

MATHÉMATIQUES
&
APPLICATIONS

Directeurs de la collection :
G. Allaire et M. Benaïm

57

MATHÉMATIQUES & APPLICATIONS

Comité de Lecture / Editorial Board

GRÉGOIRE ALLAIRE
CMAP, École Polytechnique, Palaiseau
allaire@cmapx.polytechnique.fr

MICHEL BENAÏM
Mathématiques, Univ. de Neuchâtel
michel.benaim@unine.ch

THIERRY COLIN
Mathématiques, Univ. de Bordeaux 1
colin@math.u-bordeaux1.fr

MARIE-CHRISTINE COSTA
CEDRIC, CNAM, Paris
costa@cnam.fr

GÉRARD DEGREZ
Inst. Von Karman, Louvain
degrez@vki.ac.be

JEAN DELLA-DORA
LMC, IMAG, Grenoble
jean.della-dora@imag.fr

JACQUES DEMONGEOT
TIMC, IMAG, Grenoble
jacques.demongeot@imag.fr

FRÉDÉRIC DIAS
CMLA, ENS Cachan
dias@cmla.ens-cachan.fr

NICOLE EL KAROUI
CMAP, École Polytechnique Palaiseau
elkaroui@cmapx.polytechnique.fr

MARC HALLIN
Stat. & R.O., Univ. libre de Bruxelles
mhallin@ulb.ac.be

LAURENT MICLO
LATP, Univ. de Provence
laurent : miclo@latp.univ-mrs.fr

HUYEN PHAM
Proba. et Mod. Aléatoires, Univ. Paris 7
pham@math.jussieu.fr

VALÉRIE PERRIER
LMC, IMAG, Grenoble
valerie.perrier@imag.fr

DOMINIQUE PICARD
Proba. et Mod. Aléatoires, Univ. Paris 7
picard@math.jussieu.fr

ROBERT ROUSSARIE
Topologie, Univ. de Bourgogne, Dijon
roussari@satie.u-bourgogne.fr

CLAUDE SAMSON
INRIA Sophia-Antipolis
claudesamson@sophia.inria.fr

BERNARD SARAMITO
Maths Appl., Univ. de Clermont 2
saramito@ucfma.univ-bpclermont.fr

ANNICK SARTENAER
Mathématique, Univ. de Namur
annick.sartenaer@fundp.ac.be

ZHAN SHI
Probabilités, Univ. Paris 6
zhan@proba.jussieu.fr

SYLVAIN SORIN
Equipe Comb. et Opt., Univ. Paris 6
sorin@math.jussieu.fr

JEAN-MARIE THOMAS
Maths Appl., Univ. de Pau
Jean-Marie.Thomas@univ-pau.fr

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

JEAN-PHILIPPE VIAL
HEC, Univ. de Genève
jean-philippe.vial@hec.unige.ch

BERNARD YCART
LMC, IMAG, Grenoble
bernard.ycart@imag.fr

ENRIQUE ZUAZUA
Matemáticas, Univ. Autónoma de Madrid
enrique.zuazua@uam.es

Directeurs de la collection :
G. ALLAIRE et M. BENAÏM

Instructions aux auteurs :

Les textes ou projets peuvent être soumis directement à l'un des membres du comité de lecture avec copie à G. ALLAIRE OU M. BENAÏM. Les manuscrits devront être remis à l'Éditeur sous format \LaTeX 2 ϵ .

Jean-François Delmas
Benjamin Jourdain

Modèles aléatoires

Applications aux sciences de l'ingénieur et du vivant

 Springer

Jean-François Delmas
Benjamin Jourdain

École Nationale des Ponts et Chaussées
CERMICS
6 et 8, avenue Blaise Pascal
Cité Descartes - Champs-sur-Marne
77455 Marne-la-Vallée Cedex 2
France
e-mail: delmas@cermics.enpc.fr;
jourdain@cermics.enpc.fr

Library of Congress Control Number: 2006924181

Mathematics Subject Classification (2000): 60J10, 60J27, 60J80, 60K20, 60G70,
49L20, 62F12, 62N05, 90C15, 90B22, 90B25, 90B05, 92D20, 92D25, 92B15

ISSN 1154-483X
ISBN-10 3-540-33282-0 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-33282-4 Springer Berlin Heidelberg New York

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.

Springer est membre du Springer Science+Business Media

© Springer-Verlag Berlin Heidelberg 2006

springer.com

Imprimé aux Pays-Bas

Imprimé sur papier non acide 3100/SPi - 5 4 3 2 1 0 -

À Ginger, Vickie et Gautier
À Erwan et Alexy

Préface

Ce livre reprend les notes d'un cours que nous enseignons à l'École Nationale des Ponts et Chaussées depuis l'année 2000. L'objectif de ce cours est de montrer comment des modèles aléatoires élémentaires permettent d'apporter des éléments de réponse intéressants et de se forger une intuition sur des problèmes concrets. Nous abordons plusieurs thèmes traditionnels des métiers d'ingénieur : algorithmes d'optimisation, gestion des approvisionnements, dimensionnement de files d'attente, fiabilité et dimensionnement d'ouvrages à l'aide des lois de valeurs extrêmes. Nous avons également choisi de regarder des problématiques plus récentes, voire en cours de développement. Ainsi, nous présentons des résultats sur l'étude de l'ADN : recherche de séquences exceptionnelles, de zones homogènes et estimation du taux de mutation. Nous proposons aussi une introduction aux phénomènes de coagulation qui interviennent dans la croissance des molécules de polymères ou d'aérosols. Les outils probabilistes que nous présentons pour modéliser les phénomènes considérés dans ce livre, comme les chaînes de Markov, les processus de renouvellement ou les lois de valeurs extrêmes, sont des outils généraux utilisés dans bien d'autres domaines des sciences de l'ingénieur, des sciences du vivant, de la physique, de la finance mathématique ou de l'assurance.

Le prérequis pour la lecture de ce livre est la maîtrise du contenu d'un cours d'initiation aux probabilités en première année d'école d'ingénieurs ou en troisième année du cycle Licence. Nous avons choisi d'utiliser des outils élémentaires afin que le livre soit accessible aux étudiants qui ne se destinent pas aux mathématiques : en particulier, nous ne recourons pas à la notion de martingale.

Nous avons séparé les modèles que nous présentons en deux grandes classes faisant chacune l'objet d'une partie : les modèles discrets et les modèles continus.

La première partie débute par un chapitre général sur les chaînes de Markov à temps discret et leur comportement asymptotique en temps long. Dans le chapitre 2, nous présentons l'algorithme du recuit simulé pour résoudre le problème du voyageur de commerce : trouver le trajet le plus court joignant

un certain nombre de villes. L'avantage de cet algorithme stochastique est qu'il peut s'adapter facilement à des problèmes complexes d'optimisation plus généraux. Le chapitre 3 traite de l'optimisation d'un stock de pièces de rechange par un gestionnaire qui s'approvisionne auprès d'un fournisseur et doit répondre aux demandes de ses clients. La résolution du problème de minimisation des coûts repose sur la théorie du contrôle des chaînes de Markov qui est présentée à cette occasion. Le chapitre 4 aborde la modélisation des évolutions de population et le calcul des probabilités d'extinction à l'aide des processus de Galton-Watson. Ce chapitre, où la fonction génératrice constitue l'outil privilégié, peut être lu de manière isolée. Dans le chapitre 5, nous nous intéressons à la détection du sens de lecture d'un ADN circulaire quand on connaît seulement la succession des nucléotides. Pour retrouver l'information cachée du sens de lecture, nous considérons un modèle de chaîne de Markov cachée, et nous présentons puis utilisons l'algorithme Espérance-Maximisation (EM). Le chapitre 6 est consacré à la détection des séquences exceptionnellement rares ou fréquentes dans l'ADN et qui, de ce fait, sont susceptibles d'avoir une signification biologique. Dans ce but, nous comparons, pour des séquences de quelques nucléotides, le nombre d'occurrences constaté avec le nombre d'occurrences théorique prédit quand on modélise la succession des nucléotides par une chaîne de Markov. Enfin, le chapitre 7 traite de l'estimation du taux de mutation de l'ADN à partir des différences entre les séquences d'ADN observées chez les individus, dans le cadre du modèle d'évolution de population de Wright-Fisher et du processus de coalescence associé.

Le chapitre 8, qui débute la deuxième partie, présente une construction des chaînes de Markov à temps continu puis étudie leur comportement en temps long. Dans le chapitre 9, nous nous intéressons aux modèles de files d'attente et plus particulièrement au dimensionnement du nombre de serveurs. Lorsque les durées entre les temps d'arrivée des clients et les temps de service suivent des lois exponentielles, l'évolution au cours du temps du nombre de clients dans la file d'attente constitue une chaîne de Markov à temps continu. Chacun des chapitres 10 et 11 peut être lu de façon isolée. Le premier est consacré à une introduction à la fiabilité, c'est-à-dire à la modélisation des durées de bon fonctionnement des matériels. Nous y présentons les processus de renouvellement avec lesquels nous étudions des stratégies de maintenance préventive. Dans le chapitre 11, nous développons la théorie des lois de valeurs extrêmes qui permet d'estimer des probabilités d'événements rares. Cette théorie a de nombreuses applications dans l'étude des risques, comme le dimensionnement des digues contre les crues exceptionnelles. Enfin, le chapitre 12 traite des équations de coagulation et de fragmentation discrètes qui interviennent en astronomie, en physique et en chimie. Nous explicitons, dans des cas particuliers, l'expression analytique de la solution de ces équations. Nous proposons ensuite deux algorithmes qui permettent en toute généralité d'approcher cette solution et qui consistent à simuler des chaînes de Markov à temps continu.

Dans la troisième partie, l'appendice A reprend sans démonstration les définitions et les résultats de base correspondant à un cours d'initiation aux

probabilités. Les autres appendices fournissent la démonstration de résultats qui dépassent a priori le cadre d'un tel cours et sont utilisés dans plusieurs chapitres du livre.

Remerciements

Nous tenons à remercier Vidal Cohen qui dans le courant de l'année 1999 nous a invité à proposer un cours sur les modèles aléatoires à l'École Nationale des Ponts et Chaussées. L'intérêt des étudiants qui ont suivi ce cours depuis sa création nous a motivé pour rédiger un polycopié. Nous remercions Gilles Pagès qui nous a suggéré de transformer ce polycopié en livre, et notre éditeur, Bernard Ycart, ainsi que les deux rapporteurs anonymes dont les critiques constructives nous ont beaucoup aidé à améliorer la première version du livre. Nous tenons à remercier Aurélien Alfonsi, Jean-Stéphane Dhersin et Julien Guyon pour leur relecture approfondie de la version finale des différents chapitres. Nous devons beaucoup à Jean-Philippe Chancelier pour son aide précieuse concernant l'utilisation des logiciels Latex et surtout Scilab (<http://www.scilab.org/>) avec lequel nous avons réalisé les simulations et les illustrations du livre. Nous remercions également Bernard Lapeyre, Jacques Daniel, Sylvie Berte, Khadija Elouali ainsi que tous nos collègues du CERMICS pour l'ambiance de travail sympathique et stimulante qui règne au sein de ce laboratoire.

Enfin, nous voulons particulièrement remercier nos familles pour leur soutien durant la rédaction de ce livre ainsi que pour le bonheur qu'elles nous apportent jour après jour.

Champs sur Marne,
Octobre 2005

*Jean-François Delmas
Benjamin Jourdain*

Table des matières

partie I Modèles discrets

1	Chaînes de Markov à temps discret	3
1.1	Définition et propriétés	4
1.2	Chaîne trace, états absorbants	8
1.3	Probabilités invariantes, réversibilité	10
1.4	Chaînes irréductibles, chaînes apériodiques	11
1.5	Théorème ergodique	12
1.6	Théorème central limite	23
	Références	29
2	Recuit simulé	31
2.1	Condition de Doeblin et convergence des chaînes de Markov	33
2.2	Algorithme de Metropolis	35
2.3	Le recuit simulé	37
2.3.1	Mesures de Gibbs	37
2.3.2	Un résultat partiel	40
2.3.3	Résultats théoriques	44
2.4	Le problème du voyageur de commerce	45
	Références	50
3	Gestion des approvisionnements	51
3.1	Le modèle probabiliste de gestion de stock	52
3.1.1	Le modèle à une période de temps	52
3.1.2	Le modèle dynamique de gestion de stock	57
3.2	Éléments de contrôle de chaînes de Markov	58
3.2.1	Description du modèle	58
3.2.2	Évaluation du coût associé à une stratégie	60
3.2.3	Équations d'optimalité	62
3.2.4	Application au recrutement : le problème de la secrétaire	66

3.3	Résolution du problème dynamique de gestion de stock	71
3.3.1	Gestion sans coût fixe d'approvisionnement	72
3.3.2	Gestion avec coût fixe	74
3.3.3	Délai de livraison	80
3.4	Conclusion	84
	Références	84
4	Le processus de Galton-Watson	87
4.1	Étude du phénomène d'extinction	89
4.1.1	Caractérisation de la probabilité d'extinction η	92
4.1.2	Vitesse d'extinction	94
4.2	Lois limites	97
4.2.1	Le cas surcritique	98
4.2.2	Le cas sous-critique	101
4.2.3	Le cas critique	105
4.3	Réduction de variance dans les cas sous-critique ou critique . . .	108
4.4	Loi de la population totale	111
	Références	119
5	Recherche de zones homogènes dans l'ADN	121
5.1	Chaînes de Markov cachées	123
5.2	L'estimateur du maximum de vraisemblance (EMV)	125
5.2.1	Définitions et exemples	125
5.2.2	Convergence de l'EMV dans un modèle simple	129
5.3	Présentation générale de l'algorithme EM	132
5.4	Mise en œuvre de l'algorithme EM	135
5.4.1	L'étape espérance : étape E	135
5.4.2	L'étape maximisation : étape M	142
5.5	Convergence de l'EMV pour les chaînes de Markov cachées . . .	143
5.6	Autres exemples d'application de l'algorithme EM	151
5.6.1	Le mélange	151
5.6.2	Données censurées	156
5.7	Conclusion	158
	Références	160
6	Séquences exceptionnelles dans l'ADN	163
6.1	Fluctuations du nombre d'occurrences d'un mot	165
6.2	Une autre approche asymptotique	171
6.3	Une troisième approche asymptotique	179
6.3.1	«Loi des petits nombres» ou loi de Poisson	179
6.3.2	«Loi des petits nombres» pour le nombre d'occurrences . . .	181
6.4	Un autre modèle pour la séquence d'ADN	186
6.5	Conclusion	189
	Références	194

7	Estimation du taux de mutation de l'ADN	195
7.1	Le modèle d'évolution de population	196
7.2	Étude de l'arbre phylogénique	197
7.2.1	Temps d'apparition de l'ancêtre de deux individus	197
7.2.2	Temps d'apparition de l'ancêtre de r individus	199
7.2.3	Processus de Kingman et commentaires	203
7.3	Le modèle de Wright-Fisher	203
7.4	Modélisation des mutations	206
7.4.1	Estimation du taux de mutation I	208
7.4.2	Estimation du taux de mutation II	212
7.4.3	Conclusion sur l'estimation du taux de mutation	216
	Références	216

partie II Modèles continus

8	Chaînes de Markov à temps continu	221
8.1	Construction des chaînes de Markov à temps continu	222
8.1.1	Construction	222
8.1.2	Propriété de Markov	224
8.2	Semi-groupe, générateur infinitésimal	226
8.3	Comportement asymptotique	230
8.4	Processus de Poisson	235
	Références	237
9	Files d'attente	239
9.1	Introduction	240
9.1.1	Modélisation des files d'attente	240
9.1.2	Présentation des files $M/M/K$	241
9.2	Étude des files à un serveur : $M/M/1$	243
9.2.1	Probabilité invariante	243
9.2.2	Temps passé dans la file d'attente : client virtuel	246
9.2.3	Temps passé dans la file d'attente : client réel	247
9.3	Étude des files à K serveurs : $M/M/K$	251
9.3.1	Probabilité invariante	251
9.3.2	Temps passé dans la file d'attente : client virtuel	253
9.4	Réseaux de Jackson	257
9.4.1	Modèle et propriétés	257
9.4.2	Files en tandem, processus de sortie	261
9.5	Explosion et récurrence des files $M/GI/1$	262
	Références	264

10 Éléments de fiabilité	265
10.1 Introduction à la fiabilité	266
10.1.1 Mesures de performance	266
10.1.2 Taux de défaillance	267
10.1.3 Taux de défaillance monotone, lois NBU	270
10.2 Simulation	274
10.2.1 Inversion du taux de défaillance cumulé	274
10.2.2 Méthode des pannes fictives	275
10.3 Étude de stratégies de maintenance	277
10.3.1 Éléments de renouvellement	278
10.3.2 Remplacement suivant l'âge	282
10.3.3 Remplacement préventif par bloc	285
10.3.4 Comparaisons entre les remplacements suivant l'âge et par bloc	288
10.3.5 Durées entre pannes non identiquement distribuées	292
10.4 Éléments de fiabilité des systèmes complexes	295
10.4.1 Fonction de structure, coupes	295
10.4.2 Calcul de la disponibilité	297
10.4.3 Facteurs d'importance	299
Références	300
11 Lois de valeurs extrêmes	303
11.1 Statistique d'ordre, estimation des quantiles	307
11.2 Exemples de convergence du maximum renormalisé	316
11.3 Limites des maximums renormalisés	320
11.4 Domaines d'attraction	325
11.4.1 Caractérisations générales	325
11.4.2 Domaines d'attraction des lois de Fréchet et Weibull ...	327
11.5 Estimation du paramètre de la loi de valeurs extrêmes	329
11.5.1 Estimateur de Pickand	329
11.5.2 Estimateur de Hill	332
11.6 Estimation des quantiles extrêmes	336
11.6.1 À l'aide de l'estimateur de Pickand	337
11.6.2 À l'aide de l'estimateur de Hill	340
11.7 Conclusion	340
Références	341
12 Processus de coagulation et fragmentation	343
12.1 Équations de coagulation discrètes	344
12.1.1 Définition et propriétés des solutions	344
12.1.2 Solutions explicites pour les noyaux constant, additif et multiplicatif	348
12.2 Coagulation et fragmentation discrètes	365
12.3 Chaînes de Markov à temps continu associées	367
12.3.1 Le processus de Marcus-Lushnikov	368

12.3.2 Le processus de transfert de masse.....	375
Références	383

partie III Appendice

A Rappels de probabilités	387
A.1 Variables aléatoires	387
A.1.1 Espace de probabilité	387
A.1.2 Variables aléatoires	388
A.1.3 Espérance	388
A.1.4 Convergence des espérances.....	390
A.1.5 Indépendance	391
A.1.6 Variance	391
A.1.7 Fonction caractéristique	392
A.1.8 Transformée de Laplace	393
A.1.9 Probabilités conditionnelles	393
A.2 Lois usuelles	394
A.2.1 Lois discrètes usuelles	394
A.2.2 Lois à densité usuelles	396
A.2.3 Simulation.....	398
A.3 Convergence et théorèmes limites	400
A.3.1 Convergence de variables aléatoires	400
A.3.2 Loi forte des grands nombres	402
A.3.3 Théorème central limite	402
A.3.4 Intervalles de confiance.....	403
Références	405
B Une variante du théorème central limite	407
C Fonction de répartition et quantile.....	411
D Convergence en variation sur un espace discret	415
E Étude d'une équation différentielle ordinaire	417
Références	421
Index	429