
Contents

| | | |
|----------|---------------------------------------------------------------------------------------------------|----|
| 1 | Finite Mixture Modeling | 1 |
| 1.1 | Introduction | 1 |
| 1.2 | Finite Mixture Distributions | 3 |
| 1.2.1 | Basic Definitions | 3 |
| 1.2.2 | Some Descriptive Features of Finite Mixture Distributions | 5 |
| 1.2.3 | Diagnosing Similarity of Mixture Components..... | 9 |
| 1.2.4 | Moments of a Finite Mixture Distribution | 10 |
| 1.2.5 | Statistical Modeling Based on Finite Mixture Distributions | 11 |
| 1.3 | Identifiability of a Finite Mixture Distribution | 14 |
| 1.3.1 | Nonidentifiability Due to Invariance to Relabeling the Components..... | 15 |
| 1.3.2 | Nonidentifiability Due to Potential Overfitting | 17 |
| 1.3.3 | Formal Identifiability Constraints..... | 19 |
| 1.3.4 | Generic Identifiability | 21 |
| 2 | Statistical Inference for a Finite Mixture Model with Known Number of Components | 25 |
| 2.1 | Introduction | 25 |
| 2.2 | Classification for Known Component Parameters | 26 |
| 2.2.1 | Bayes' Rule for Classifying a Single Observation..... | 26 |
| 2.2.2 | The Bayes' Classifier for a Whole Data Set | 27 |
| 2.3 | Parameter Estimation for Known Allocation | 29 |
| 2.3.1 | The Complete-Data Likelihood Function | 29 |
| 2.3.2 | Complete-Data Maximum Likelihood Estimation | 30 |
| 2.3.3 | Complete-Data Bayesian Estimation of the Component Parameters | 31 |
| 2.3.4 | Complete-Data Bayesian Estimation of the Weights.... | 35 |
| 2.4 | Parameter Estimation When the Allocations Are Unknown ... | 41 |
| 2.4.1 | Method of Moments | 42 |

| | | |
|----------|--------------------------------------------------------------------------------------------------------------|-----------|
| 2.4.2 | The Mixture Likelihood Function | 43 |
| 2.4.3 | A Helicopter Tour of the Mixture Likelihood Surface for Two Examples | 44 |
| 2.4.4 | Maximum Likelihood Estimation | 49 |
| 2.4.5 | Bayesian Parameter Estimation | 53 |
| 2.4.6 | Distance-Based Methods | 54 |
| 2.4.7 | Comparing Various Estimation Methods | 54 |
| 3 | Practical Bayesian Inference for a Finite Mixture Model with Known Number of Components | 57 |
| 3.1 | Introduction | 57 |
| 3.2 | Choosing the Prior for the Parameters of a Mixture Model | 58 |
| 3.2.1 | Objective and Subjective Priors | 58 |
| 3.2.2 | Improper Priors May Cause Improper Mixture Posteriors | 59 |
| 3.2.3 | Conditionally Conjugate Priors | 60 |
| 3.2.4 | Hierarchical Priors and Partially Proper Priors | 61 |
| 3.2.5 | Other Priors | 62 |
| 3.2.6 | Invariant Prior Distributions | 62 |
| 3.3 | Some Properties of the Mixture Posterior Density | 63 |
| 3.3.1 | Invariance of the Posterior Distribution | 63 |
| 3.3.2 | Invariance of Seemingly Component- Specific Functionals | 64 |
| 3.3.3 | The Marginal Posterior Distribution of the Allocations | 65 |
| 3.3.4 | Invariance of the Posterior Distribution of the Allocations | 67 |
| 3.4 | Classification Without Parameter Estimation | 68 |
| 3.4.1 | Single-Move Gibbs Sampling | 69 |
| 3.4.2 | The Metropolis–Hastings Algorithm | 72 |
| 3.5 | Parameter Estimation Through Data Augmentation and MCMC | 73 |
| 3.5.1 | Treating Mixture Models as a Missing Data Problem | 73 |
| 3.5.2 | Data Augmentation and MCMC for a Mixture of Poisson Distributions | 74 |
| 3.5.3 | Data Augmentation and MCMC for General Mixtures | 76 |
| 3.5.4 | MCMC Sampling Under Improper Priors | 78 |
| 3.5.5 | Label Switching | 78 |
| 3.5.6 | Permutation MCMC Sampling | 81 |
| 3.6 | Other Monte Carlo Methods Useful for Mixture Models | 83 |
| 3.6.1 | A Metropolis–Hastings Algorithm for the Parameters | 83 |
| 3.6.2 | Importance Sampling for the Allocations | 84 |
| 3.6.3 | Perfect Sampling | 85 |

- 3.7 Bayesian Inference for Finite Mixture Models Using Posterior Draws 85
 - 3.7.1 Sampling Representations of the Mixture Posterior Density 85
 - 3.7.2 Using Posterior Draws for Bayesian Inference 87
 - 3.7.3 Predictive Density Estimation 89
 - 3.7.4 Individual Parameter Inference 91
 - 3.7.5 Inference on the Hyperparameter of a Hierarchical Prior 92
 - 3.7.6 Inference on Component Parameters 92
 - 3.7.7 Model Identification 94

- 4 **Statistical Inference for Finite Mixture Models Under Model Specification Uncertainty** 99
 - 4.1 Introduction 99
 - 4.2 Parameter Estimation Under Model Specification Uncertainty 100
 - 4.2.1 Maximum Likelihood Estimation Under Model Specification Uncertainty 100
 - 4.2.2 Practical Bayesian Parameter Estimation for Overfitting Finite Mixture Models 103
 - 4.2.3 Potential Overfitting 105
 - 4.3 Informal Methods for Identifying the Number of Components 107
 - 4.3.1 Mode Hunting in the Mixture Posterior 108
 - 4.3.2 Mode Hunting in the Sample Histogram 109
 - 4.3.3 Diagnosing Mixtures Through the Method of Moments 110
 - 4.3.4 Diagnosing Mixtures Through Predictive Methods 112
 - 4.3.5 Further Approaches 114
 - 4.4 Likelihood-Based Methods 114
 - 4.4.1 The Likelihood Ratio Statistic 114
 - 4.4.2 AIC, BIC, and the Schwarz Criterion 116
 - 4.4.3 Further Approaches 117
 - 4.5 Bayesian Inference Under Model Uncertainty 117
 - 4.5.1 Trans-Dimensional Bayesian Inference 117
 - 4.5.2 Marginal Likelihoods 118
 - 4.5.3 Bayes Factors for Model Comparison 119
 - 4.5.4 Formal Bayesian Model Selection 121
 - 4.5.5 Choosing Priors for Model Selection 122
 - 4.5.6 Further Approaches 123

| | | |
|----------|-----------------------------------------------------------------------------------------------------------------------|-----|
| 5 | Computational Tools for Bayesian Inference for Finite Mixtures Models Under Model Specification Uncertainty .. | 125 |
| 5.1 | Introduction | 125 |
| 5.2 | Trans-Dimensional Markov Chain Monte Carlo Methods | 125 |
| 5.2.1 | Product-Space MCMC | 126 |
| 5.2.2 | Reversible Jump MCMC | 129 |
| 5.2.3 | Birth and Death MCMC Methods | 137 |
| 5.3 | Marginal Likelihoods for Finite Mixture Models | 139 |
| 5.3.1 | Defining the Marginal Likelihood | 139 |
| 5.3.2 | Choosing Priors for Selecting the Number of Components | 141 |
| 5.3.3 | Computation of the Marginal Likelihood for Mixture Models | 143 |
| 5.4 | Simulation-Based Approximations of the Marginal Likelihood | 143 |
| 5.4.1 | Some Background on Monte Carlo Integration | 143 |
| 5.4.2 | Sampling-Based Approximations for Mixture Models .. | 144 |
| 5.4.3 | Importance Sampling | 146 |
| 5.4.4 | Reciprocal Importance Sampling | 147 |
| 5.4.5 | Harmonic Mean Estimator | 148 |
| 5.4.6 | Bridge Sampling Technique | 150 |
| 5.4.7 | Comparison of Different Simulation-Based Estimators | 154 |
| 5.4.8 | Dealing with Hierarchical Priors | 159 |
| 5.5 | Approximations to the Marginal Likelihood Based on Density Ratios | 159 |
| 5.5.1 | The Posterior Density Ratio | 159 |
| 5.5.2 | Chib's Estimator | 160 |
| 5.5.3 | Laplace Approximation | 164 |
| 5.6 | Reversible Jump MCMC Versus Marginal Likelihoods? | 165 |
| 6 | Finite Mixture Models with Normal Components | 169 |
| 6.1 | Finite Mixtures of Normal Distributions | 169 |
| 6.1.1 | Model Formulation | 169 |
| 6.1.2 | Parameter Estimation for Mixtures of Normals | 171 |
| 6.1.3 | The Kiefer–Wolfowitz Example | 174 |
| 6.1.4 | Applications of Mixture of Normal Distributions | 176 |
| 6.2 | Bayesian Estimation of Univariate Mixtures of Normals | 177 |
| 6.2.1 | Bayesian Inference When the Allocations Are Known | 177 |
| 6.2.2 | Standard Prior Distributions | 179 |
| 6.2.3 | The Influence of the Prior on the Variance Ratio | 179 |
| 6.2.4 | Bayesian Estimation Using MCMC | 180 |
| 6.2.5 | MCMC Estimation Under Standard Improper Priors .. | 182 |

| | | |
|----------|--------------------------------------------------------------------------------------|------------|
| 6.2.6 | Introducing Prior Dependence Among the Components | 185 |
| 6.2.7 | Further Sampling-Based Approaches | 187 |
| 6.2.8 | Application to the Fishery Data | 188 |
| 6.3 | Bayesian Estimation of Multivariate Mixtures of Normals | 190 |
| 6.3.1 | Bayesian Inference When the Allocations Are Known | 190 |
| 6.3.2 | Prior Distributions | 192 |
| 6.3.3 | Bayesian Parameter Estimation Using MCMC | 193 |
| 6.3.4 | Application to Fisher's Iris Data | 195 |
| 6.4 | Further Issues | 195 |
| 6.4.1 | Parsimonious Finite Normal Mixtures | 195 |
| 6.4.2 | Model Selection Problems for Mixtures of Normals | 199 |
| 7 | Data Analysis Based on Finite Mixtures | 203 |
| 7.1 | Model-Based Clustering | 203 |
| 7.1.1 | Some Background on Cluster Analysis | 203 |
| 7.1.2 | Model-Based Clustering Using Finite Mixture Models | 204 |
| 7.1.3 | The Classification Likelihood and the Bayesian MAP Approach | 207 |
| 7.1.4 | Choosing Clustering Criteria and the Number of Components | 210 |
| 7.1.5 | Model Choice for the Fishery Data | 216 |
| 7.1.6 | Model Choice for Fisher's Iris Data | 218 |
| 7.1.7 | Bayesian Clustering Based on Loss Functions | 220 |
| 7.1.8 | Clustering for Fisher's Iris Data | 224 |
| 7.2 | Outlier Modeling | 224 |
| 7.2.1 | Outlier Modeling Using Finite Mixtures | 224 |
| 7.2.2 | Bayesian Inference for Outlier Models Based on Finite Mixtures | 225 |
| 7.2.3 | Outlier Modeling of Darwin's Data | 226 |
| 7.2.4 | Clustering Under Outliers and Noise | 227 |
| 7.3 | Robust Finite Mixtures Based on the Student- <i>t</i> Distribution | 230 |
| 7.3.1 | Parameter Estimation | 230 |
| 7.3.2 | Dealing with Unknown Number of Components | 233 |
| 7.4 | Further Issues | 233 |
| 7.4.1 | Clustering High-Dimensional Data | 233 |
| 7.4.2 | Discriminant Analysis | 235 |
| 7.4.3 | Combining Classified and Unclassified Observations | 236 |
| 7.4.4 | Density Estimation Using Finite Mixtures | 237 |
| 7.4.5 | Finite Mixtures as an Auxiliary Computational Tool in Bayesian Analysis | 238 |

| | | |
|----------|----------------------------------------------------------------------------------|-----|
| 8 | Finite Mixtures of Regression Models | 241 |
| 8.1 | Introduction | 241 |
| 8.2 | Finite Mixture of Multiple Regression Models | 242 |
| 8.2.1 | Model Definition | 242 |
| 8.2.2 | Identifiability | 243 |
| 8.2.3 | Statistical Modeling Based on Finite Mixture of Regression Models | 246 |
| 8.2.4 | Outliers in a Regression Model | 249 |
| 8.3 | Statistical Inference for Finite Mixtures of Multiple Regression Models | 249 |
| 8.3.1 | Maximum Likelihood Estimation | 249 |
| 8.3.2 | Bayesian Inference When the Allocations Are Known | 250 |
| 8.3.3 | Choosing Prior Distributions | 252 |
| 8.3.4 | Bayesian Inference When the Allocations Are Unknown | 253 |
| 8.3.5 | Bayesian Inference Using Posterior Draws | 254 |
| 8.3.6 | Dealing with Model Specification Uncertainty | 255 |
| 8.4 | Mixed-Effects Finite Mixtures of Regression Models | 256 |
| 8.4.1 | Model Definition | 256 |
| 8.4.2 | Choosing Priors for Bayesian Estimation | 256 |
| 8.4.3 | Bayesian Parameter Estimation When the Allocations Are Known | 257 |
| 8.4.4 | Bayesian Parameter Estimation When the Allocations Are Unknown | 258 |
| 8.5 | Finite Mixture Models for Repeated Measurements | 259 |
| 8.5.1 | Pooling Information Across Similar Units | 260 |
| 8.5.2 | Finite Mixtures of Random-Effects Models | 260 |
| 8.5.3 | Choosing the Prior for Bayesian Estimation | 265 |
| 8.5.4 | Bayesian Parameter Estimation When the Allocations Are Known | 265 |
| 8.5.5 | Practical Bayesian Estimation Using MCMC | 267 |
| 8.5.6 | Dealing with Model Specification Uncertainty | 269 |
| 8.5.7 | Application to the Marketing Data | 270 |
| 8.6 | Further Issues | 273 |
| 8.6.1 | Regression Modeling Based on Multivariate Mixtures of Normals | 273 |
| 8.6.2 | Modeling the Weight Distribution | 274 |
| 8.6.3 | Mixtures-of-Experts Models | 274 |
| 9 | Finite Mixture Models with Nonnormal Components | 277 |
| 9.1 | Finite Mixtures of Exponential Distributions | 277 |
| 9.1.1 | Model Formulation and Parameter Estimation | 277 |
| 9.1.2 | Bayesian Inference | 278 |
| 9.2 | Finite Mixtures of Poisson Distributions | 279 |

| | | |
|-----------|-----------------------------------------------------------------------------------------|------------|
| 9.2.1 | Model Formulation and Estimation | 279 |
| 9.2.2 | Capturing Overdispersion in Count Data | 280 |
| 9.2.3 | Modeling Excess Zeros | 282 |
| 9.2.4 | Application to the Eye Tracking Data | 283 |
| 9.3 | Finite Mixture Models for Binary and Categorical Data | 286 |
| 9.3.1 | Finite Mixtures of Binomial Distributions | 286 |
| 9.3.2 | Finite Mixtures of Multinomial Distributions | 288 |
| 9.4 | Finite Mixtures of Generalized Linear Models | 289 |
| 9.4.1 | Finite Mixture Regression Models for Count Data | 290 |
| 9.4.2 | Finite Mixtures of Logit and Probit Regression Models | 292 |
| 9.4.3 | Parameter Estimation for Finite Mixtures of GLMs | 293 |
| 9.4.4 | Model Selection for Finite Mixtures of GLMs | 294 |
| 9.5 | Finite Mixture Models for Multivariate Binary and Categorical Data | 294 |
| 9.5.1 | The Basic Latent Class Model | 295 |
| 9.5.2 | Identification and Parameter Estimation | 296 |
| 9.5.3 | Extensions of the Basic Latent Class Model | 297 |
| 9.6 | Further Issues | 298 |
| 9.6.1 | Finite Mixture Modeling of Mixed-Mode Data | 298 |
| 9.6.2 | Finite Mixtures of GLMs with Random Effects | 299 |
| 10 | Finite Markov Mixture Modeling | 301 |
| 10.1 | Introduction | 301 |
| 10.2 | Finite Markov Mixture Distributions | 301 |
| 10.2.1 | Basic Definitions | 302 |
| 10.2.2 | Irreducible Aperiodic Markov Chains | 304 |
| 10.2.3 | Moments of a Markov Mixture Distribution | 308 |
| 10.2.4 | The Autocorrelation Function of a Process Generated by a Markov Mixture Distribution | 310 |
| 10.2.5 | The Autocorrelation Function of the Squared Process | 311 |
| 10.2.6 | The Standard Finite Mixture Distribution as a Limiting Case | 312 |
| 10.2.7 | Identifiability of a Finite Markov Mixture Distribution | 313 |
| 10.3 | Statistical Modeling Based on Finite Markov Mixture Distributions | 314 |
| 10.3.1 | The Basic Markov Switching Model | 314 |
| 10.3.2 | The Markov Switching Regression Model | 315 |
| 10.3.3 | Nonergodic Markov Chains | 316 |
| 10.3.4 | Relaxing the Assumptions of the Basic Markov Switching Model | 316 |

| | | |
|-----------|-----------------------------------------------------------------------------------------------|-----|
| 11 | Statistical Inference for Markov Switching Models | 319 |
| 11.1 | Introduction | 319 |
| 11.2 | State Estimation for Known Parameters | 319 |
| 11.2.1 | Statistical Inference About the States | 320 |
| 11.2.2 | Filtered State Probabilities | 320 |
| 11.2.3 | Filtering for Special Cases | 323 |
| 11.2.4 | Smoothing the States | 324 |
| 11.2.5 | Filtering and Smoothing for More General Models | 326 |
| 11.3 | Parameter Estimation for Known States | 327 |
| 11.3.1 | The Complete-Data Likelihood Function | 327 |
| 11.3.2 | Complete-Data Bayesian Parameter Estimation | 329 |
| 11.3.3 | Complete-Data Bayesian Estimation of the Transition Matrix | 329 |
| 11.4 | Parameter Estimation When the States are Unknown | 330 |
| 11.4.1 | The Markov Mixture Likelihood Function | 330 |
| 11.4.2 | Maximum Likelihood Estimation | 333 |
| 11.4.3 | Bayesian Estimation | 334 |
| 11.4.4 | Alternative Estimation Methods | 334 |
| 11.5 | Bayesian Parameter Estimation with Known Number of States | 335 |
| 11.5.1 | Choosing the Prior for the Parameters of a Markov Mixture Model | 335 |
| 11.5.2 | Some Properties of the Posterior Distribution of a Markov Switching Model | 336 |
| 11.5.3 | Parameter Estimation Through Data Augmentation and MCMC | 337 |
| 11.5.4 | Permutation MCMC Sampling | 340 |
| 11.5.5 | Sampling the Unknown Transition Matrix | 340 |
| 11.5.6 | Sampling Posterior Paths of the Hidden Markov Chain | 342 |
| 11.5.7 | Other Sampling-Based Approaches | 345 |
| 11.5.8 | Bayesian Inference Using Posterior Draws | 345 |
| 11.6 | Statistical Inference Under Model Specification Uncertainty | 346 |
| 11.6.1 | Diagnosing Markov Switching Models | 346 |
| 11.6.2 | Likelihood-Based Methods | 346 |
| 11.6.3 | Marginal Likelihoods for Markov Switching Models | 347 |
| 11.6.4 | Model Space MCMC | 348 |
| 11.6.5 | Further Issues | 348 |
| 11.7 | Modeling Overdispersion and Autocorrelation in Time Series of Count Data | 348 |
| 11.7.1 | Motivating Example | 348 |
| 11.7.2 | Capturing Overdispersion and Autocorrelation Using Poisson Markov Mixture Models | 349 |
| 11.7.3 | Application to the Lamb Data | 351 |

12 Nonlinear Time Series Analysis Based on Markov Switching Models 357

12.1 Introduction 357

12.2 The Markov Switching Autoregressive Model 358

 12.2.1 Motivating Example 358

 12.2.2 Model Definition 360

 12.2.3 Features of the MSAR Model 362

 12.2.4 Markov Switching Models for Nonstationary Time Series 363

 12.2.5 Parameter Estimation and Model Selection 365

 12.2.6 Application to Business Cycle Analysis of the U.S. GDP Data 365

12.3 Markov Switching Dynamic Regression Models 371

 12.3.1 Model Definition 371

 12.3.2 Bayesian Estimation 371

12.4 Prediction of Time Series Based on Markov Switching Models 372

 12.4.1 Flexible Predictive Distributions 372

 12.4.2 Forecasting of Markov Switching Models via Sampling-Based Methods 374

12.5 Markov Switching Conditional Heteroscedasticity 375

 12.5.1 Motivating Example 375

 12.5.2 Capturing Features of Financial Time Series Through Markov Switching Models 377

 12.5.3 Switching ARCH Models 378

 12.5.4 Statistical Inference for Switching ARCH Models 380

 12.5.5 Switching GARCH Models 383

12.6 Some Extensions 384

 12.6.1 Time-Varying Transition Matrices 384

 12.6.2 Markov Switching Models for Longitudinal and Panel Data 385

 12.6.3 Markov Switching Models for Multivariate Time Series 386

13 Switching State Space Models 389

13.1 State Space Modeling 389

 13.1.1 The Local Level Model with and Without Switching 389

 13.1.2 The Linear Gaussian State Space Form 391

 13.1.3 Multiprocess Models 393

 13.1.4 Switching Linear Gaussian State Space Models 393

 13.1.5 The General State Space Form 394

13.2 Nonlinear Time Series Analysis Based on Switching State Space Models 396

 13.2.1 ARMA Models with and Without Switching 396

XVIII Contents

| | | |
|----------|----------------------------------------------------------------------------------------|-----|
| 13.2.2 | Unobserved Component Time Series Models | 397 |
| 13.2.3 | Capturing Sudden Changes in Time Series | 398 |
| 13.2.4 | Switching Dynamic Factor Models | 400 |
| 13.2.5 | Switching State Space Models as a Semi-Parametric Smoothing Device | 401 |
| 13.3 | Filtering for Switching Linear Gaussian State Space Models | 401 |
| 13.3.1 | The Filtering Problem | 402 |
| 13.3.2 | Bayesian Inference for a General Linear Regression Model | 402 |
| 13.3.3 | Filtering for the Linear Gaussian State Space Model | 404 |
| 13.3.4 | Filtering for Multiprocess Models | 406 |
| 13.3.5 | Approximate Filtering for Switching Linear Gaussian State Space Models | 406 |
| 13.4 | Parameter Estimation for Switching State Space Models | 410 |
| 13.4.1 | The Likelihood Function of a State Space Model | 411 |
| 13.4.2 | Maximum Likelihood Estimation | 412 |
| 13.4.3 | Bayesian Inference | 412 |
| 13.5 | Practical Bayesian Estimation Using MCMC | 415 |
| 13.5.1 | Various Data Augmentation Schemes | 416 |
| 13.5.2 | Sampling the Continuous State Process from the Smoother Density | 417 |
| 13.5.3 | Sampling the Discrete States for a Switching State Space Model | 420 |
| 13.6 | Further Issues | 421 |
| 13.6.1 | Model Specification Uncertainty in Switching State Space Modeling | 421 |
| 13.6.2 | Auxiliary Mixture Sampling for Nonlinear and Nonnormal State Space Models | 422 |
| 13.7 | Illustrative Application to Modeling Exchange Rate Data | 423 |
| A | Appendix | 431 |
| A.1 | Summary of Probability Distributions | 431 |
| A.1.1 | The Beta Distribution | 431 |
| A.1.2 | The Binomial Distribution | 432 |
| A.1.3 | The Dirichlet Distribution | 432 |
| A.1.4 | The Exponential Distribution | 433 |
| A.1.5 | The F-Distribution | 433 |
| A.1.6 | The Gamma Distribution | 434 |
| A.1.7 | The Geometric Distribution | 435 |
| A.1.8 | The Multinomial Distribution | 435 |
| A.1.9 | The Negative Binomial Distribution | 435 |
| A.1.10 | The Normal Distribution | 436 |
| A.1.11 | The Poisson Distribution | 437 |
| A.1.12 | The Student- <i>t</i> Distribution | 437 |
| A.1.13 | The Uniform Distribution | 438 |

| | |
|---------------------------------------|-----|
| A.1.14 The Wishart Distribution | 438 |
| A.2 Software | 439 |
| References | 441 |
| Index | 481 |