

# Markov Chain Monte Carlo in Practice

Edited by

**W.R. Gilks**

Medical Research Council Biostatistics Unit  
Cambridge  
UK

**S. Richardson**

French National Institute for Health and Medical Research  
Vilejuif  
France

and

**D.J. Spiegelhalter**

Medical Research Council Biostatistics Unit  
Cambridge  
UK



**CHAPMAN & HALL**

London · Glasgow · Weinheim · New York · Tokyo · Melbourne · Madras

# Contents

---

<b>Contributors</b>	<b>xv</b>
<b>1 Introducing Markov chain Monte Carlo</b>	<b>1</b>
<i>W. R. Gilks, S. Richardson and D. J. Spiegelhalter</i>	
1.1 Introduction	1
1.2 The problem	2
1.2.1 Bayesian inference	2
1.2.2 Calculating expectations	3
1.3 Markov chain Monte Carlo	4
1.3.1 Monte Carlo integration	4
1.3.2 Markov chains	5
1.3.3 The Metropolis-Hastings algorithm	5
1.4 Implementation	8
1.4.1 Canonical forms of proposal distribution	8
1.4.2 Blocking	12
1.4.3 Updating order	12
1.4.4 Number of chains	13
1.4.5 Starting values	13
1.4.6 Determining burn-in	14
1.4.7 Determining stopping time	15
1.4.8 Output analysis	15
1.5 Discussion	16
<b>2 Hepatitis B: a case study in MCMC methods</b>	<b>21</b>
<i>D. J. Spiegelhalter, N. G. Best, W. R. Gilks and H. Inskip</i>	
2.1 Introduction	21
2.2 Hepatitis B immunization	22
2.2.1 Background	22
2.2.2 Preliminary analysis	22
2.3 Modelling	25
2.3.1 Structural modelling	25
2.3.2 Probability modelling	27
2.3.3 Prior distributions	27

2.4	Fitting a model using Gibbs sampling	28
2.4.1	Initialization	29
2.4.2	Sampling from full conditional distributions	29
2.4.3	Monitoring the output	31
2.4.4	Inference from the output	34
2.4.5	Assessing goodness-of-fit	34
2.5	Model elaboration	36
2.5.1	Heavy-tailed distributions	36
2.5.2	Introducing a covariate	37
2.6	Conclusion	40
	Appendix: BUGS	42
<b>3</b>	<b>Markov chain concepts related to sampling algorithms</b>	<b>45</b>
	<i>G. O. Roberts</i>	
3.1	Introduction	45
3.2	Markov chains	45
3.3	Rates of convergence	48
3.4	Estimation	49
3.4.1	Batch means	50
3.4.2	Window estimators	50
3.5	The Gibbs sampler and Metropolis–Hastings algorithm	51
3.5.1	The Gibbs sampler	51
3.5.2	The Metropolis–Hastings algorithm	54
<b>4</b>	<b>Introduction to general state-space Markov chain theory</b>	<b>59</b>
	<i>L. Tierney</i>	
4.1	Introduction	59
4.2	Notation and definitions	60
4.3	Irreducibility, recurrence and convergence	62
4.3.1	Irreducibility	62
4.3.2	Recurrence	63
4.3.3	Convergence	64
4.4	Harris recurrence	65
4.5	Mixing rates and central limit theorems	67
4.6	Regeneration	70
4.7	Discussion	71
<b>5</b>	<b>Full conditional distributions</b>	<b>75</b>
	<i>W. R. Gilks</i>	
5.1	Introduction	75
5.2	Deriving full conditional distributions	75
5.2.1	A simple example	75
5.2.2	Graphical models	77
5.3	Sampling from full conditional distributions	78

5.3.1	Rejection sampling	79
5.3.2	Ratio-of-uniforms method	80
5.3.3	Adaptive rejection sampling	82
5.3.4	Metropolis–Hastings algorithm	84
5.3.5	Hybrid adaptive rejection and Metropolis–Hastings	85
5.4	Discussion	86
<b>6</b>	<b>Strategies for improving MCMC</b>	<b>89</b>
	<i>W. R. Gilks and G. O. Roberts</i>	
6.1	Introduction	89
6.2	Reparameterization	90
6.2.1	Correlations and transformations	90
6.2.2	Linear regression models	92
6.2.3	Random-effects models	93
6.2.4	Nonlinear models	96
6.2.5	General comments on reparameterization	97
6.3	Random and adaptive direction sampling	98
6.3.1	The hit-and-run algorithm	98
6.3.2	Adaptive direction sampling (ADS)	99
6.4	Modifying the stationary distribution	101
6.4.1	Importance sampling	101
6.4.2	Metropolis-coupled MCMC	103
6.4.3	Simulated tempering	104
6.4.4	Auxiliary variables	105
6.5	Methods based on continuous-time processes	108
6.6	Discussion	110
<b>7</b>	<b>Implementing MCMC</b>	<b>115</b>
	<i>A. E. Raftery and S. M. Lewis</i>	
7.1	Introduction	115
7.2	Determining the number of iterations	116
7.3	Software and implementation	118
7.4	Output analysis	119
7.4.1	An example	120
7.5	Generic Metropolis algorithms	121
7.5.1	An example	124
7.6	Discussion	127
<b>8</b>	<b>Inference and monitoring convergence</b>	<b>131</b>
	<i>A. Gelman</i>	
8.1	Difficulties in inference from Markov chain simulation	131
8.2	The risk of undiagnosed slow convergence	132
8.3	Multiple sequences and overdispersed starting points	135
8.4	Monitoring convergence using simulation output	136

PSRF!

8.5	Output analysis for inference	139
8.6	Output analysis for improving efficiency	140
<b>9</b>	<b>Model determination using sampling-based methods</b>	<b>145</b>
	<i>A. E. Gelfand</i>	
9.1	Introduction	145
9.2	Classical approaches	146
9.3	The Bayesian perspective and the Bayes factor	148
9.4	Alternative predictive distributions	149
	9.4.1 Cross-validation predictive densities	150
	9.4.2 Posterior predictive densities	151
	9.4.3 Other predictive densities	151
9.5	How to use predictive distributions	151
9.6	Computational issues	154
	9.6.1 Estimating predictive densities	154
	9.6.2 Computing expectations over predictive densities	155
	9.6.3 Sampling from predictive densities	156
9.7	An example	157
9.8	Discussion	158
<b>10</b>	<b>Hypothesis testing and model selection</b>	<b>163</b>
	<i>A. E. Raftery</i>	
10.1	Introduction	163
10.2	Uses of Bayes factors	165
10.3	Marginal likelihood estimation by importance sampling	167
10.4	Marginal likelihood estimation using maximum likelihood	170
	10.4.1 The Laplace–Metropolis estimator	170
	10.4.2 Candidate’s estimator	172
	10.4.3 The data-augmentation estimator	173
10.5	Application: how many components in a mixture?	176
	10.5.1 Gibbs sampling for Gaussian mixtures	176
	10.5.2 A simulated example	177
	10.5.3 How many disks in the Galaxy?	180
10.6	Discussion	181
	Appendix: S-PLUS code for the Laplace–Metropolis estimator	186
<b>11</b>	<b>Model checking and model improvement</b>	<b>189</b>
	<i>A. Gelman and X.-L. Meng</i>	
11.1	Introduction	189
11.2	Model checking using posterior predictive simulation	189
11.3	Model improvement via expansion	192
11.4	Example: hierarchical mixture modelling of reaction times	193
	11.4.1 The data and the basic model	193
	11.4.2 Model checking using posterior predictive simulation	196

11.4.3 Expanding the model	196
11.4.4 Checking the new model	198
<b>12 Stochastic search variable selection</b>	<b>203</b>
<i>E. I. George and R. E. McCulloch</i>	
12.1 Introduction	203
12.2 A hierarchical Bayesian model for variable selection	204
12.3 Searching the posterior by Gibbs sampling	207
12.4 Extensions	209
12.4.1 SSVS for generalized linear models	209
12.4.2 SSVS across exchangeable regressions	210
12.5 Constructing stock portfolios with SSVS	211
12.6 Discussion	213
<b>13 Bayesian model comparison via jump diffusions</b>	<b>215</b>
<i>D. B. Phillips and A. F. M. Smith</i>	
13.1 Introduction	215
13.2 Model choice	216
13.2.1 Example 1: mixture deconvolution	216
13.2.2 Example 2: object recognition	218
13.2.3 Example 3: variable selection in regression	219
13.2.4 Example 4: change-point identification	220
13.3 Jump-diffusion sampling	221
13.3.1 The jump component	222
13.3.2 Moving between jumps	226
13.4 Mixture deconvolution	226
13.4.1 Dataset 1: galaxy velocities	228
13.4.2 Dataset 2: length of porgies	228
13.5 Object recognition	233
13.5.1 Results	233
13.6 Variable selection	235
13.7 Change-point identification	235
13.7.1 Dataset 1: Nile discharge	236
13.7.2 Dataset 2: facial image	236
13.8 Conclusions	238
<b>14 Estimation and optimization of functions</b>	<b>241</b>
<i>C. J. Geyer</i>	
14.1 Non-Bayesian applications of MCMC	241
14.2 Monte Carlo optimization	241
14.3 Monte Carlo likelihood analysis	244
14.4 Normalizing-constant families	245
14.5 Missing data	249
14.6 Decision theory	251

14.7 Which sampling distribution?	251
14.8 Importance sampling	253
14.9 Discussion	255
<b>15 Stochastic EM: method and application</b>	<b>259</b>
<i>J. Diebolt and E. H. S. Ip</i>	
15.1 Introduction	259
15.2 The EM algorithm	260
15.3 The stochastic EM algorithm	261
15.3.1 Stochastic imputation	261
15.3.2 Looking at the plausible region	262
15.3.3 Point estimation	263
15.3.4 Variance of the estimates	264
15.4 Examples	264
15.4.1 Type-I censored data	264
15.4.2 Empirical Bayes probit regression for cognitive diagnosis	268
<b>16 Generalized linear mixed models</b>	<b>275</b>
<i>D. G. Clayton</i>	
16.1 Introduction	275
16.2 Generalized linear models (GLMs)	276
16.3 Bayesian estimation of GLMs	277
16.4 Gibbs sampling for GLMs	278
16.5 Generalized linear mixed models (GLMMs)	279
16.5.1 Frequentist GLMMs	279
16.5.2 Bayesian GLMMs	280
16.6 Specification of random-effect distributions	281
16.6.1 Prior precision	281
16.6.2 Prior means	283
16.6.3 Intrinsic aliasing and contrasts	283
16.6.4 Autocorrelated random effects	286
16.6.5 The first-difference prior	287
16.6.6 The second-difference prior	288
16.6.7 General Markov random field priors	289
16.6.8 Interactions	289
16.7 Hyperpriors and the estimation of hyperparameters	291
16.8 Some examples	292
16.8.1 Longitudinal studies	293
16.8.2 Time trends for disease incidence and mortality	293
16.8.3 Disease maps and ecological analysis	294
16.8.4 Simultaneous variation in space and time	296
16.8.5 Frailty models in survival analysis	296
16.9 Discussion	298

<b>17 Hierarchical longitudinal modelling</b>	<b>303</b>
<i>B. P. Carlin</i>	
17.1 Introduction	303
17.2 Clinical background	305
17.3 Model detail and MCMC implementation	306
17.4 Results	309
17.5 Summary and discussion	315
<b>18 Medical monitoring</b>	<b>321</b>
<i>C. Berzuini</i>	
18.1 Introduction	321
18.2 Modelling medical monitoring	322
18.2.1 Nomenclature and data	322
18.2.2 Linear growth model	323
18.2.3 Marker growth as a stochastic process	324
18.3 Computing posterior distributions	327
18.3.1 Recursive updating	327
18.4 Forecasting	329
18.5 Model criticism	330
18.6 Illustrative application	330
18.6.1 The clinical problem	330
18.6.2 The model	332
18.6.3 Parameter estimates	332
18.6.4 Predicting deterioration	333
18.7 Discussion	335
<b>19 MCMC for nonlinear hierarchical models</b>	<b>339</b>
<i>J. E. Bennett, A. Racine-Poon and J. C. Wakefield</i>	
19.1 Introduction	339
19.2 Implementing MCMC	341
19.2.1 Method 1: Rejection Gibbs	342
19.2.2 Method 2: Ratio Gibbs	343
19.2.3 Method 3: Random-walk Metropolis	343
19.2.4 Method 4: Independence Metropolis–Hastings	344
19.2.5 Method 5: MLE/prior Metropolis–Hastings	344
19.3 Comparison of strategies	344
19.3.1 Guinea pigs data	345
19.4 A case study from pharmacokinetics-pharmacodynamics	348
19.5 Extensions and discussion	350
<b>20 Bayesian mapping of disease</b>	<b>359</b>
<i>A. Mollié</i>	
20.1 Introduction	359
20.2 Hypotheses and notation	360



20.3	Maximum likelihood estimation of relative risks	360
20.4	Hierarchical Bayesian model of relative risks	363
20.4.1	Bayesian inference for relative risks	363
20.4.2	Specification of the prior distribution	364
20.4.3	Graphical representation of the model	366
20.5	Empirical Bayes estimation of relative risks	369
20.5.1	The conjugate gamma prior	369
20.5.2	Non-conjugate priors	370
20.5.3	Disadvantages of EB estimation	371
20.6	Fully Bayesian estimation of relative risks	371
20.6.1	Choices for hyperpriors	371
20.6.2	Full conditional distributions for Gibbs sampling	372
20.6.3	Example: gall-bladder and bile-duct cancer mortality	375
20.7	Discussion	376
<b>21</b>	<b>MCMC in image analysis</b>	<b>381</b>
	<i>P. J. Green</i>	
21.1	Introduction	381
21.2	The relevance of MCMC to image analysis	382
21.3	Image models at different levels	383
21.3.1	Pixel-level models	383
21.3.2	Pixel-based modelling in SPECT	385
21.3.3	Template models	389
21.3.4	An example of template modelling	391
21.3.5	Stochastic geometry models	392
21.3.6	Hierarchical modelling	393
21.4	Methodological innovations in MCMC stimulated by imaging	394
21.5	Discussion	395
<b>22</b>	<b>Measurement error</b>	<b>401</b>
	<i>S. Richardson</i>	
22.1	Introduction	401
22.2	Conditional-independence modelling	403
22.2.1	Designs with individual-level surrogates	403
22.2.2	Designs using ancillary risk-factor information	406
22.2.3	Estimation	407
22.3	Illustrative examples	408
22.3.1	Two measuring instruments with no validation group	408
22.3.2	Influence of the exposure model	411
22.3.3	Ancillary risk-factor information and expert coding	412
22.4	Discussion	414
<b>23</b>	<b>Gibbs sampling methods in genetics</b>	<b>419</b>
	<i>D. C. Thomas and W. J. Gauderman</i>	

23.1	Introduction	419
23.2	Standard methods in genetics	419
23.2.1	Genetic terminology	419
23.2.2	Genetic models	421
23.2.3	Genetic likelihoods	423
23.3	Gibbs sampling approaches	424
23.3.1	Gibbs sampling of genotypes	425
23.3.2	Gibbs sampling of parameters	429
23.3.3	Initialization, convergence, and fine tuning	431
23.4	MCMC maximum likelihood	432
23.5	Application to a family study of breast cancer	434
23.6	Conclusions	437
<b>24</b>	<b>Mixtures of distributions: inference and estimation</b>	<b>441</b>
	<i>C. P. Robert</i>	
24.1	Introduction	441
24.1.1	Modelling via mixtures	441
24.1.2	A first example: character recognition	442
24.1.3	Estimation methods	443
24.1.4	Bayesian estimation	445
24.2	The missing data structure	446
24.3	Gibbs sampling implementation	448
24.3.1	General algorithm	448
24.3.2	Extra-binomial variation	449
24.3.3	Normal mixtures: star clustering	449
24.3.4	Reparameterization issues	452
24.3.5	Extra-binomial variation: continued	455
24.4	Convergence of the algorithm	455
24.5	Testing for mixtures	456
24.5.1	Extra-binomial variation: continued	458
24.6	Infinite mixtures and other extensions	459
24.6.1	Dirichlet process priors and nonparametric models	459
24.6.2	Hidden Markov models	461
<b>25</b>	<b>An archaeological example: radiocarbon dating</b>	<b>465</b>
	<i>C. Litton and C. Buck</i>	
25.1	Introduction	465
25.2	Background to radiocarbon dating	466
25.3	Archaeological problems and questions	469
25.4	Illustrative examples	470
25.4.1	Example 1: dating settlements	470
25.4.2	Example 2: dating archaeological phases	473
25.4.3	Example 3: accommodating outliers	476
25.4.4	Practical considerations	477

## 25.5 Discussion

## Index