

Méthodes de Monte-Carlo avec R

Springer

Paris

Berlin

Heidelberg

New York

Hong Kong

Londres

Milan

Tokyo

Christian P. Robert
George Casella

Méthodes de Monte-Carlo avec R

Traduit de l'anglais par

Joachim Robert, Robin Ryder, Julyan Arbel, Pierre Jacob
et Brigitte Plessis

 Springer

Christian P. Robert

Ceremade – Université Paris-Dauphine
Institut Universitaire de France et CREST
Place du Maréchal-de-Lattre-de-Tassigny
75775 Paris Cedex 16
France

George Casella

Department of Statistics
University of Florida
103 Griffin-Floyd Hall
Gainesville FL 3261-8545
USA

ISBN : 978-2-8178-0180-3 Springer Paris Berlin Heidelberg New York
ISSN : 2112-8294

© Springer-Verlag France, 2011

Traduction de l'édition en langue anglaise : Christian P. Robert & George Casella, *Introducing Monte Carlo Methods With R*
Copyright © 2010 Springer Science + Business Media LLC
Tous droits réservés

Imprimé en France

Springer-Verlag est membre du groupe Springer Science + Business Media

Cet ouvrage est soumis au copyright. Tous droits réservés, notamment la reproduction et la représentation, la traduction, la réimpression, l'exposé, la reproduction des illustrations et des tableaux, la transmission par voie d'enregistrement sonore ou visuel, la reproduction par microfilm ou tout autre moyen ainsi que la conservation des banques de données. La loi française sur le copyright du 9 septembre 1965 dans la version en vigueur n'autorise une reproduction intégrale ou partielle que dans certains cas, et en principe moyennant le paiement des droits. Toute représentation, reproduction, contrefaçon ou conservation dans une banque de données par quelque procédé que ce soit est sanctionnée par la loi pénale sur le copyright.

L'utilisation dans cet ouvrage de désignations, dénominations commerciales, marques de fabrique, etc. même sans spécification ne signifie pas que ces termes soient libres de la législation sur les marques de fabrique et la protection des marques et qu'ils puissent être utilisés par chacun.

La maison d'édition décline toute responsabilité quant à l'exactitude des indications de dosage et des modes d'emploi. Dans chaque cas il incombe à l'utilisateur de vérifier les informations données par comparaison à la littérature existante.

Maquette de couverture: Jean-François Montmarché



Collection Pratique R

dirigée par Pierre-André Cornillon
et Eric Matzner-Løber

Département MASS
Université Rennes-2-Haute-Bretagne
France

Comité éditorial :

Eva Cantoni
Département d'économétrie
Université de Genève
Suisse

Vincent Goulet
École d'actuariat
Université Laval
Canada

Philippe Grosjean
Département d'écologie
numérique des milieux aquatiques
Université de Mons
Belgique

Nicolas Hengartner
Los Alamos National Laboratory
USA

François Husson
Département Sciences de l'ingénieur
Agrocampus Ouest
France

Sophie Lambert-Lacroix
Département IUT STID
Université Pierre Mendès France
France

Déjà paru dans la même collection :

Régression avec R
Pierre-André Cornillon, Eric Matzner-Løber, 2011

À paraître dans la même collection :

Séries temporelles avec R
Yves Aragon, 2011

To our parents, who taught us much in many ways.

REMERCIEMENTS

Nous sommes profondément reconnaissants à nos collègues et amis pour l'aide qu'ils nous ont prodiguée, en particulier à Ed George pour ses commentaires sur l'orientation générale du livre et sur quelques exercices spécifiques; Jim Hobert et Fernando Quintana pour des discussions formatrices sur la version Monte-Carlo de l'algorithme EM; Alessandra Iacobucci pour avoir repéré à temps une erreur fatale; Jean-Michel Marin pour nous avoir permis de recycler le premier chapitre de *Bayesian Core* (Marin & Robert, 2007) en notre introduction à R et pour de nombreux conseils sur R et sur la pédagogie du livre; Antonietta Mira pour les erreurs signalées lors d'une session d'une conférence MCMC à Warwick; François Perron pour son invitation (de Christian Robert) à Montréal, qui a donné à celui-ci la fenêtre nécessaire pour finir le Chapitre 8 (et l'opportunité de faire de la cascade de glace à Québec!), ainsi qu'à François Perron et Clémentine Trimont pour avoir testé la version bêta du livre des points de vue respectifs du professeur et de l'étudiante; Martyn Plummer pour ses réponses à nos questions sur `coda`; Jeff Rosenthal pour des échanges illuminants sur `amcmc`; Dimitris Rizopoulos pour son Exercice 7.10; et Phil Spector de Berkeley pour l'aide anonyme fournie par ses notes et transparents sur R disponibles sur le Web, maintenant publiées dans Spector (2009). Les commentaires des deux rapporteurs ont été particulièrement utiles dans l'étape finale de rédaction. Nous sommes aussi reconnaissants à John Kimmel de Springer, New York, pour ses conseils et son efficacité, mais aussi pour avoir créé la série Use R! chez Springer, fournissant ainsi une tribune pour le développement du langage R. Enfin, au nom de la communauté R tout entière, nous remercions les professeurs Gentleman et Ihaka pour avoir développé le langage R et pour l'avoir fait dans le cadre de l'open-source.

Sur la traduction

Cette version française de *Introduction to Monte Carlo Methods with R* a été réalisée par Joachim Robert (Chapitres 2 à 4), Robin Ryder (Chapitres 5 et 7), Julyan Arbel (Chapitre 6), Pierre Jacob (Chapitre 8) et Brigitte Plessis (Chapitre 4), sous la supervision du premier auteur. Nous sommes reconnaissants à ces traducteurs pour avoir produit une version quasi définitive en quelques mois. Merci aussi à Mathilde Bouriga pour sa relecture. Nous apprécions également l'effort fourni par Springer Paris pour lancer la collection Pratique R, ainsi que le soutien de Charles Ruelle, d'Eric Matzner-Løber et de Pierre-André Cornillon.

Sceaux et Gainesville
7 décembre 2010

Christian P. Robert et George Casella

AVANT-PROPOS

“After that, it was down to attitude.”

Ian Rankin

Black & Blue

Le but de ce livre est de fournir une porte d'entrée autonome dans le domaine des méthodes de calcul de type Monte-Carlo. Tout d'abord, il ne doit pas être pris comme une annexe technique de notre précédent livre, *Monte Carlo Statistical Methods*, dont la seconde édition est parue en 2004. Ce nouveau livre vise un autre but, qui est de rendre un public plus vaste à même de programmer en R les méthodes de Monte-Carlo en le guidant dans ses premiers pas. Par conséquent, les aspects du langage R sont centraux dans ce livre, dont le contenu est clairement différent de celui de *Monte Carlo Statistical Methods*. Evidemment, la vision globale du champ des méthodes simulées n'en demeure pas moins la même, mais nous avons supprimé ici toute insertion dans les aspects théoriques de ces méthodes.

L'insistance sur le côté *pratique* de cette introduction a en effet pour but de faire en sorte qu'un public composé d'abord d'étudiants en statistique, biostatistique, traitement du signal, économétrie, etc., devienne assez familier avec les méthodes de simulation pour les utiliser dans son travail quotidien. Le livre est aussi conçu pour les scientifiques de toutes les disciplines, attirés par la versatilité des outils de Monte-Carlo. Il peut de plus être employé dans des cours de statistique plus traditionnels, en tant que rapide introduction, *via* R, aux méthodes modernes de calcul statistique, par exemple en fin de Licence comme c'est le cas à l'Université Paris-Dauphine.

Le choix d'un langage de programmation comme R, au lieu d'alternatives plus efficaces comme Matlab ou C et de logiciels plus élaborés comme BUGS, est fondé sur sa simplicité d'apprentissage et sur sa versatilité. Les lecteurs peuvent aisément élaborer des expériences par eux-mêmes dans le langage qu'ils ou elles préfèrent à partir des exemples fournis dans ces pages. (Bien entendu, ce livre peut servir comme supplément à des livres plus traditionnels comme nos livres *Le Choix Bayésien* (Robert, 2006) et *Monte Carlo Statistical Methods* (Robert & Casella, 2004).) Enfin, ce livre peut aussi être vu comme un compagnon, plutôt que comme un concurrent, du livre R de Jim Albert *Bayesian Computation with R* (Albert, 2009). De fait, envisagé en tandem, ces deux livres fournissent une introduction parfaite aux aspects empiriques des méthodes de Monte-Carlo et de modélisation bayésienne.

Dès cette première page, nous insistons sur le fait que, dans une perspective de production (c'est-à-dire dans l'utilisation de méthodes de Monte-Carlo avancées pour l'analyse de jeux de données complexes), R ne peut être recommandé comme langage de programmation par défaut. Cependant, l'expertise et l'intuition accumulées lors de la lecture de ce livre devraient hautement faciliter le passage à un autre langage de programmation.

Contrairement aux usages standard, nous avons choisi d'inclure les exercices aux endroits les plus appropriés du texte, plutôt que d'attendre systématiquement la fin de chaque chapitre. Ce choix n'est pas toujours compris mais nous avons au moins deux raisons. D'une part, les exercices peuvent servir à établir des résultats ou des procédures nécessaires dans la suite du chapitre mais pénibles à construire. D'autre part, ces exercices signalent au lecteur qu'une étape de révision ou tout du moins de réflexion est la bienvenue avant d'accéder au thème suivant. Les exercices peuvent donc servir d'auto-tests, d'autant que les exercices à numéro impair sont corrigés sur Internet (voir nos pages personnelles ou le site de Springer). Des exercices supplémentaires sont aussi fournis à la fin de chaque chapitre.

Table des matières

Avant-propos	xi
1 Préliminaires	1
1.1 Introduction	2
1.2 La librairie <code>mcs</code>	3
1.3 Distributions de probabilité prédéfinies en \mathbb{R}	5
1.4 Quelques mots sur le bootstrap	6
1.5 Exercices supplémentaires	9
2 Génération de variables aléatoires	11
2.1 Introduction	12
2.1.1 La simulation uniforme	12
2.1.2 Méthode de la transformation inverse	14
2.2 Méthodes de transformation générales	16
2.2.1 Un générateur aléatoire normal	17
2.2.2 Distributions discrètes	19
2.2.3 Représentations par mélange	21
2.3 Méthodes de rejet	22
2.4 Exercices supplémentaires	28
3 Intégration de Monte-Carlo	33
3.1 Introduction	34
3.1.1 Intégration numérique	34
3.2 L'intégration classique de Monte-Carlo	37
3.3 Echantillonnage préférentiel	41
3.3.1 Un changement arbitraire de mesure de référence	42
3.3.2 Rééchantillonnage préférentiel	47
3.3.3 Sélection de la loi instrumentale	51
3.4 Exercices supplémentaires	59
4 Contrôler et accélérer la convergence	63
4.1 Introduction	64

4.2	Evaluation de la variation	65
4.3	Variance asymptotique de l'échantillonnage préférentiel	67
4.4	Taille effective et perplexité d'un échantillon	72
4.5	Contrôles simultanés	74
4.6	Rao–Blackwellisation et déconditionnement	80
4.7	Méthodes d'accélération	84
4.7.1	Simulations corrélées	84
4.7.2	Variables antithétiques	87
4.7.3	Variables de contrôle	89
4.8	Exercices supplémentaires	95
5	Optimisation par les méthodes de Monte-Carlo	99
5.1	Introduction	100
5.2	Méthodes d'optimisation numérique	101
5.3	Recherche stochastique	104
5.3.1	Une solution basique	104
5.3.2	Méthodes de gradient stochastique	110
5.3.3	Recuit simulé	113
5.4	Approximation stochastique	120
5.4.1	Optimisation d'approximations de Monte-Carlo	121
5.4.2	Modèles à données manquantes et démarginalisation	124
5.4.3	L'algorithme EM	125
5.4.4	EM par Monte-Carlo	130
5.5	Exercices supplémentaires	137
6	Algorithmes de Metropolis–Hastings	141
6.1	Introduction	142
6.2	Aperçu sur la théorie des chaînes de Markov	142
6.3	Algorithmes de Metropolis–Hastings élémentaires	144
6.3.1	Un algorithme MCMC générique	145
6.3.2	Algorithme de Metropolis–Hastings indépendant	149
6.4	Sélection des lois de proposition	156
6.4.1	Marches aléatoires	156
6.4.2	Lois de proposition alternatives	160
6.4.3	Choix de modèles	163
6.5	Taux d'acceptation	166
6.6	Exercices supplémentaires	169
7	Echantillonneurs de Gibbs	173
7.1	Introduction	174
7.2	L'échantillonneur de Gibbs à deux étapes	174
7.3	L'échantillonneur de Gibbs à plusieurs étapes	179
7.4	Données manquantes et variables latentes	183
7.5	Structures hiérarchiques	195

7.6	Autres considérations	197
7.6.1	Reparamétrisation	198
7.6.2	Rao–Blackwellisation	200
7.6.3	Metropolis dans Gibbs et stratégies hybrides	204
7.6.4	Distributions a priori impropres	206
7.7	Exercices supplémentaires	209
8	Contrôle et adaptation des algorithmes MCMC	211
8.1	Introduction	212
8.2	Que contrôler et pourquoi?	212
8.2.1	Convergence vers la loi stationnaire	212
8.2.2	Convergence de moyennes	214
8.2.3	Approcher les échantillons iid	214
8.2.4	La librairie <code>coda</code>	215
8.3	Contrôler la convergence vers la stationnarité	216
8.3.1	Diagnostics graphiques	216
8.3.2	Tests non paramétriques de stationnarité	219
8.3.3	Analyse spectrale	221
8.4	Contrôler la convergence des moyennes	225
8.4.1	Diagnostics graphiques	225
8.4.2	Variances intra et inter	227
8.4.3	Taille effective d'échantillon	229
8.4.4	Moyennes par lots de taille fixée	231
8.5	MCMC adaptatif	233
8.5.1	Précautions d'emploi	233
8.5.2	La librairie <code>amcmc</code>	237
8.6	Exercices supplémentaires	240
	Bibliographie	243
	Index	249
	Index des commandes	255