Contents

Preface xiii

1 Introduction 1
- 1.1 Purpose 1
- 1.2 Background 2
  - 1.2.1 The EM algorithm 3
  - 1.2.2 Markov chain Monte Carlo 3
- 1.3 Why analysis by simulation? 4
- 1.4 Looking ahead 6
  - 1.4.1 Scope of the rest of this book 6
  - 1.4.2 Knowledge assumed on the part of the reader 7
  - 1.4.3 Software and computational details 7
- 1.5 Bibliographic notes 8

2 Assumptions 9
- 2.1 The complete-data model 9
- 2.2 Ignorability 10
  - 2.2.1 Missing at random 10
  - 2.2.2 Distinctness of parameters 11
- 2.3 The observed-data likelihood and posterior 11
  - 2.3.1 Observed-data likelihood 11
  - 2.3.2 Examples 13
  - 2.3.3 Observed-data posterior 17
- 2.4 Examining the ignorability assumption 20
  - 2.4.1 Examples where ignorability is known to hold 20
  - 2.4.2 Examples where ignorability is not known to hold 22
  - 2.4.3 Ignorability is relative 23
- 2.5 General ignorable procedures 23
  - 2.5.1 A simulated example 24

2.5.2 Departures from ignorability 26
2.5.3 Notes on nonignorable alternatives 28
2.6 The role of the complete-data model 29
2.6.1 Departures from the data model 29
2.6.2 Inference treating certain variables as fixed 31

3 EM and data augmentation **37**
3.1 Introduction 37
3.2 The EM algorithm 37
3.2.1 Definition 37
3.2.2 Examples 41
3.2.3 EM for posterior modes 46
3.2.4 Restrictions on the parameter space 46
3.2.5 The ECM algorithm 49
3.3 Properties of EM 51
3.3.1 Stationary values 51
3.3.2 Rate of convergence 55
3.3.3 Example 59
3.3.4 Further comments on convergence 61
3.4 Markov chain Monte Carlo 68
3.4.1 Gibbs sampling 69
3.4.2 Data augmentation 70
3.4.3 Examples of data augmentation 73
3.4.4 The Metropolis-Hastings algorithm 78
3.4.5 Generalizations and hybrid algorithms 79
3.5 Properties of Markov chain Monte Carlo 80
3.5.1 The meaning of convergence 80
3.5.2 Examples of nonconvergence 80
3.5.3 Rates of convergence 83

4 Inference by data augmentation **89**
4.1 Introduction 89
4.2 Parameter simulation 90
4.2.1 Dependent samples 90
4.2.2 Summarizing a dependent sample 93
4.2.3 Rao-Blackwellized estimates 98
4.3 Multiple imputation 104
4.3.1 Bayesianly proper multiple imputations 105
4.3.2 Inference for a scalar quantity 107
4.3.3 Inference for multidimensional estimands 112
4.4 Assessing convergence 118
4.4.1 Monitoring convergence in a single chain 119

4.4.2 Monitoring convergence with parallel chains 126
4.4.3 Choosing scalar functions of the parameter 128
4.4.4 Convergence of posterior summaries 131
4.5 Practical guidelines 134
4.5.1 Choosing a method of inference 135
4.5.2 Implementing a parameter-simulation experiment 136
4.5.3 Generating multiple imputations 138
4.5.4 Choosing an imputation model 139
4.5.5 Further comments on imputation modeling 143

5 Methods for normal data **147**
5.1 Introduction 147
5.2 Relevant properties of the complete-data model 148
5.2.1 Basic notation 148
5.2.2 Bayesian inference under a conjugate prior 150
5.2.3 Choosing the prior hyperparameters 154
5.2.4 Alternative parameterizations and sweep 157
5.3 The EM algorithm 163
5.3.1 Preliminary manipulations 163
5.3.2 The E-step 164
5.3.3 Implementation of the algorithm 166
5.3.4 EM for posterior modes 170
5.3.5 Calculating the observed-data loglikelihood 173
5.3.6 Example: serum-cholesterol levels of heart-attack patients 175
5.3.7 Example: changes in heart rate due to marijuana use 178
5.4 Data augmentation 181
5.4.1 The I-step 181
5.4.2 The P-step 183
5.4.3 Example: cholesterol levels of heart-attack patients 185
5.4.4 Example: changes in heart rate due to marijuana use 189

6 More on the normal model **193**
6.1 Introduction 193
6.2 Multiple imputation: example 1 193
6.2.1 Cholesterol levels of heart-attack patients 193
6.2.2 Generating the imputations 194
6.2.3 Complete-data point and variance estimates 194

9.3.2 Likelihood inference for restricted models 344
9.3.3 Bayesian inference 346
9.4 Algorithms for incomplete mixed data 348
9.4.1 Predictive distributions 348
9.4.2 EM for the unrestricted model 352
9.4.3 Data augmentation 355
9.4.4 Algorithms for restricted models 357
9.5 Data examples 359
9.5.1 St. Louis Risk Research Project 359
9.5.2 Foreign Language Attitude Scale 367
9.5.3 National Health and Nutrition Examination Survey 372

10 Further topics **379**
10.1 Introduction 379
10.2 Extensions of the normal model 379
10.2.1 Restricted covariance structures 379
10.2.2 Heavy-tailed distributions 380
10.2.3 Interactions 380
10.2.4 Semicontinuous variables 381
10.3 Random-effects models 382
10.4 Models for complex survey data 383
10.5 Nonignorable methods 384
10.6 Mixture models and latent variables 384
10.7 Coarsened data and outlier models 385
10.8 Diagnostics 386

Appendices
A Data examples 387
B Storage of categorical data 395
C Software 399

References **401**

Index **415**

6.2.4 Combining the estimates 197
6.2.5 Alternative choices for the number of imputations 197
6.3 Multiple imputation: example 2 200
6.3.1 Predicting achievement in foreign language study 200
6.3.2 Applying the normal model 202
6.3.3 Exploring the observed-data likelihood and posterior 204
6.3.4 Overcoming the problem of inestimability 206
6.3.5 Analysis by multiple imputation 208
6.4 A simulation study 211
6.4.1 Simulation procedures 212
6.4.2 Complete-data inferences 214
6.4.3 Results 216
6.5 Fast algorithms based on factored likelihoods 218
6.5.1 Monotone missingness patterns 218
6.5.2 Computing alternative parameterizations 220
6.5.3 Noniterative inference for monotone data 223
6.5.4 Monotone data augmentation 226
6.5.5 Implementation of the algorithm 229
6.5.6 Uses and extensions 234
6.5.7 Example 236

7 Methods for categorical data 239
7.1 Introduction 239
7.2 The multinomial model and Dirichlet prior 240
7.2.1 The multinomial distribution 240
7.2.2 Collapsing and partitioning the multinomial 243
7.2.3 The Dirichlet distribution 247
7.2.4 Bayesian inference 250
7.2.5 Choosing the prior hyperparameters 251
7.2.6 Collapsing and partitioning the Dirichlet 255
7.3 Basic algorithms for the saturated model 257
7.3.1 Characterizing an incomplete categorical dataset 257
7.3.2 The EM algorithm 260
7.3.3 Data augmentation 264
7.3.4 Example: victimization status from the National Crime Survey 267
7.3.5 Example: Protective Services Project for Older Persons 272

7.4 Fast algorithms for near-monotone patterns 275
7.4.1 Factoring the likelihood and prior density 275
7.4.2 Monotone data augmentation 279
7.4.3 Example: driver injury and seatbelt use 282

8 Loglinear models 289
8.1 Introduction 289
8.2 Overview of loglinear models 289
8.2.1 Definition 289
8.2.2 Eliminating associations 292
8.2.3 Sufficient statistics 294
8.2.4 Model interpretation 295
8.3 Likelihood-based inference with complete data 297
8.3.1 Maximum-likelihood estimation 297
8.3.2 Iterative proportional fitting 298
8.3.3 Hypothesis testing and goodness of fit 302
8.3.4 Example: misclassification of seatbelt use and injury 303
8.4 Bayesian inference with complete data 305
8.4.1 Prior distributions for loglinear models 305
8.4.2 Inference using posterior modes 307
8.4.3 Inference by Bayesian IPF 308
8.4.4 Why Bayesian IPF works 312
8.4.5 Example: misclassification of seatbelt use and injury 318
8.5 Loglinear modeling with incomplete data 320
8.5.1 ML estimates and posterior modes 320
8.5.2 Goodness-of-fit statistics 322
8.5.3 Data augmentation and Bayesian IPF 324
8.6 Examples 325
8.6.1 Protective Services Project for Older Persons 325
8.6.2 Driver injury and seatbelt use 328

9 Methods for mixed data 333
9.1 Introduction 333
9.2 The general location model 334
9.2.1 Definition 334
9.2.2 Complete-data likelihood 336
9.2.3 Example 338
9.2.4 Complete-data Bayesian inference 339
9.3 Restricted models 341
9.3.1 Reducing the number of parameters 341