ processus s'exécutent en même temps. Comment dimensionner le tableau pour que la probabilité d'avoir à mettre (au moins) un processus en attente soi au plus de 10%.*

Solution On suppose donc que le nombre de processus à traiter (à chaque moment) est donné par une loi de Poisson de paramètre λ. Ce nombre, k, doit être tel que la probabilité de dépasser la taille du tableau, N soit au plus de 10%. En d'autres mots, on doit avoir $\sum_{k>N} P(X = k) < 10\%$, ou encore $\sum_{k\leq N} P(X = k) \geq 90\%$. La solution numérique exige de recourir à une table de la loi de Poisson (ou encore de simuler la loi avec le paramètre λ). Ainsi, lorsque $\lambda = 10$ un tableau de taille $N = 14$ suffit ; avec $\lambda = 5$, on peut prendre $N = 8$. $\square$

Exercice 5.17 Tickets *Le nombre de tickets postés sur un serveur du service informatique au cours de la première heure de la journée est environ de 10. Utilisez la loi de Poisson pour calculer la probabilité qu'en un jour normal, il n'y ait aucun ticket qui soit posté pendant les 90 première minute de la première heure.*

Solution Nous allons considérer trois variables aléatoires X, Y, Z, chacune suivant une loi de Poisson de paramètres différents. La variable X est de paramètre $\lambda = 10$ où l'unité de temps sous-jacente est l'heure. La variable Y est de paramètre $\lambda = 15$ et l'unité de temps sous-jacente est de 90 minutes. finalement, Z est de paramètre $\lambda = 5$ et l'unité de temps sous-jacente est la demi-heure. Chacune de ces variables modélise le phénomène étudiée, mais sur des intervalles de temps différents.

La valeur cherchée est $P(Y = 0)$. On a $P(Y = 0) = \sum_{k\geq 0} P(Y = 0|X = k)P(X = k) = P(Y = 0|X = 0)P(X = 0)$ puisque l'intersection $Y_0 \cap X = k$ est vide sauf si $X = 0$. On a donc :

$$P(Y = 0) = P(Y = 0|X = 0)P(X = 0) = P(Z = 0) = e^{-5} \sim 0.00673794$$

$\square$

Exercice 5.18 *Un central téléphonique possède L lignes. On estime à 1200 le nombre de personnes susceptibles d'appeler le standard sur une journée de 8 heures, la durée des appels étant de deux minutes en moyenne. On note X la variable aléatoire égale au nombre de personnes en train de téléphoner à un instant donné.*

- *Montrer que l'on est en droit d'approcher la distribution de X par une loi de Poisson. Donner ses paramètres.*

- *On suppose $L = 3$. Calculer la probabilité d'encombrement à un instant donné, à savoir $P(X > L)$.*
- *Quelle doit être la valeur minimale de L pour qu'à un instant donné, la probabilité d'encombrement ne dépasse pas $0,1$.*

Solution On s'intéresse ici au nombre d'appels réceptionnés au standard sur un intervalle de temps. C'est ce que permet de modéliser la loi de Poisson. ON peut prendre soit une loi de paramètre $\lambda = 1200$ associé à une période de temps de 8 heures, ou une loi T de paramètre $\lambda = 5$ sur une période de temps de 2 minutes.

On a donc ici, $P(T = 0) = e^{-5} \sim 0.006737947$, $P(T = 1) = 0.033689735$, $P(T = 2) = 0.08422433749$, $P(T = 3) = 0.1403738958$, $P(T = 4) = 0.1754673698$, etc.

On a encombrement de la centrale dès lors que le nombre d'appels, T, excède le nombre de lignes disponibles L. Or, $P(T > L) = \sum_{k \geq L} e^{-\lambda} \lambda^k / k!$. Lorsque $L = 3$, $P(T > 3) = 1 - \big(P(T = 0) + P(T = 1) + P(T = 2) + P(T = 3)\big) \sim 1 - 0.265 = 0.735$. Cette probabilité est inférieure à 0.1 dès lors que $L = 8$.

□

Exercice 5.19 Clients *Les clients arrivent dans une banque à un rythme moyen de 40 clients par heure qu'on modélise par une loi de Poisson de paramètre 40 (avec l'unité de temps qui est l'heure). Sachant que 30 client sont arrivés à la première heure, quelle est la probabilité pour qu'il y en ait 60 qui arrivent dans les premières 90 minutes ?*

5.3 Variance

La moyenne donne une première indication du comportement d'une variable aléatoire (ou de la distribution de ses valeurs). La variance vient compléter cette information et nous renseigne sur l'étalement des valeurs autour de la moyenne : les valeurs sont-elles en moyenne proches ou éloignées de la moyenne ?

Deux variables aléatoires de même moyenne peuvent différer au niveau de la variance. Pensez par exemple à une variable aléatoire de Rademacher de paramètre $p = 1/2$, et dont l'espérance est nulle (voir l'exercice 3.8). Une variable qui vaut 10^6 ou 10^{-6} avec même probabilité a elle aussi une espérance nulle, mais varie sur un tout autre domaine.

La traduction littérale de la définition que nous avons donnée de la variance nous amène à :

$$\mathbb{V}(X) = \mathbb{E}\big(\big(X - \mathbb{E}(X)\big)^2\big) \tag{5.2}$$

La variance de la variable X sera aussi parfois notée σ_X^2. L'*écart-type* est la racine carrée de la variance, et est notée $\sigma_X = \sqrt{\sigma_X^2}$.

Exercice 5.20 Casino – encore *Quel est la variance du gain du joueur de dés au casino ?*

Exercice 5.21 *Soit X une variable aléatoire. Observez que la variable $X - \mathbb{E}(X)$ est de même variance que X.*

Exercice 5.22 Loi binomiale *Calculer la variance d'une variable suivant la loi binomiale de paramètres (N, p).*

Exercice 5.23 Poisson *Calculez l'espérance et la variance de la distribution de Poisson de paramètre λ.*

Exercice 5.24 Algorithme *Un algorithme reçoit en entrée une liste d'entiers positifs. Sa complexité est fonction du nombre d'entiers pairs p de la liste et vaut $f(p) = p^2 + 1$. Décrivez le comportement de l'algorithme (sa complexité moyenne et sa variance en temps d'exécution).*

5.4 Somme et produit de variables aléatoires

Soient deux variables aléatoires indépendantes X_1, X_2 à valeurs dans $\mathbb{R}$. Comme ces deux variables sont indépendantes, on a :

$$P(X_1 + X_2 = z) = \sum_{x+y=z} P(X_1 = x)P(X_2 = y)$$

$$P(X_1 X_2) = \sum_{xy=z} P(X_1 = x)P(X_2 = y)$$

Par conséquent, on peut montrer :

Exercice 5.25

$$\mathbb{E}(X_1 + X_2) = \mathbb{E}(X_1) + \mathbb{E}(X_2) \tag{5.3}$$

$$\mathbb{E}(X_1 X_2) = \mathbb{E}(X_1)\mathbb{E}(X_2) \tag{5.4}$$

lorsque les variables X_1 et X_2 sont indépendantes.

De même, si X_1, …, X_n sont des variables aléatoires indépendantes à valeurs dans $\mathbb{R}$, alors :

$$\sigma^2_{X_1+\cdots+X_n} = \sigma^2_{X_1} + \cdots + \sigma^2_{X_n} \tag{5.5}$$

Exercice 5.26 Démonstration *Il suffit de démontrer l'identité (5.5) lorsque $n = 2$.*

Exercice 5.27 Scalaire *Soit une variable aléatoire X à valeurs dans $\mathbb{R}$ et $\alpha \in \mathbb{R}$ positif. On considère la variable aléatoire αX. Montrez :*

$$\mathbb{E}(\alpha X) = \alpha \mathbb{E}(X), \quad \mathbb{V}(\alpha X) = \alpha^2 \mathbb{V}(X)$$

Comparez ce résultat aux identités (5.3) et (5.5).

Exercice 5.28 *Soit X une variable aléatoire correspondant à un évènement de Bernouilli de paramètre p. Calculez son espérance et sa variance. En déduire directement le calcul de l'espérance et de la variance d'une variable aléatoire suivant une loi binomiale de paramètres N et p.*

Solution La variable X prend deux valeurs 0 ou 1 avec probabilité $1-p$ et p respectivement. Son espérance est donc $\mathbb{E}(X) = 0 \cdot (16p) + 1 \cdot p = p$. Sa variance est $(0 - \mathbb{E}(X))^2 \cdot (1-p) + (1 - \mathbb{E}(X))^2 \cdot p = p(1-p)$.

Une variable $S_N = X_1 + \cdots + X_N$ de loi binomiale de paramètres N et p est une somme de N variables X_i ($i = 1, \ldots, N$) de Bernouilli indépendantes de même paramètre p. On déduit le calcul de l'espérance et de la variance de S_N des équations (5.3) et (5.5). $\square$

Chapitre 6
Lois des grands nombres

L'espérance (la moyenne) nous donne déjà une première information sur le comportement d'une variable aléatoire. La variance nous informe sur la distribution des valeurs de la variable autour de sa valeur moyenne. La loi des grands nombres vient préciser ces informations, en terme probabilistes.

Théorème 6.0.1 Inégalité de Markov *Soit $Z : \Omega \to \mathbb{R}$ une variable aléatoire non négative. Alors :*

$$P(Z \geq a) \leq \frac{\mathbb{E}(Z)}{a} \tag{6.1}$$

Inégalité de Chebyshev *En particulier, si $X : \Omega \to \mathbb{R}$ est une variable aléatoire alors pour tout $\epsilon > 0$, on a :*

$$P(|X - \mathbb{E}(X)| \geq \epsilon) \leq \frac{\sigma_X^2}{\epsilon^2} \tag{6.2}$$

La preuve est relativement simple. On considère la variable aléatoire indicatrice $1_{Z \geq a}$. Cette variable vaut $1_{Z\geq a}(\omega) = 1$ exactement lorsque $Z(\omega) \geq a$. Par conséquent, $\forall \omega \in \Omega, Z(\omega) \geq a 1_{Z\geq a}$. Par suite, puisque l'espérance est une fonction monotone, $\mathbb{E}(Z) \geq a\mathbb{E}(1_{Z\geq a}) = aP(Z \geq a)$.

Pour obtenir la seconde identité, on applique (6.1) à la variable aléatoire $|X - \mathbb{E}(X)|^2$ et $a = \epsilon^2$. $\square$

La loi des grands nombres établit un lien entre probabilité et fréquence empirique. Dans le jeu de pile ou face, avec probabilité 1/2 d'avoir pile ou face, la loi nous assure que :

$$\lim_{n\to\infty} P(\frac{S_n}{n} - \frac{1}{2}) = 1$$

le nombre de fois où l'on observe "pile" (la valeur de la variable S_n) tend vers 1/2 si l'on joue un grand nombre de fois. On le déduit de l'identité de Chebyshev. En effet, soit $X_1, \ldots, X_n$ des variables aléatoires indépendantes et identiquement distribuées. Considérons la variables aléatoire $(X_1+\cdots+X_n)/n$, dont l'espérance est égale à $\mathbb{E}(X_1)$ (pourquoi ?) et la variance est égale à $nV(X_1)$ (pourquoi ?). L'identité de Chebyshev nous donne :

$$P(\Big|\frac{X_1+\cdots+X_n}{n}-\mathbb{E}(X_1)\Big|\geq\epsilon)\leq\frac{\sigma^2_{X_1}}{n\epsilon^2} \tag{6.3}$$

puisque $\mathbb{E}(\frac{X_1+\cdots+X_n}{n}) = \mathbb{E}(X_1)$ et $\mathbb{V}(\frac{X_1+\cdots+X_n}{n}) = \frac{\mathbb{V}(X_1)}{n}$ en vertu des identités (5.3) et (5.5) et de l'exercice 5.27. Par suite :

$$\lim_{n\to\infty}P(\Big|\frac{X_1+\cdots+X_n}{n}-\mathbb{E}(X_1)\Big|\geq\epsilon)=0. \tag{6.4}$$

C'est en ce sens que la moyenne empirique tend vers la moyenne probabiliste. Cette identité est la *loi faible des grands nombres.*

La *loi forte des grands nombres* affirme que :

$$\lim_{n\to\infty}\frac{X_1+\cdots+X_n}{n}=\mathbb{E}(X_1) \tag{6.5}$$

selon laquelle la variable aléatoire du membre gauche tend vers la variable constante de valeur $\mathbb{E}(X_1)$ – encore faudrait-il préciser ici ce que nous entendons par là puisqu'il s'agit de convergence en probabilité, mais nous ne nous étendrons pas sur ce résultat.

Exercice 6.1 *On modélise le nombre de téléchargements d'un fichier sur une heure par une variable aléatoire d'espérance 50.*

- *Estimer la probabilité que le nombre de téléchargements dépasse 75 ?*
- *On sait de plus que la variance du nombre de téléchargements est de 25. Estimer la probabilité que le nombre de téléchargements sur une heure soit compris entre 40 et 60 ?*

Exercice 6.2 *Dans une population de 30 000 000 individus, la proportion d'individus présentant le caractère C est $p=0.4$. On prélève un échantillon de taille $N=1600$ et on note X le nombre d'individus de l'échantillon présentant le caractère C.*

- *Minorer la probabilité des événements :*

$$0.30\leq\frac{X}{N}\leq0.50\quad 0.35\leq\frac{X}{N}\leq0.45\quad 0.38\leq\frac{X}{N}\leq0.42$$

- *Calculer la longueur minimale L de la fourchette telle que :*

$$P(0.40-\frac{L}{2}\leq\frac{X}{N}\leq0.40+\frac{L}{2})\geq0.95$$

Solution On modélise le comptage du nombre d'individus portant le caractère C par une variable aléatoire binomiale de paramètre $p=0.4$. Cette variable X est donc la somme d'évènements de Bernouilli $X=\sum_{i=1}^{N}X_i$ de paramètre p (avec $N=1600$). En vertu de la loi des grands nombres, on a :

$$P(|\frac{\sum_{i=1}^{N} X_i}{N} - \mathbb{E}(X_1)| \geq \epsilon) \leq \frac{\sigma_{X_1}^2}{N\epsilon^2}$$

Or, comme on a $\mathbb{E}(X_1) = p = 0.4$ et $\sigma_{X_1}^2 = p(1-p) = 0.24$, on voit que la probabilité de chacun des évènements peut être évaluée à l'aide de valeurs différentes pour ϵ. Dans le premier, on prend $\epsilon = 0.1$ et on trouve :

$$P(0.30 \leq \frac{X}{N} \leq 0.50) = P(|\frac{X}{N} - \mathbb{E}(X_1)| \geq 0.1) \leq \frac{0.24}{1600 \times 0.01} = 0.015$$

Pour déterminer le nombre L (la fourchette) on observe que : :

$$\begin{aligned} P(0.4 - \frac{L}{2} \leq \frac{X}{N} \leq 0.4 + \frac{L}{2}) &= P(|\frac{X}{N} - 0.4| \leq \frac{L}{2}) \\ &= 1 - P(|\frac{X}{N} - 0.4| \geq \frac{L}{2}) \\ &\geq 1 - \frac{\sigma_{X_1}^2}{N(\frac{L}{2})^2} \\ &= 1 - \frac{0.24}{1600(\frac{L}{2})^2} \end{aligned}$$

Il faut donc que L satisfasse $\frac{0.24}{1600(\frac{L}{2})^2} \leq 0.05$, c'est-à-dire $L \geq 0.1096$. $\square$

Exercice 6.3 *On effectue des tirages avec remise dans une urne contenant deux boules blanches et quatre boules bleues. Quelqu'un affirme que, bien évidemment, vous aurez tiré une boule blanche une fois sur trois. Combien de tirages faut-il effectuer pour que cette affirmation soit correcte : que l'on s'écarte de la valeur prédite (1/3) d'au plus 0.2 avec probabilité au plus 1/100 ?*

Exercice 6.4 *Un fournisseur d'accès à Internet met en place un point local d'accès, qui dessert 5000 abonnés. A instant donné, chaque abonné a une probabilité égale à 20% d'être connecté. Les comportements des abonnés sont supposés indépendants les uns des autres.*

- *On note X la variable aléatoire égale au nombre d'abonnés connectés à un instant t. Quelle est la loi de X ? Quelle est son espérance, son écart-type ?*
- *Le fournisseur d'accès souhaite savoir combien de connexions simultanées le point d'accès doit pouvoir gérer pour que sa probabilité d'être saturé soit inférieure à 2,5%.*

Exercice 6.5 *On lance une pièce équilibrée. En utilisant l'inégalité de Chebychev (6.2), estimer le nombre de lancers nécessaires pour que la fréquence de Pile observé au jeu de Pile ou Face soit comprise entre 0.4 et 0.6 avec une probabilité au moins égale à 0.9.*

6.1 Simulation et méthodes Monte Carlo

Les méthodes de Monte Carlo sont une alternative souvent efficace aux méthodes numériques (pour calculer ou approximer des fonctions). Elles sont aussi conceptuellement faciles à appréhender et à implémenter. Prenons un exemple pour illustrer notre propos.

Considérez la figure 3.4 et imaginons que l'on souhaite calculer l'aire de l'ensemble A. La difficulté vient de ce que le contour de l'ensemble peut difficilement être décrit par une fonction analytique (un cercle, une ellipse, etc.). Du point de vue informatique, on peut imaginer que la courbe est donnée sous la forme d'une ligne brisée – cela pourrait nous permettre de réduire le problème par un calcul, fastidieux, d'aires de petits triangles. On peut procéder différemment.

Il est facile de calculer l'aire du rectangle englobant Ω. On peut ensuite raisonner de manière probabiliste : si on choisit “au hasard” un point (x, y) dans le rectangle, la probabilité qu'il appartienne à l'ensemble A est donnée par le ratio des aires (aire A / aire Ω). On peut donc procéder comme suit. Convenons que le coin inférieur gauche du rectangle est à l'origine, et que ses dimensions sont $a \times b$ (base $\times$ hauteur).

– On effectue un tirage aléatoire uniforme de points dans $[0, a] \times [0, b]$;
– On teste l'appartenance des points à l'ensemble A ;
– On calcule le ratio (nombre de points dans A) / nombre total de points.

Les tirages aléatoires des coordonnées (x, y) se font de manière indépendantes et selon des lois uniforme sur chaque intervalle. Plus le nombre de tirage est grand, plus l'évaluation de l'aire sera précise. La précision de cette approximation dépend aussi de la précision du contour de l'ensemble A.

Formalisons l'approche que nous venons d'introduire. La méthode de Monte Carlo revient en réalité à évaluer l'espérance mathématique d'une variable aléatoire. Dans le cas qui nous préoccupe, la variable aléatoire X (associée à l'ensemble A) est défini sur l'ensemble des points dans $[0, a] \times [0, b]$ et vaut 1 ou 0 selon qu'un point (x, y) est ou non dans l'ensemble A. On peut alors considérer des variables aléatoires $X_1, \ldots, X_N$ (les tirages aléatoires des n points dans le rectangle) et de considérer l'approximation :

$$E(X) = \frac{1}{N}(X_1 + \cdots + X_N).$$

Puisque l'on approxime une quantité à l'aide de variables aléatoires, la loi des grands nombres nous permet de maîtriser l'erreur commise. Ainsi, on voit que l'erreur dépend fortement de la variance de la loi de probabilités des variables utilisées. On peut donc avoir intérêt à utiliser une loi de variance la plus petite possible. On verra plus loin comment estimer l'erreur commise à l'aide d'une loi gaussienne (section 8.3).

Chapitre 7
Probabilités et simulation, génération aléatoire

Vous utiliserez pour ces exercices l'environnement ou le langage de programmation et la librairie de votre choix. Des exemples typiques sont :

– Java et la librairie/classe java.util.Random ou encore java.math (méthode random()) ;
– le langage C++ comprend aussi les librairies nécessaires à la simulation de loi de probabilités ;
– un tableur (Excel/OpenOffice) permet aussi de faire certaines choses [BMPS98] [BPSD07].

Exercice 7.1 Loi uniforme sur $[0, 1]$ **– Tirage aléatoire d'entiers** *Ecrivez un court programme effectuant un tirage aléatoire uniforme dans l'intervalle réel* $[0, 1]$.

Effectuez un tirage à plusieurs reprises. La suite calculée est-elle invariablement la même ? Y a t-il moyen d'obtenir la même suite lors de deux exécutions de ce même programme (pensez à positionné le germe – seed, en anglais – de la suite).

Exercice 7.2 Pile ou face *Utilisez le tirage uniforme dans* $[0, 1]$ *pour simuler un tirage de Bernouilli avec probabilité* $0 < p < 1$.

Exercice 7.3 Rademacher *Utilisez le tirage uniforme dans* $[0, 1]$ *pour simuler une loi de Rademacher, qui vaut 1 avec probabilité* p *et -1 avec probabilité* $1 - p$.

Exercice 7.4 Tirage aléatoire d'entiers *Ecrivez un court programme qui tire au hasard des entiers dans un intervalle* $[a, b]$. *Assurez-vous du caractère uniforme de la loi sous-jacente.*

Relancez votre programme à plusieurs reprises. La suite calculée est-elle invariablement la même ? Y a t-il moyen d'obtenir la même suite lors de deux exécutions de ce même programme (pensez à positionné le germe – seed, en anglais – de la suite).

Exercice 7.5 Bandit manchot *Ecrivez un programme qui simule le jeu du bandit manchot décrit au début de la section 5. Observez le gain du jouer converger vers l'espérance mathématique prédit par la théorie.*

Exercice 7.6 Nombres et chapeaux *Ecrivez un programme qui simule le jeu du casino où le croupier dissimule deux nombres sous des chapeaux (exercice 3.17).*

7.1 Génération aléatoire

Nous avons vu à la section précédente comment simuler une loi de probabilité. Les besoins de la simulation peuvent parfois être un peu différent et exiger de générer aléatoirement une structure de données. On peut par exemple avoir à générer aléatoirement un ensemble de k éléments parmi N, ou une permutation de N éléments.

Imaginez que vous ayez à mettre au point une simulation pour tester et valider un système de gestion du trafic urbain, tenter de valider un scénario d'équilibrage des charges, etc. Vous allez vouloir générer un afflux du trafic en certains points du réseau au départ de la simulation. En d'autres mots, vous allez par exemple vouloir sélectionner, au hasard, k points du réseau parmi N. Ou encore, vous souhaiterez générer des entiers n_1, n_2, …, n_k dont la somme fait N. Ce problème n'est pas anodin. En mettant une procédure ad hoc pour construire une telle structure, on est pas assuré de pouvoir générer toutes les structures possibles, et même si c'est le cas, on n'est pas assuré de le faire de manière équiprobable.

Exercice 7.7 Générer des entiers *En machine, les entiers sont décrits par des vecteurs de bits de taille N (= 32 ou 64). On peut donc penser générer aléatoirement des entiers en jouant à pile ou face N fois et en prenant le résultat du jeu comme résultat. Une suite générée ainsi forme t-elle une suite aléatoire d'entiers ?*

Ecrivez un court programme qui tire au hasard un nombre entiers de l'ensemble $\{0, 2^N - 1\}$ de cette façon ($N = 32$ ou 64 selon l'architecture de la machine utilisée).

Exercice 7.8 Générer des ensembles *Etant donné une liste ordonnée de N éléments distincts e_1, …, e_N on peut coder un sous-ensemble F de $E = \{e_1, \ldots, e_N\}$ par un vecteur $(b_1, \ldots, b_N)$ de N bits 0 ou 1 où $b_i = 1$ si et seulement si l'élément $e_i \in F$. Proposez une procédure pour générer aléatoirement de manière équiprobable un sous-ensemble. Implémentez-là.*

Exercice 7.9 Graphe aléatoire *Un* graphe simple *G est donné par un ensemble de* sommets *V et d'*arêtes *E où $E \subset V \times V$. Une arête $e \in E$ est une paire de sommets $\{u, v\}$. Erdös et Renyi ont proposé un modèle de graphe* *aléatoire qui consiste à effectuer pour chaque paire* $\{u, v\}$ *un tirage de Bernouilli avec probabilité* p.

Soit Ω *l'ensemble de tous les graphes simples sur un ensemble* V *et considérons la variable aléatoire* $X : \Omega \to \mathbb{N}$ *donnant le nombre d'arêtes d'un graphe* G.

- *Quelle est la distribution de la variable* X.
- *Quel est le nombre moyen d'arêtes d'un graphe aléatoire selon le modèle Erdös-Renyi ?*
- *Donnez une implémentation de la procédure de tirage aléatoire d'un graphe.*

Exercice 7.10 Générer un arbre étiqueté *Cet exercice se penche sur la génération aléatoire uniforme d'un arbre étiquetés. Un tel arbre* $T = (V, E =$ *est un graphe connexe non-orienté, sans cycle, sur l'ensemble* $V = \{1, \ldots, N\}$. *Ces structures sont très présentes en informatiques et dans des problèmes divers de calculs de flots, de couverture de graphes, etc.*

Travailler directement sur la structure de graphe pour tenter de générer une telle structure n'est pas simple. En effet, comment déterminer les arêtes $e \in E$ *à ajouter ? comment s'assurer que l'ajout d'une arête ne provoque l'apparition de cycle , comment s'assurer de générer un graphe connexe ?*

La solution passe par un codage de la structure en un objet plus simple, au moins pour ce qui concerne la génération aléatoire. On peut coder un arbre par une séquence de Prüfer ; c'est une séquence $S = (s_0, \ldots, s_{N-2})$ *de* $N - 2$ *entiers choisis parmi* $\{1, \ldots, N\}$. *On l'obtient d'un arbre étiqueté en appliquant la méthode suivante :*

- *(à répéter tant qu'il reste plus de deux sommets dans l'arbre courant* T*)*
- *identifier la feuille* v *de l'arbre courant ayant le numéro minimum ;*
- *ajouter à la suite* S *le seul sommet* s adjacent *à* v *dans l'arbre* T *courant ;*
- *enlever de l'arbre* T *courant le sommet* v *et l'arête* incidente *à* v.

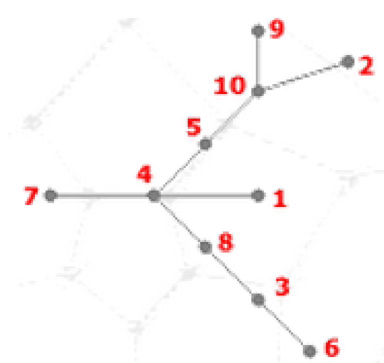


Fig. 7.1 Un arbre étiqueté.

Simulez l'algorithme de codage d'un arbre à partir de l'arbre de la figure 7.1.

A l'inverse, on peut construire l'arbre correspondant à une séquence de Prüfer en suivant le procédé suivant :

- S *est donnée, on initialise* I *à* $\{1, \ldots, N\}$.
- *A répéter tant qu'il reste des éléments dans* S *et plus de deux éléments dans* I
- *identifier le plus petit élément* i *de* I *n'apparaissant pas dans la suite* S *;*
- *relier par une arête de* T *le sommet* i *avec le sommet s correspondant au premier élément de la suite* S *;*
- *enlever* i *de* I *et* s *de* S.
- *Les deux éléments qui restent dans* I *à la fin de l'algorithme constituent les extrémités de la dernière arête à ajouter à* T.

Calculez l'arbre associée à la séquence de Prüfer $S = 4, 10, 3, 8, 4, 4, 5, 10$.

On transforme donc ainsi le problème de la génération aléatoire d'un arbre étiqueté en celui de la génération aléatoire d'une séquence de $N - 2$ *entiers indépendants. Ecrivez et implémentez maintenant un algorithme qui génère aléatoirement et de manière uniforme un arbre étiquetés sur* $\{1, \ldots, N\}$.

Exercice 7.11 Générer aléatoirement une permutation *Une permutation est une liste de* N *entiers distincts choisis parmi* $\{1, \ldots, N\}$. *En d'autres mots, c'est une liste non ordonnée de ces entiers, "dans n'importe quel ordre". Générer aléatoirement et uniformément une telle liste peut être utile pour tester un algorithme. Pensez aux algorithmes de tri : en faisant tourner l'algorithme sur un bon nombre de permutations choisies au hasard, on peut évaluer empiriquement ses performances. Pensez à un programme qui doit traiter des tâches dans un ordre qui n'est pas déterminé au départ et qui peut varier aléatoirement : il faut bien pouvoir simuler ces ordres d'arrivée des tâches pour le tester.*

Comment donc générer une permutation en s'assurant que chacune ait la même probabilité d'apparaître ? On pourrait penser à une méthode naïve : on choisit au hasard N *entiers dans* $\{1, \ldots, N\}$, *en s'assurant que chacun n'apparait qu'une seule fois (si on tire un entier déjà dans la séquence, on le jette et on recommence). Mais est-on bien certain dans ces conditions d'avoir un tirage uniforme ?*

Essayez. Implémentez cet algorithme naïf et observez la distribution empirique des permutations (pour de petites valeurs de N*).*

Le tirage est bien uniforme, essentiellement parce que les tirages des entiers sont indépendants. L'inconvénient de ce procédé est sa complexité. On peut en effet être amener à rejeter souvent des tirages d'entiers. En effet, la probabilité de tirer à la i*ème position un nombre déjà présent dans la liste est* i/N, *et croit donc à mesure que l'on construit la permutation. Sauriez-vous montrez que la complexité de cet algorithme est de l'ordre de* $N \log N$ *?*

Exercice 7.12 Permutations et tables d'inversion *Nous allons maintenant voir un autre algorithme, plus efficace, pour générer une permutation.*

Une permutation peut être codée par une table d'inversion *(on parle aussi de son code de Lehmer), qui décrit la position relative des entiers dans la liste. La table associée à la permutation identité* $1, 2, \ldots, N$ *est* $(0, 0, \ldots, 0)$ *puisque tout est bien en ordre. Etant donnée une permutation* $\sigma = \sigma_1\sigma_2 \ldots \sigma_N$*, le nombre d'inversions* b_i *de* σ_i *est le nombre d'entiers* σ_j *qui se trouve à sa gauche (*$j < i$*) et qui sont plus grands (*$\sigma_j > \sigma_i$*).*

Donnez la table d'inversion de la permutation $N, \ldots, 2, 1$*. Donnez la table d'inversion de* $1, 3, 5, \ldots, N-1, 2, 4, \ldots, N$ *(pour* N *pair par exemple).*

Une suite $b = b_1, b_2, \ldots, b_{N-1}$ *d'entiers* $0 \leq b_i \leq N - i$ *détermine une unique permutation. On commence par considérer la séquence réduite à l'entier* N *et on ajoute à la suite les entiers* $N-1, N-2, \ldots, 2, 1$ *en plaçant* i *à la position* b_i *(la position à l'extrême gauche est la position 0).*

Donnez la permutation ($N = 5$*) associée à la table d'inversion* $3, 1, 1, 1$*. Que peut-on dire des permutations dont la table d'inversion commence par* $N - 1$ *?*

La génération aléatoire uniforme sur $\{1, \ldots, N\}$ *d'une permutation peut donc se réduire à une série de* $N-1$ *tirages aléatoires indépendants d'entiers (chacun sur un intervalle différents).*

Exercice 7.13 Partitions d'ensemble *Une partition de l'ensemble* $V = \{1, 2, \ldots, N\}$ *est un découpage en sous-ensemble* $V_1, V_2, \ldots, V_k \subset V$ *disjoints – c'est-à-dire que* $V_i \cap V_j = \emptyset$ *pout toute paire d'indices* $i \neq j$ *– et tels que* $V_1 \cup V_2 \cup \cdots \cup V_k = V$*. La partition est souvent notée* $\pi = \{V_1,^l dots, V_k\}$ *et* k *est son nombre de* blocs.

Le nombre de partitions en k *d'un ensemble à* N *éléments est égal au* nombre de Stirling $S(N, k)$*. La somme* $\sum_{k=1}^{N} S(N, k) = B_N$ *est égal au* nombre de Bell*. On peut calculer les nombre de Stirling à l'aide la récurrence :*

$$S(N, k) = S(N-1, k-1) + kS(N-1, k) \tag{7.1}$$

soumis aux conditions initiales $S(N, 1) = S(N, N) = 1$*. La récurrence s'interprète facilement : soit l'élément* N *est seul dans un bloc. Sinon, suffit de construire une partition sur* $\{1, \ldots, N-1\}$ *et de choisir au hasard un bloc (parmi les* k *blocs) dans lequel on l'insère. Les nombres* $S(N, k)/B_N$ *nous donnes donc la distribution de probabilité de la variable rendant compte du nombre de blocs d'une partition.*

On peut générer une partition en reprenant le schéma suggéré par la récurrence (7.1). Cela dit, il faut avoir préalablement choisi le nombre k *de blocs qu'aura cette partition. Cela exige donc de choisir au hasard le nombre* k*,* mais selon la distribution des valeurs $S(N, k)$.

Précisez cet algorithme et implémentez-le.

7.2 Méthodes à rejet

Supposons que l'on sache simuler une loi uniforme sur un ensemble A : on sait choisir aléatoirement un élément de A de manière équiprobable à l'aide d'un algorithme $\mathcal{A}$. Alors on peut utiliser cet algorithme pour simuler une loi uniforme sur un sous-ensemble $B \subset A$. Il suffit de générer des éléments de A jusqu'à tomber sur un élément de B.

Cette méthode simule bien une variable aléatoire X de loi uniforme sur l'ensemble B. Il nous faut montrer que la probabilité de tirer aléatoirement un élément $b \in B$ est égale à $1/|B|$. Or, on a probabilité $1/|A|$ de tomber sur $b \in B$ dès le premier tirage, puis probabilité $(1 - |B|/|A|)1/|A|$ d'y arriver au second tirage, etc. on a probabilité de réussir dès le premier tirage, puis probabilité $(|A| - |B|)/|A| \cdot |B|/|A|$ de réussir au second tirage, etc. Ainsi, $P(X = b) = \sum_{k \geq 1}(1 - |B|/|A|)^{k-1}1/|A| = 1/|B|$ montrant ainsi que le tirage aléatoire est bien uniforme sur B.

Remarque 7.2.1 *Nous avons déjà implicitement soulignée l'usage de cette méthode lors du tirage aléatoire de permutations (exercice 7.11, page 58). La méthode "naïve" que nous présentions tire au hasard un entier d'un sous-ensemble $A \subset \{1, \ldots, N\}$ en utilisant une méthode à rejet.*

Cette méthode se généralise au cas des lois continues. Simuler une variable aléatoire X de densité f_X revient à tirer un point au hasard sous le graphe de f_X et retourner l'abcisse de ce point. On peut donc, au passage, avoir à utiliser la méthode à rejet pour tirer au hasard (uniformément) un point sous le graphe de f_X.

Exercice 7.14 *Ecrivez un court programme qui permet de générer aléatoirement un points dans un disque de rayon R donné.*

Chapitre 8
Variable aléatoire réelle, densité de probabilité

Nous avons jusqu'à maintenant vu des exemples et résultats concernant les variables aléatoires discrètes, c'est-à-dire qui prennent des valeurs entières dans $\mathbb{N}$ ou $\mathbb{Z}$. Il est toutefois possible, et parfois nécessaire, d'étendre la notion de probabilité au cas des variables prenant une valeur dans $\mathbb{R}$. Imaginons qu'il s'agissent d'étudier la variation de température d'un processus, ou la variation de prix d'une action sur un marché, etc. Le domaine des valeurs ne peut alors être restreint à un ensemble fini ou aux entiers naturels.

On se trouve alors face à une difficulté, puisque la probabilité d'observer une valeur particulière est, en réalité, nulle. En effet, quelle est la probabilité d'observer la valeur d'une action à un prix donné ? d'observer une température précise ? Il est plus naturel de chercher à connaître la probabilité que la valeur observée se trouve dans un intervalle de valeurs. Cela nous amène à considérer la notion de *densité de probabilité*.

8.1 Densité de probabilité et fonction de répartition

Soit $(\Omega, \mathcal{U}, P)$ un espace de probabilité continu et X une variable aléatoire sur $\mathbb{R}$. La probabilité des évènements sur Ω exige alors nombre de précautions que nous passerons sous silence. Soulignons toutefois que la mesure de probabilité se calcule à l'aide d'une *intégrale*, en exigeant d'abord l'existence d'une *densité de probabilité*, c'est-à-dire une fonction f_X définie sur $\mathbb{R}$ et telle que :

$$\int_{-\infty}^{+\infty} f_X(x)dx = 1.$$

On peut voir cette équation comme le pendant continu de la condition exigeant que la somme des probabilités des épreuves $\omega \in \Omega$ soit égales à 1.

La densité f_X nous permet alors de calculer les probabilités associées à la variable X :

$$P(X \leq b) = \int_{-\infty}^{b} f_X(x) \tag{8.1}$$

Cette définition permet de calculer, par exemple, la probabilité $P(a \leq X \leq b) = \int_a^b f_X(x)$ ou encore $P(a \leq X \leq b) = P(X \leq b) - P(X \leq a)$. La notion d'espérance mathématique doit, elle aussi, être revue :

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x f_X(x) dx.$$

La fonction de répartition associée à une densité de probabilité f_X est la fonction F_X donnée par $F_X(b) = P(X \leq b) = \int_{-\infty}^{b} f_X(x)$.

La fonction de densité et la fonction de répartition sont évidemment liées par la relation $\frac{d}{dx} F_X = f_X$, c'est-à-dire qu'on obtient f_X en dérivant F_X, puisqu'on intègre f_X pour obtenir F_X.

8.2 Lois continues classiques

8.2.1 Loi de distribution uniforme

La *loi de distribution uniforme* sur l'intervalle $[a, b)$ est une variable aléatoire réelle X de densité de probabilité $f_X(x) = \frac{1}{b-a} 1_{[a,b)}$ où $1_{[a,b)}$ est la fonction indicatrice associé à l'intervalle $[a, b)$. On dit alors que X suit une loi uniforme sur l'intervalle $[a, b)$.

En réalité, on connaît déjà cette loi. C'est celle qui sous-tend l'appel à la fonction random() qui retourne un nombre réel aléatoire de l'intervalle $[0, 1)$. Cette fonction fait ce qu'on attend d'elle : elle choisit au hasard un nombre réel parmi tous les nombres réels de l'intervalle [0, 1). Cet énoncé n'est pas anodin n et pourrait soulever quelques interrogations : comment peut-on faire un tirage aléatoire d'un nombre réel en machine alors qu'on ne dispose que d'un nombre fini de représentations en machine ? La question n'est en effet pas simple, et nous la passerons sous silence. Nous considérerons que cette loi uniforme est une brique de base pour toutes nos constructions informatiques (et mathématiques).

Notez au passage que la variable aléatoire $X_{a,b}$ uniforme sur $[a, b)$ se déduit de la variable $X_{0,1}$ uniforme sur $[0, 1)$ par multiplication $X_{a,b} = bX_{0,1} - a$ (c'est bien ainsi que l'on peut procéder pour implémenter la variable aléatoire uniforme sur $[a, b)$).

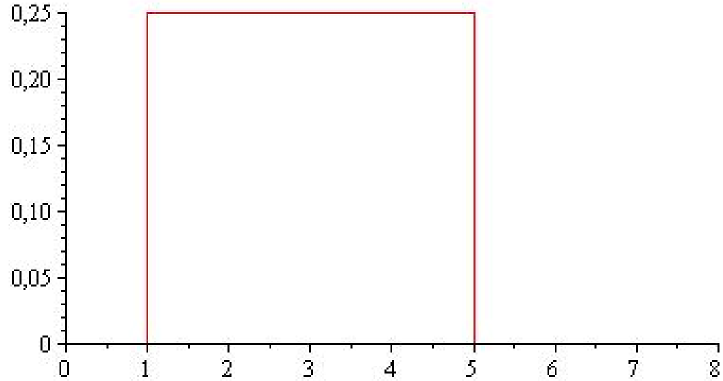


Fig. 8.1 La loi uniforme correspond à une distribution constante sur tout l'intervalle $[a,b)$ (avec ici $a = 1, b = 5$), l'aire sous la courbe étant égale à 1.

Exercice 8.1 *Vérifiez que $f_X(x) = \frac{1}{b-a}1_{[a,b)}$ est bien une densité. Calculez sa fonction de répartition (qui donne $P(X \leq x)$). Calculez l'espérance mathématique et la variance de la loi uniforme.*

Exercice 8.2 *Toutes les vingt minutes un bus se présente à un arrêt. Un usager peut arriver à cet arrêt au hasard, à tout instant à cet arrêt. On estime que le temps d'attente X d'un usager suit une loi uniforme sur l'intervalle $[0,20]$.*

- *Donnez la fonction de densité et la fonction de répartition de la variable X.*
- *Montrez que la probabilité qu'un usager attende le bus moins de cinq minutes satisfait est de 50%.*
- *Quelle est la probabilité qu'un usager attende plus de 18 minutes ?*
- *Quel est le temps d'attente moyen ?*

Solution La fonction de densité et la fonction de répartition d'une loi uniforme sur l'intervalle $[0,20]$ sont :

$$f(t) = t \begin{cases} 0 & \text{si } t < 0 \\ \frac{1}{20} & \text{si } t \in [0,20] \\ 0 & \text{si } t > 20 \end{cases}$$

$$F(t) = \int_{-\infty}^{t} f(x)dx = \begin{cases} 0 & \text{si } t < 0 \\ \frac{t}{20} & \text{si } t \in [0,20] \\ 1 & \text{si } t > 20 \end{cases}$$

L'usager attendra moins de cinq minutes est $P(X \leq 5) = F(5) = 0,25$. La probabilité qu'il attende plus de 18 minutes est $P(X > 18) = 1 - P(X \leq$

$18) = 1 - F(18) = 1 - \frac{18}{20} = 0,1$. Le temps moyen d'attente est $\mathbb{E}(X) = \int_{-\infty}^{+\infty} tf(t)dt = \int_0^{20} \frac{t}{20} = 10$. $\square$

8.2.2 Loi de distribution non uniforme et fonction de répartition inverse

Remarque 8.2.1 *Nous avons eu affaire à des cas de variables aléatoires continues non-uniformes sans le souligner explicitement. Considérons en effet le cas d'une variable aléatoire discrète X à valeurs dans un ensemble fini V, et notons $p_v = P(X = v)$. La variables X est alors complètement définie par la suite des probabilités $(p_v)_{v \in V}$. Pour simuler la variable X il nous faut choisir un état $v \in V$ en respectant les probabilités d'obtenir chacun des états $v \in V$.*

On le fait en tirant aléatoirement un nombre réel $p \in [0, 1)$ et en déterminant à partir de ce p l'un des états comme suit : on découpe l'intervalle $[0, 1)$ en $|V|$ intervalles chacun de longueur p_v. Le nombre p tombe alors dans un unique intervalle et détermine ainsi l'état $v \in V$ (la probabilité de tomber en plein sur une extrémité d'un intervalle est nulle).

La dernière remarque n'est pas étrangère à un procédé général. Soit une variable aléatoire X admettant une fonction de densité de probabilité f_X. On définit la fonction de répartition (associée à X) en posant $F(x) = \int_{-\infty}^{x} f_X(t)dt$. On a donc $F(x) = P(X \leq x)$.

On pose $F^{-1}(t) = \inf\{x|F(x) \geq t\}$. Il nous fait prendre quelques précautions avec F^{-1} puisque F n'est pas nécessairement injective ; elle est toutefois monotone. En d'autres mots, la fonction est croissante mais peut passer par des paliers. Un palier est un intervalle $[x, x+\epsilon]$ sur lequel la fonction est constante : tous les points de cet intervalle on même valeur par F. La valeur $F^{-1}(t)$ est donc le point le plus à gauche du palier.

Proposition 1. *Si U est une variable aléatoire de loi uniforme sur l'intervalle $[0, 1)$, alors la variable $F^{-1}(U)$ suit la même loi que X.*

En d'autres mots, il suffit de pouvoir simuler la loi uniforme sur $[0, 1)$ pour pouvoir simuler n'importe quelle autre variable aléatoire, à condition de pouvoir calculer l'inverse F^{-1} de la fonction de répartition. Remarquez que c'est exactement ce qui est proposé à la remarque 8.2.1, où la fonction de répartition est alors une fonction en escalier.

Pour montrer que $Y = F^{-1}(U)$ suit la même loi que X, on montre qu'elles ont même fonction de répartition. On calcule :

$$\begin{aligned} P(Y \leq t) &= P(F^{-1}(U) \leq t) \\ &= P(F \circ F^{-1}(U) \leq F(t) \qquad (8.2) \\ &= P(U \leq F(t)) = \int_0^{F(t)} 1dx = F(t) \qquad (8.3) \end{aligned}$$

Notez que le passage à l'égalité (8.2) utilise le fait que la fonction F est monotone croissante. La dernière égalité (8.3) utilise le fait que la densité de la loi uniforme U est la fonction égale à 1 sur l'intervalle [0, 1] et égale à 0 partout ailleurs. La démonstration de ce résultat indique un élément important de la méthode. Il suffit en effet que la fonction de répartition F admettent un inverse à droite, c'est-à-dire que $F \circ F^{-1}$ soit la fonction identité, et pas nécessairement à gauche (cela n'est de toute façon pas possible si la fonction admet un palier puisqu'elle n'est alors pas injective).

Exercice 8.3 *Considérons la fonction de densité* triangulaire *:*

$$f(t) = \begin{cases} \mathit{0} & \mathit{si} \quad t < 0 \\ 4t & \mathit{si} \quad 0 \leq t < 1/2 \\ 4 - 4t & \mathit{si} \quad 1/2 \leq t < 1 \\ \mathit{0} & \mathit{sinon} \end{cases}$$

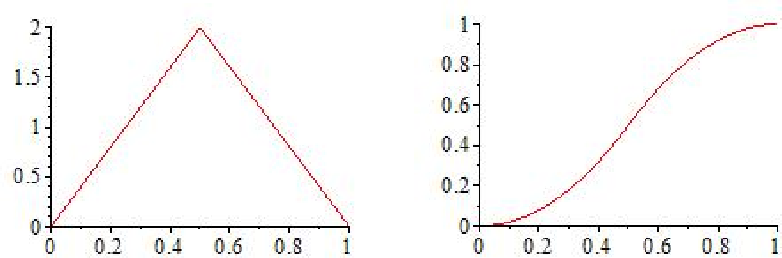


Fig. 8.2 Cette courbe décrit la loi de densité non-uniforme d'une variable aléatoire. La probabilité d'observer une valeur près de 0 ou près de 1 est plus faible que pour les valeurs situées au centre de l'intervalle $[0, 1)$ et suit une progression linéaire croissante sur $[0, 1/2)$ et décroissante sur $[1/2, 1)$. La courbe de droite décrit la fonction de répartition associée.

Calculez la fonction de répartition, et la fonction de répartition inverse. Utilisez ce résultat pour produire le code d'une fonction effectuant un tirage aléatoire selon la densité triangulaire.

Solution La fonction de répartition se calcule en intégrant la fonction de densité. On trouve :

$$F(t) = \begin{cases} 0 & \text{si} \quad t < 0 \\ 2t^2 & \text{si} \quad 0 \le t < 1/2 \\ -2t^2 + 4t - 1 & \text{si} \quad 1/2 \le t < 1 \\ 0 & \text{sinon} \end{cases}$$

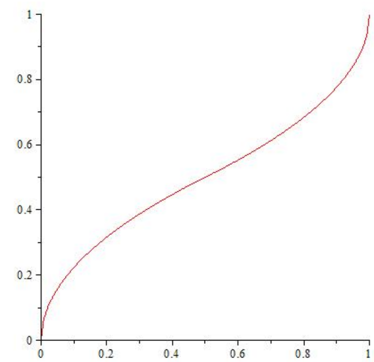


Fig. 8.3 Cette courbe décrit la fonction de répartition inverse F^{-1}. On le constate aussi facilement puisque la courbe s'obtient par une réflexion dans l'axe diagonale de la courbe de la fonction de répartition.

La fonction de répartition inverse F^{-1} est donc la fonction $F^{-1}(t)$ telle que $F \circ F^{-1}(t) = t$. Il suffit d'étudier la fonction F^{-1} sur l'intervalle $[0,1)$ et de l'étendre ailleurs en posant $F^{-1}(t) = 0$ si $t \notin [0,1)$. On doit donc avoir, sur $[0,1/2)$, l'identité $2 \cdot F^{-1}(t)^2 = t$ et par suite, $F^{-1}(t) = \frac{\sqrt{t}}{\sqrt{2}}$. Sur $[1/2,1)$, on doit avoir $-2 \cdot F^{-1}(t)^2 + 4 \cdot F^{-1}(t) - 1 = t$. On trouve finalement :

$$F^{-1}(t) = \begin{cases} 0 & \text{si} \quad t < 0 \\ \sqrt{\frac{t}{2}} & \text{si} \quad 0 \le t < 1/2 \\ 1 - \sqrt{\frac{1-t}{2}} & \text{si} \quad 1/2 \le t < 1 \\ 0 & \text{sinon} \end{cases}$$

Observez que cette fonction inverse est un inverse à droite et à gauche. Cela vient du fait que la fonction de répartition est bijective, puisque strictement croissante – et par conséquent, la fonction inverse F^{-1} est, elle aussi, bijective.

La procédure à construire consiste à générer aléatoirement un nombre réel $x \in [0,1)$ (à l'aide de la fonction `random()`, par exemple), puis de calculer $F^{-1}(x)$. $\square$

Exercice 8.4 *On considère maintenant la fonction de densité*

$$f(t) = \begin{cases} \textit{0} \quad \textit{si} \quad t < 0 \\ \sqrt{t} \quad \textit{si}\ 0 \leq t < 1 \\ \textit{0} \quad \textit{si} \quad t \geq 1 \end{cases}$$

Calculez la fonction de répartition associée, la fonction de répartition inverse pour produire une procédure permettant de simuler une variable aléatoire de densité $f(t)$.

Solution La fonction de répartition associé est :

$$F(t) = \begin{cases} 0 \quad \text{si} \quad t < 0 \\ t^{3/2} \quad \text{si}\ 0 \leq t < 1 \\ 0 \quad \text{si} \quad t \geq 1 \end{cases}$$

La fonction de répartition inverse est $F^{-1}(x) = x^{2/3}$. La procédure qui permet de générer des nombres réels suivant la densité f consiste donc à générer aléatoirement un nombre $x \in [0, 1)$ et à calculer $F^{-1}(x) = x^{2/3}$. (Note : on calcule une puissance fractionnaire en faisant $x^{2/3} = e^{2/3 \log x}$.) □

8.2.3 Méthodes à rejet

Nous avons déjà vu une méthode à rejet pour effectuer un tirage aléatoire uniforme (section 7.2). Mais qu'en est-il si l'on souhaite simuler une fonction de densité non-uniforme ?

Plus précisément, soit une fonction de densité $f(t)$ que l'on souhaite simuler. Supposons que l'on dispose d'une autre fonction de densité $g(t)$ que l'on sait simuler et qui est telle que $f(t) \geq M \cdot g(t)$. Pour générer aléatoirement un nombre Y selon la densité f, on peut procéder à l'aide d'un méthode à rejet :

– On génère un nombre X selon la densité g
– On génère un nombre aléatoire U selon la densité uniforme sur $[0, 1)$
– On retourne $Y = X$ si $U \leq \frac{f(X)}{M \cdot g(X)}$, sinon on recommence

Exercice 8.5 *Comment pourrait-on simuler une variable aléatoire dont la loi de densité est $f(t) = 6t(1-t)$ pour $t \in [0, 1)$ et $f(t) = 0$ sinon.*

La fonction de répartition associé à cette densité est $F(t) = \int_{-\infty}^{t} 6x(1-x)dx = t^2(3-2t)$

Solution On a $f(t) = 6t(1-t) \leq 1.5$. On est donc dans la situation décrite précédemment, où $g(t)$ est la distribution uniforme sur $[0, 1)$ et $M = 1.5$. □

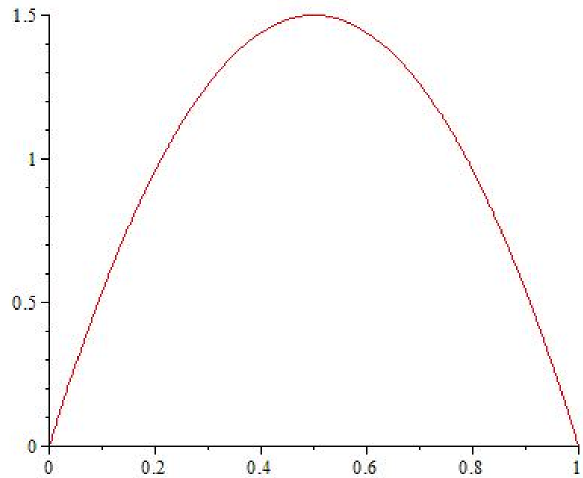


Fig. 8.4 Cette courbe décrit la loi de densité non-uniforme d'une variable aléatoire. La probabilité d'observer une valeur près de 0 ou près de 1 est plus faible que pour les valeurs situées au centre de l'intervalle $[0, 1)$.

8.2.4 Loi exponentielle

Il existe des phénomènes ou il n'y a pas de vieillissement ou d'usure. Il s'agit en general de phénomènes accidentels. Pour ces phénomènes, la probabilité, pour un objet d'être encore en vie ou de ne pas tomber en panne avant un délai donné sachant que l'objet est en bon état à un instant t, ne depend pas de t. Par exemple, pour un programme, la probabilité qu'il lève une exception ne dépend pas du moment considéré (sauf à modifier le code et à recompiler . . .).

Par définition, on dit qu'une durée de vie est sans "usure" si la probabilité d'un "accident" à l'instant t ne dépend pas de t. La durée de vie d'un élément est une variable aléatoire T, à valeurs dans $[0, +\infty)$ pour laquelle, t étant un réel positif, l'évènement $T \geq t$ signifie que l'élément est toujours en activité à l'instant t.

Par définition, en remarquant que $T \geq 0$ est l'évènement certain et en supposant que $P(T \geq t) \neq 0$, la variable aléatoire T représente une durée de vie sans usure si et seulement si :

$$\forall s > 0, P(T \geq t + s|T \geq t) = P(T \geq s|T \geq 0) = P(T \geq s) \tag{8.4}$$

Or,

$$P(T \geq t+s|T \geq t) = \frac{P(T \geq t+s)}{P(T \geq t)} \tag{8.5}$$

Finalement, on obtient en combinant (8.4) et (8.5) :

$$\frac{P(T \geq t+s)}{P(T \geq t)} = P(T \geq s).$$

Si on voit la probabilité $P(T \geq t)$ comme une fonction de t, en posant $G(t) = P(T \geq t)$, on obtient pour $t \geq 0$ et $s \geq 0$: $G(t+s) = G(s) \times G(t)$ avec $G(0) = 1$. La fonction G étant non nulle, et en supposant G dérivable, on montre que les seules fonctions satisfaisant cette condition sont de la forme : $G(t) = e^{kt}$, où $k \in \mathbb{R}$.

Or, $G(t)$ est la probabilité d'un évènement donc, $\forall t \geq 0$, on a : $0 \leq G(t) \leq 1$. Par conséquent $k \leq 0$. On pose alors $k = -\lambda$ avec $\lambda > 0$ et on écrit :

$$P(T \geq t) = G(t) = e^{-\lambda t}$$

De plus, $P(T < t) = 1 - P(T \geq t) = 1 - G(t)$, donc $P(T < t) = 1 - e^{-\lambda t}$.

Ceci définit la fonction de répartition de la variable aléatoire T, que l'on notera F. Ainsi, en dérivant F, on obtient une fonction f, qui est la densité de probabilité de la variable aléatoire T. C'est-à-dire que l'on a $f(t) = F'(t) = 0 - (-\lambda e^{-\lambda t}) = \lambda e^{-\lambda t}$.

Cette démonstration nous amène donc à définir la loi exponentielle

Définition 2. Soit λ un réel strictement positif. La fonction f définie sur $\mathbb{R}$ par :

$$f(t) = \begin{cases} \lambda e^{-\lambda t} & \text{si } t \geq 0 \\ 0 & \text{si } t < 0 \end{cases} \tag{8.6}$$

est la densité d'une loi de probabilité appelée *loi exponentielle de paramètre* λ.

Mentionnons quelques propriétés de cette loi :

- La fonction f est continue et positive sur $[0, +\infty)$
- On a $\lim_{t\to\infty} \int_0^t f(x)dx = 1$
- Si T est une variable aléatoire représentant une durée de vie et suivant une loi exponentielle, alors sa fonction de repartition est donnée par :

  $$F(t) = P(T \leq t) = P([0,t]) = \int_0^t f(x)dx = 1 - e^{-\lambda t},$$

  pour $t \geq 0$.
- Le calcul de l'espérance donne $E(T) = \frac{1}{\lambda}$, ce qui correspond a la durée moyenne de vie.
- Le calcul de la variance est égale à $\frac{1}{\lambda^2}$. L'écart-type est donc égale à la moyenne.

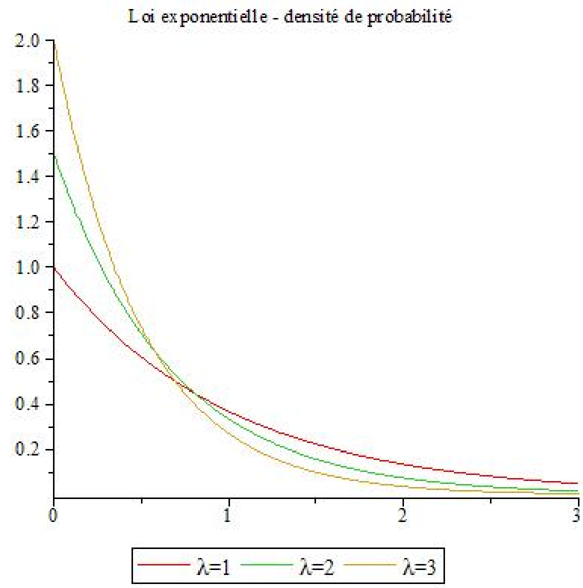


Fig. 8.5 La densité de probabilité de la loi exponentielle décroit “de manière exponentielle” lorsque $t \to +\infty$, à un rythme plus ou moins fort selon la valeur de λ.

Remarque 8.2.2 *Les variables aléatoires décrivant une durée de vie sans usure suivent sont modélisées par une loi exponentielle. L'étude qui précède nous montre que la loi d'un phénomène de nature totalement aléatoire peut être modélisée par une fonction exponentielle.*

Exercice 8.6 *On modélise le temps de réponse (en secondes) à un terminal relié à un ordinateur par une variable aléatoire X qui suit une loi exponentielle de paramètre $0,3$.*
- *Donnez la fonction de densité de X.*
- *Calculez la probabilité que le temps de réponse se situe entre 4 et 8 secondes.*
- *Quelle est la probabilité que le temps de réponse soit supérieur à la demi-seconde ?*
- *Peut-on estimer le temps minimum de réponse pour la moitié des appels ?*

Exercice 8.7 *Un standardiste prétend qu'elle reçoit en moyenne 2 appels par minutes. Le patron étonné, voudrait établir que le nombre d'appels par minute est inférieur à 2. Pour ce faire, il procède de la façon (peu recommandable) suivante. A un moment aléatoire de la journée, il entre chez le standardiste et attend l'arrivée du prochain appel. Il note son temps d'attente : $X = 3$ minutes. Calculer la probabilité de cet évènement, $P(X \geq 3)$, en* *supposant que le nombre moyen d'appels est réellement 2. A t-on là un argument pour declarer que le nombre moyen d'appels est inférieur à 2 ?*

Solution La variable X est de loi exponentielle de paramètre $\lambda = 2$ (puisque la durée entre deux appels est, selon ce que prétend le standardiste, en moyenne de $1/\lambda$ minute, soit 30 secondes). Ainsi, $P(X \geq 3) = 1 - P(X < 3) = 1 - F(3) = e^{-3\lambda} = e^{-6} = 0,0024$. Il est donc hautement improbable d'avoir à attendre 3 minutes entre deux appels si le temps moyen d'attente est réellement de 30 secondes. Le calcul de l'évènement $P(X \geq 2) \sim 0,02$ va aussi dans le même sens. □

Exercice 8.8 *Appliquer le principe de la fonction de répartition inverse pour donner un procédé de génération de nombre aléatoire réel selon la loi exponentielle.*

Solution La fonction de densité $f : [0, +\infty) - > to[0, 1)$ de la loi exponentielle est $f(t) = \lambda e^{-\lambda t}$ et sa fonction de répartition $F : [0, 1) \to [0, +\infty)$ est $F(t) = 1 - e^{-\lambda t}$ pour $t \in [0, +\infty)$ (et 0 ailleurs). La fonction inverse est donc $F^{-1}(t) = -\frac{1}{\lambda} \log 1 - t$. On petu donc simuler la loi exponentielle en procédant comme suit :

- On génère un nombre aléatoire U selon la densité uniforme sur $[0, 1)$
- On retourne le nombre $X = -\frac{1}{\lambda} \log U$ (note : les variables U et $(1 - U)$ sont toutes deux uniformes sur $[0, 1)$

□

Exercice 8.9 *On s'intéresse à la durée de vie, exprimée en semaines, d'un composant électronique. On modélise cette situation par une loi de probabilité* p *de durée de vie* sans vieillissement *définie sur l'intervalle* $[0, +\infty)$. *La probabilité que le composant ne soit plus en état de marche au bout de* t *semaines est :* $P(T < t) = 1 - e^{-\lambda t}$ *où* T *est la variable aléatoire indiquant la durée de vie d'un composant et* λ *est une constante qu'il reste à déterminer.*

Une étude statistique, montrant qu'environ 50% d'un lot important de ces composants sont encore en état de marche au bout de 200 semaines, permet de poser $P(T \in [0, 200)) = 0, 5$.

- *Montrer que* $\lambda = \ln 2/200$.
- *Quelle est la probabilité qu'un de ces composants pris au hasard ait une durée de vie supérieure à 300 semaines ?*

Solution On fait ici l'hypothèse que la durée de vie d'un composant T (en nombre de semaines) suit une loi exponentielle. La probabilité d'une durée de vie d'un temps au plus t est donc de la forme $P(T \leq t) = 1 - e^{-\lambda t}$. Il nous faut déterminer λ. Or, on sait qu'après 200 semaines 50% des composants sont toujours fonctionnels, ce que l'on peut modéliser par $P(T \geq 200) = 0, 5$. Par conséquent, $P(T \geq 200) = 1 - P(T < 200) = e^{-200\lambda} = 1/2$. On trouve donc bien $\lambda = \ln 2/200$.

Ainsi, on peut évaluer la probabilité d'une durée de vie supérieur à 300 jours, qui est donnée par $P(T \geq 300) = e^{-300 \cdot \ln 2/200} \sim 0, 3536$. □

Exercice 8.10 *Un fabricant estime qu'un composant (d'un ordinateur) a une durée de vie moyenne de 1080 jours. On suppose que la variable T représentant la durée de vie de la pièce suit une loi exponentielle.*
- *Donnez les fonction de densité et de répartition de la variable T.*
- *Donnez la probabilité que le composant soit encore en fonctionnement au bout de deux ans.*
- *Donnez la probabilité que le composant fonctionne 5 ans sachant qu'il est en fonction depuis déjà 2 ans.*
- *Au bout de combien de temps peut-on s'attendre à ce que 10% des composants soient en panne ?*

Exercice 8.11 *Quelle est la probabilité qu'une variable de loi exponentielle prenne une valeur située à plus d'un écart-type à droite de sa moyenne ?*

Solution La moyenne d'une variable X suivant une loi exponentielle de paramètre λ est $1/\lambda$. Sa variance est égale à $1/\lambda^2$, son écart-type est donc égal à $1/\lambda$. On cherche à calculer $P(X \geq 2 \cdot \frac{1}{\lambda})$, qui est égal à $e^{-2} = 0,1353353$. $\square$

8.2.4.1 Simuler une loi de Poisson

Nous avons vu précédemment comment simuler une variable aléatoire dont la densité suit une loi exponentielle. ON peut mettre à profit ce procédé pour simuler une loi de Poisson.

Le résultat que l'on exploite est le suivant. Soit X_i $(i = 1, 2, \ldots)$ des variables aléatoires dont les densités sont des lois exponentielles de paramètres λ. Soit N une variable aléatoire qui suit une loi de Poisson de paramètre λ. Alors on a :

$$P(N = k) = P(X_1 + \cdots + X_k \geq 1 < X_1 + \cdots + X_k + X_{k+1}) \tag{8.7}$$

On peut donc simuler une loi de Poisson comme suit :
- On initialise $S = 0$, $k = 0$
- Tant que S < 1
  - On génère un nombre aléatoire X selon une loi exponentielle de paramètre λ
  - On l'ajoute à S
  - On incrémente k
- On retourne k

Exercice 8.12 *Ecrivez un court programme qui simule une loi de Poisson suivant ce procédé. Dressez un histogramme de fréquences pour vérifier empiriquement que votre programme est valide. Comparez le temps d'exécution de cette procédure à l'algorithme "naïf" (Cet algorithme consiste à gé-*

nérer un nombre aléatoire U choisi uniformément dans $[0, 1)$ et à retourner le plus petit indice k tel que $\sum_{i=0}^{k} e^{-\lambda}\lambda^i/i! < U \leq \sum_{i=0}^{k+1} e^{-\lambda}\lambda^i/i!$.)

8.3 Loi de distribution de Gauss (loi normale)

Une variable aléatoire X est dite *gaussienne* de moyenne μ et de variance σ^2 si elle admet la densité de probabilité :

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} \tag{8.8}$$

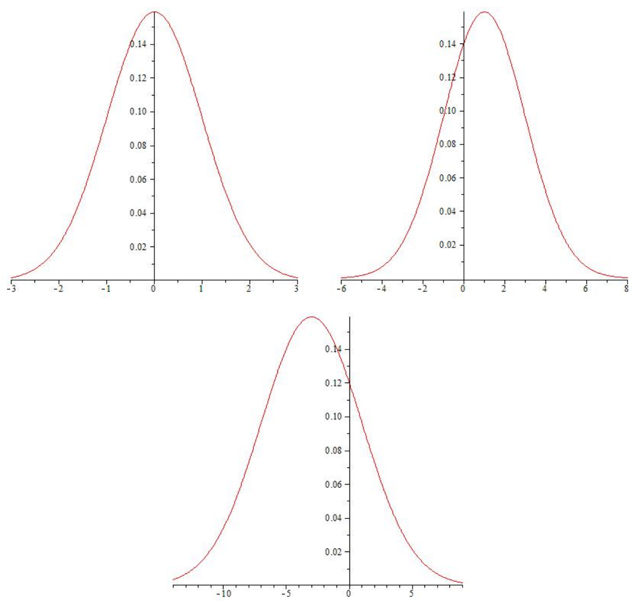


Fig. 8.6 La densité de probabilité d'une loi gaussienne prend la forme d'une "courbe en cloche". Elle est centrée en la moyenne μ et s'étale plus ou moins de manière symétrique autour de la moyenne selon la valeur de σ. Ici, en haut à gauche $\mu = 0, \sigma = 1$, en haut à droite $\mu = 1, \sigma = 2$, en bas $\mu = -3, \sigma = 4$

Cette distribution s'observe dans un nombre étonnant de phénomène en sciences : elles apparaissent dans les équations de diffusion en physique ; pour décrire les orbites moléculaires en chimie ; elles sont utilisées pour lisser le signal et calculer des représentations multi-niveaux des images numériques. Mais aussi dans le domaine du vivant et en sciences sociales (les caractères morphologiques dans une population suivent souvent une distribution gaussienne). Sa place dans le domaine des probabilités et des statistiques est centrale.

La fonction de répartition de la loi de densité de Gaussienne n'admet pas de forme analytique close et se résume à la définition :

$$F_X(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} dx$$

ce qui explique qu'on ait souvent recourt à des tables pour évaluer cette fonction. Par conséquent, on ne possède pas non plus de forme close pour la fonction de répartition inverse. On peut cependant simuler cette variable de manière relativement efficace.

8.3.0.2 Simuler une loi gaussienne

Soient deux nombres aléatoires U_1, U_2 choisis uniformément dans l'intervalle $(0,1)$ (notez qu'on prend soin de ne pas inclure ni 0, ni 1). Désignons par R et θ les quantités :

$$R = \sqrt{-\log U_1}, \quad \theta = 2\pi U_2$$

alors les variables aléatoires :

$$Z_1 = R\cos(\theta), \quad Z_2 = R\sin(\theta)$$

suivent toutes deux une loi normales de moyenne nulle et de variance unitaire. Les variables :

$$T_1 = \mu + \sigma Z_1, \quad T_2 = \mu + \sigma Z_2$$

suivent toutes deux une loi normale de moyenne $\mu = 0$ et de variance $\sigma^2 = 1$.

Cette distribution joue aussi un rôle essentiel en théorie des probabilités et en statistiques, comme l'illustre le résultat suivant :

Théorème Central Limite

C'est ce résultat qui donne une place prédominante à la loi normale dans des études de terrain. En effet, on étudie souvent comment se comporte la valeur moyenne d'observations faites sur des individus d'une population, et on fait l'hypothèse que ces observations sont indépendantes et de même loi de probabilité.

Ce résultat permet aussi d'évaluer l'erreur commise lors d'un calcul d'approximation par des méthodes probabilistes (dites de Monte Carlo).

Partie II
Processus stochastiques

Chapitre 9
Introduction

Ce chapitre utilise les notations et le calcul matriciel. La section 10 fait quelques rappels utiles. Une loi de probabilités ou un variable aléatoire modélise une situation au caractère statique : une expérience se déroule et on cherche à décrire les observations possibles et leurs probabilités.

Ce que nous appellerons un processus stochastique affiche une ambition, celle de modéliser un phénomène dynamique où la situation varie, où les probabilités d'observer tel ou tel autre évènement évolue dans le temps. Le défi est alors de pouvoir prédire, en termes probabilistes, ce qui pourra se passer. Commençons par donner des exemples de situations qu'il est utile de modéliser ainsi.

Deux groupes industriels se partagent un marché. Au vu de l'évolution de ce marché et/ou des stratégies marketing des deux entreprises, il semble que l'une des entreprises puissent prendre à son concurrent une part de ce marché. Mais cette situation est en constante évolution, et à chaque moment un client potentiel peut basculer de l'autre côté pour changer les équilibres. Imaginons que l'on puisse chiffrer ce changement d'équilibre et définir la part de marché qu'une entreprise A prend à B et inversement. Peut-on prédire comment les choses vont évoluer ? La réponse à cette question peut être utile pour décider de l'opportunité d'investir dans des outils cherchant à rompre avec la situation ? Y a t-il un meilleur moment pour agir ? y a t-il un moment à partir duquel il serait trop tard ?

Un hôpital admet des patients de manière continu et observe comment ce flux de personnes se retrouvent à un moment ou un autre dans ses autres services, de pneumologie, de soins intensifs, etc. En formulant cette situation en termes probabilistes (il y a telle probabilités pour qu'un patient admis à l'hôpital soit dirigé vers tel service, qu'un patient admis dans tel service passe à tel autre service, etc.), peut-on décide d'un état d'équilibre ou de déséquilibre ? le nombre de lits par services est-il adapté au flux de patients ?

Un système admet des ressources/tâches en entrée, qu'il traite, avant de les libérer pour ensuite se tourner vers les ressources/tâches mises en attente de traitement. En formulant la situation en termes probabilistes, peut-on prédire la quantité de ressource consommée/produite après un temps donné ? peut-on s'assurer de la fluidité du processus ou peut-on prédire son blocage et à quel horizon ?

Ce sont de tels scénarios qu'il s'agit d'aborder, en tentant de donner des outils de modélisation et de calculs.

Chapitre 10
Calcul matriciel, graphes et chemins

Ce chapitre reprend des éléments du calcul matriciel qui sont utiles pour le cours. Il s'agit essentiellement d'une notation compacte pour désigner les fonctions de transition des chaînes de Markov, comme nous allons le voir dans les chapitre suivants. Par ailleurs, nous n'aborderons qu'un ensemble minimal des propriétés du calcul matriciel. Essentiellement, nous soulignons ici combien le produit de matrice répond au calcul de parcours dans un graphe, qui correspond lui-même à la simulation d'une chaîne de Markov.

Il est utile de penser à une distribution de probabilité $(\pi(i))_{i\in V}$ comme à un *vecteur* colonne et aux éléments de V comme à un *espace d'états*. La valeur $\pi(i)$ indique alors la probabilité de se trouver à l'état $i \in V$.

Une matrice sur V est un tableau $|V| \times |V|$ noté $P = (p_{i,j})_{i,j\in V}$. On peut voir une telle matrice comme un opérateur qui prend en entrée un vecteur et retourne un vecteur $\pi' = P \cdot \pi$. Le calcul se fait en multipliant chaque ligne de la matrice par le vecteur π, retournant ainsi les coordonnées du nouveau vecteur π' :

$$\pi'(i) = \sum_{j\in V} p_{i,j}\pi(j). \tag{10.1}$$

On représente ce calcul à l'aide d'un *graphe* (orienté) dont les sommets correspondent aux états de V. Une arête $i \to j$ relie deux sommets du graphe lorsque $p_{i,j} > 0$. Le calcul de l'équation (10.1) se "lit" sur le graphe : on passe du sommet i au sommet j avec probabilité $p_{i,j}$.

Si on pense au vecteur π comme donnant la probabilité de se trouver sur un état $j \in V$, alors le vecteur π' donne la probabilité de se trouver sur le sommet i *venant d'un sommet* j (où l'on pouvait se trouver selon la distribution de probabilité π).

De la même manière, on peut multiplier la matrice P par elle-même et obtenir la matrice $P^2 = P\times = (r_{i,j})_{i,j\in V}$ où $r_{i,j} = \sum_{k\in V} p_{i,k}p_{k,j}$. Ce calcul s'interprète naturellement en terme de chemins dans le graphe. La valeur

$r_{i,j}$ cumule les probabilité d'aller de i à j en passant par un état transitoire k.

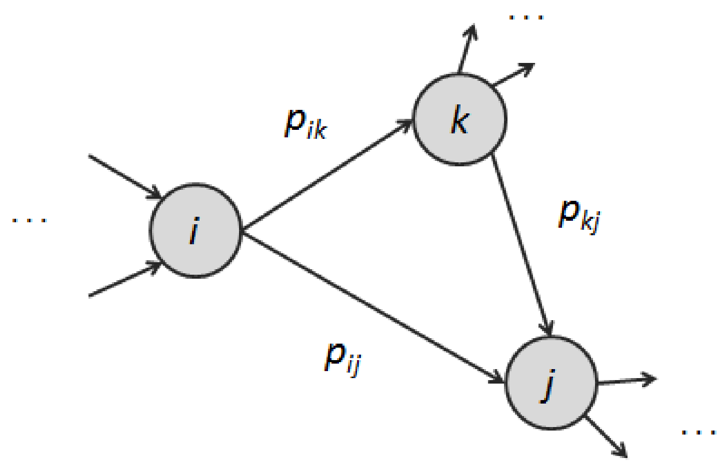


Fig. 10.1 Chaque sommet du graphe (orienté, pondéré) représente un état du système. Les arcs du graphe décrivent les transitions entre les états du système. Les poids des arcs correspondent aux probabilités de réaliser ces transitions.

Ce calcul, et son interprétation, passe aux ordres supérieurs. Les valeurs de la matrice $P^n = (r_{i,j}^{(n)})_{i,j\in V}$ donne les probabilités de passer d'un état i à un état j en suivant un chemin de longueur n dans le graphe.

Exercice 10.1 Matrice et graphe *Soit la matrice :*

$$\begin{bmatrix} \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \end{bmatrix}$$

Représentez graphiquement le graphe correspondant.

Exercice 10.2 *Si on multiplie la matrice P de l'exercice précédent par elle-même un grand nombre de fois, on trouve :*

$$\lim_{n\to\infty} P^n = \begin{bmatrix} \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \\ \frac{1}{4} & 0 & \frac{1}{2} & 0 & \frac{1}{4} \\ 0 & 0 & 1 & 0 & 0 \\ \frac{1}{4} & 0 & \frac{1}{2} & 0 & \frac{1}{4} \\ \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \end{bmatrix}$$

A nouveau, représentez graphiquement le graphe de transition associé à cette matrice. Comparez-le avec le graphe de la matrice de départ.

Exercice 10.3 Convergence *Soit la matrice :*

$$\begin{bmatrix} 0 & 1 & 0 \\ 1-p & 0 & p \\ 0 & 1 & 0 \end{bmatrix}$$

- *Calculez* P^2
- *Montrez que* $P^4 = P^2$
- *Calculez* P^n*, pour tout* $n \geq 1$

Exercice 10.4 Nilpotence *Soit la matrice :*

$$M = \begin{bmatrix} -1 & a & a \\ 1 & -1 & 0 \\ -1 & 0 & -1 \end{bmatrix}$$

- *Calculer* $M + I$, $(M + I)^2$, $(M + I)^3$.
- *En déduire la valeur de* $(M + I)^n$ *(utilisez la formule du binôme de Newton et exploiter le résultat précédent).*

Exercice 10.5 Matrice singulière *Soit la matrice :*

$$M = \begin{bmatrix} 1-x & x \\ y & 1-y \end{bmatrix}$$

On écrit cette matrice sous la forme $M = A + I$.
- *Montrer (calculer) que* $A^2 = -A(x + y)$
- *En déduire, par récurrence, que* $A^n = (-1)^n A(x + y)^n$
- *En déduire la valeur de* M^n *(utilisez la formule du binôme de Newton et exploiter le résultat précédent.*

Chapitre 11
Chaînes de Markov

Les chaînes de Markov proposent une approche intuitive pour modéliser les phénomènes dynamiques et stochastiques. On imagine disposer d'un certain nombre d'états (dans lequel se trouve un système, par exemple) et on décrit les probabilités de passer d'un état à un autre. En faisant une simple hypothèse sur la mémoire du système, on arrive à calculer de manière analytique les probabilités de se retrouver en tel ou tel autre état du système en se projetant dans le temps.

11.1 Chaîne de Markov à deux états

Avant même d'en donner une définition, nous allons tout d'abord nous pencher sur les chaînes les plus simples, celles où il n'y a que deux états. Elles permettent déjà de proposer des modèles simples, et d'exposer les fondements de cette théorie.

On note par 0 et 1 les deux états possibles d'un système. On peut penser, par exemple, qu'il s'agit de sonder périodiquement un canal/capteur pour savoir s'il émet/capte ou s'il est en attente d'émission/réception. On peut penser à un système biologiques dont on suit l'évolution, et où l'on note l'activité ou la non activité d'un gène (une protéine), par exemple.

Le système est donc observé à des instants discrets $n = 0, 1, 2, \ldots$ Nous utiliserons les variables aléatoires X_n qui permettent d'indiquer la probabilité qu'au temps n, le système soit dans l'état 0 ou 1. Nous posons, pour tout $n \geq 0$,

$$P(X_{n+1} = 1 | X_n = 0) = p \tag{11.1}$$
$$P(X_{n+1} = 0 | X_n = 1) = q \tag{11.2}$$

où p, q sont des probabilités, $0 \leq p, q \leq 1$. On en déduit :

$$P(X_{n+1} = 0 | X_n = 0) = 1 - p \tag{11.3}$$
$$P(X_{n+1} = 1 | X_n = 1) = 1 - q \tag{11.4}$$

Les probabilités de changement d'état du système sont donc constante dans le temps puisqu'elles ne dépendent pas de n. Ce type de système est appelé une chaîne de Markov *homogène* (puisque les probabilités des équations (11.3) (11.2) (11.1) (11.4) ne dépendent pas de n). On peut représenter ce système par un diagramme (Fig. 11.1).

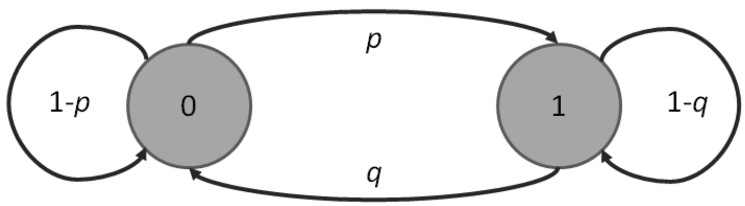


Fig. 11.1 Chaque sommet du graphe (orienté, pondéré) représente un état du système. Les arcs du graphe décrivent les transitions entre les états du système. Les poids des arcs correspondent aux probabilités des équations (11.1) à (11.4).

On déduit des équations (11.1) à (11.4) les valeurs de $P(X_n = 0)$ (et $P(X_n = 1)$. En effet, on a :

$$\begin{aligned}
P(X_{n+1} = 0) &= P(X_{n+1} = 0 \,\&\, X_n = 0) + P(X_{n+1} = 0 \,\&\, X_n = 1) \\
&= P(X_n = 0)P(X_{n+1} = 0 | X_n = 0) + P(X_n = 1)(X_{n+1} = 0 | X_n = 1) \\
&= (1-p)P(X_n = 0) + q(1 - P(X_n = 0)) \\
&= (1-p-q)P(X_n = 0) + q \\
&= (1-p-q)^{n+1}P(X_0 = 0) + q\sum_{k=0}^{n}(1-p-q)^k \\
&= (1-p-q)^{n+1}P(X_0 = 0) + \frac{q}{p+q}\big(1 - (1-p-q)^{n+1}\big)
\end{aligned}$$

Si $0 < p, q < 1$, on a $|1-p-q| < 1$ et alors $\lim_{n\to\infty}(1-p-q)^n = 0$ et

$$\lim_{n\to\infty} P(X_n = 0) = \frac{q}{p+q}. \tag{11.5}$$

Exercice 11.1 *Reprenez le calcul précédent pour montrer :*

$$\lim_{n\to\infty} P(X_n = 1) = \frac{p}{p+q}. \tag{11.6}$$

On en déduit que la distribution de probabilité la plus “plausible” pour X_n est :

$$P(X_n = 0) = \frac{q}{p+q} \qquad P(X_n = 1) = \frac{p}{p+q}, \tag{11.7}$$

et dans ce cas, la distribution est stationnaire puisque toutes les variables X_n auront alors même distribution.

Exercice 11.2 *Montrer ce dernier énoncé, c'est-à-dire qu'en effet,* $P(X_0 = 0) = \frac{q}{p+q}$ *et* $P(X_0 = 1) = \frac{p}{p+q}$ *entraîne* $P(X_n) = P(X_0)$ *pour tout* $n \geq 0$.

Exercice 11.3 *Donnez la matrice associé au graphe de la figure 11.1 qui décrit les transitions de ce système. (Voir l'exercice 10.5.)*

11.1.1 Condition de Markov

Définition 3. Une chaîne de Markov est donc donné au départ par une suite de variables aléatoires $(X_n)_{n\geq 0}$, pour lesquelles on suppose disposer de toutes les probabilités conditionnelles $P(X_{n+1}|X_n)$. La variable aléatoire X_n donne les probabilités de se trouver, au temps n, sur les probabilités $P(X_n = x)$ de se trouver sur les différents états $x \in V$ de la chaîne. On peut voir ces états comme les sommets d'un graphe $G = (V, E)$ dont les arêtes sont les transitions de la chaînes : les possibilités de passer d'un état à un autre. Les probabilités conditionnelles $P(X_{n+1}|X_n)$ donnent les probabilités d'emprunter ces arêtes.

A ces ingrédients s'ajoute la condition de Markov qui stipule que :

$$P(X_{n+1} = x_{n+1}|X_n = x_n, \ldots, X_1 = x_1, X_0 = x_0) = P(X_{n+1}|X_n). \tag{11.8}$$

En d'autres mots, la probabilité de se retrouver à l'état x_{n+1} au temps $n+1$ ne dépend *que* de l'état x_n auquel l'on se trouvait au temps n.

Si les probabilités $P(X_{n+1}|X_n)$ ne dépendent pas de n (et reste identique pour tout $n \geq 0$), on dit que la chaîne de Markov est *homogène*. Sauf, exception, nous ne considérerons que le cas homogène dans ce cours.

La condition de Markov (11.8) entraîne que l'on est en mesure de calculer la distribution jointe des variables aléatoires X_0, X_1, ..., X_n puisque :

Exercice 11.4 Distribution jointe

$$P(X_n = x_n, \ldots, X_1 = x_1, X_0 = x_0) = P(X_{n+1}|X_n)\cdots P(X_1|X_0)P(X_0 = x_0). \tag{11.9}$$

(Le cas $n = 1$ est trivial. Procédez par induction et utilisez la condition de Markov Eq. (11.8)).

Exercice 11.5 *On revient à la chaîne de Markov à deux états de la section précédente.*

a) Calculez :

$$P(X_1 = 0|X_0 = 0, X_2 = 0) = \frac{(1-p)^2}{(1-p)^2 + pq}$$

b) Calculez :

$$P(X_1 \neq X_2) = \pi_0(0)(1-p-q)(p-q) + q(1+p-q)$$

Exercice 11.6 Marche aléatoire sur les entiers et chemins dans le plan *On considère des variables aléatoires ξ_0, $\xi_1, \ldots$ à valeurs entières dans $\mathbb{Z}$, et de même fonction de densité f. Soit X_0 une variable aléatoire à valeurs entières indépendante des ξ_i. On définit une suite de variables aléatoires $(X_n)_{n\geq 0}$ en posant $X_n = X_0 + \xi_0 + \xi_1 + \cdots + \xi_n$.*

- *Quel est l'espace des états de cette chaîne ?*
- *Simulez la suite X_n*
- *Représentez graphiquement une simulation de X_n par un chemin dans le plan $(0, X_0), (1, X_1), \ldots, (n, X_n), \ldots$*

La suite (X_n) est appelée une marche aléatoire sur les entiers de $\mathbb{Z}$.

Exercice 11.7 Chaîne d'Ehrenfest *On imagine avoir d boules étiquetées $1, 2, \ldots, d$ qui sont, au départ, réparties entre deux urnes U_1 et U_2. On modifie la répartition des boules dans les urnes de la manière suivante : on choisit (uniformément) au hasard un entier i parmi $1, 2, \ldots, d$ et on fait passer alors la boule étiquetée i de l'urne dans laquelle elle se trouve dans l'autre urne.*

Décrivez la chaîne de Markov définie par cette procédure, donnez son espace d'état et les probabilités des transitions.

Solution Les états de cette chaîne de Markov peuvent être décrit comme des couples d'entiers positifs ou nuls (i, j) tels que $d = i + j$. En réalité, on peut faire plus simple : il suffit de noter le nombre de boules qui sont dans l'urne U_1, les états étant alors décrit par un entier positif ou nul $0 \leq i \leq d$. On peut passer de l'état i à l'état $i-1$ ou $i+1$ avec probabilité i/d et $d - i/d$ respectivement.

En ordonnant les états naturellement de 0 à d, on peut décrire les transitions de cette chaîne de manière matricielle :

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ \frac{1}{d} & 0 & \frac{d-1}{d} & 0 & 0 & 0 \\ 0 & \frac{2}{d} & 0 & \frac{d-2}{d} & 0 & 0 \\ 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 0 & 0 & \frac{d-1}{d} & 0 & \frac{1}{d} \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

□

Exercice 11.8 Marche aléatoire sur un graphe *On considère un graphe simple $G = (V, E)$ (non orienté, sans boucle). On note $\deg_G(u)$ le degré d'un sommet $u \in V$. C'est son nombre de voisins $|N_G(u)|$, où l'ensemble des voisins de $u \in V$ est l'ensemble des sommets qui lui sont incidents : $N_G(u) = \{v \in V | \{u, v\} \in E\}$.*

On forme la chaîne de Markov dont les états sont les sommets de G. Les transitions sont donnés par le processus suivants. Si on se trouve à l'état $u \in V$, alors on va sur un état $v \in N_G(u)$ avec probabilité $\deg_G(u)$. En d'autres mots, on choisit au hasard et uniformément un voisin de $u \in V$ et on s'y déplace.

Simulez la marche aléatoire sur un graphe en prenant pour $G = (V, E)$:
- *une chaîne de longueur finie ;*
- *une étoile (un sommet central et N sommets qui ont ce sommet central comme unique voisin).*

Prenez soin de noter le nombre de fois qu'un sommet est visité au cours de la marche, et comparez cette valeur au nombre d'arête du graphe.

11.2 Fonction de transition et matrices

Il est opportun d'avoir une vision "en image" des chaînes de Markov. On peut penser aux états de la chaîne comme aux sommets d'un graphe. La chaîne étant une suite de variables aléatoires, une simulation de la chaîne correspond à se déplacer sur la chaîne en choisissant, à l'étape n, le voisin du sommet en fonction de la distribution de probabilité de X_n.

Considérons une chaîne ayant un nombre fini d'états et dont les probabilités de transition sont décrites par la matrice :

$$\begin{bmatrix}
0 & 0.552 & 0.034 & 0.345 & 0 & 0.069 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.656 & 0 & 0 & 0.281 & 0 & 0 & 0 & 0.063 \\
0 & 0 & 0.692 & 0.091 & 0 & 0.206 & 0 & 0 & 0.011 & 0 \\
0 & 0 & 0.023 & 0.592 & 0.090 & 0.136 & 0 & 0 & 0.159 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.909 & 0 & 0 & 0.091 \\
0 & 0 & 0.007 & 0.027 & 0.010 & 0.854 & 0 & 0 & 0.095 & 0.007 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.703 & 0.297 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.882 & 0.118 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}$$

Exercice 11.9 *Dessinez le graphe correspondant à cette matrice.*

Cette matrice contient toutes les probabilités conditionnelles $P(X_{n+1} = y | X_n = x) = p(x, y)$, où x, y sont des états de la chaîne, et qui ne dépendent donc pas de n si la chaîne est homogène.

Pour l'anecdote : les entrées de cette matrice décrivent la probabilité pour un patient d'un centre hospitalier, de passer d'un service à un autre. Les services correspondent ici aux lignes/colonnes et dans l'ordre : Admission, Urgence, Unité coronarienne, Investigation, Chirurgie, Thérapeutique, Soins intensifs, Soins, Congé, Décès. On comprend donc pourquoi les deux dernières lignes ne contiennent que des zéros [1].

Exercice 11.10 *On reprend la chaîne de Markov qui effectue une marche aléatoire sur un graphe simple. Décrivez la matrice de transition de cette chaîne lorsque le graphe est :*
- *une chaîne*
- *une étoile*

Exercice 11.11 Boules et boîtes *On dispose de d boules noires et d boules rouges (on pourrait tout aussi bien avoir des processus associés à des utilisateurs, etc.). Les boules sont placées soit dans une première boîte B_1, soit dans une autre boîte B_2. La distribution des boules noires et rouges dans les boîtes B_1 et B_2 constitue un état du système. Le système peut passer d'un état à un autre de la manière suivante : une boule est choisie au hasard de chacune des boîtes ; puis ces deux boules sont échangées. La variable X_0 donne la distribution des boules dans les boîtes au départ, et pour $n \geq 1$, la variable X_n donne la distribution dans les états suivants.*
- *Donnez une description de ce système. Définit-il une chaîne de Markov ?*
- *Donnez sa fonction de transition et représentez-le graphiquement.*

Solution Désignons par k (avec $1 \leq k \leq 2d-1$) le nombre de boules qui sont placées dans la boîte B_1. La boîte B_2 en contiendra donc $2d-k$. Puisqu'à chaque étape on échange une boule de chacune des boîtes, ces nombres restent invariants. Ce sont les nombres n_R et n_N de boules rouges et noires dans la boîte B_1 qui varieront. Les nombres de boules rouges et noires qui sont dans la boîte B_2 varient aussi mais ces quantités se déduisent de n_R et n_N. De même, puisque le nombre total de boules dans la boîte B_1 est k, la quantité n_N se déduit de n_R.

Un état du système est donc complètement décrit par un entier n_R avec $0 \leq n_R \leq \min\{k, d\}$.

Supposons que les boîtes soient dans la configuration n_R. Lorsqu'on choisit une paire de boules (!), l'une de la boîte B_1, l'autre de la boîte B_2, on peut choisir deux boules rouges, deux boules noires ou une boule de chaque couleur. Les probabilités de chacun de ces choix dépend de l'entier n_R. Le choix de boules de la même couleur laisse le système dans le même état ; le choix de boules de couleurs différentes fait évoluer l'état du système. Notez que le nombre de boules noires dans la boîte B_1 est $k-n_R$;

1. Cet exemple est tiré du livre de Pierre Leroux (1983). Algèbre linéaire – Une approche matricielle. Modulo éditeur.

le nombre de boules rouges dans la boîte B_2 est $d-n_R$; le nombre de boules noires dans la boîte B_2 est $d-(k-n_R)$.

- La probabilité du choix de deux boules rouges est de $\frac{n_R}{k} \times \frac{d-n_R}{2d-k}$ (le produit des probabilités du choix d'une boule rouge dans chacune des boîtes indépendamment).
- La probabilité du choix de deux boules noires est de $\frac{k-n_R}{k} \times \frac{d-(k-n_R)}{2d-k}$ (le produit des probabilités du choix d'une boule rouge dans chacune des boîtes indépendamment).
- La probabilité du choix d'une boule rouge en B_1 et d'une boule noire en B_2 est $\frac{n_R}{k} \times \frac{d-(k-n_R)}{2d-k}$. On passe alors dans l'état $n_R - 1$
- La probabilité du choix d'une boule noire en B_1 et d'une boule rouge en B_2 est $\frac{k-n_R}{k} \times \frac{d-(k-n_R)}{2d-k}$. On passe alors dans l'état $n_R + 1$.

La probabilité de la boucle est donc $\frac{n_R}{k} \times \frac{d-n_R}{2d-k} + \frac{k-n_R}{k} \times \frac{d-(k-n_R)}{2d-k}$

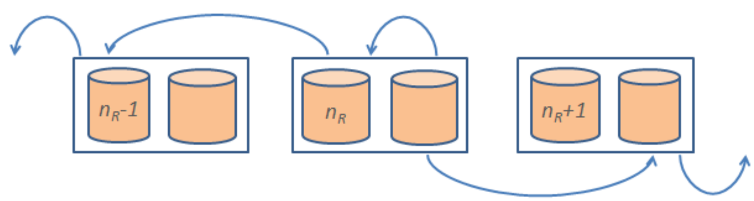


Fig. 11.2 Les transitions consistent à diminuer ou augmenter le nombre de boules rouges dans la boîte B_1.

□

Exercice 11.12 Condition de Markov *Montrez que la condition de Markov entraîne :* $P(X_{n+1} = x, X_n = y | X_{n-1} = z) = P(X_{n+1} = x | X_n = y) P(X_n = y | X_{n-1} = z)$.

Ainsi, on a :

$$
\begin{aligned}
& P(X_{n+1} = y_1, X_{n+m} = y_m, \ldots, | X_0 = x_0, X_1 = x_1, \ldots, X_n = x) \\
&= P(X_{n+1} = y_1, X_{n+m} = y_m, \ldots, | X_n = x) \\
&= P(X_{n+m} = y_m | X_{n+m-1} = y_{m-1}) \\
&\qquad \cdots P(X_{n+2} = y_2 | X_{n+1} = y_1) P(X_{n+1} = y_1 | X_n = x)
\end{aligned}
$$

En d'autres mots, la probabilité de passer de l'état $y_0 = x$ à l'état y_m en m étapes et en passant par les états $y_1, \ldots, y_{m-1}$ est donné par le produit des probabilités conditionnelles correspondantes. Si on ne s'intéresse qu'aux états de départ $y_0 = x$ et d'arrivée y_m, et que l'on s'autorise à passer par tous les chemins intermédiaires possibles, on trouve :

$$
\begin{aligned}
&P(X_{n+m} = y_m | X_n = y_0) \\
= &\sum y_1, \dots, y_{m-1} P(X_{n+m} = y_m | X_{n+m-1} = y_{m-1}) \\
&\quad \dots P(X_{n+2} = y_2 | X_{n+1} = y_1) P(X_{n+1} = y_1 | X_n = x)
\end{aligned}
$$

ce que l'on peut exprimer de manière matricielle par l'équation :

$$X_{n+m} = X_n P^m.$$

Exercice 11.13 *Une compagnie aérienne étudie la réservation sur l'un de ses vols. Une place donnée, libre le jour d'ouverture de la réservation, voit son état évoluer chaque jour jusqu'à la fermeture de la réservation de la manière suivante :*

- *si la place est réservée le jour k, elle le sera encore le jour $k + 1$ avec la probabilité 9/10*
- *si la place est libre le jour k, elle sera réservée le jour $k + 1$ avec la probabilité 4/10*

Pour k entier positif, on note r_k la probabilité que la place soit réservée le jour k. On suppose que $r_0 = 0$.

a) Exprimer r_{k+1} en fonction de r_k

b) En déduire l'expression explicite de r_k en fonction de k et calculer $\lim_{n\to\infty} r_n$

Solution On définit une suite de variables aléatoires X_k ($k \geq 0$) qui donne la probabilité que la place soit libre. On convient que $X_k = 0$ correspond à "place libre" et $X_k = 1$ au cas contraire. Par hypothèse, on a $P(X_{k+1} = 0 | X_k = 0) = 6/10$ et $P(X_{k+1} = 0 | X_k = 1) = 1/10$. De même, $P(X_{k+1} = 1 | X_k = 0) = 4/10$ et $P(X_{k+1} = 1 | X_k = 1) = 9/10$. On est donc ramener au calcul des probabilités de la chaîne de Markov à deux états avec des valeurs particulières pour p et q. Les équations (11.5) et (11.6) nous permettent de conclure à $\lim_{n\to\infty} P(X_n = 1) = \frac{4/10}{1/10+4/10} = 4/5$.

On peut écrire une fonction qui permet de simuler un pas de la marche. Il lui faut connaître l'état X_n de la chaîne à l'instant n, et la fonction choisit aléatoirement le prochain état, selon le modèle : elle effectue un tirage de Bernouilli dont les paramètres dépend de l'état X_n. On convient que l'état de départ est $X_0 = 0$ (la place libre), puis on peut itérer la marche pour valider le résultat théorique.

De manière plus intéressante, on peut écrire dans un fichier l'évolution des fréquences de visites des états de la chaîne pour visualiser le rythme auquel le système converge vers les valeurs prédites par la théorie.

Pour la fonction Bernouilli, voir l'exercice 3.7. Le logiciel `gnuplot` peut facilement produire une (des) courbe(s) pour visualiser l'évolution des fréquences écrite dans un fichier, comme le fait la fonction `freqSystemeResa`.

□

```
# on gere les resa de maniere independante
# on applique ici la "regle" de la chaine
# state = 0 correspond à place libre
# on applique ici la "regle" de la chaine
# state = 0 correspond à place libre
def transition(state):
    if state == 0:
        return Bernouilli(0.4)
    else:
        return Bernouilli(0.9)
```

Fig. 11.3 Implémentation de la fonction de transition de la chaîne de Markov modélisant le système de réservation de places.

```
# simulation du systeme de reservation de places
# le parametre m est le nombre de pas à effectuer
# c-a-d le nombre de jours jusqu'a cloture de vente des places
# n est le nombre de palces libres au depart
# simulation du systeme de reservation de places
# le parametre m est le nombre de pas à effectuer
# c-a-d le nombre de jours jusqu'a cloture de vente des places
def systemeResa(m):
    # place = 0 ou 1 (libre ou reservee)
    places = creerTableau(m, 0)
    for j in range(1, m):
        places[j] = transition(places[j-1])
    return places

def freqSystemeResa(places, file):
    fd = open(file, 'w')
    fd.write("#N\tR\tL\n")
    str = ""
    R = 0
    L = 0
    for i in range(len(places)):
        if places[i] == 0:
            L += 1
        else:
            R += 1

        str = "%d\t%d\t%d\n" % (i, R, L)
        fd.write(str)
    fd.close()
    return True
```

Fig. 11.4 Simulation d'une marche aléatoire sur la chaîne de Markov modélisant le système de réservation de places. L'écriture sur fichier permet ensuite d'exploiter les résultats de la simulation, par exemple pour produire des courbes d'évolution du taux de remplissage dans le temps.

Exercice 11.14 *Soit* $a \in (0, 1/2)$ *un nombre réel.*

Dans une bourse de valeurs, un titre donné peut monter, *rester* stable *ou* baisser.

Dans un modèle mathématique, on considère que :

– *le premier jour le titre est stable*
– *si un jour n le titre monte, le jour $n+1$; il montera avec probabilité $1-2a$, restera stable avec probabilité a et baissera avec probabilité a*
– *si un jour n le titre est stable, le jour $n+1$ il montera avec probabilité a, restera stable avec probabilité $1-2a$ et baissera avec probabilité a*
– *si un jour n le titre baisse, le jour $n+1$ il montera avec probabilité a, restera stable avec probabilité a et baissera avec la probabilité $1-2a$*

On note M_n (resp. S_n, resp. B_n) l'événement "le titre donné monte" (resp. reste stable, resp. baisse) le jour n. On pose $p_n = P(M_n)$, $q_n = P(S_n)$ et $r_n = P(B_n)$.

a) *Donnez une chaîne de Markov qui décrit cette situation, représentez-la graphiquement.*
b) *Que vaut $p_n + q_n + r_n$? En déduire l'expression de r_n en fonction de p_n et q_n,*
c) *Expliciter p_{n+1} (resp. q_{n+1}) en fonction de p_n, q_n, r_n,*
d) *En déduire p_n, q_n puis r_n,*
e) *Donner la limite de ces trois suites et interpréter le résultat.*

Exercice 11.15 La ruine du joueur *Un joueur qui joue au casino démarre sa soirée avec un certain capital (en €) et que les probabilités de gagner et perdre (1 €) sont p et $1-p$. Si le joueur perd tout son argent, il cesse de jouer. Décrivez ce jeu à l'aide d'une chaîne de Markov et donnez-en la fonction de transition.*

Simulez cette chaîne de Markov. Donnez une représentation graphique d'une simulation de la chaîne.

Exercice 11.16 *Pour représenter le passage d'une molécule de phosphore dans un écosystème, nous considérerons quatre états possibles :*
– *la molécule est dans le sol,*
– *la molécule est dans l'herbe,*
– *la molécule a été absorbée par du bétail,*
– *la molécule est sortie de l'écosystème.*

La matrice de transition (illustré par le graphe plus haut) est la suivante :

$$P = \begin{bmatrix} \frac{3}{5} & \frac{3}{10} & 0 & \frac{1}{10} \\ \frac{1}{10} & \frac{2}{5} & \frac{1}{2} & 0 \\ \frac{3}{4} & 0 & \frac{1}{5} & \frac{1}{20} \\ \frac{1}{100} & 0 & 0 & \frac{99}{100} \end{bmatrix}$$

Notez que cette matrice de transition correspond au système d'équations :

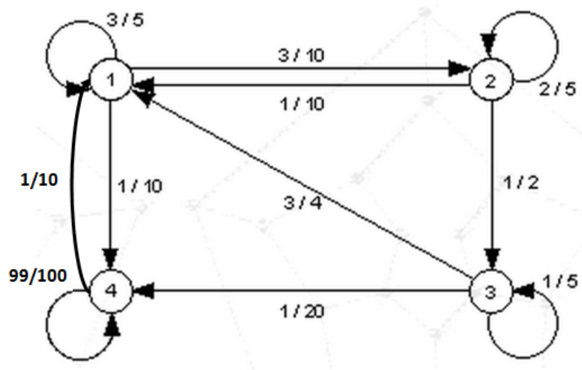


Fig. 11.5 Chaque sommet du graphe représente un état de la molécule. Les arcs du graphe décrivent les probabilités des transitions entre les états.

$$
\begin{aligned}
S_{k+1} &= 0.6S_k + 0.1H_k + 0.75B_k + 0.01O_k \\
H_{k+1} &= 0.3S_k + 0.4H_k \\
B_{k+1} &= 0.5H_k + 0.2B_k \\
O_{k+1} &= 0.1S_k + 0.05B_k + 0.99O_k
\end{aligned}
$$

Simulez cette chaîne et formulez une hypothèse quant à la distribution d'un million de molécules disséminées au départ dans l'herbe.

Exercice 11.17 *Deux pièces A et B sont reliées entre elles par une porte ouverte. Seule la pièce B possède une issue vers l'extérieur. Une guêpe initialement dans la pièce A voudrait sortir à l'air libre. Son trajet obéit aux règles suivantes :*

- *Lorsqu'elle est en A au temps $t = n$, alors au temps $t = n + 1$, elle reste en A avec une probabilité égale à 1/3 ou elle passe en B avec une probabilité égale à 2/3,*
- *Lorsqu'elle est en B au temps $t = n$, alors au temps $t = n + 1$, elle retourne en A avec une probabilité égale à 1/4, ou elle reste en B avec une probabilité égale à 1/2, ou elle sort à l'air libre avec une probabilité égale à 1/4.*

Au temps $t = 0$, la guêpe est dans la pièce A. Lorsqu'elle est sortie, elle ne revient plus.

a) Modélisez cette situation à l'aide d'une chaîne de Markov $(X_n)_{n\geq 0}$, représentez graphiquement la chaîne,

b) Calculez explicitement les distributions de probabilités des variables X_0, X_1 X_2,

c) Exprimez $P(X_{n+1} = A)$ et $P(X_{n+1} = B)$ en fonction de $P(X_n = A)$ et $P(X_n = B)$ (avec des notations évidentes), *d)* *Vérifiez que la suite* $\frac{6}{10}P(X_n = A) - \frac{3}{10}P(X_n = B)$ *est constante,*

e) *Vérifiez que la suite* $\frac{4}{10}P(X_n = A) + \frac{3}{10}P(X_n = B)$ *est géométrique de raison* $\frac{5}{6}$*,*

f) *En déduire l'expression de* $P(X_n = A)$ *et* $P(X_n = B)$*,*

g) *Montrer que pour* $n \geq 2$*,* $P(X_n = S) = \frac{1}{4}\sum_{k=1}^{n-1} P(X_{n-1} = B)$*. En déduire* $P(X_n = S)$*.*

Solution On peut modéliser cette situation par une chaîne de Markov à trois états (A, B, S) (Figure 11.6). La matrice de transition de cette chaîne est :

$$M = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 1 \end{bmatrix}$$

La guêpe se trouve initialement dans la pièce A, on a donc $X_0 = (1, 0, 0)$ (en prenant soin d'ordonner les états par A, B, S). Le vecteur d'état X_1 s'obtient de X_0 en appliquant la matrice $X_1 = X_0 \cdot M = (1/3, 2/3, 0)$. De même on calcule $X_2 = (5/18, 5/9, 1/6)$. Soit le vecteur d'état $X_n = (a_n, b_n, s_n)$, on a alors $X_{n+1} = (1/3a_n + 1/4b_n, 2/3a_n + 1/2b_n, 1/4b_n + s_n)$ (où $a_n = P(X_n = A)$, $b_n = P(X_n = B)$ et $s_n = P(X_n = S)$). On peut donc écrire $a_{n+1} = 1/3a_n + 1/4b_n$, $b_{n+1} = 2/3a_n + 1/2b_n$, $s_{n+1} = 1/4b_n + s_n$.

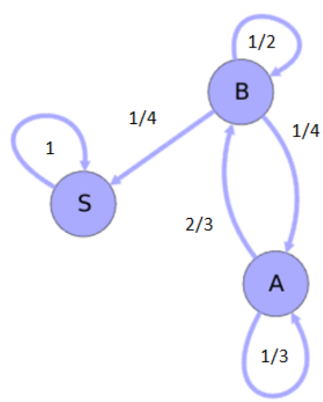


Fig. 11.6 Chaque sommet du graphe représente une pièce où peut se trouver la guêpe.

On vérifie que $\frac{6}{10}P(X_n = A) - \frac{3}{10}P(X_n = B) = \frac{6}{10}a_n - \frac{3}{10}b_n$ est constante pour $n \geq 1$ par récurrence. On le vérifie pour $n = 1$. On calcule :

$$\frac{6}{10}a_{n+1} - \frac{3}{10}b_{n+1} = \frac{6}{10}(1/3a_n + 1/4b_n) - \frac{3}{10}(2/3a_n + 1/2b_n) = 0$$

On utilise les notations $a_n = P(X_n = A)$, $b_n = P(X_n = B)$. On vérifie que $\frac{4}{10}a_1 + \frac{3}{10}b_1 = 1/3$.

$$\begin{aligned}\frac{4}{10}a_{n+1} + \frac{3}{10}b_{n+1} &= \frac{4}{10}(\frac{1}{3}a_n + \frac{1}{4}b_n) + \frac{3}{10}(\frac{2}{3}a_n + \frac{1}{2}b_n)\\ &= \frac{5}{6}(\frac{4}{10}a_n + \frac{3}{10}b_n)\end{aligned}$$

La récurrence nous donne donc $\frac{4}{10}a_{n+1} + \frac{3}{10}b_{n+1} = \frac{5}{6}^n * 1/3$. Les deux équations considérées nous permettent donc de déduire que $a_{n+1} = \frac{5}{6}^n * 1/3$, par conséquent $\lim_{n\to+\infty} P(X_n = A) = 0$. De même, on a $b_{n+1} = 2a_{n+1} = P(X_n = B) = \frac{5}{6}^n * 2/3$ et $\lim_{n\to+\infty} P(X_n = B) = 0$ (la guêpe sortira donc des pièces A et B avec certitude – à un moment donné).

On a $P(X_{n+1} = S) = 1 - P(X_{n+1} = A) - P(X_{n+1} = B) = 1 - \frac{5}{6}^n$, et on est certain que la guêpe finira par sortir à l'extérieur puisque $\lim_{n\to+\infty} P(X_n = S) = 1$. $\square$

Exercice 11.18 *Une boite A contient deux jetons portant le numéro 0 et une boite B contient deux jetons portant le numéro 1. On modifie la situation par des tirages aléatoires répétés : on tire au hasard un jeton dans chaque boîte et on les échange. On s'intéresse à la somme des jetons contenus dans l'urne A à l'instant $t = n$.*

a) *Modélisez cette situation par une chaîne de Markov $(X_n)_{n\geq 0}$, représentez-là graphiquement,*
1. *Calculer les distributions des variables X_0 et X_1,*
c) *Exprimer $P(X_{n+1} = 0)$ (resp. $P(X_{n+1} = 1)$; resp. $P(X_{n+1} = 2)$) en fonction de $P(X_n = 0)$, $P(X_{n+1} = 1)$, $P(X_{n+1} = 2)$*
d) *Montrer que $P(X_{n+2} = 1) = \frac{1}{2}P(X_{n+1} = 1) + \frac{1}{2}P(X_n = 1)$ $(n \geq 0)$*
e) *En déduire l'expression de $P(X_n = 1)$ en fonction de n puis celle de $P(X_n = 0)$ et de $P(X_n = 2)$,*
f) *Déterminer les limites de ces trois suites et interprétez les résultats.*

Exercice 11.19 *Deux urnes U_1 et U_2 contiennent des boules blanches et noires. L'urne U_1 contient 13 boules blanches et 5 boules noires, l'urne U_2 contient 10 boules blanches et 12 boules noires. On effectue un premier tirage dans une urne choisie au hasard et on remet la boule obtenue dans son urne d'origine. Si l'on obtient une boule blanche (resp. noire), le 2ème tirage se fait dans U_1 (resp. U_2). Au $i$ème tirage, si la boule obtenue est blanche (resp. noire), le $(i + 1)$ème tirage se fait dans U_1 (resp. U_2).*

a) *Modélisez cette situation par une chaîne de Markov $(X_n)_{n\geq 0}$*

Nous allons chercher à décrire les probabilités de transition de cette chaîne (en fonction des évènements B_i = "on obtient une boule blanche au $i$ème tirage").

1. *Calculer* $P(B_1)$ *et* $P(B_2)$
2. *Exprimer* $P(B_n)$ *en fonction de* $P(B_{n-1})$
3. *Montrer que la suite* $P(B_n)$ *converge et déterminer sa limite. Interprétez le résultat.*

11.3 Distribution stationnaire

Le principal intérêt d'une chaîne de Markov, hormis la possibilité de simuler un système, est de pouvoir prédire le comportement de ce système. On peut voir la chaîne comme un dispositif permettant de calculer une distribution de probabilité sur les états de la chaîne : la variable X_N indique la probabilité de se retrouver sur chacun des états. Lorsque les bonnes conditions sont réunies, les variables X_N convergent vers une distribution limite.

Définition 4. Soit une chaîne de Markov de matrice de transition P. Une distribution de probabilité π (sur l'ensemble des états de la chaîne) est *stationnaire* lorsqu'on a $\pi P = \pi$.

Nous avons déjà vu un exemple de telle distribution pour la chaîne de Markov à deux états à l'équation (11.7) (page 85, voir aussi les exercices 11.13, 11.14).

Exercice 11.20 *Montrer que la distribution équiprobable* $(1/3, 1/3, 1/3)$ *est stationnaire pour le système modélisant la bourse de valeurs de l'exercice 11.14, comme le prédit la limite calculée dans cet exercice.*

De même, la distribution $(0, 0, 1)$ *est stationnaire pour le système modélisant le comportement de la guêpe de l'exercice 11.17.*

Théorème 1. *Toute chaîne de Markov (toute matrice de transition) possède une distribution stationnaire.*

La technicité de la démonstration nous empêche de la donner ici. La preuve fournit toutefois des indications sur la manière d'obtenir une distribution stationnaire. Soit en effet la matrice de transition d'une chaîne de Markov et soit une distribution de probabilité initiale π_0 (sur les états de la chaîne). On peut construire une suite de distribution ν_k $(k \geq 0)$:

$$\nu_k = \frac{1}{k}(\pi_0 + P\pi_0 + P^2\pi_0 + \cdots + P^k\pi_0)$$

La distribution stationnaire prédite par le théorème est une limite d'une sous-suite bien choisie de la suite $(\nu_k)_{k\geq 0}$. Dans les cas qui nous intéresse, la suite elle-même converge vers une distribution stationnaire.

Exercice 11.21 *Reprenez l'implémentation de la chaîne modélisant le système de réservation de la compagnie aérienne. Il s'agit d'une instance de la chaîne de Markov à deux états. Validez empiriquement le résultat du théorème 1.*

Exercice 11.22 *On revient au graphe en étoile. Partant d'une distribution quelconque (mais non stationnaire), formez une suite de distribution qui converge vers la distribution stationnaire (voir exercice 11.27).*

Exercice 11.23 *Considérons une chaîne de Markov à trois états, dont les transitions sont décrites par la matrice :*

$$P = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/4 & 1/2 & 1/4 \\ 1/6 & 1/3 & 1/2 \end{bmatrix}$$

Calculez la distribution stationnaire de cette chaînes.

Exercice 11.24 *On suppose que la population d'arbres d'une forêt suit un modèle de Markov (simpliste), et que les arbres appartiennent à l'une des catégories suivantes : les arbres de bas âge (de la naissance jusqu'à leur 15ème année), les arbres jeunes (de 15 à 30 ans), les arbres mûrs ((de 30 à 45 ans) et les arbres vieux (de plus de 45 ans). Un arbre croît donc au fil du temps et passe d'une catégorie à une autre. Chaque année, un part des arbres d'une catégorie passeront à la catégorie suivante. Certains des arbres parmi les plus âgés mourront.*

Ces proportions répondent donc au système d'équations :

$$\begin{aligned} B_{k+1} &= d_B B_k + d_J J_k + d_M M_k + d_V V_k \\ J_{k+1} &= (1 - d_B) B_k \\ M_{k+1} &= (1 - d_J) J_k \\ V_{k+1} &= (1 - d_M) M_k + (1 - d_V) V_k \end{aligned}$$

(avec des notations évidentes, Bébé, Jeune, Mûrs, Vieux) où les coefficients d_B, d_J, d_M *et* d_V *déterminent comment évolue la population de la forêt.*

- *Donnez une représentation graphique de cette chaîne.*
- *On pose* $d_B = 0.1$, $d_J = 0.2$, $d_M = 0.3$ *et* $d_V = 0.4$. *Donnez la matrice de transition.*
- *Ecrivez un court programme qui permet de simuler une marche aléatoire dans cette chaîne, partant d'un état donné.*

Imaginons qu'au départ 50 000 arbres soient plantés.

- *Modifiez votre programme de manière à pouvoir simuler l'évolution de la forêt au cours du temps. En d'autres mots la distribution des arbres par catégories d'âge après 15 ans, après 30 ans, après 45 ans, par exemple.*

– *Déterminez cette évolution par calcul matriciel (utilisez Maple, par exemple).*
– *Vérifiez que la distribution* $\frac{1}{3.88}(1, 0.9, 0.72, 1.26)$ *est stationnaire, de manière analytique (calcul matriciel) et expérimentalement (avec votre programme).*

Solution La chaîne peut être décrite par le graphe orienté et valué de la figure 11.7 :

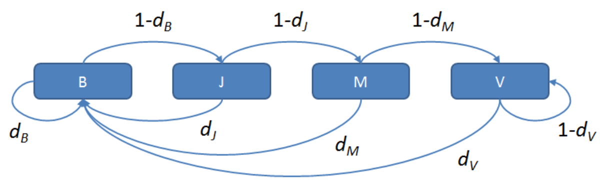


Fig. 11.7 Chaque sommet du graphe représente un état de la molécule. Les arcs du graphe décrivent les probabilités des transitions entre les états.

Notez la dualité entre le système d'équations et le graphe : les probabilités des arcs entrants en un sommet du graphe apparaissent dans l'équation le concernant. La matrice de transition de ce système est :

$$P = \begin{bmatrix} dB & 1-dB & 0 & 0 \\ dJ & 0 & 1-dJ & 0 \\ dM & 0 & 0 & 1-dM \\ dV & 0 & 0 & 1-dV \end{bmatrix} = \begin{bmatrix} 0.1 & 0.9 & 0 & 0 \\ 0.2 & 0 & 0.8 & 0 \\ 0.3 & 0 & 0 & 0.7 \\ 0.4 & 0 & 0 & 0.6 \end{bmatrix}$$

On peut simuler une transition du système à l'aide d'une fonction qui dépend des paramètres du modèle $dB, dJ, \ldots$ et de l'état où se trouve le système. Comme chaque sommet possède deux arcs sortant, un simple tirage de Bernouilli permet de simuler le déplacement sur le prochain état de la chaîne. $\square$

Exercice 11.25 *Le département chargé du développement commercial adopte une stratégie qui définit quatre catégories de clients :*
– *les clients qui sont prospectés (P) selon une procédure rapide (téléphone, salons) et dont une partie revient vers la société ; on compte parmi ceux-ci les clients qui viennent spontanément (recommandation d'un proche ou d'une connaissance) ;*
– *les clients qui donnent suite et entrent dans une discussion plus approfondie (A) des services offerts par la société (échange de documentation technique, premier(s) rendez-vous) ;*
– *les clients dont la démarche est engagée (E) (élaboration de devis spécifique) ;*

- *les clients ayant passé commande (C) ;*
- *à chacun de ces phases, un certain nombre de clients interrompent leur démarche (I) ; on estime toutefois qu'une part de ces clients reviennent vers la société ultérieurement.*

On suppose que l'évolution semaine après semaine *de ces catégories de clients suit de l'activité de la société selon un modèle de Markov (simpliste). Au fil des démarches, certains clients passent de simples prospects vers les classes plus "prometteuses". Les proportions de ces catégories de clients sont dictées par ces équations :*

$$\begin{aligned} P_{k+1} &= 0.1P_k + 0.2A_k + 0.3E_k + 0.4C_k + 0.1I_k \\ A_{k+1} &= 0.3P_k + 0.1I_k \\ E_{k+1} &= 0.4A_k \\ C_{k+1} &= 0.5E_k + 0.5C_k \\ I_{k+1} &= 0.6P_k + 0.4A_k + 0.2E_k + 0.1C_k + 0.8I_k \end{aligned} \tag{11.10}$$

*avec des notations évidentes (**P**rospect, **A**pprofondis, **E**ngagés, **C**ommande, **I**nterruption).*

1. *Comment peut-on interpréter la dernière équation ?*
2. *Donnez une représentation graphique de cette chaîne.*
3. *Donnez sa matrice de transition.*
4. *La chaîne est donnée par une suite de variables aléatoires* $(X_N)_{N\geq 0}$ *? Décrivez la variable* X_N *(pour* N *quelconque). Comment peut-on interpréter les valeurs qu'elle prend ?*
5. *Quels sont les états récurrents, transitoires, absorbants de la chaîne ?*
6. *Imaginons qu'à l'occasion d'une campagne se déroulant sur un salon, doublé d'une campagne de marketing téléphonique, la société sollicite 10 000 clients. Comment peut-on utiliser le modèle pour faire une projection chiffrée à un an du devenir de ces clients ?*
7. *Vérifiez que la distribution* $(0.13, 0.11, 0.04, 0.04, 0.68)$ *est quasiment stationnaire (explicitez le calcul).*
8. *Comment peut-on interpréter ce résultat ?*

Solution

$$M = \begin{bmatrix} 0.1 & 0.3 & 0 & 0 & 0.6 \\ 0.2 & 0 & 0.4 & 0 & 0.4 \\ 0.3 & 0 & 0 & 0.5 & 0.2 \\ 0.4 & 0 & 0 & 0.5 & 0.1 \\ 0.1 & 0.1 & 0 & 0 & 0.8 \end{bmatrix}$$

□

Exercice 11.26 *Le jeu de l'oie est un jeu de société de parcours où l'on déplace des pions en fonction des résultats de deux dés. Traditionnellement, le jeu de l'oie comprend 63 cases disposées en spirale enroulée vers l'intérieur et comportant un certain nombre de pièges. Le but est d'arriver le premier à la dernière case. Ce jeu,* joué à un seul joueur, *peut être modélisé à l'aide d'une chaîne de Markov* $(X_N)_{N\geq 0}$.

1. *Quels sont les états de la chaîne ?*
2. *Quelles sont typiquement les transitions depuis un état de la chaîne ?*
3. *Quels sont ses états récurrents ? transitoires ? absorbants ?*
4. *Quelle est la distribution initiale* X_0 *?*
5. *Cette chaîne admet-elle une distribution stationnaire ? quelle est-elle ?*

Exercice 11.27 *Reprenons l'exemple de la marche aléatoire sur un graphe* $G = (V, E)$ *dont les états sont les sommets de* V*. Considérons la distribution de probabilité* $\pi(u) = \frac{\deg_G(u)}{\sum_{v\in V}\deg_G(v)}$*. En d'autres mots la probabilité* $\pi(u)$ *de se retrouver sur un sommet (un état)* $u \in V$ *est proportionnelle à son degré. Montrez que cette distribution est stationnaire.*

Solution En effet, nous avons calculé à l'exercice 11.8 la matrice de transition dont l'entrée en position ligne/colonne (u, v) est $1/\deg_G(u)$, à condition que V soit un voisin de u, c'est-à-dire lorsque $v \in N_G(u)$. On a donc :

$$\begin{aligned}
(\pi P)_u &= \sum_{v\in N_G(u)} \pi(u)P_{(u,v)} \\
&= \sum_{v\in N_G(u)} \frac{\deg_G(u)}{\sum_{w\in V}\deg_G(w)}\frac{1}{\deg_G(u)} \\
&= \sum_{v\in N_G(u)} \frac{1}{\sum_{w\in V}\deg_G(w)} \\
&= \frac{|N_G(u)|}{\sum_{w\in V}\deg_G(w)} = \frac{\deg_G(u)}{\sum_{w\in V}\deg_G(w)} = \pi(u)
\end{aligned}$$

□

11.3.1 Etats transitoires, récurrents, absorbants

L'exemple de la marche aléatoire sur un graphe $G = (V, E)$ (voir exercice 11.27) est particulier : les états de la chaîne coincident avec l'ensemble des sommets du graphe. La marche aléatoire consiste à se déplacer dans le graphe, comme si un jeton se déplaçait de sommet en sommet en suivant les arêtes du graphe (avec des probabilités induites du degré des sommets). La distribution stationnaire indique la probabilité d'observer le jeton sur l'état $v \in V$ de se trouver sur un état donné (comme si on soulevait un couvercle par surprise pour l'observer).

Mais que se passe t-il plus généralement lorsqu'on se déplace sur un espace d'états d'une chaîne de Markov quelconque ? La question est plus difficile, puisque les probabilités ne sont pas nécessairement liées cette fois au degré des états de la chaîne (si on voit la chaîne comme un graphe). Cette question requière d'examiner de plus près *comment* on se déplace entre les états de cette chaîne. Certains états peuvent avoir un effet décisif sur la capacité de la chaîne à visiter certains sous-espaces d'états.

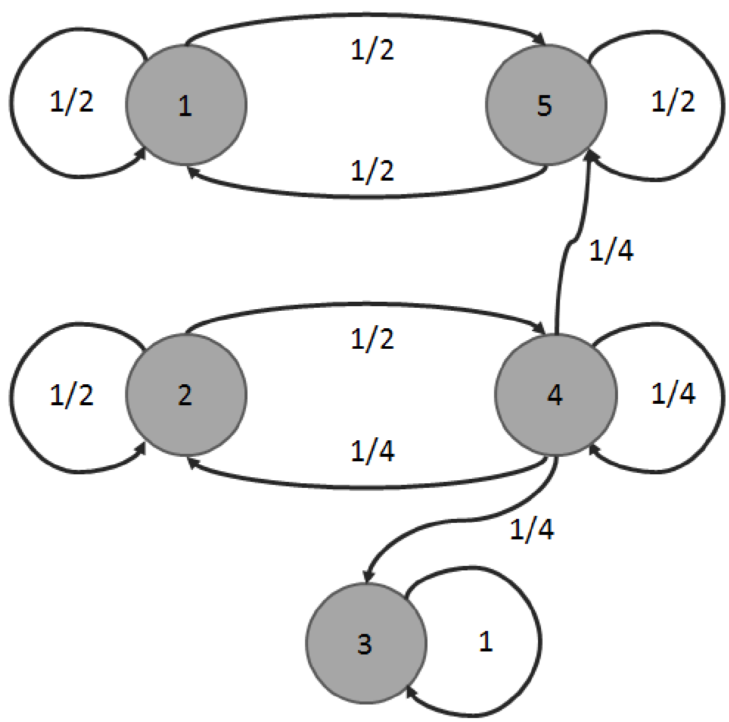


Fig. 11.8 Tous les états de cette chaîne ne sont pas indifféremment accessibles. Par exemple, lorsque l'on quitte l'état 4 pour aller vers l'état 3 ou 5, on ne revient plus jamais vers les états 1 et 2.

Considérons par exemple la chaîne de Markov implicitement définie par le graphe de la figure 11.8. Tous les états de cette chaîne ne sont pas indifféremment accessibles. Par exemple, lorsque l'on quitte l'état 4 pour aller vers l'état 3 ou 5, on ne revient plus jamais vers les états 1 et 2. En d'autres mots, on ne se déplace pas librement parmi *tous* les états de cette chaîne, certains sommets peuvent devenir inaccessible à partir d'un moment, d'autres ont un effet absorbants (comme le sommet 3 de la figure).

11.3.1.1 Temps d'attente

Il faut donc s'intéresser au temps que l'on peut mettre à visiter un état de la chaîne (en termes probabilistes). On définit :

$$T_v = \min\{n > 0 | P(X_n = v) > 0\}$$

lorsque l'ensemble $\{n|P(X_n = v) > 0\}$ n'est pas vide (en d'autres mots, lorsqu'il est possible d'atteindre l'état $v \in V$, et $T_v = \infty$ sinon.

Notation Nous noterons par $P_v(-)$ la probabilité d'évènements définie en termes d'une chaîne de Markov démarrant sur l'état $v \in V$. En d'autres mots, pour laquelle $X_0 = \mathbf{1}_{\{v\}}$ est l'indicatrice (voir définition 1) de l'ensemble réduit à $\{v\}$.

Ainsi, $P_v(T_u = m)$ désigne la probabilité que l'état $u \in V$ soit visité au temps m si l'on démarre depuis l'état $v \in V$. On peut donc écrire :

$$P^n(v, u) = \sum_{m=1}^{n} P_v(T_u = m) P^{n-m}(u, u) \tag{11.11}$$

D'une certaine manière, cette équation partage tous les chemins allant de v à u selon le moment auquel ils visitent le sommet u pour la première fois.

Définition 5. On dit qu'un état $v \in V$ est *absorbant* si $P(v, v) = 1$. En d'autres mots, la chaîne ne quitte plus cet état dès lors qu'elle le visite.

Exercice 11.28 *Le modèle de comportement de la guêpe (exercice 11.17) comporte un état absorbant. La chaîne de Markov des figures 11.5 et 11.8 comportent des états absorbants.*

Considérer, pour $u, v \in V$ donnés, la probabilité $P_v(T_u < \infty)$: c'est la probabilité d'atteindre effectivement l'état u en démarrant depuis v. Cette probabilité – que nous ne calculerons pas – nous permet de distinguer les états les uns des autres. Nous avons en particulier $P_v(T_u < \infty) = 0$ lorsque v est un état absorbant, sauf si $v = u$ auquel cas nous avons trivialement $P_v(T_v < \infty) = 1$ puisque $P(v, v) = 1$.

Définition 6. Il est toutefois possible d'avoir $P_v(T_v < \infty) = 1$ même lorsque v n'est pas un état absorbant. Cela signifie que la probabilité de revenir visiter un état après l'avoir quitté est certaine. On y reviendra même une infinité de fois. On dit alors que l'état $v \in V$ est *récurrent* : on y revient de manière récurrente.

Exercice 11.29 *Les états de la chaîne de Markov à deux états présenté en ouverture de ce chapitre sont tous deux récurrents. La chaîne de Markov de la figure 11.8 comportent des états récurrents.*

Tous les états du modèle de réservation de places de l'exercice 11.13, comme pour celui de l'évolution de la valeur d'un titre en bourse (exercice 11.14) sont récurrents.

Définition 7. Reste le cas où $P_v(T_v < \infty) < 1$, c'est-à-dire que la probabilité de *ne plus jamais revenir en l'état* v après l'avoir quitté n'est pas nulle. Cet état est alors dit *transitoire.*

Exercice 11.30 *Les états correspondant aux pièces où peut se trouver la guêpe (exercice 11.17) sont transitoires. Les deux états en milieu du diagramme de la figure 11.8 sont transitoires.*

Comme on peut s'en douter, le fait que certains états soient absorbants, transitoires ou récurrents est déterminant lorsqu'il s'agit de prédire le comportement d'un système modélisé par une chaîne de Markov. L'existence d'un état absorbant donne lieu à une distribution stationnaire univalué (une distribution dont une composante est égale à 1). Cela n'exclut pas l'existence d'autres distribution stationnaire – puisque le choix des états de départ (de la distribution initiale X_0) importe.

Les états transitoires ont un effet sur les valeurs de la distribution stationnaire, et aussi sur la vitesse de convergence de la chaîne. Typiquement, une marche sur la chaîne de Markov finira par quitter ces états et ne plus y revenir (avec une probabilité non nulle). La dynamique de la chaîne se joue parmi les états récurrents :

Théorème 11.3.1 *L'ensemble des états récurrents* $\mathcal{R}$ *d'une chaîne de Markov se décomposent de manière unique en composantes fortement connexes* $\mathcal{R} = C_1 \cup C_2 \cup \cdots \cup C_k$ *: ce sont des sous-ensembles dans lesquels il est toujours possible d'aller d'un état* $u \in V$ *à un état* $v \in V$ *pour tout choix d'états* $u, v \in C_k$.

Ce théorème ne dit rien de sorcier si on a en tête la représentation d'une chaîne de Markov par un graphe. Les composantes dont il est question dans le théorème s'obtiennent en décomposant le graphe.

Le graphe sous-jacent à une chaîne de Markov est un graphe orienté. Une composante fortement connexe est précisément un ensemble de sommets bien connectés : d'un sommet, on peut trouver un chemin pour aller vers n'importe quel autre.

Définition 8. Il peut arriver parfois que le graphe associé à une chaîne de Markov soit symétrique : pour toute paire d'états $u, v \in V$, le graphe comporte les deux arc $(u, v), (v, u) \in E$ *et ces deux transitions* (arcs) *ont même probabilité.* On dit alors que la chaîne de Markov est *symétrique.* On verra que cela a une incidence sur la distribution stationnaire.

Exercice 11.31 File d'attente *Imaginons des clients se présentant à un comptoir pour être servi (dans un supermarché, par exemple). On se donne un intervalle de temps convenable pour "mesurer" l'arrivée des clients (disons, une minute). Si on est à servir un client, les autres clients arrivant au comptoir forment une* file d'attente.

On suppose qu'à chaque intervalle de temps, s'il y a des clients à servir, on sert un client. Soit $(\xi_n)_{n\geq 0}$ *les variables aléatoires donnant le nombre de clients arrivant au temps* n*. On supposera que les variables* ξ_n *sont à valeurs entières et de même distribution. On désigne par* X_0 *le nombre de clients qui se trouvent au comptoir au départ, et par* X_n *le nombre de clients se trouvant au comptoir au temps* n*. Si* $X_n = 0$ *alors* $X_{n+1} = \xi_{n+1}$ *et si* $X_n \geq 1$ *alors* $X_{n+1} = X_n + \xi_{n+1} - 1$.

- *Décrivez la chaîne de Markov définie par ce processus de file d'attente. Quel est son espace d'états ? Décrivez les probabilités des transitions (en fonction de la distribution des variables* ξ_n*).*
- *Discutez de la nature des états de la chaîne (absorbants, récurrents, transitoires) en fonction de* λ.
- *Simulez cette chaîne, en prenant soin de paramétrer* λ*. Testez différentes valeurs de* λ.

Exercice 11.32 *Un individu vit dans un milieu où il est susceptible d'attraper une maladie par piqûre d'insecte. Il peut être dans l'un des trois états suivants : immunisé (I), malade (M), non malade et non immunisé (S). D'un mois à l'autre, son état peut changer selon les règles suivantes :*

- *étant immunisé, il peut le rester avec une probabilité 0.9 ou passer à l'état S avec une probabilité 0.1 ;*
- *étant dans l'état S, il peut le rester avec une probabilité 0.5 ou passer à l'état M avec une probabilité 0.5 ;*
- *étant malade, il peut le rester avec une probabilité 0.2 ou passer à l'état I avec une probabilité 0.8.*

Tracez un graphe probabiliste pour décrire cette situation et écrivez la matrice de transition. Calculez l'état de probabilité de l'individu au bout de trois mois, de six mois, d'un an, de deux ans, pour chacune des situations suivantes :

- *au départ, il est immunisé (I) ;*
- *au départ, il est non malade et non immunisé (S) ;*
- *au départ, il est malade (M).*

Pouvez-vous donner des éléments sur la proportion d'individus malades dans la population étudiée ?

Exercice 11.33 Parts de marché *Une entreprise* K *possède 25% des parts de marché dans son secteur d'activité. Des enquêtes clientèles indiquent que la part de clients qui lui restent fidèles d'année en année est de 88%, mais que 12% des autres clients passent à la compétition. On estime de plus que 85% des clients des compétiteurs leur restent fidèles, mais que l'entreprise* K *est en mesure d'attirer 15% de ces clients. En faisant l'hypothèse que ces* *tendances se maintiennent, on peut modéliser la situation par une chaîne de Markov :*

- *Décrivez la chaîne de Markov : donnez ses états et ses transitions.*
- *Calculez (calcul probabiliste) la répartition des parts de marché entre K et ses compétiteurs dans 2 ans.*
- *Quelle sera la tendance à long terme ?*

Exercice 11.34 *On dispose, dans une maison individuelle, de deux systèmes de chauffage, l'un de base, et l'autre d'appoint. On dira qu'on est dans l'état 1 si seul le chauffage de base fonctionne, et dans l'état 2 si les deux systèmes fonctionnent. Si un jour on est dans l'état 1, on estime qu'on y reste le lendemain avec une probabilité 1/2 ; en revanche, si on est dans l'état 2, le lendemain la maison est chaude, et l'on passe à l'état 1 avec une probabilité 3/4. Soit X_n l'état du système au jour numéro n.*

- *Expliquez pourquoi $(X_n)_{n\geq 0}$ peut être modélisé par une chaîne de Markov homogène. Quel est son espace d'états ? Déterminer sa matrice de transition et son graphe.*
- *On pose $p_n = P(X_n = 1)$. Déterminer une relation de récurrence entre p_n et p_{n+1}, puis exprimer p_n en fonction de p_0. Que vaut $\lim_{n\to\infty} p_n$?*
- *Sachant qu'on est dans l'état 1 un dimanche, trouver la probabilité d'être dans le même état le dimanche suivant ?*
- *Montrer que si un jour on se trouve dans l'état 1 avec une probabilité 3/5, alors il en est de même tous les jours qui suivent.*
- *Chaque journée dans l'état 1 coûte 1,5€, et dans l'état 2 coûte 2€. Chaque transition de l'état 1 à l'état 2 ou inversement coûte 0,5€. Calculer le coût moyen d'une journée dans la situation précédente.*

Exercice 11.35 Bruit qui court *Une personne pose une question dont la réponse ne peut être que ((oui)) ou ((non)). Cette réponse est donnée à travers N intermédiaires. Chaque intermédiaire a la même probabilité $p \in [0, 1]$ de transmettre le message correctement.*

- *Modélisez la transmission du message par une chaîne de Markov homogène $(X_1, X_2, \ldots)$, dont on précisera la matrice de transition P. Donnez les caractéristiques de cette chaîne : classes de récurrence, périodicité le cas échéant. Qu'en déduisez-vous en ce qui concerne une mesure invariante associée à la chaîne ?*
- *Ecrivez (pseudo-code) une fonction (de N) simulant une trajectoire $(X_1, X_2, \ldots, X_k)$ de la chaîne.*
- *Essayez de voir quelle est la vitesse de convergence de la chaîne. A cet effet, faites varier p, et représentez le nombre d'itérations nécessaires pour chaque p pour ne plus distinguer la loi de X_n (donnée par P_n) de sa loi limite. Vos résultats sont-ils cohérents avec les résultats sur le taux de mixage et le temps de mixage selon lequel cette convergence a lieu avec une vitesse $|1 - 2p_n|$?*

Exercice 11.36 La fin du pétrole *Le patron d'une station essence* SuperPetrole *s'inquiète de l'ouverture d'une station concurrente* Global *tout près de* *son commerce. Bien que 80% des automobilistes s'arrêtent chez lui pour faire le plein, il observe une tendance à la baisse : après une semaine, 25% de ses clients s'arrêtent maintenant chez le concurrent. Décidé à garder sa place, il mène pendant une semaine une campagne publicitaire et observe qu'une personne sur deux abandonnent* Global *pour venir chez lui. Il est décidé à poursuivre son matraquage publicitaire.*

- *Modélisez cette situation par une chaîne de Markov.*
- *En supposant que ces données de sondage restent identique dans les semaines à venir, comment la clientèle se partagera t-elle entre ces deux stations concurrentes dans deux semaines ?*
- *Comment les choses évolueront-elles si les tendances se maintiennent ?*

11.4 Chaînes de Markov et optimisation

Nombre de problèmes n'admettent pas de solution générale, c'est-à-dire sans poser de conditions ou limiter la taille des données d'entrée (la taille maximale du graphe à traiter, par exemple). La théorie de la complexité dresse un tableau précis de cette situation en définissant des classes de problèmes selon l'existenc eou non d'algorithmes pour les résoudre, ou pour vérifier si on en a une solution. Il demeure toutefois important de pouvoir calculer des solutions approchées à ces problèmes à l'aide d'heuristique : des algorithmes qui tentent de trouver une solution optimale sans pouvoir garantir y arriver. Nous abordons dans ce chapitre quelques-uns de ces problèmes et regardons comment les chaînes de Markov permettent d'explorer l'espace des solutions pour “choisir” une solution approchée.

Exercice 11.37 Coloriage *On suppose disposer d'une carte géographique concernant N pays, $S = s_1, ..., s_N$, et d'un* graphe de voisinage *sur S. On note $i \leftrightarrow j$ si i et j partagent une frontière commune (on dit qu'ils sont* voisins*). On veut réaliser un* coloriage *de la carte avec K couleurs $F = c_1, ..., c_K$. Un coloriage est décrit par une affectation $\boldsymbol{x} = (x_1, ..., x_N)$ d'une couleur x_i à chaque pays s_i. On cherche à réaliser un coloriage minimisant :*

$$U(x) = \sum_{i \leftrightarrow j} 1_{x_i = x_j}$$

(où $1_{x_i=x_j}$ est la fonction indicatrice qui vaut 1 si x_i et x_j sont de la même couleur, et 0 sinon. La somme indique donc le nombre de paire de pays voisins qui sont de la même couleur. Un coloriage correct correspond à $U(x) = 0$: des couleurs différentes pour des pays voisins.

Considérons une chaîne de Markov où les états correspondent à tous les coloriages possibles de la carte. Combien y a t-il d'états dans cette chaîne ?

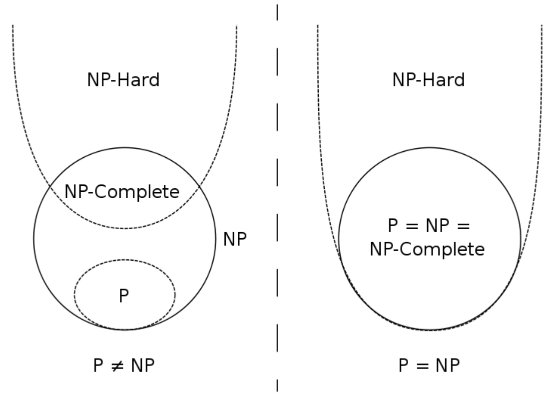


Fig. 11.9 Schéma décrivant la hiérarchie des classes de complexité. Les problèmes exigeant d'être affrontés à l'aide d'heuristiques se trouve dans la zone NP du diagramme.

On considère la fonction de transition suivante : un pays i est choisi au hasard dans S. On propose le changement $x_i \to y_i$, où y_i est choisi au hasard uniformément sur $F \setminus \{x_i\}$ (et il n'y a pas de changement ailleurs).

- *Expliciter cette fonction de transition. (Préciser les valeurs possibles de probabilités de transition.)*
- *Étudier l'irréductibilité de cette chaîne.*

Etant donné un pays x_i on note $N_F(x_i)$ l'ensemble des couleurs des pays voisins de x_i. On modifie maintenant la fonction de transition qui consiste à modifier la couleur $x_i \to y_i$ en choisissant y_i au hasard uniformément parmi $F \setminus N_F(x_i)$.

- *Discuter l'impact de cette modification sur la fonction de transition.*

Cette idée s'étend à tout graphe $G = (V, E)$ pour lequel on cherche à colorier les sommets. Implémentez les deux versions de cette chaîne.

Solution Les états de la chaîne de Markov considérée correspondent aux coloriages possibles de la carte. Un coloriage n'est autre chose qu'une fonction $S \to F$. Or, il existe $|F|^{|S|}$ telles fonctions (ici, K^N puisque $|S| = N$ et $|F| = K$).

Une transition de la chaîne consiste à modifier une composante du vecteur $(x_1, \dots, x_N)$. On choisit i aléatoirement avec probabilité $1/N$, puis une couleur parmi $F \setminus \{x_i\}$ avec probabilité $1/(K-1)$. On a donc $N(K-1)$ transitions possibles toutes avec même probabilité. Notez aussi que la chaîne est symétrique : la probabilité de passer de $\mathbf{x}$ à $\mathbf{y}$ est la même que celle qui permet de passer se $\mathbf{x}$ à $\mathbf{x}$. Dans cette chaîne, il existe toujours un chemin allant d'un coloriage $\mathbf{x}$ à un coloriage $\mathbf{y}$ puisque l'on peut passer de l'un à l'autre en modifiant les composants une à une. Ainsi, tous les états sont récurrents et la chaîne ne compte qu'une seule composante fortement connexe. Par conséquent, la distribution stationnaire de la chaîne est la distribution uniforme. Par conséquent, effectuer une marche aléatoire sur la chaîne suffisamment longtemps revient à choisir un coloriage aléatoire sur le graphe.

La modification apportée à la fonction de transition fait perdre à la chaîne son caractère symétrique. En effet, soit $\mathbf{x} = (x_1, \ldots, x_i, \ldots, x_N)$ un coloriage et $x'_i \in F \setminus N_F(x_i)$ induisant un coloriage $\mathbf{x}'$. Or, on a $x_i \in N_F(x_i)$ dès lors que x_i est de la même couleur que l'un de ses voisins. Le passage de $\mathbf{x}'$ à $\mathbf{x}$ est donc impossible puisqu'on ne peut pas ensuite réaffecté la couleur x_i à s_i. Par ailleurs, il se peut que l'ensemble $F \setminus N_F(s_i)$. La condition proposée ne résoud donc pas cette situation. On pourrait dans ce cas, soit prescrire un retour sur $\mathbf{x}$, avec comme conséquence de faire de $\mathbf{x}$ un état absorbant. Ou alors de modifier aléatoirement la couleur de l'un des pays pour "débloquer" la situation (et visiter d'autres solutions potentielles). $\square$

Exercice 11.38 Voyageur *Un voyageur de commerce doit visiter $N+1$ villes V_1, V_1, …, V_N. Il part de V_1, y revient à la fin de sa tournée, et passe une et une seule fois par chacune de ces villes. On désigne une tournée par une suite $V_1 = x_1 \to x_1 \to \cdots \to x_N \to x_{N+1} = V0$, qui correspond à une permutation de l'ensemble V_1, …, V_N. Le problème classique du voyageur de commerce consiste à trouver un ordre de visite des villes qui minimise la distance parcourue (on suppose connaître les distances $d(V_i, V_j)$ entre les villes) lors d'une tournée $\boldsymbol{x} = (x_1, \ldots, x_N)$:*

$$U(\boldsymbol{x}) = \sum_{i=0}^{N} d(x_i, x_{i+1})$$

On forme une chaîne de Markov sont les états sont les tournées possibles (combien y a t-il d'états ?). On considère trois stratégies pour définir les transitions entre les états.

- Q_1 – Echange de deux villes *: on choisit au hasard deux villes x_i et x_j différentes dans le parcours $\boldsymbol{x}$, et on les échange pour obtenir un nouveau parcours $\boldsymbol{x}'$.*
- Q_2 – Echange de deux villes consécutives *: on choisit au hasard une ville x_i (avec $1 \leq i < N$) et on échange x_i et $x_i{+}1$ pour obtenir un nouveau parcours $\boldsymbol{x}'$.*
- Q_3 – Retournement d'un segment *: on choisit au hasard deux villes différentes x_i et x_j (avec $1 \leq i < j < N$) et on inverse dans le parcours x le segment $x_i \to \cdots \to x_j$ qui mène x_i à x_j pour obtenir un nouveau parcours $\boldsymbol{x}'$.*

Étudier la symétrie et l'irréductibilité de ces transitions. Décrivez comment simuler les transitions Q_i ($i = 1, 2, 3$) avec Q_j ($i \neq j$).

- *On modifie maintenant la chaîne en posant sur les transitions précédentes une condition supplémentaire : on ne passera de x à x' seule-*

ment si $U(\boldsymbol{x}') \leq U(\boldsymbol{x})$. *Discutez de l'impact de cette modification de la chaîne.*

Implémentez cette chaîne de Markov (vous recueillerez préalablement les distances kilométriques qui séparent des villes du territoire français ou européen) et ces trois fonctions de transitions. Comparez les vitesses de convergence.

Solution Les états de la chaîne de Markov sont toutes les listes ordonnées possibles des N villes, qui sont au nombre de $N!$. Nous allons étudier les trois fonctions de transition séparément.

$Q1$ Partant d'une tournée, on peut aller vers $\binom{N}{2}$ autres tournées avec même probabilité $1/\binom{N}{2}$. Cette fonction de transition est symétrique puisqu'on peut revenir en arrière en permutant les deux même villes à nouveau.

$Q2$ Partant d'une tournée, on peut aller vers $N-1$ autres tournées avec même probabilité $1/(N-1)$. Cette fonction de transition est elle aussi symétrique puisqu'il suffit de permuter à nouveau ces villes pour revenir en arrière.

$Q3$ Partant d'une tournée, on peut aller vers $\binom{N}{2}$ autres tournées avec même probabilité $1/\binom{N}{2}$. Cette fonction de transition est symétrique puisqu'on peut revenir en arrière en permutant les deux même villes à nouveau.

Notez que l'enchaînement de transitions permutant deux villes consécutives permet d'aller d'une tournée vers n'importe quelle autre. Or, ce type de transition est présent dans chacun des cas. Les états de la chaîne sont donc récurrents, quel que soit la fonction de transition adoptée, la chaîne est irréductible et la distribution stationnaire est la distribution uniforme. Le choix de la fonction de transition détermine ici la vitesse de convergence de la chaîne (sa capacité à visiter l'ensemble de la chaîne, et de manière uniforme). $\square$

Exercice 11.39 Allocation *On dispose d'un budget global de* L *€ et de* N *demandes de financement de* L_i *€ chacune* ($i = 1, \ldots, N$)*. Comme* $L < L_1 + \cdots + L_N$, *il faudra choisir les demandes à satisfaire. On considère les sous-ensembles* $A \subset \{1, \ldots, N\}$ *tels que* $\sum_{\in A} L_i \leq L$ *et on considère l'objectif* $U(A) = L - \sum_{\in A} L_i$ *qu'il s'agit de minimiser (on veut utiliser une masse budgétaire maximale).*

On considère une chaîne de Markov dont les états sont les sous-ensembles A *et la fonction de transition* $Q(A, B)$ *suivante :*

– *Choisir au hasard uniformément l'entier* i *parmi* $\{1, \ldots, N\}$ *;*
– *Les seuls changements* $A \to B$ *sont*
  (a) *Si* $i \in A$, *on passe à* $B = A \setminus \{i\}$
  (b) *Si* $i \notin A$, *on passe à* $B = A \cup \{i\}$ *quand c'est possible (si* $U(B) \geq 0$*) ; sinon, on reste en* A

Expliciter cette fonction de transition. Montrer que cette fonction de transition est irréductible et symétrique.

Solution Les états de cette chaîne correspondent à des sous-ensembles dont le budget total associé ne dépasse pas B. Le nombre d'états est donc au plus égal à 2^N, et peut vraisemblablement être plus petit selon la contrainte imposé par le budget global B à distribuer.

La fonction de transition permet de passer de A à $A' = A \setminus \{i\}$ (avec $i \in A$) avec probabilité $1/N$. Il permet de passer de A à $A' = A \cup \{i\}$ (avec $i \notin A$) avec probabilité $1/N$ lorsque $U(A') \leq B$. La probabilité de rester en A est égale à k/N où $k = |\{i | U(A \cup \{i\}) > B\}|$ (c'est le nombre d'ajout à A qui font passer au-delà du budget global). Cette fonction est symétrique : si je peux ajouter (enlever) i à A je peux ensuite le retirer (ajouter).

La chaîne est irréductible, puisqu'on peut aller de n'importe quelle solution vers n'importe quelle autre. En effet, on peut toujours construire une solution depuis l'ensemble vide. Tous les états sont donc récurrents et la chaîne est irréductible (elle possède une seule composante fortement connexe). Par conséquent, effectuer une marche aléatoire sur cette chaîne suffisamment longtemps revient à choisir une solution au hasard (sans s'assurer d'être optimal). □

Exercice 11.40 Chaîne de Markov et tirage aléatoire *On considère l'ensemble de tous les graphes simples (non orienté, sans boucle) sur* $\{1, \ldots, N\}$. *On se donne une règle pour passer d'un graphe* $G = (V, E)$ *à un autre* $G' = (V, E')$ *:*

– *On tire au hasard uniformément deux entiers distincts* $1 \leq i, j \leq N$
– *Si* $\{i, j\} \in E$ *alors on supprime l'arête du graphe, en d'autres mots* $E' = E \setminus \{\{i, j\}\}$
– *Si* $\{i, j\} \notin E$ *alors on ajoute l'arête au graphe, en d'autres mots* $E' = E \cup \{\{i, j\}\}$

Quels sont les états de cette chaîne de Markov. Combien compte t-elle d'états ? Sauriez-vous décrire la matrice de transition ? Discutez de la symétrie et de l'irréductibilité de la chaîne.

11.5 Méthodes Monte Carlo et chaînes de Markov

Les méthodes de Monte Carlo sont une alternative souvent efficace aux méthodes numériques (pour calculer ou approximer des fonctions). Elles sont aussi conceptuellement faciles à appréhender et à implémenter. Nous allons exploiter certaines propriétés des chaînes de Markov, vues comme des dispositifs permettant d'explorer des espaces de solutions permettant ainsi de calculer des solutions acceptables à des problèmes d'optimisation complexes.

11.5.1 *Chaînes de Markov et marche aléatoire uniforme*

Soit une chaîne de Markov $(X_i)_{i\geq 0}$ sur un espace d'états V et de matrice de transition M. Soit π une distribution de probabilité sur V. On dira que π (par rapport à M) satisfait la condition d'équilibre si, pour tout états $u, v \in V$ on a :

$$\pi(u)M_{u,v} = \pi(v)M_{v,u} \tag{11.12}$$

Proposition 2. *Soit une distribution π qui satisfait la condition d'équilibre par rapport à une chaîne de Markov $(X_i)_{i\geq 0}$ sur un espace d'états V et de matrice de transition M. Alors π est stationnaire.*

En effet, on calculons les composantes du vecteur πM et montrons qu'il coïncide avec π :

$$\begin{aligned}(\pi M)_u &= \sum_{v\in V} \pi_v M_{v,u} \\ &= \sum_{v\in V} \pi_u M_{u,v} \\ &= \pi_u \sum_{v\in V} M_{u,v} = \pi_u \qquad \square\end{aligned}$$

Corollaire 11.5.1 *Soit une chaîne de Markov $(X_i)_{i\geq 0}$ irréductible et de matrice de transition M symétrique. Alors la distribution uniforme est stationnaire (et unique).*

L'unicité de la distribution vient de l'irréductibilité de la chaîne (son graphe sous-jacent est fortement connexe). On vérifie que la distribution uniforme satisfait la condition d'équilibre. Cela est trivialement vrai puisque les termes $\pi(u)$ et $\pi(v)$ de l'équation (11.12) sont, dans ce cas, égaux. Le résultat suit donc de la symétrie de la matrice de transition, puisqu'alors $M_{u,v} = M_{v,u}$. $\square$

Exemple 11.5.1 *On en déduit une procédure simple pour parcourir un espace de probabilité V et rechercher une solution, dans V, à un problème donné. Marche aléatoire parmi les sous-ensembles d'un ensemble E. De manière équivalente, il s'agit de parcourir l'espace des n-uplets à valeurs* $\{0, 1\}$.

11.5.1.1 Problème du sac à dos

Le problème du sac à dis est célèbre à la fois par sa simplicité et par sa complexité. Il illustre bien le paradoxe typique des problèmes d'optimisation. Toute procédure à caractère déterministe, capable d'approcher la solution optimale dans des cas spécifiques, est voué à l'échec en tant que solution universelle. Le problème peut être formulé comme suit. On doit placer dans un sac à dos, de contenance maximale μ (en poids) finie, des objets dont on connaît le poids et la valeur. Le problème consiste à placer des objets dans le sac en cherchant à maximiser la valeur totale des objets qu'il contient.

On possède donc des objets $u \in T$, dont on connaît le poids $\kappa_u \in \mathbb{R}^+$ et la valeur $\eta_u \in \mathbb{R}^+$. Une solution σ au problème est un sous-ensemble d'objets $U \subset T$ tel que $\sum_{u \in U} \kappa_u \leq \mu$. On cherche à maximiser $\sum_{u \in U} \eta_u$ sous ces conditions.

Approche naïve : parcourir l'ensemble de toutes les solutions, évaluer la valeur du sac et garder la meilleure solution possible. Inconvénient : temps de calcul éventuellement long, dû à la grande taille de l'espace à fouiller. Malgré cet inconvénient, comment procéder pour construire un tel procédé ? Chaîne de Markov, transition symétrique dans l'espace. On choisit $u \in T$ au hasard, si σ contient u, on le retire du sac. S'il ne le contient pas, on l'y met à condition que la contenance le permette, sinon on ne fait rien. Matrice de transition ?

11.5.2 *L'algorithme de Metropolis*

Approche plus maline : on voudrait influer sur la probabilité de trouver une "bonne solution". On aimerait par exemple que la probabilité d'une solution soit liée à sa valeur, qu'elle soit de $\pi(\sigma) = C_\beta^{-1} e^{\beta \sum_{u \in \sigma} \eta_u}$. Il s'agit bien là d'une distribution de probabilité à condition de normaliser les valeurs prises par π, ce que l'on peu faire en ajustant C et β.

Metropolis. Soit Ω un espace de probabilité et M une matrice de transition symétrique sur cet espace. Soit π une distribution telle que $\pi(\omega) > 0$ pour tout $\omega \in \Omega$. On peut définir une chaîne de Markov sont la distribution stationnaire est π.

– On modifie la chaîne : on effectue les transitions selon la matrice M (choix d'un état voisin).
– On accepte la transition $\sigma_1 \to \sigma_2$ avec probabilité $\alpha = \min\{1, \frac{\pi(\sigma_2}{\pi(\sigma_1)}\}$ (soit on va vers le nouvel état, soit on reste sur place).

Remarque 11.5.1 *Nous avons indiqué qu'il nous fallait normaliser les valeurs proposées pour la distribution cible* π*, mais son utilisation montre que la normalisation n'a pas d'effet sur l'algorithme puisque le ratio* $\frac{\pi(\sigma_2}{\pi(\sigma_1)}$ *annule son effet.*

Chapitre 12
Chaînes de Markov à temps continu

Nous allons maintenant nous pencher sur le cas où le temps sous-jacent aux chaînes de Markov est continu. Ce cadre, jumelé aux conditions de Markov, impose nombre de contraintes qu'il nous faut décrire.

On peut voir un processus de Markov comme une variable aléatoire X qui permet de passer d'un état à un autre. Le temps étant continu, cette variable dépend donc du temps t et peut être vu comme une fonction du temps $X(t)$.

$$X(t) = \begin{cases} v_0, & 0 \leq t < \tau_1 \\ v_1, & \tau_1 \leq t < \tau_2 \\ v_2, & \tau_2 \leq t < \tau_3 \\ \vdots \end{cases}$$

Ainsi, le système décrit par la variable $X(t)$ reste dans l'état v_i pendant un certain temps, avant de passer à l'état suivant v_{i+1}. Les temps $\tau_0 = 0$, τ_1, τ_2, ... sont ceux auxquels le système effectue un *saut* vers le prochain état. Le système reste donc dans l'état v_i *pendant* un temps $\tau_{i+1} - \tau_i$.

On ne sait pour l'instant rien du temps pendant lequel le système reste dans un certain état, et celui-ci peut varier. On définit donc, pour chaque état v_i, une fonction de répartition (voir section 8.1) F_{v_i} qui évalue donc la probabilité $F_{v_i}(t) = P(\tau_i < t)$. En d'autres termes, $F_{v_i}(t)$ donne la probabilité que le système reste dans l'état v_i pendant un temps au plus égal à t.

Nous reprenons ici les notations introduites à la section 11.3.1.1, et notons par $P_u(-)$ la probabilité d'un évènement sachant que le système est dans l'état $u \in V$. Par exemple, $P_u(X(t) = v)$ donne la probabilité de se trouver à l'état v au temps t étant donné que l'on se trouvait à l'état u tout juste avant.

On se donne aussi des probabilités de transition m_{uv} qui donne les probabilités de passage de l'état u à un autre état v. On a donc $\sum_{u \in V} m_{uv} = 1$. De plus, on suppose que ces probabilité sont indépendantes du temps, de sorte que $P_v(X(\tau_1) = u \cap \tau_1 < t) = m_{uv}F_v(t)$. En d'autres mots, une fois que le système passe dans l'état u, il se comporte comme s'il avait débuté dans cet état. On peut donc supposé avoir une distribution de probabilité initiale π_0 sur l'ensemble des états de la chaîne ($\sum_v \pi_0(v) = 1$) et on peut donc écrire $P(X(t) = v) = \sum_u \pi_0(u)P_u(X(t) = v)$.

Nous avions résolu le calcul de cette dernière égalité dans le cas des chaînes de Markov discrètes puisqu'alors la probabilité peut être calculée en fonction des temps précédant le temps t. En d'autres mots, on souhaite pouvoir calculer la probabilité $P(X(t_1) = v_1, \ldots, X(t_n) = v_n)$ avec $t_1 < \ldots < t_n$. On est amené – comme dans le cas discret – à poser la *condition de Markov*, pour $0 \leq s_1 \leq \cdots \leq s_n \leq s \leq t$:

$$P(X(t) = v|X(s_1) = u_1, \ldots, X(s_n) = u_n, X(s) = u) = P_u(X(t) = v)$$

En d'autres mots, comme dans le cas discret, l'état v dans lequel passe le système ne dépend que de l'état u qui précède immédiatement et ignore les états $v_1, \ldots, v_n$ qui ont été traversés auparavant.

12.1 Condition de Markov et loi exponentielle

On peut montrer – mais cela sort du cadre de ce cours – que la condition de Markov implique que l'on a nécessairement :

$$P_v(\tau > t + s|\tau > s) = P_v(\tau > t)$$

On peut reformuler cette condition en terme de la fonction de répartition F_v que l'on s'est donné et on trouve alors la condition qui mène à définir la distribution exponentielle (voir section 8.2.4). Les temps pendant lequel la chaîne demeure sur un même état avant de sauter suit donc une loi exponentielle, de fonction de densité f_v (associée à la fonction de répartition F_v) de paramètre q_v :

$$f_v(t) = \begin{cases} q_v e^{-q_v t} & \text{si } t \geq 0 \\ 0 & \text{sinon} \end{cases}$$

et par suite $F_v(t) = 1 - e^{-q_v t}$.

Désignons maintenant par $P_{uv}(t)$ la probabilité que le système se trouve à l'état v au temps t étant donné qu'il démarre à l'état u. En d'autres mots, $P_{uv}(t) = P_u(X(t) = v)$. On a donc aussi, $\sum_{v \in V} P_{uv}(t) = 1$ et $P(X(t) = v) = \sum_{u \in V} \pi_0(u)P_{uv}(t)$. Observons aussi que :

$$P_u(X(t) = w \cap X(t + s) = v) = P_{uw}(t)P_{wv}(t + s)$$

12.1.1 *Backward et forward équations.*

Ce sont deux équations fondamentales de la théorie des chaînes de Markov continues. Il est important de noter que la fonction $P_{uv}(t)$ est bien une fonction réelles d'une variable t, qui est continue et différentiable (nous ne nous attarderons pas sur les bonnes hypothèses et propriétés de cette fonction qui justifie les calculs à suivre). On note $q_{uv} = P'_{uv}(0)$ le *taux de variation* de la probabilité de passage de l'état u à l'état v au temps $t = 0$. Les équations backward et forward sont :

$$P'_{uv}(t) = \sum_{w \in V} q_{uw} P(_{wv}(t) \tag{12.1}$$

$$P'_{uv}(t) = \sum_{w \in V} P(_{uw}(t) q_{wv} \tag{12.2}$$

Nous aurons à manipuler ces équations pour des systèmes particuliers, et auront à résoudre certaines équations différentielles à l'aide du résultat suivant :

12.1.2 *Système à deux états*

On se donne un système à deux états 0 et 1, et on suppose connues les taux de transitions au temps $t = 0$ $q_{01} = \lambda$ et $q_{10} = \mu$. On peut réécrire les équations backward et forward pour ce système.

$$P'_{00}(t) = -\lambda P_{00}(t) + \lambda P_{10}(t) \tag{12.3}$$

$$P'_{10}(t) = \mu P_{00}(t) - \mu P_{10}(t) \tag{12.4}$$

d'où on obtient :

$$\frac{d}{dt}(P_{00}(t) - P_{10}(t)) = -(\lambda + \mu)(P_{00}(t) - P_{10}(t))$$

Or, l'unique fonction solution de l'équation

$$\frac{d}{dt} f(t) = c f(t) \tag{12.5}$$

est la fonction exponentielle $f(t) = e^{ct}$ (à une constante multiplicative près). Ainsi, la fonction $P_{00}(t) - P_{10}(t) = e^{-(\lambda+\mu)t}$ est la solution à notre équation. Par suite, on trouve :

$$P'_{00}(t) = -\lambda e^{-(\lambda+\mu)t}$$

et donc :

$$
\begin{aligned}
P_{00}(t) &= P_{00}(0) + \int_0^t P'_{00}(s)ds \\
&= 1 - \int_0^t \lambda e^{-(\lambda+\mu)s} ds \\
&= 1 - \frac{\lambda}{\lambda+\mu}(1 - e^{-(\lambda+\mu)t}) \\
&= \frac{\mu}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu} e^{-(\lambda+\mu)t}
\end{aligned}
$$

On utilise (12.3) pour trouver

$$
P_{10}(t) = \frac{\mu}{\lambda+\mu} + \frac{\mu}{\lambda+\mu} e^{-(\lambda+\mu)t}
$$

Un calcul similaire permet d'obtenir :

$$
\begin{aligned}
P_{01}(t) &= \frac{\lambda}{\lambda+\mu} - \frac{\lambda}{\lambda+\mu} e^{-(\lambda+\mu)t} \\
P_{11}(t) &= \frac{\lambda}{\lambda+\mu} + \frac{\mu}{\lambda+\mu} e^{-(\lambda+\mu)t}
\end{aligned}
$$

On a :

$$
\begin{aligned}
P(X(t)=0) &= \pi_0(0)P_{00}(t) + \pi_0(1)P_{10}(t) \\
&= \frac{\mu}{\lambda+\mu} + (\pi_0(0) - \frac{\mu}{\lambda+\mu}) e^{-(\lambda+\mu)t} \\
P(X(t)=1) &= \pi_0(0)P_{01}(t) + \pi_0(1)P_{11}(t) \\
&= \frac{\lambda}{\lambda+\mu} + (\pi_0(1) - \frac{\lambda}{\lambda+\mu}) e^{-(\lambda+\mu)t}
\end{aligned}
$$

et lorsque $t \to +\infty$, on trouve :

$$
P(X(t)=0) = \frac{\mu}{\lambda+\mu}, \qquad P(X(t)=1) = \frac{\lambda}{\lambda+\mu}
$$

que l'on peut comparer à la distribution stationnaire du système discret homologue de celui-ci.

12.2 Description matricielle

Revenons au système continu à deux états et observons que l'on peut le décrire de manière matricielle :

$$\begin{bmatrix} P'_{00}(t) & P'_{01}(t) \\ P'_{10}(t) & P'_{11}(t) \end{bmatrix} = \begin{bmatrix} -\lambda & \lambda \\ \mu & -\mu \end{bmatrix} \begin{bmatrix} P_{00}(t) & P_{01}(t) \\ P_{10}(t) & P_{11}(t) \end{bmatrix}$$

En d'autres mots, on a une matrice $\mathbb{P}(t)$ dont les entrées sont des fonctions qui dépendent d'une variable t, et une matrice $\mathbb{M}$ de transition faisant intervenir les quantités q_{uv} liées par l'équation matricielle $\frac{d}{dt}\mathbb{P}(t) = \mathbb{M}\mathbb{P}(t)$. De manière analogue à l'équation différentielle (12.5), l'unique solution à notre système d'équation différentielles est :

$$\mathbb{P}(t) = \mathbb{P}(0)e^{\mathbb{M}t}$$

où l'on doit préciser le calcul de l'exponentielle d'une matrice $e^{\mathbb{A}}$. Notez déjà que dans notre cas, la matrice $\mathbb{P}(0)$ est la matrice identité en vertu de la définition même des fonctions $P_{uv}(t)$, et que la solution cherchée est simplement $\mathbb{P}(t) = e^{\mathbb{M}t}$.

Le cas général suit des équations forward (12.2) (respectivement backward (12.1)) $[P'_{uv}(t)] = [q_{uv}]\,[P'_{uv}(t)]$, ou de manière matricielle $\mathbb{P}'(t) = \mathbb{M}\mathbb{P}(t)$.

La solution à cette équation passe donc par le calcul de l'exponentielle $e^{\mathbb{M}t}$. Or, on obtient le résultat de calcul de l'exponentielle $e^{\mathbb{A}}$ en passant par le développement en série de Taylor de la fonction exponentielle :

$$e^{\mathbb{A}} = \sum_{k=0}^{\infty} \frac{A^k}{k!}$$

et exige donc de calculer les puissances successives de la matrice. Observons que ce calcul est relativement simple lorsque la matrice $\mathbb{A}$ est diagonale puisqu'alors on a :

$$\exp(\begin{bmatrix} \lambda_1 & \cdots & \\ & \ddots & \\ & \cdots & \lambda_k \end{bmatrix}) = \begin{bmatrix} e^{\lambda_1} & \cdots & \\ & \ddots & \\ & \cdots & e^{\lambda_k} \end{bmatrix}$$

La solution à notre problème consiste donc à trouver une forme diagonale pour la matrice $\mathbb{A}$. Cette forme s'obtient par une transformation de la matrice de départ par conjugaison. Il nous faut trouver une matrice $\mathbb{U}$ et son inverse $\mathbb{U}^{-1}$ de manière à avoir :

$$\mathbb{U}^{-1}\mathbb{A}\mathbb{U} = \begin{bmatrix} \lambda_1 & \cdots & \\ & \ddots & \\ & \cdots & \lambda_k \end{bmatrix}$$

ou de manière équivalente $\mathbb{A} = \mathbb{U}\mathbb{D}\mathbb{U}^{-1}$ (où $\mathbb{D}$ désigne la matrice diagonale). Cela résout le calcul de la puissance de la matrice $\mathbb{A}$ puisqu'alors $\mathbb{A}^n = (\mathbb{U}\mathbb{D}\mathbb{U}^{-1})^n = \mathbb{U}\mathbb{D}^n\mathbb{U}^{-1}$. Par suite, on en déduit $e^{\mathbb{A}} = \mathbb{U}e^{\mathbb{D}}\mathbb{U}^{-1}$.

12.2.1 Le cas à deux états

Exercice 12.1 *On revient à la description matricielle du système à deux états.*

– *On considère la matrice :*

$$\mathbb{U} = \begin{bmatrix} -\frac{\lambda}{\mu} & 1 \\ 1 & 1 \end{bmatrix}$$

Calculez la matrice inverse $\mathbb{U}^{-1}$. *En d'autres trouvez les valeurs* a, b, c, d *telle que :*

$$\begin{bmatrix} -\frac{\lambda}{\mu} & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

– *Vérifiez que la matrice :*

$$\mathbb{U}^{-1} \begin{bmatrix} -\lambda & \lambda \\ \mu & -\mu \end{bmatrix} \mathbb{U}$$

est une matrice diagonale $\mathbb{D}$.

– *Déduisez-en les fonctions* $P_{uv}(t)$.

– *Déduisez ensuite les fonctions* $P(X(t) = u)$ *(pour* $u = 0, 1$*) fonction d'une distribution initiale* π_0.

Solution On a :

$$\mathbb{U}^{-1} = \begin{bmatrix} -\frac{\mu}{\lambda+\mu} & \frac{\mu}{\lambda+\mu} \\ \frac{\mu}{\lambda+\mu} & \frac{\lambda}{\lambda+\mu} \end{bmatrix}$$

$$\mathbb{U}^{-1} \begin{bmatrix} -\lambda & \lambda \\ \mu & -\mu \end{bmatrix} \mathbb{U} = \begin{bmatrix} -(\lambda+\mu) & 0 \\ 0 & 0 \end{bmatrix} = \mathbb{D}$$

On en déduit :

$$\mathbb{P}(t) = e^{\begin{bmatrix} -\lambda & \lambda \\ \mu & -\mu \end{bmatrix} t} = e^{(\mathbb{U}\mathbb{D}\mathbb{U}^{-1})t} = \mathbb{U}e^{\mathbb{D}t}\mathbb{U}^{-1} = \begin{bmatrix} -\lambda & \lambda \\ \mu & -\mu \end{bmatrix} \begin{bmatrix} e^{-[\lambda+\mu)} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -\frac{\mu}{\lambda+\mu} & \frac{\mu}{\lambda+\mu} \\ \frac{\mu}{\lambda+\mu} & \frac{\lambda}{\lambda+\mu} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\lambda e^{-(\mu+\lambda)t}}{\mu+\lambda} & -\frac{\lambda e^{-(\mu+\lambda)t}}{\mu+\lambda} \\ -\frac{\mu e^{-(\mu+\lambda)t}}{\mu+\lambda} & \frac{\mu e^{-(\mu+\lambda)t}}{\mu+\lambda} \end{bmatrix}$$

Notez que la solution à l'équation différentielle est unique à une constante près, qu'il nous faut ajuster ici de manière à respecter la condition $P_{uv}(0) = 1$. Notre solution est donc :

$$\mathbb{P}(t) = \begin{bmatrix} \frac{\mu}{\mu+\lambda} + \frac{\lambda e^{-(\mu+\lambda)t}}{\mu+\lambda} & \frac{\lambda}{\mu+\lambda} - \frac{\lambda e^{-(\mu+\lambda)t}}{\mu+\lambda} \\ \frac{\mu}{\mu+\lambda} - \frac{\mu e^{-(\mu+\lambda)t}}{\mu+\lambda} & \frac{\lambda}{\mu+\lambda} + \frac{\mu e^{-(\mu+\lambda)t}}{\mu+\lambda} \end{bmatrix}$$

qui nous donne, en fonction d'une distribution initiale π_0 :

$$\begin{aligned} P(X(t) = 0) &= \pi_0(0)P_{00}(t) + \pi_0(1)P_{10}(t) \\ &= \pi_0(0)(\frac{\mu}{\mu+\lambda} + \frac{\lambda e^{-(\mu+\lambda)t}}{\mu+\lambda}) + \pi_0(1)(\frac{\mu}{\mu+\lambda} - \frac{\mu e^{-(\mu+\lambda)t}}{\mu+\lambda}) \\ P(X(t) = 1) &= \pi_0(0)P_{01}(t) + \pi_0(1)P_{11}(t) \\ &= \pi_0(0)(\frac{\lambda}{\mu+\lambda} - \frac{\lambda e^{-(\mu+\lambda)t}}{\mu+\lambda}) + \pi_0(1)(\frac{\lambda}{\mu+\lambda} + \frac{\mu e^{-(\mu+\lambda)t}}{\mu+\lambda}) \end{aligned}$$

□

12.3 Une chaîne continue à trois états

On considère maintenant une chaîne continue à trois états dont le graphe sous-jacent est un graphe complet (un peu à l'image de l'exercice 11.14 décrivant un modèle naïf des mouvements des cours en bourses).

Exercice 12.2 – *Donnez le système matriciel (continu) décrivant ce système donnez la matrice $\mathbb{M}$). On supposera pour l'instant les valeurs q_{uv} indéterminée et on fixera les probabilités de transition entre les états à 1/2.*

- *On fixe maintenant les valeurs q_{uu} à 1 (ces paramètres sont les paramètres des lois exponentielles sous-jacentes aux états de la chaîne). Donnez la matrice de transition.*
- *On considère la matrice :*

$$\mathbb{U} = \begin{bmatrix} 1 & -1 & -1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

Montrez (par simple calcul) que son inverse est la matrice :

$$\begin{bmatrix} 1/3 & 1/3 & 1/3 \\ -1/3 & -1/3 & 2/3 \\ -1/3 & 2/3 & -1/3 \end{bmatrix}$$

– *Montrez que la matrice* $\mathbb{U}^{-1}\mathbb{M}\mathbb{U}$ *est une matrice diagonale* $\mathbb{D}$.
– *Déduisez-en la matrice* $\mathbb{P}(t)$.
– *Calculez les fonctions* $P(X(t) = u)$ *pour les trois états* u *de la chaîne.*

Solution On a :

$$\frac{d}{dt}\mathbb{P}(t) = \begin{bmatrix} -1 & 1/2 & 1/2 \\ 1/2 & -1 & 1/2 \\ 1/2 & 1/2 & -1 \end{bmatrix} \mathbb{P}(t)$$

On a :

$$\mathbb{U}^{-1}U = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -3/2 & 0 \\ 0 & 0 & -3/2 \end{bmatrix}$$

On en déduit la solution :

$$\mathbb{P}(t) = \mathbb{U}e^{\mathbb{D}t}\mathbb{U}^{-1} = \begin{bmatrix} 1/3 + 2/3\,e^{-3/2\,t} & 1/3 - 1/3\,e^{-3/2\,t} & 1/3 - 1/3\,e^{-3/2\,t} \\ 1/3 - 1/3\,e^{-3/2\,t} & 1/3 + 2/3\,e^{-3/2\,t} & 1/3 - 1/3\,e^{-3/2\,t} \\ 1/3 - 1/3\,e^{-3/2\,t} & 1/3 - 1/3\,e^{-3/2\,t} & 1/3 + 2/3\,e^{-3/2\,t} \end{bmatrix}$$

On calcule les fonctions $P(X(t) = u)$ (pour $u = 0, 1, 2$) en fonction d'une distribution initiale π_0 :

$$\begin{aligned} P(X(t) = 0) &= \pi_0(0)P_{00}(t) + \pi_0(1)P_{10}(t) + \pi_0(2)P_{20}(t) \\ &= \pi_0(0)(1/3 + 2/3\,e^{-3/2\,t}) + \pi_0(1)(1/3 - 1/3\,e^{-3/2\,t}) + \pi_0(2)(1/3 - 1/3\,e^{-3/2\,t}) \\ &= 1/3 + (2/3\pi_0(0) - 1/3\pi_0(1) - 1/3\pi_0(2))e^{-3/2t} \\ &\quad \dots \end{aligned}$$

$\square$