

Sylvie Méléard

Modèles aléatoires en Ecologie et Evolution



Mathématiques et Applications

Directeurs de la collection:
J. Garnier et V. Perrier

77

MATHÉMATIQUES & APPLICATIONS

Comité de Lecture 2012–2015/Editorial Board 2012–2015

Rémi ABGRALL
Inst. Math., University of Zurich, CH
remi.abgrall@math.uzh.ch

Grégoire ALLAIRE
CMAP, École Polytechnique, Palaiseau, FR
gregoire.allaire@polytechnique.fr

Michel BENAÏM
Inst. Math., Univ. de Neuchâtel, CH
michel.benaim@unine.ch

Maitine BERGOUNIOUX
MAPMO, Université d'Orléans, FR
maitine.bergounioux@univ-orleans.fr

Thierry COLIN
Inst. Math., Université Bordeaux 1, FR
colin@math.u-bordeaux1.fr

Marie-Christine COSTA
UMA, ENSTA, Paris, FR
marie-christine.costa@ensta.fr

Arnaud DEBUSSCHE
ENS Cachan, Bruz, FR
arnaud.debussche@bretagne.ens-cachan.fr

Isabelle GALLAGHER
Inst. Math. Jussieu, Univ. Paris 7, FR
gallagher@math.jussieu.fr

Josselin GARNIER
Lab. Proba. et Mod. Aléatoires, Univ. Paris 7, FR
garnier@math.univ-paris-diderot.fr

Stéphane GAUBERT
INRIA, École Polytechnique, Palaiseau, FR
stephane.gaubert@inria.fr

Emmanuel GOBET
CMAP, École Polytechnique, Palaiseau, FR
emmanuel.gobet@polytechnique.edu

Raphaèle HERBIN
CMI LATP, Université d'Aix-Marseille, FR
raphaele.herbin@latp.univ-mrs.fr

Marc HOFFMANN
CEREMADE, Université Paris-Dauphine, FR
hoffmann@ceremade.dauphine.fr

Claude LE BRIS
CERMICS, ENPC, Marne la Vallée, FR
lebris@cermics.enpc.fr

Sylvie MÉLÉARD
CMAP, École Polytechnique, Palaiseau, FR
sylvie.meleard@polytechnique.edu

Felix OTTO
MPI MIS Leipzig, GE
felix.otto@mis.mpg.de

Valérie PERRIER
Lab. Jean-Kunzmann, ENSIMAG, Grenoble, FR
valerie.perrier@imag.fr

Philippe ROBERT
INRIA Rocquencourt, Le Chesnay, FR
philippe.robert@inria.fr

Pierre ROUCHON
Automatique et Systèmes, École Mines, Paris, FR
pierre.rouchon@ensmp.fr

Bruno SALVY
INRIA, LIP - ENS Lyon, FR
bruno.salvy@inria.fr

Annick SARTENAER
Dépt. Mathématiques, Univ. Namur, Namur, BE
annick.sartenaer@fundp.ac.be

Eric SONNENDRÜCKER
MPI für Plasmaphysik, Garching, GE
eric.sonnendruecker@ipp.mpg.de

Alain TROUVÉ
CMLA, ENS Cachan, FR
trouve@cmla.ens-cachan.fr

Cédric VILLANI
IHP, Paris, FR
villani@math.univ-lyon1.fr

Enrique ZUAZUA
UAM, Madrid, ES
enrique.zuazua@uam.es

Directeurs de la collection:
J. GARNIER et V. PERRIER

Sylvie Méléard

Modèles aléatoires en Ecologie et Evolution

Sylvie Méléard
École Polytechnique CMAP
Palaiseau Cedex
France

ISSN 1154-483X ISSN 2198-3275 (electronic)
Mathématiques et Applications
ISBN 978-3-662-49454-7 ISBN 978-3-662-49455-4 (eBook)
DOI 10.1007/978-3-662-49455-4

Library of Congress Control Number: 2016934437

Mathematics Subject Classification (2010): 60J10, 60J70, 60H10, 60J75, 60J80, 60J85, 92D15, 92D25, 92D40

© Springer-Verlag Berlin Heidelberg 2016

Tous droits de traduction, de reproduction et d'adaptation réservés pour tous pays.

La loi du 11 mars 1957 interdit les copies ou les reproductions destinées à une utilisation collective. Toute représentation, reproduction intégrale ou partielle faite par quelque procédé que ce soit, sans le consentement de l'auteur ou de ses ayants cause, est illicite et constitue une contrefaçon sanctionnée par les articles 425 et suivants du Code pénal.

Imprimé sur papier non acide

This Springer imprint is published by Springer Nature
The registered company is Springer-Verlag GmbH Berlin Heidelberg

En mémoire de ma mère, biologiste passionnée,

A Pierre-Louis

Ce livre s'adresse aux élèves de master en mathématiques appliquées ou aux biologiste théoriciens qui souhaitent étoffer leur connaissance des outils probabilistes de modélisation. Il est né d'un cours donné aux élèves de troisième année de l'Ecole Polytechnique mais il a au fil des années pris une ampleur qui dépasse largement le cours. Son but est de donner au lecteur des outils rigoureux permettant la modélisation de phénomènes biologiques soumis à fluctuations aléatoires. Il se focalise sur les modèles stochastiques construits à partir des comportements individuels. Qu'il me soit permis de remercier très chaleureusement ceux qui ont contribué, par leur enthousiasme, leur passion pour ce domaine scientifique et de nombreuses discussions éclairantes, à la genèse de cet ouvrage : Vincent Bansaye, Sylvain Billiard, Nicolas Champagnat, Pierre Collet, Camille Coron, Régis Ferrière, Christophe Giraud, Pierre-Henri Gouyon, Carl Graham, Amaury Lambert, Chi Viet Tran, Amandine Véber.

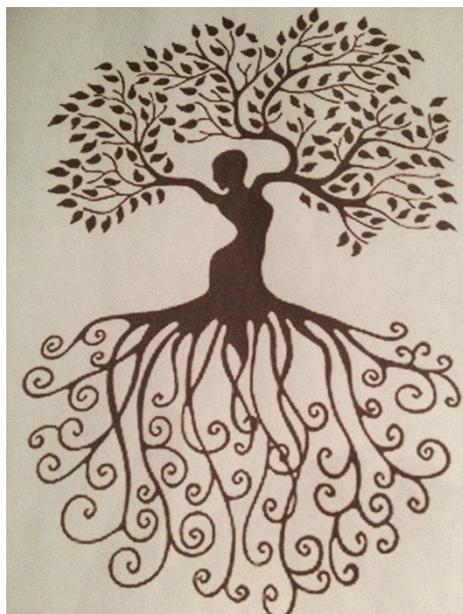


Table des matières

1	Introduction	1
1.1	Introduction du cours	2
1.2	Importance de la modélisation	3
1.3	Modélisation mathématique en écologie et évolution	4
1.3.1	Ecologie et évolution	4
1.3.2	Dispersion et colonisation	5
1.3.3	Dynamique des populations	6
1.3.4	Génétique des populations	6
1.3.5	Quelques termes de vocabulaire	7
2	Populations spatiales et temps discret	9
2.1	Marches aléatoires et chaînes de Markov	9
2.2	Etude des temps de passage	15
2.2.1	Temps d'arrêt et propriété de Markov forte	15
2.2.2	Loi du premier temps de retour en 0 d'une marche aléatoire simple	17
2.3	Réurrence et transience - Théorèmes ergodiques	20
2.4	Marches aléatoires absorbées ou réfléchies	28
2.4.1	Barrières absorbantes	28
2.4.2	Barrières réfléchissantes	30
2.5	Martingales à temps discret	31
2.6	Exercices	35
3	Dynamique de population en temps discret	39
3.1	Chaînes de Markov de vie et de mort	40
3.2	Le processus de Bienaymé-Galton-Watson	43
3.2.1	Définition	43
3.2.2	Résultats élémentaires	45
3.2.3	Le comportement à l'infini	46

3.2.4	Cas sous-critique : Analyse fine de l'extinction	53
3.3	Relation entre processus de Bienaymé-Galton-Watson et modèle de généalogie.	57
3.4	Comportement quasi-stationnaire	58
3.4.1	Distribution quasi-stationnaire et limite de Yaglom	58
3.4.2	Distributions quasi-stationnaires pour une chaîne de BGW	61
3.5	Extension 1 : Les chaînes densité-dépendantes	64
3.6	Extension 2 : Chaîne de BGW avec immigration	66
3.7	Extension 3 : le processus de BGW multitype	69
3.8	Exercices	76
4	Mouvement brownien et processus de diffusion	81
4.1	Convergence fini-dimensionnelle de marches aléatoires renormalisées	82
4.2	Processus aléatoire et mouvement brownien	85
4.3	Quelques propriétés du mouvement brownien	89
4.4	Propriété de Markov et mouvement brownien	93
4.5	Martingales à temps continu et temps d'arrêt	96
4.5.1	Martingales à temps continu	96
4.5.2	Inégalités fondamentales et comportement à l'infini	99
4.5.3	Le théorème d'arrêt	102
4.5.4	Applications au mouvement brownien	105
4.6	Intégrales stochastiques et EDS	110
4.6.1	Intégrales stochastiques	111
4.6.2	Equations différentielles stochastiques (EDS)	115
4.6.3	Système différentiel stochastique	120
4.6.4	Propriété de Markov d'une solution d'EDS	121
4.6.5	Formule d'Itô	122
4.6.6	Générateur - Lien avec les équations aux dérivées partielles	125
4.6.7	Applications aux temps d'atteinte de barrières	128
4.7	Equations différentielles stochastiques pour l'étude des populations	130
4.7.1	Equation de Feller	130
4.7.2	Equation de Feller logistique	133
4.7.3	Processus de Ornstein-Uhlenbeck	135
4.7.4	Autres exemples de déplacements spatiaux	136
4.7.5	Processus de Wright-Fisher	137
4.8	Exercices	138

5	Processus de population en temps continu	143
5.1	Processus markovien de saut	144
5.2	Un prototype : le processus de Poisson	147
5.2.1	Définition d'un processus de Poisson	147
5.2.2	Propriété de Markov forte	151
5.2.3	Comportement asymptotique d'un processus de Poisson	153
5.2.4	Processus de Poisson composé	155
5.3	Générateur d'un processus markovien de saut	156
5.3.1	Le générateur infinitésimal	156
5.3.2	Chaîne de Markov incluse	160
5.4	Processus de branchement en temps continu	162
5.4.1	Définition et propriété de branchement	162
5.4.2	Equation pour la fonction génératrice	165
5.4.3	Critère de non-explosion	166
5.4.4	Equation de moments - Probabilité et temps d'extinction	167
5.4.5	Le cas binaire	168
5.4.6	Extensions	170
5.5	Processus de naissance et mort	171
5.5.1	Définition et critère de non-explosion	171
5.5.2	Equations de Kolmogorov et mesure invariante	175
5.5.3	Critère d'extinction - Temps d'extinction	176
5.6	Approximations continues	181
5.6.1	Approximations déterministes - Equations malthusienne et logistique	181
5.6.2	Approximation stochastique - Stochasticité démographique, Equ-	
	ation de Feller	185
5.6.3	Les modèles de proie-prédateur, système de Lotka-Volterra	188
5.7	Exercices	189
6	Processus d'évolution génétique	203
6.1	Un modèle idéalisé de population infinie : le modèle de Hardy-Weinberg . .	203
6.2	Population finie : le modèle de Wright-Fisher	204
6.2.1	Processus de Wright-Fisher	204
6.2.2	Distribution quasi-stationnaire pour un processus de Wright-Fisher	211
6.2.3	Processus de Wright-Fisher avec mutation	213
6.2.4	Processus de Wright-Fisher avec sélection	214
6.3	Modèles démographiques de diffusion	215

- 6.3.1 Diffusion de Wright-Fisher 215
- 6.3.2 Diffusion de Wright-Fisher avec mutation ou sélection 216
- 6.3.3 Autre changement d'échelle de temps 219
- 6.4 La coalescence : description des généalogies 219
 - 6.4.1 Asymptotique quand N tend vers l'infini : le coalescent de Kingman 220
 - 6.4.2 Le coalescent avec mutation 226
 - 6.4.3 Loi du nombre d'allèles distincts, formule d'Ewens 229
 - 6.4.4 Le point de vue processus de branchement avec immigration 233
- 6.5 Exercices 234

- 7 Quelques développements modernes en Ecologie-Evolution 237**
 - 7.1 Survie et croissance de métapopulations réparties sur un graphe 237
 - 7.1.1 Première approche : processus de Galton-Watson multitype 237
 - 7.1.2 Deuxième approche - Chaîne de Markov sur un graphe 238
 - 7.2 Abondance en environnement aléatoire 241
 - 7.3 Etude de l'invasion d'un mutant dans une grande population résidente à l'équilibre 243
 - 7.4 Un modèle stochastique pour l'auto-incompatibilité des plantes à fleurs . . . 245
 - 7.5 Modélisation d'une population diploïde 249
 - 7.5.1 Le processus de naissance et mort multitype 249
 - 7.5.2 Approximations en grande population 252
 - 7.6 Arbres généalogiques de populations sexuées 256
 - 7.6.1 Modèle de Wright-Fisher diploïde avec recombinaison. 257
 - 7.6.2 Nombre de générations pour arriver à l'ancêtre commun 261

- Bibliographie 265**

Chapitre 1

Introduction

After years, I have deeply regretted that I did not proceed far enough at least to understand something of the great leading principles of mathematics : for men thus endowed seem to have an extra-sense. (Darwin, Autobiography).

Darwin, dans son fameux ouvrage "*De l'origine des espèces au moyen de la sélection naturelle, ou la Préservation des races favorisées dans la lutte pour la vie*", paru en 1859, révolutionne la biologie par sa théorie de l'évolution et de la sélection naturelle. En voici l'essence, extraite de son livre : "*Comme il naît beaucoup plus d'individus de chaque espèce qu'il n'en peut survivre et que par conséquent il se produit souvent une lutte pour la vie, il s'ensuit que tout être qui varie, même légèrement, d'une façon qui lui est profitable, dans les conditions complexes et quelquefois variables de la vie, a une plus grande chance de survivre. Cet être est ainsi l'objet d'une sélection naturelle. En vertu du principe si puissant de l'hérédité, toute variété ainsi choisie aura tendance à se multiplier sous sa nouvelle forme modifiée.*"

Dans ce livre, nous n'avons pas la prétention d'aborder dans sa globalité la modélisation de ces phénomènes complexes mais plutôt celle de donner les outils mathématiques de base pour la modélisation de chaque étape, telles des briques apportant chacune leur modeste contribution à l'édifice.

1.1 Introduction du cours

People who have mathematical computation and statistical skills, and establish a collaboration and work on real biological problems, have the chance of doing some very, very significant things for human welfare. (Jaroslav Stark, Imperial College -Science, 2004)

La biologie des populations va d'études très microscopiques, comme la recherche de séquences sur un brin d'ADN, l'étude des échanges moléculaires dans une cellule, l'évolution de tumeurs cancéreuses, l'invasion de parasites dans une cellule, à des problèmes beaucoup plus macroscopiques concernant des comportements de grands groupes d'individus et leurs interactions (extinction de populations, équilibre des écosystèmes, invasion d'une population par une autre, métapopulations), ou des problèmes de génétique de populations (recherche d'ancêtres communs à plusieurs individus dans une espèce, phylogénies). A tous les niveaux, l'aléatoire intervient, et les modèles stochastiques sont utilisés pour décrire des phénomènes biologiques à chaque échelle du vivant. Même si la population semble présenter un certain nombre de caractéristiques déterministes, elle est composée d'individus dont le comportement peut être soumis à une grande variabilité. Ainsi, chaque individu se déplace dans une direction différente, chaque bactérie a son propre mécanisme de division cellulaire, chaque réplication de l'ADN peut engendrer une mutation. Cette variabilité individuelle est une idée fondamentale de la biologie évolutive et en particulier de Darwin. En effet, même si dans de nombreuses situations, la taille de la population peut être suffisamment grande pour que le système biologique puisse être résumé par un modèle déterministe, dans la plupart des cas, le phénomène biologique, au moins dans une phase de son développement, est déterminé par un petit nombre d'individus ou est fortement influencé par les fluctuations environnementales, spatiales ou temporelles. Par exemple, les fluctuations aléatoires peuvent être déterminantes au stade initial d'une épidémie, dans le développement d'un processus cellulaire ou en cas d'apparition d'un individu mutant. La démarche du probabiliste consiste à décrire les comportements individuels et à en déduire des informations au niveau de la population. Cela permet ainsi, à partir d'une description microscopique précise, de prédire des comportements macroscopiques de manière rigoureuse, en prenant en compte les fluctuations aléatoires.

Le but de ce livre est de définir et étudier une grande gamme d'outils probabilistes qui peuvent apporter une meilleure compréhension de certains phénomènes en biologie des populations. L'hypothèse fondamentale des modèles introduits est que la population a un comportement *markovien* : son comportement aléatoire dans le futur ne dépendra de son passé que par l'information que donne son état présent. Cette hypothèse est largement admise par les biologistes, même si c'est une approximation de la réalité. La dynamique de la population est décrite par un processus stochastique, c'est-à-dire une fonction aléatoire du temps. Le temps peut être discret ($n \in \mathbb{N}$). Il décrira alors la succession des différentes générations ou des reproductions périodiques, saisonnières ou annuelles ou bien il sera une discrétisation du temps continu. Les outils probabilistes de base seront alors les

chaînes de Markov à temps discret. Trois modèles classiques illustreront ce propos : les marches aléatoires pour décrire les déplacements spatiaux d'un individu, les processus de Bienaymé-Galton-Watson qui modélisent la dynamique d'une population, le processus de Wright-Fisher qui décrit une généalogie. Mais le temps peut aussi être le temps physique, un temps $t \in \mathbb{R}_+$. Nous serons alors amenés à considérer des processus en temps continu, à savoir des familles de variables aléatoires indexées par le temps continu. Les processus décrivant les comportements individuels seront alors, soit des processus continus du temps modélisant par exemple le mouvement spatial et désordonné de petits organismes, dont le prototype est le mouvement brownien, soit des processus discontinus décrivant les naissances et morts d'individus, tels le processus de Poisson ou les processus de branchement ou les processus de naissance et mort en temps continu, que nous étudierons en détail. Nous développerons en particulier les outils du calcul stochastique et la notion d'équation différentielle stochastique.

Quand la taille de la population est très grande, il devient difficile de décrire le comportement microscopique de la population, en prenant en compte chaque naissance ou mort d'individu. Nous changerons alors d'échelle de taille de la population et d'échelle de temps, pour nous ramener à des approximations plus facilement manipulables mathématiquement, sur lesquelles nous pourrions développer résultats théoriques et calculs. Dans certaines échelles, nous obtiendrons des approximations déterministes, qui ont été historiquement les premières introduites pour décrire les dynamiques de population, comme l'équation logistique ou les systèmes de Lotka-Volterra. Dans d'autres échelles, nous obtiendrons des approximations aléatoires définies comme solutions de certaines équations différentielles stochastiques, telles les équations de Feller dans le cas de la dynamique des populations ou les équations de Wright-Fisher dans le cas de la génétique des populations. Des calculs sur ces processus de diffusion permettront d'en déduire un certain nombre d'informations sur le comportement des populations. Nous définirons également, à partir du modèle de Wright-Fisher, un processus permettant la modélisation des généalogies, le coalescent de Kingman, qui est devenu dans ces dernières années fondamental en génétique des populations.

La dernière partie de cet ouvrage présente quelques applications récentes des outils mathématiques que nous avons développés à des problématiques biologiques.

1.2 Importance de la modélisation

A vital next step will be to promote the training of scientists with expertise in both mathematics and biology. (Science 2003).

Le but est d'utiliser un modèle mathématique pour mieux comprendre l'évolution temporelle, ou "dynamique", d'un phénomène biologique. Les systèmes biologiques sont extrêmement complexes : des problèmes multi-échelles, des interactions multiples, pas de lois de conservation. Il faut donc beaucoup simplifier le système biologique pour obtenir un modèle accessible mathématiquement, en se focalisant sur le phénomène biologique que

L'on cherche à comprendre. Une difficulté de cette démarche est donc d'obtenir un bon compromis entre le réalisme biologique du modèle et la faisabilité des calculs, et ce travail ne peut se faire qu'en très bonne concertation, entre biologistes et mathématiciens. Le modèle mathématique permet alors de pouvoir quantifier numériquement certains phénomènes et de pouvoir prédire certains comportements : par exemple, montrer qu'une certaine population va s'éteindre et calculer le temps moyen d'extinction ou savoir comment elle va envahir l'espace. Il est important de se poser la question de la justification du modèle. Dans le cas où il est possible d'obtenir des données observées pour le phénomène d'intérêt, une étape ultérieure sera de valider le modèle par une approche statistique. Nous n'aborderons pas cette question dans cet ouvrage.

L'intérêt d'un modèle réside aussi dans son "universalité". Des problèmes biologiques très différents (par exemple par les échelles de taille : gènes - cellules - bactéries - individus - colonies) peuvent avoir des comportements aléatoires similaires (en terme de reproduction, mort, migration) et être étudiés par des modèles analogues. Un même objet mathématique pourra être utilisé à des fins biologiques différentes. Ainsi, une marche aléatoire simple pourra modéliser la dynamique de la taille d'une population sur \mathbb{N} ou un déplacement sur \mathbb{Z} . Dans le premier cas, si cette marche atteint 0 (extinction de la population), elle y restera (sans immigration). Dans le deuxième cas, le marcheur pourra repasser une infinité de fois par 0.

Les modèles développés dans ce livre sont "élémentaires" au sens où les processus étudiés sont à valeurs dans des espaces de dimension finie. Prendre en compte la structure d'une population, suivant un certain nombre de variables telles que l'âge, un ou plusieurs phénotypes, le génotype, la position spatiale, etc., exigent des modèles en dimension infinie et nécessitent des outils mathématiques beaucoup plus lourds, que nous n'aborderons pas ici.

1.3 Modélisation mathématique en écologie et évolution

Notre but est d'étudier les modèles probabilistes de base qui apparaissent en écologie, en dynamique des populations et en génétique des populations. Tous ces problèmes contribuent à l'étude de la biodiversité. Nous allons préciser ici les questions biologiques qui nous intéressent et sous-tendent l'organisation du livre.

1.3.1 Ecologie et évolution

L'écologie est l'étude des interactions entre les êtres vivants (animaux, végétaux, micro-organismes, ...) et avec le milieu qui les entoure et dont ils font eux-mêmes partie, comme par exemple leur habitat et l'environnement. L'écologie étudie en particulier la hiérarchie complexe des écosystèmes, les mécanismes biologiques associés à l'extinction des espèces,

la dynamique de la biodiversité, l'adaptation des populations. À son développement contribuent les disciplines plus spécialisées de la dynamique et de la génétique des populations. La dynamique des populations a depuis longtemps fait l'objet de modélisations mathématiques. Les premières études quantitatives de populations sont dues à Malthus (1798), puis à Verhulst (1838). De leurs travaux émerge l'idée que la croissance de la population est limitée par celle des ressources et par son environnement. Cette idée prend toute sa force dans la théorie développée par Darwin (1859). Celui-ci défend le concept fondamental de sélection naturelle et de modification des espèces vivantes au cours du temps. Cette théorie de l'évolution est fondée sur l'idée simple que les individus les mieux adaptés à leur environnement ont les descendance les plus importantes et sont sélectionnés. L'adaptation découle de la variabilité individuelle qui apparaît dans les mécanismes de reproduction (mutation, recombinaison) ou de transfert et par les interactions entre les individus et leur environnement. Les interactions peuvent être de multiples formes. On peut citer la compétition des individus d'une espèce pour le partage des ressources, les liens dans un réseau trophique (chaîne alimentaire) comme par exemple les interactions proie-prédateur ou les interactions hôte-parasite, la lutte des femelles pour les mâles.

L'idée de variabilité individuelle est à la base de la modélisation probabiliste.

1.3.2 Dispersion et colonisation

Dans ce cours, nous allons nous intéresser aux processus probabilistes modélisant les dynamiques spatiales de populations. Ces processus modélisent la manière dont les individus se déplacent ou dispersent leur progéniture. Ils peuvent être développés à toutes les échelles biologiques, des virus aux mammifères. La structuration spatiale de la population est le résultat des comportements individuels de dispersion, comme par exemple la progression d'un troupeau, l'extension d'une épidémie, la dispersion des grains de pollen, oeufs, graines, etc. Ces déplacements peuvent être dus à la recherche de nourriture, au besoin de trouver un partenaire, à la fuite devant un prédateur ou à la nécessité de coloniser des zones propices à la survie ou à la reproduction. Dans tous les cas, la structure spatiale s'associe pour la population, au fait de minimiser sa probabilité d'extinction. Elle est donc fondamentalement liée à l'évolution démographique de la population et à sa répartition génétique. Durant les vingt dernières années, de nombreux travaux ont été développés dans ces directions. Ils nécessitent l'introduction d'objets mathématiques sophistiqués, mais les éléments de base de ces modèles sont développés dans ce livre.

Les Chapitres 2 et 4 étudient des processus stochastiques markoviens qui permettent de modéliser les déplacements individuels, en temps discret ou en temps continu. L'individu peut évoluer dans un espace discret, \mathbb{Z} , \mathbb{Z}^2 ou \mathbb{Z}^3 , ou dans un espace continu, \mathbb{R} , \mathbb{R}^2 ou \mathbb{R}^3 , ou dans un sous-domaine de ces espaces. Le choix de la dimension est lié à la nature du problème biologique, la dimension 1 pour un déplacement le long d'une route, d'une rivière, d'un brin d'herbe ou d'un brin d'ADN, la dimension 2 pour un déplacement sur le sol, une paroi, un fond marin, la dimension 3 pour un déplacement dans tout l'espace.

Soit à cause de frontières naturelles (mer, chaîne de montagne, membrane cellulaire), ou de limitation de la zone de ressources ou de pièges, les populations se déplacent souvent dans des domaines fermés et le comportement au bord du domaine est important. Il peut être absorbant si les individus meurent ou restent bloqués à son contact ou réfléchissant si les individus rebroussement chemin. Les probabilités d'atteinte de barrières seront étudiées en détail.

1.3.3 Dynamique des populations

La dynamique des populations étudie la répartition et le développement quantitatif de populations d'individus, asexués ou sexués. Elle s'intéresse aux mécanismes d'auto-régulation des populations, au problème de l'extinction d'une population ou à l'existence d'un éventuel état stationnaire ou quasi-stationnaire. Elle étudie également les interactions entre différentes espèces, comme les liens de prédation entre proies et prédateurs ou plus généralement la structuration des réseaux trophiques ou les modèles de coopération ou (et) de compétition.

Les modèles de populations non structurées que nous allons voir dans les Chapitres 3 et 5 ont l'avantage de la simplicité, mais sont bien sûr très naïfs pour rendre compte de la diversité biologique. Néanmoins, ils mettent en évidence des comportements variés, en fonction des paramètres démographiques et nous donnent des moyens de calcul : calcul de la probabilité de persistance d'une population ou de son extinction et du temps moyen d'extinction. Nous pourrions également évaluer la taille d'une population en temps long, si celle-ci persiste et en étudier la composition entre types, dans le cas où les individus peuvent être de types différents.

Nous mettrons en évidence deux grandes classes de modèles : ceux qui reposent sur des lois de reproduction générales mais ne supposent pas d'interaction entre les individus, appelés processus de branchement, et ceux pour lesquels la loi de reproduction est plus simple mais qui tiennent compte de la compétition entre individus, appelés processus de naissance et mort. Les outils d'étude de ces objets sont différents.

1.3.4 Génétique des populations

Pour comprendre l'évolution des espèces, il faut pouvoir comprendre les mécanismes internes des généalogies à l'intérieur d'une espèce. La génétique des populations a pris son essor à partir des travaux de Mendel (1822-1884), qui fût à l'origine des règles de retransmission génétique (communément appelées lois de Mendel), définissant la manière dont les gènes se transmettent de génération en génération. Elle a été initialement développée mathématiquement par les biologistes Fisher, Haldane et Wright entre 1920 et 1940. La génétique des populations se concentre sur l'étude de la fluctuation des allèles (les différentes versions d'un gène) au cours du temps dans les populations d'individus d'une même

espèce. Ces fluctuations peuvent être dues à l'influence de la dérive génétique (la variabilité due à l'aléa des événements individuels de naissance et de mort), aux mutations, à la sélection naturelle et aux migrations.

Dans le Chapitre 6, nous présenterons les modèles classiques décrivant la dynamique de l'information génétique des individus au cours du temps, en supposant que la taille de la population observée reste constante. Ce point de vue est le point de vue historique de Wright et Fisher et reste toujours d'actualité pour les généticiens des populations. Il revient à se placer d'emblée dans un état stationnaire du point de vue de la dynamique des populations. Nous aborderons également le point de vue inverse : reconstruire les lignées généalogiques d'un groupe d'individus observés, en remontant dans le temps l'histoire de ces individus. En changeant l'échelle de temps, nous mettrons en évidence un objet limite d'une nouvelle nature : un processus stochastique qui permet de reconstruire les ancêtres communs successifs d'un groupe d'individus, appelé le coalescent de Kingman.

Une fois ces briques élémentaires posées, il devient possible d'aborder la modélisation de la dynamique des individus en prenant en compte l'information génétique qu'ils contiennent et les interactions qu'ils développent. Ces questions sont au coeur de la recherche fondamentale qui se développe aujourd'hui entre mathématiques appliquées et modélisation de la biodiversité. C'est pourquoi, plutôt que de donner des exemples biologiques trop simplistes, nous avons préféré développer en détail, dans le Chapitre 7, des résultats obtenus dans des articles de recherche récents, motivés par des questions biologiques actuelles.

1.3.5 Quelques termes de vocabulaire

- Une cellule biologique est dite **haploïde** lorsque les chromosomes qu'elle contient sont en un seul exemplaire. Le concept est généralement opposé à **diploïde**, terme désignant les cellules avec des chromosomes en double exemplaire. Chez les humains et la plupart des animaux, la reproduction génétique met en jeu une succession de phases diploïdes (phases dominantes) et haploïdes (phases de formation des gamètes).
- Un **gamète** est une cellule reproductrice de type haploïde qui a terminé la méiose.
- Un **locus**, en génétique, est un emplacement précis sur le chromosome. Il peut contenir un gène.
- Les **allèles** sont des versions différentes de l'information génétique codée sur un locus. Par exemple, les types "ridés" et "lisses" des pois, dans les expériences de Mendel, correspondent à des allèles distincts.
- L'**avantage sélectif** ou **fitness** d'un allèle est une mesure qui caractérise son aptitude à se transmettre, et qui dépend ainsi de l'aptitude qu'il confère à son porteur à se reproduire et survivre.

Chapitre 2

Populations spatiales et temps discret

*The mathematics is not there till we put it there. Sir Arthur Eddington (1882 - 1944),
The Philosophy of Physical Science.*

Dans ce chapitre, nous modélisons une dynamique aléatoire par un processus aléatoire $(X_n)_{n \in \mathbb{N}}$ indexé par le temps discret. Nous introduirons deux classes de modèles qui prennent en compte des structures de dépendance différentes et qui généralisent les suites de variables aléatoires indépendantes : les chaînes de Markov et les martingales à temps discret. Ces deux classes se rejoignent dans l'étude des marches aléatoires, qui constituent un prototype de ce chapitre. Une motivation importante sera la modélisation de dynamiques spatiales mais nous verrons également des modèles de dynamiques de taille de population.

2.1 Marches aléatoires et chaînes de Markov

Nous rappelons les principales propriétés des marches aléatoires et des chaînes de Markov à temps discret, en privilégiant des problématiques propres aux modèles de populations. Pour plus de détails, nous renvoyons aux ouvrages suivants, [36], [62] et [30].

Le modèle aléatoire le plus simple pour décrire le déplacement au hasard d'un individu est celui d'une marche aléatoire sur le réseau \mathbb{Z}^d . Le temps est discret. L'individu, à partir d'un point $x = (x_1, \dots, x_d)$ peut aller vers un autre point du réseau avec une certaine probabilité. Les déplacements successifs de l'individu sont indépendants les uns des autres. Plus précisément, on note X_n la position de l'individu à l'instant n , et $Z_n = X_n - X_{n-1}$ son n -ième déplacement.

Définition 2.1.1 La suite de variables aléatoires $(X_n)_{n \in \mathbb{N}}$ est appelée marche aléatoire sur le réseau \mathbb{Z}^d si

$$X_n = X_0 + \sum_{k=1}^n Z_k,$$

et les déplacements successifs $Z_k \in \mathbb{Z}^d$ sont indépendants et de même loi. Si chaque déplacement ne peut se faire que vers l'un de ses proches voisins, la marche aléatoire est dite simple. Si de plus les déplacements vers chacun des voisins immédiats se font avec la même probabilité $\frac{1}{2d}$, la marche est dite symétrique.

La figure 2.1 montre des simulations de marches aléatoires simples symétriques en dimensions 1 et 2. Dans le premier cas, la marche peut avancer ou reculer d'une amplitude 1 avec probabilité $\frac{1}{2}$. Dans le deuxième cas, l'individu pourra sauter vers l'un de ses voisins immédiats avec probabilité uniforme $\frac{1}{4}$.

Une marche aléatoire vérifie les deux propriétés fondamentales suivantes :

- Pour tous n et k dans \mathbb{N} , la variable aléatoire $X_{n+k} - X_n$ est indépendante de X_n, X_{n-1}, \dots, X_0 .
- Pour tous n et k , $X_{n+k} - X_n$ a même loi que $X_k - X_0$.

Les accroissements de la marche aléatoire sont donc indépendants et stationnaires.

Définition 2.1.2 Nous dirons plus généralement qu'un processus $(X_n)_{n \in \mathbb{N}}$ est un processus à accroissements indépendants et stationnaires (noté plus simplement PAIS) si pour tous entiers $0 \leq n_1 < n_2 < \dots < n_k$, les variables aléatoires $(X_{n_1}, X_{n_2} - X_{n_1}, \dots, X_{n_k} - X_{n_{k-1}})$ sont indépendantes et si $X_{n_2} - X_{n_1}$ a même loi que $X_{n_2 - n_1} - X_0$.



FIGURE 2.1 – Marches aléatoires en dimension 1 et en dimension 2

Remarque 2.1.3 Dans cet ouvrage, nous allons essentiellement nous limiter aux marches aléatoires en dimension un qui peuvent en particulier décrire des déplacements spatiaux ou des dynamiques de taille de population. Dans la suite de ce chapitre, l'espace d'état considéré sera \mathbb{Z} .

Les marches aléatoires introduites ci-dessus sont un cas particulier simple de chaînes de Markov. La propriété de Markov décrit une propriété de nombreux phénomènes aléatoires, pour lesquels l'évolution aléatoire future ne dépend du passé qu'à travers l'état du processus au temps présent. Dans la suite du cours, nous verrons un grand nombre de processus vérifiant la propriété de Markov.

Définition 2.1.4 Une suite de variables aléatoires $(X_n)_{n \in \mathbb{N}}$ à valeurs dans \mathbb{Z} satisfait la propriété de Markov si pour tous $n \in \mathbb{N}$, $j, i_0, \dots, i_n \in \mathbb{Z}$, tels que $\mathbb{P}(X_n = i_n, \dots, X_0 = i_0) > 0$, on a

$$\mathbb{P}(X_{n+1} = j | X_n = i_n, \dots, X_0 = i_0) = \mathbb{P}(X_{n+1} = j | X_n = i_n). \quad (2.1)$$

La loi conditionnelle de X_{n+1} sachant X_0, \dots, X_n est égale à sa loi conditionnelle sachant X_n .

Si cette loi ne dépend pas de n , on dit que la chaîne de Markov est homogène (en temps).

La loi d'une chaîne de Markov homogène est caractérisée par ses probabilités de transition

$$P_{i,j} = \mathbb{P}(X_1 = j | X_0 = i) \quad i, j \in \mathbb{Z},$$

et par sa condition initiale :

$$\mathbb{P}(X_0 = i_0, \dots, X_n = i_n) = \mathbb{P}(X_0 = i_0) P_{i_0, i_1} \dots P_{i_{n-1}, i_n}.$$

Ultérieurement, nous noterons \mathbb{P}_i la loi de la chaîne de Markov issue de l'état i : pour tout événement A ,

$$\mathbb{P}_i(A) = \mathbb{P}(A | X_0 = i).$$

Définition 2.1.5 Les probabilités de transition $(P_{i,j})_{i,j \in \mathbb{Z}}$ définissent une matrice, appelée matrice de transition, à coefficients positifs et dont la somme de chaque ligne est égale à 1 : pour tous $i, j \in \mathbb{Z}$,

$$P_{i,j} \geq 0 \quad ; \quad \sum_{j \in \mathbb{Z}} P_{i,j} = 1.$$

Introduisons la suite des tribus $\mathcal{F}_n = \sigma(X_0, \dots, X_n)$ engendrées par la chaîne $(X_n)_n$. La tribu \mathcal{F}_n est engendrée par les ensembles de la forme $\{(X_0, X_1, \dots, X_n) = (i_0, \dots, i_n)\}$ pour tous i_0, \dots, i_n entiers relatifs. Elle décrit l'information donnée par le processus jusqu'à l'instant n . Les tribus \mathcal{F}_n sont croissantes (pour l'inclusion) comme l'information qui croît au cours du temps. La suite $(\mathcal{F}_n)_n$ s'appelle la filtration engendrée par la chaîne de Markov $(X_n)_n$.

Remarque 2.1.6 La suite $(X_n)_n$ satisfait la propriété de Markov si et seulement si pour toute fonction f bornée sur \mathbb{Z} , et pour tout entier naturel n , on a

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

En effet, il suffit de vérifier que pour tous i_0, \dots, i_n ,

$$\begin{aligned} \mathbb{E}(f(X_{n+1})|(X_0, \dots, X_n) = (i_0, \dots, i_n)) &= \sum_{j \in \mathbb{Z}} f(j) \mathbb{P}(X_{n+1} = j | (X_0, \dots, X_n) = (i_0, \dots, i_n)) \\ &= \sum_{j \in \mathbb{Z}} f(j) \mathbb{P}(X_{n+1} = j | X_n = i_n) = \mathbb{E}(f(X_{n+1})|X_n = i_n). \end{aligned}$$

Exemples :

1) *Marche aléatoire simple dans \mathbb{Z} .* Les déplacements Z_n sont indépendants et équidistribués de loi chargeant $\{-1, 1\}$ avec probabilités respectives $1-p$ et p . La marche aléatoire simple est une chaîne de Markov homogène de matrice de transition $(P_{i,j})$ vérifiant

$$P_{i,i+1} = p ; P_{i,i-1} = 1-p ; P_{i,j} = 0 \text{ si } j \neq i+1, i-1.$$

Si $p = \frac{1}{2}$, la marche aléatoire est symétrique.

Ce modèle peut par exemple décrire le déplacement vertical d'une particule de planc-ton dans l'océan Atlantique, où la profondeur moyenne est d'environ 3300 m. Si nous considérons sa position à des temps discrets, nous pouvons supposer que cette particule se déplace verticalement comme une marche aléatoire unidimensionnelle. La profondeur est suffisamment grande par rapport à la taille de la cellule pour que le déplacement de celle-ci soit considéré comme possible sur tout \mathbb{Z} .

2) *Modèle de dynamique des ressources.* A chaque instant $n \geq 0$, une quantité R de ressources est créée et une quantité aléatoire Z_n en est consommée par un certain groupe d'individus. On suppose que les variables aléatoires $(Z_n)_n$ sont indépendantes et de même loi à valeurs entières et l'on note $p_k = \mathbb{P}(Z_1 = k)$ pour tout entier k . Si l'on note par X_n le niveau des ressources à l'instant n , on obtient la dynamique suivante :

$$X_n = (X_{n-1} + R - Z_n)^+ \quad \text{pour tout } n \geq 1.$$

La suite $(X_n)_n$ forme alors une chaîne de Markov homogène à valeurs dans \mathbb{N} , de matrice de transition $(P_{i,j})$ vérifiant pour tous $i, j \in \mathbb{N}$,

$$P_{i,0} = \mathbb{P}(Z_n \geq i + R) = \sum_{k \geq i+R} p_k ; P_{i,j} = \mathbb{P}(i + R - Z_n = j) = p_{R+i-j} \text{ si } j \leq R + i - 1 ;$$

$$P_{i,j} = 0 \text{ sinon.}$$

3) *Processus de branchement.* Ces modèles ont été introduits pour modéliser les générations successives d'une population. Ils seront développés au Chapitre 3. Désignons par

X_n la taille d'une population à l'instant n . Notons par Y_i^n la variable aléatoire à valeurs entières représentant le nombre de descendants de chaque individu $i \leq X_n$ au temps n . La dynamique de la taille de la population est décrite pour $n \geq 0$ par

$$X_{n+1} = \sum_{i=1}^{X_n} Y_i^n,$$

où la somme est nulle sur $\{X_n = 0\}$. On suppose de plus que les variables aléatoires $(Y_i^n, i, n \in \mathbb{N})$ sont indépendantes et de de même loi. La suite $(X_n)_n$ forme alors une chaîne de Markov homogène à valeurs dans \mathbb{N} , de matrice de transition $(P_{i,j})$ vérifiant

$$P_{0,0} = 1 ; P_{i,j} = \mathbb{P}\left(\sum_{k=1}^i Y_k^0 = j\right) \text{ pour tous } i, j \in \mathbb{N}, i \geq 1.$$

Dans cette exemple, la matrice de transition est difficile à utiliser et nous verrons au Chapitre 3 qu'il est plus simple d'étudier la fonction génératrice de X_n .

4) *Processus de naissance et mort.* Voici un autre modèle markovien de taille de population indexé par le temps discret. Désignons comme précédemment par X_n la taille d'une population à l'instant n . A chaque instant n , un unique individu peut naître ou mourir. Ses probabilités de naissance et mort dépendent de la taille de la population à cet instant. Ainsi, conditionnellement à $\{X_n = i\}$, les probabilités de naissance et mort seront respectivement b_i et d_i . Nous supposons que $b_0 = d_0 = 0$ et que $d_i > 0$ pour $i \geq 1$. Le processus prend ses valeurs dans \mathbb{N} sauf si l'on impose une contrainte sur la taille de la population, par exemple que celle-ci ne puisse pas dépasser le seuil N . Dans ce cas nous supposons également que $b_N = 0$. Cette contrainte peut-être une manière de modéliser une compétition des individus pour le partage des ressources : si il y a trop d'individus, ceux-ci utilisent la totalité des ressources pour survivre et n'ont plus les moyens énergétiques pour se reproduire. Les probabilités de transition du processus de naissance et mort sont égales à

$$P_{i,i+1} = b_i ; P_{i,i-1} = d_i ; P_{i,i} = 1 - b_i - d_i ; P_{i,j} = 0 \text{ si } i \geq 1, j \neq i+1, i-1, i.$$

De plus, $P_{0,0} = 1, P_{0,j} = 0$ pour $j \neq 0$ et dans le cas d'une population de taille au plus N , nous aurons également $P_{N,N+1} = b_N = 0$.

Revenons à la situation générale où nous considérons une chaîne de Markov homogène quelconque et pour $n \in \mathbb{N}^*$, notons par $(P_{i,j}^{(n)})$ la matrice définie par

$$P_{i,j}^{(n)} = \mathbb{P}(X_n = j \mid X_0 = i). \tag{2.2}$$

Proposition 2.1.7 *La matrice $P^{(n)}$ est égale au produit de matrices P^n .*

Preuve. Montrons cette propriété par récurrence sur n . Elle est clairement vraie pour $n = 1$. Supposons qu'elle soit satisfaite au temps $n - 1$, c'est-à-dire que $P^{(n-1)} = P^{n-1}$. Nous avons

$$\begin{aligned}
 P_{i,j}^{(n)} &= \mathbb{P}(X_n = j \mid X_0 = i) = \sum_{k \in \mathbb{Z}} \mathbb{P}(X_n = j, X_{n-1} = k \mid X_0 = i) \\
 &= \sum_{k \in \mathbb{Z}} \mathbb{P}(X_n = j \mid X_{n-1} = k, X_0 = i) \mathbb{P}(X_{n-1} = k \mid X_0 = i) \\
 &= \sum_{k \in \mathbb{Z}} \mathbb{P}(X_n = j \mid X_{n-1} = k) \mathbb{P}(X_{n-1} = k \mid X_0 = i) \quad \text{par la propriété de Markov} \\
 &= \sum_{k \in \mathbb{Z}} P_{k,j} P_{i,k}^{(n-1)} = \sum_{k \in \mathbb{Z}} P_{k,j} P_{i,k}^{n-1} \quad \text{par la propriété de récurrence} \\
 &= P_{i,j}^n.
 \end{aligned}$$

□

Les relations entre les états de la chaîne de Markov engendrent une classification pour ces états.

On dira que l'état j est atteint depuis l'état i s'il existe $n > 0$ tel que $P_{i,j}^{(n)} > 0$. Cette relation est notée $i \rightarrow j$. Si $i \rightarrow j$ et $j \rightarrow i$, on dit que les états i et j communiquent, et l'on note $i \leftrightarrow j$: il existe $n > 0$ et $n' > 0$ tels que $P_{i,j}^{(n)} > 0$ et $P_{j,i}^{(n')} > 0$. Il est facile de montrer que cette relation entre i et j définit une relation d'équivalence.

Définition 2.1.8 *Les classes d'équivalence de la relation de communication $i \leftrightarrow j$ sont appelées les classes de la chaîne de Markov.*

Si il n'y a qu'une seule classe, la chaîne de Markov est dite irréductible.

Une classe C est dite fermée s'il est impossible de passer de C au complémentaire de C en une étape : $P_{i,j} = 0$ pour $i \in C$ et $j \notin C$.

Exemple 2.1.9 Modèle markovien de substitution de base (Allman, Rhodes [3] p. 141). Nous allons décrire un modèle d'évolution moléculaire grâce à une chaîne de Markov à 4 états. Chaque site dans une séquence d'ADN est l'une des 4 bases A, G, C, T (Adénine, Guanine, Cytosine, Thymine), choisie aléatoirement suivant les probabilités p_A, p_G, p_C, p_T telles que $p_A + p_G + p_C + p_T = 1$. Ainsi, le vecteur $P_0 = (p_A, p_G, p_C, p_T)$ décrit la distribution ancestrale des bases dans la séquence d'ADN. Nous supposons que les sites sont indépendants les uns des autres. Supposons de plus que ces bases peuvent muter et qu'une seule mutation peut avoir lieu par génération. Le but est d'étudier l'évolution de la chaîne de Markov qui décrit l'état d'un site de la séquence d'ADN à valeurs dans $\{A, G, C, T\}$ et de distribution initiale P_0 . Le modèle d'évolution moléculaire de Jukes Cantor [43] est un cas particulier supposant que les transitions décrivant les probabilités de mutation d'une

base à une autre dans la structure d'ADN sont toutes égales à un certain nombre $\frac{\alpha}{3}$, avec $0 < \alpha < 1$. Ainsi la matrice de transition vaut

$$P = \begin{pmatrix} 1 - \alpha & \frac{\alpha}{3} & \frac{\alpha}{3} & \frac{\alpha}{3} \\ \frac{\alpha}{3} & 1 - \alpha & \frac{\alpha}{3} & \frac{\alpha}{3} \\ \frac{\alpha}{3} & \frac{\alpha}{3} & 1 - \alpha & \frac{\alpha}{3} \\ \frac{\alpha}{3} & \frac{\alpha}{3} & \frac{\alpha}{3} & 1 - \alpha \end{pmatrix}.$$

La chaîne de Markov de matrice de transition P est irréductible. Remarquons également que la somme des éléments de chaque colonne est égale à 1 (on a déjà cette propriété pour les lignes). Dans ce cas, la matrice de transition est dite doublement stochastique. Cela reste vrai pour toutes ses puissances. Si la répartition ancestrale des bases est uniforme, alors elle restera uniforme. En effet, calculons par exemple la probabilité d'avoir une base de Guanine au temps n . Par une analogie entre $\{A, G, C, T\}$ et $\{1, 2, 3, 4\}$, cette probabilité vaut

$$Q_n^G = \sum_{i=1}^4 (P_0)_i P_{i,2}^{(n)} = \frac{1}{4} \sum_{i=1}^4 P_{i,2}^{(n)} = \frac{1}{4}.$$

En revanche, si la répartition initiale vaut $(\frac{1}{3}, \frac{1}{6}, \frac{1}{6}, \frac{1}{3})$, alors au temps $n = 3$,

$$Q_3^G = \sum_{i=1}^4 (P_0)_i P_{i,2}^{(3)} = \frac{1}{3} P_{1,2}^{(3)} + \frac{1}{6} P_{2,2}^{(3)} + \frac{1}{6} P_{3,2}^{(3)} + \frac{1}{3} P_{4,2}^{(3)}.$$

On peut vérifier facilement que $P_{2,2}^{(3)} = (1 - \alpha)^3 + \alpha^2(1 - \alpha) + 2\frac{\alpha^3}{9}$ et que donc pour $i \neq 2$, $P_{i,2}^{(3)} = \frac{1}{3}(1 - P_{2,2}^{(3)})$. Ainsi, $Q_3^G = \frac{1}{6}(\frac{5}{3} - \frac{2}{3}P_{2,2}^{(3)})$.

2.2 Etude des temps de passage

2.2.1 Temps d'arrêt et propriété de Markov forte

Il est important de connaître la chaîne de Markov aux temps discrets n , mais certains temps aléatoires vont aussi s'avérer fondamentaux, en particulier ceux dont la connaissance est liée à la dynamique du processus. Considérons une chaîne de Markov $(X_n)_n$. Les temps successifs de passage de ce processus à un état particulier i sont essentiels. On les définit ainsi :

$$\begin{aligned} T_i &= \inf\{n \geq 1; X_n = i\} \\ T_i^1 &= T_i; \quad T_i^{k+1} = \inf\{n > T_i^k; X_n = i\}, \quad \text{pour tout } k \geq 1, \end{aligned} \quad (2.3)$$

avec la convention habituelle que l'infimum de l'ensemble vide vaut $+\infty$. La suite $(T_i^k)_k$ décrit les temps de passage successifs de la chaîne en l'état i .

Si $X_0 = i$ alors T_i est le temps de premier retour à l'état i et T_j , ($j \neq i$), est le temps de premier passage en l'état j . La loi de T_i est alors donnée par les $f_{i,i}^{(n)} = \mathbb{P}(T_i = n \mid X_0 = i)$, pour tout $n \geq 1$ (n peut éventuellement être infini). Remarquons que $f_{i,i}^{(1)} = P_{i,i}$. On verra dans le paragraphe suivant une méthode de calcul de ces lois.

Remarquons que l'observation de X_0, \dots, X_n , c'est-à-dire du processus jusqu'à l'instant n , permet de décider si T_i vaut n ou non, s'il est plus petit ou plus grand que n , puisque par exemple,

$$\{T_i = n\} = \{X_1 \neq i\} \cap \dots \cap \{X_{n-1} \neq i\} \cap \{X_n = i\}.$$

En d'autres termes, nous avons

$$\{T_i = n\}, \{T_i < n\}, \{T_i > n\} \in \mathcal{F}_n.$$

Nous allons nous intéresser plus généralement à tous les temps aléatoires qui vérifient cette propriété, appelés temps d'arrêt.

Définition 2.2.1 *On appelle temps d'arrêt (relatif à la filtration $(\mathcal{F}_n)_n$) une variable aléatoire T à valeurs dans $\mathbb{N} \cup \{+\infty\}$ qui vérifie que pour tout $n \in \mathbb{N}$,*

$$\{T = n\} \in \mathcal{F}_n. \tag{2.4}$$

Exemple 2.2.2 Les temps aléatoires T_i sont des temps d'arrêt. Plus généralement, pour tout $A \subset \mathbb{Z}$, le temps d'atteinte de A par la chaîne de Markov, défini par

$$T_A = \inf\{n \geq 0; X_n \in A\}$$

est un temps d'arrêt. En effet,

$$\{T_A = n\} = \{X_0 \notin A\} \cap \dots \cap \{X_{n-1} \notin A\} \cap \{X_n \in A\} \in \mathcal{F}_n.$$

Le k -ième temps de passage en A est encore un temps d'arrêt. On peut le montrer par exemple par récurrence sur k en remarquant que

$$\{T_i^k = n\} = \cup_{m=0}^{n-1} \{T_i^{k-1} = m\} \cap \{X_{m+1} \notin A\} \cap \dots \cap \{X_{n-1} \notin A\} \cap \{X_n \in A\}.$$

En revanche, l'instant de dernier passage en A ,

$$L_A = \sup\{n \geq 0; X_n \in A\},$$

n'est pas un temps d'arrêt. Pour savoir si $L_A \leq n$, il faut connaître les états de la chaîne ultérieurs à n .

La chaîne de Markov satisfait encore la propriété de Markov si sa loi est conditionnée par la position de la chaîne à un temps d'arrêt. Cette propriété s'appelle la propriété de Markov forte et est fondamentale pour les applications.

Théorème 2.2.3 (Propriété de Markov forte). *Soit $(X_n)_n$ une chaîne de Markov et T un temps d'arrêt. Conditionnellement à X_T sur $\{T < \infty\}$, la suite $(X_{T+n})_{n \geq 1}$ est une chaîne de Markov indépendante de la trajectoire de $(X_n)_n$ jusqu'au temps T . Autrement dit, pour tout événement $A \subset \mathbb{Z}$ tel que $A \cap \{T = n\} \in \mathcal{F}_n$ pour tout n ,*

$$\begin{aligned} \mathbb{P}(A \cap \{X_{T+1} = i_1, \dots, X_{T+m} = i_m\} | X_T = i, T < \infty) \\ = \mathbb{P}(A | X_T = i, T < \infty) \mathbb{P}_i(X_1 = i_1, \dots, X_m = i_m). \end{aligned} \quad (2.5)$$

Preuve. Il suffit de remarquer que

$$\begin{aligned} \mathbb{P}(A \cap \{T = n\} \cap \{X_{T+1} = i_1, \dots, X_{T+m} = i_m\} | X_T = i) \\ = \mathbb{P}(A \cap \{T = n\} | X_T = i) \mathbb{P}_i(X_1 = i_1, \dots, X_m = i_m). \end{aligned}$$

□

2.2.2 Loi du premier temps de retour en 0 d'une marche aléatoire simple

Imaginons une particule de plancton se déplaçant sur une colonne verticale. A-t-elle tendance à aller vers la surface ou vers le fond, et va-t-elle y rester ? Pour cela nous pouvons étudier ses temps de passage en un point de la colonne, en fonction des probabilités de monter ou de descendre de cette particule.

Dans ce paragraphe, nous allons étudier les temps de passage en 0 d'une marche aléatoire simple $(X_n)_n$ (sur \mathbb{Z}), issue de 0 : $X_n = \sum_{k=1}^n Z_k$, et les variables aléatoires Z_k sont indépendantes et de même loi. Nous nous intéressons à la suite des temps d'arrêt où la marche se retrouve à l'origine. Pour décrire ces instants, il suffit de considérer l'instant de premier retour à l'origine, puisque les instants suivants sont des sommes de copies indépendantes de celui-ci par la propriété de Markov forte. En effet, si

$$T_0 = \inf\{n \geq 1; X_n = 0\} \in \mathbb{N}^* \cup \{\infty\},$$

désigne l'instant de premier retour en 0, (en posant $\min\{\emptyset\} = \infty$), et T_0^j désigne l'instant du j -ième retour en 0 pour la marche aléatoire, nous remarquons que sur $\{T_0^j < \infty\}$,

$$T_0^{j+1} = T_0^j + \tilde{\tau}_0,$$

où $\tilde{\tau}_0$ est le premier temps d'atteinte de 0 par la marche aléatoire $(X_{T_0^j+n})_{n \geq 0}$. L'égalité reste vraie sur $\{T_0^j = +\infty\}$. Par application de la propriété de Markov forte (2.5) au temps d'arrêt T_0^j , nous en déduisons que $\tilde{\tau}_0$ est indépendant de T_0^j et a même loi que T_0 .

Nous souhaitons connaître la loi de T_0 . Nous allons calculer sa fonction génératrice. Pour cela, introduisons, pour $n \geq 1$

$$f(n) = \mathbb{P}(T_0 = n) = \mathbb{P}(X_1 \neq 0, \dots, X_{n-1} \neq 0, X_n = 0).$$

Attention, il se peut que $\mathbb{P}(T_0 = +\infty) > 0$ (si le processus ne revient jamais en 0), auquel cas la fonction génératrice F de T_0 , définie pour $s \in [0, 1]$ par

$$F(s) = \mathbb{E}(s^{T_0}) = \sum_{n=1}^{\infty} f(n) s^n,$$

vérifie $F(1) = 1 - \mathbb{P}(T_0 = +\infty) = \mathbb{P}(T_0 < +\infty)$.

Proposition 2.2.4 *Si $(X_n)_n$ est une marche aléatoire simple issue de 0, la fonction génératrice du premier temps de retour en 0 vaut pour tout $s \in [0, 1]$,*

$$F(s) = 1 - (1 - 4pq s^2)^{1/2}.$$

Preuve. Nous introduisons la fonction auxiliaire Q définie pour $s \in [0, 1]$ par

$$Q(s) = \sum_{n=0}^{\infty} \mathbb{P}(X_n = 0) s^n,$$

que l'on calcule plus aisément. Montrons que

- (i) $Q(s) = 1 + Q(s) F(s)$.
- (ii) $Q(s) = (1 - 4pq s^2)^{-1/2}$.

D'après la formule de Bayes,

$$\mathbb{P}(X_n = 0) = \sum_{k=1}^n \mathbb{P}(X_n = 0 | T_0 = k) \mathbb{P}(T_0 = k).$$

Les variables aléatoires Z_i sont indépendantes et de même loi et

$$\mathbb{P}(X_n = 0 | T_0 = k) = \mathbb{P}(X_n - X_k = 0 | T_0 = k) = \mathbb{P}(X_{n-k} = 0).$$

Nous en déduisons que $a_n = \mathbb{P}(X_n = 0)$ est solution de

$$a_n = \sum_{k=1}^n a_{n-k} f(k),$$

et $a_0 = 1$. En multipliant par s^n et en sommant en n , puis en utilisant le théorème de Fubini, nous obtenons

$$\begin{aligned} Q(s) &= \sum_{n \geq 0} a_n s^n = 1 + \sum_{n \geq 1} \sum_{k=1}^n a_{n-k} s^{n-k} f(k) s^k \\ &= 1 + \sum_{k=1}^{\infty} f(k) s^k \sum_{n=k}^{+\infty} a_{n-k} s^{n-k} = 1 + F(s) Q(s), \end{aligned}$$