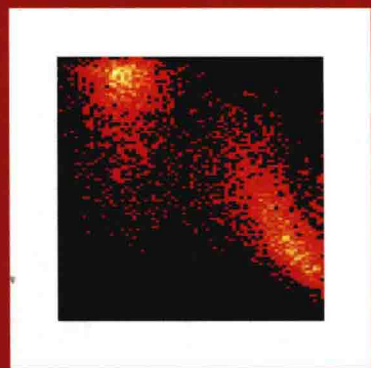


NONLINEAR MIXTURE MODELS

A Bayesian Approach

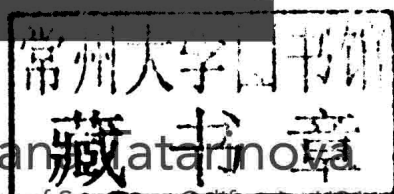
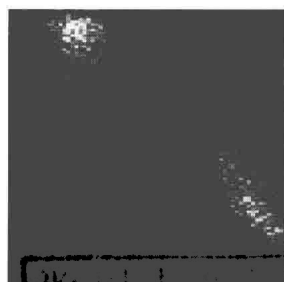


Tatiana Tatarinova
Alan Schumitzky

Imperial College Press

NONLINEAR MIXTURE MODELS

A Bayesian Approach



Tatiana Tatarnova
University of Southern California, USA

Alan Schumitzky
University of Southern California, USA

Published by

Imperial College Press
57 Shelton Street
Covent Garden
London WC2H 9HE

Distributed by

World Scientific Publishing Co. Pte. Ltd.
5 Toh Tuck Link, Singapore 596224
USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601
UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

Library of Congress Cataloging-in-Publication Data

Tatarinova, Tatiana V.

Nonlinear mixture models : a bayesian approach / Tatiana Tatarinova, University of Glamorgan, UK, Alan Schumitzky, University of Southern California, USA.

pages cm

Includes bibliographical references and index.

ISBN 978-1-84816-756-8 (hardcover : alk. paper)

1. Markov processes. 2. Bayesian statistical decision theory. 3. Nonparametric statistics.
4. Multivariate analysis. I. Schumitzky, Alan. II. Title.

QA274.7.T38 2015

519.2'33--dc23

2014038898

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

Copyright © 2015 by Imperial College Press

All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the Publisher.

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. In this case permission to photocopy is not required from the publisher.

Printed in Singapore

NONLINEAR MIXTURE MODELS

A Bayesian Approach

To our parents: Natalie and Valerii Tatarinov and Dorothy and
Abe Schumitzky

Acknowledgments

We would like to thank personally the many friends, colleagues and students who have helped to write this book through great discussion and ideas: Roger Jelliffe, Michael Neely, Jay Bartroff, David Bayard, Michael Van Guilder, Walter Yamada, David D'Argenio, Xiaoning Wang, Carolyn Lindberg, Michael Christie, Anna-Michelle Christie, Alyona Kryshenko, Hannah Garbett and Farzana Rahmann. We would also like to thank those statisticians and mathematicians whose works inspired many of the ideas in this book: Christian Robert, the late George Casella, Mathew Stephens, Jon Wakefield, Chuan Zhou, Sujit Sahu, Russell Cheng, Allain Mallet, Peter Müller, Gary Rosner, Hedibert Freitas Lopes, Sylvia Fruwirth-Schnatter and Adrian Smith.

Special thanks go to Alyona Kryshenko and Carolyn Lindberg who helped proofread the book; and to Catharina Weijman, our extremely helpful editor from Imperial College Press. Of course there are others names we should have mentioned. Please accept our regrets.

Because the authors were practically never in the same place at the same time, this book took much longer to finish than was expected. And we are greatly indebted to our spouses, Michael and Carolyn, for bearing with us and for their patience and understanding.

Finally, this book has been supported in part by NIH grants GM068968, EB005803 and EB001978, NIH-NICHD grant HD070996, Royal Society grant TG103083, and by the University of Southern California and University of South Wales.

Contents

<i>List of Tables</i>	xiii
<i>List of Figures</i>	xvii
<i>Acknowledgments</i>	xxv
1. Introduction	1
1.1 Bayesian Approach	3
1.2 Review of Applications of Mixture Models in Population Pharmacokinetics	4
1.3 Review of Applications of Mixture Models to Problems in Computational Biology	6
1.3.1 Problems with mixture models	11
1.4 Outline of the Book	11
2. Mathematical Description of Nonlinear Mixture Models	15
2.1 Fundamental Notions of Markov chain Monte Carlo	15
2.2 Nonlinear Hierarchical Models	19
2.2.1 One-component model	19
2.2.2 Mixture models	21
2.2.3 Normal mixture models	23
2.2.4 Nonlinear normal mixture models and pharmacokinetic examples	24
2.2.5 Linear normal mixture models and Eyes example .	26
2.2.6 Allocation formulation	29
2.3 Gibbs Sampling	30
2.3.1 Theoretical convergence of the Gibbs sampler . . .	32

2.3.2	Irreducibility and aperiodicity of the hybrid Gibbs–Metropolis chain	33
2.3.3	WinBUGS and JAGS	35
2.4	Prior Distributions: Linear and Nonlinear Cases	38
3.	Label Switching and Trapping	43
3.1	Label Switching and Permutation Invariance	43
3.1.1	Label switching	43
3.1.2	Permutation invariance	44
3.1.3	Allocation variables $\{Z_i\}$	46
3.1.4	An example of label switching	47
3.2	Markov Chain Convergence	48
3.2.1	Trapping states and convergence	54
3.3	Random Permutation Sampler	59
3.3.1	RPS post processing	63
3.4	Re-parametrization	65
3.5	Stephens’ Approach: Relabeling Strategies	67
3.5.1	Linear normal mixture: Eyes problem	70
4.	Treatment of Mixture Models with an Unknown Number of Components	73
4.1	Introduction	73
4.2	Finding the Optimal Number of Components Using Weighted Kullback–Leibler Distance	74
4.2.1	Distance between K-component model and collapsed (K-1)-component model	75
4.2.2	Weighted Kullback–Leibler distance for the mixture of multivariate normals	76
4.2.3	Weighted Kullback–Leibler distance for the multivariate normal mixture model with diagonal covariance matrix	78
4.2.4	Weighted Kullback–Leibler distance for the one-dimensional mixture of normals	78
4.2.5	Comparison of weighted and un-weighted Kullback–Leibler distances for one-dimensional mixture of normals	79
4.2.6	Weighted Kullback–Leibler distance for a mixture of binomial distributions	80

4.2.7	Weighted Kullback–Leibler distance for a mixture of Poisson distributions	81
4.2.8	Metric and semimetric	82
4.2.9	Determination of the number of components in the Bayesian framework	85
4.3	Stephens’ Approach: Birth–Death Markov Chain Monte Carlo	86
4.3.1	Treatment of the Eyes model	89
4.3.2	Treatment of nonlinear normal mixture models	91
4.4	Kullback–Leibler Markov Chain Monte Carlo – A New Algorithm for Finite Mixture Analysis	91
4.4.1	Kullback–Leibler Markov chain Monte Carlo with random permutation sampler	98
5.	Applications of BDMCMC, KLMCMC, and RPS	101
5.1	Galaxy Data	101
5.2	Simulated Nonlinear Normal Mixture Model	109
5.3	Linear Normal Mixture Model: Boys and Girls	112
5.4	Nonlinear Pharmacokinetics Model and Selection of Prior Distributions	121
5.5	Nonlinear Mixture Models in Gene Expression Studies	135
5.5.1	Simulated dataset	138
6.	Nonparametric Methods	143
6.1	Definition of the Basic Model	144
6.2	Nonparametric Maximum Likelihood	145
6.2.1	Bender’s decomposition	146
6.2.2	Convex optimization over a fixed grid G	147
6.2.3	Interior point methods	148
6.2.4	Nonparametric adaptive grid algorithm	149
6.2.5	Pharmacokinetic population analysis problem	151
6.3	Nonparametric Bayesian Approach	154
6.3.1	Linear model	158
6.3.2	Dirichlet distribution	159
6.3.3	Dirichlet process for discrete distributions	159
6.4	Gibbs Sampler for the Dirichlet Process	161
6.4.1	Implementation of the Gibbs sampler	162
6.5	Nonparametric Bayesian Examples	164

6.5.1	Binomial/beta model	164
6.5.2	Normal prior and linear model	166
6.5.3	One-dimensional linear case	168
6.5.4	Two-dimensional linear case	169
6.5.5	Plotting the posterior using Gibbs sampling	170
6.5.6	Galaxy dataset	170
6.6	Technical Notes	171
6.6.1	Treatment of multiplicity	172
6.7	Stick-Breaking Priors	175
6.7.1	Truncations of the stick-breaking process	176
6.7.2	Blocked Gibbs algorithm	176
6.8	Examples of Stick-Breaking	178
6.8.1	Binomial/beta: Rolling thumbtacks example	179
6.8.2	Determination of the number of mixture components in the binomial/beta model	181
6.8.3	Poisson/gamma Eye-Tracking example	186
6.8.4	Determination of the number of mixture components in Poisson/gamma case	187
6.8.5	Stephens' relabeling for Poisson/gamma case	191
6.8.6	Pharmacokinetics example	192
6.9	Maximum Likelihood and Stick-Breaking (A Connection Between NPML and NPB Approaches)	196
6.9.1	Galaxy dataset	197
7.	Bayesian Clustering Methods	199
7.1	Brief Review of Clustering Methods in Microarray Analysis	199
7.2	Application of KLMCMC to Gene Expression Time-Series Analysis	201
7.3	Kullback–Leibler Clustering	205
7.3.1	Pharmacokinetic example	209
7.4	Simulated Time-Series Data with an Unknown Number of Components (Zhou Model)	213
7.4.1	Model description	214
7.4.2	Reductive stepwise method	215
7.4.3	Zhou model using KLC algorithm	217
7.4.4	Zhou model using KLMCMC	218
7.5	Transcription Start Sites Prediction	222

7.6	Conclusions	228
Appendix A	Standard Probability Distributions	231
Appendix B	Full Conditional Distributions	233
B.1	Binomial/Beta	235
Appendix C	Computation of the Weighted Kullback– Leibler Distance	237
C.1	Weighted Kullback–Leibler Distance for a Mixture of Univariate Normals	237
C.2	Weighted Kullback–Leibler Distance for a Mixture of Multivariate Normals	239
C.3	Weighted Kullback–Leibler Distance for a Mixture of Beta Distributions	242
Appendix D	BUGS Codes	245
D.1	BUGS Code for the Eyes Model	245
D.2	BUGS Code for the Boys and Girls Example	248
D.3	BUGS Code for Thumbtack Data	249
D.4	BUGS Code for Eye-Tracking Data	251
D.5	BUGS Code for PK Example, Nonparametric Bayesian Approach with Stick-Breaking Priors	253
	<i>Bibliography</i>	255
	<i>Index</i>	267

List of Tables

2.1	WinBUGS sampling methods	36
3.1	Eyes example. Results of Gibbs sampler illustrating the switching effect	48
3.2	Eyes example. Gibbs sampler trapping problem	56
3.3	Eyes example. BOA Raftery and Lewis convergence diagnostic	59
3.4	Eyes example. BOA Geweke convergence diagnostic	59
3.5	Eyes example. Re-parametrization	66
3.6	Eyes example. Output of Stephens' algorithm	71
4.1	Eyes example. Results of Stephens' relabeling algorithm applied to the KLMCMC output	95
4.2	Eyes example. Output of the KLMCMC with random permutation sampler	98
5.1	Galaxy dataset [Postman <i>et al.</i> (1986)], containing recession velocities of 82 galaxies	103
5.2	Galaxy data. KLMCMC results with application of Stephens' relabeling algorithm to the $K = 7$ component model	104
5.3	Results of KLMCMC after relabeling for the simulated nonlinear dataset	112
5.4	Boys and Girls dataset [Potthoff and Roy (1964)], containing dental measurements for 27 children	113
5.5	Boys and Girls dataset. Reductive stepwise method. Parameters estimated by mixture model with two components and distances between components	117
5.6	Boys and Girls dataset. Reductive stepwise method. Posterior parameter values with component labels removed	117

5.7	Boys and Girls dataset. Results of application of Stephens' algorithm to the output of the reductive stepwise method . . .	118
5.8	One-compartment PK model, Case 1. Results of application of Stephens' relabeling algorithm to the $K = 2$ model, with randomization	125
5.9	One compartment PK model, Case 2. Results of application of Stephens' relabeling algorithm to the $K = 2$ model, variable V_i	127
5.10	One-compartment PK model, Case 3. Results of application of relabeling algorithm to the $K = 2$ model, bimodal V , with randomization	129
5.11	One-compartment PK model, Case 4. Results of application of relabeling algorithm to the $K = 2$ model, bimodal V , unknown σ_e	130
5.12	One-compartment PK model, Case 5. Comparison of deviance information, frequency and likelihood for various numbers of components K . Values of deviance information criterion and likelihood are normalized	134
5.13	One-compartment PK model, Case 4. Results of application of relabeling algorithm to the $K = 2$, bimodal V , unknown σ_e , with randomization	134
5.14	Simulated seed example: Results of the 500 iterations of the KLMCMC with RPS	140
6.1	Size of G_0 as a function of parameter dimension	150
6.2	Output of NPAG for the PK model	154
6.3	Final support points for NPAG method	154
6.4	Rolling thumbtacks example	179
6.5	Binomial/beta model, thumbtacks example. Determination of the number of components	184
6.6	Eye-Tracking example. Measurements	186
6.7	Eye-Tracking example. Results of Gibbs sampler	187
6.8	Eye-Tracking example. Posterior estimations of model parameters, before and after relabeling	192
7.1	Yeast sporulation time series. KLMCMC profiles conditional on $K = 7$	204
7.2	Zhou model. Means and standard deviations of parameters within clusters, as defined by the KLC method	217