Contents

Preface xi
Acknowledgements xiii

Chapter 1 Introduction: The Bayesian Method, its Benefits and Implementation 1
1.1 Advantages of the Bayes approach 1
1.2 Statements about uncertainty and Bayesian updating 2
1.2.1 Prior and likelihood 3
1.3 Using existing knowledge to specify priors 4
1.4 The Sampling perspective: estimating the densities of parameters 6
1.5 Predictions from sampling 7
1.6 The present book 8

Chapter 2 Standard Distributions: Updating, Inference and Prediction 11
2.1 Introduction 11
2.2 Continuous data models: univariate normal 12
2.2.1 Estimating a normal mean when the variance is known 12
2.2.2 Predictions and predictive distributions 13
2.2.3 Samples of $n > 1$ observations 14
2.2.4 Hypothesis tests 15
2.2.5 Interval and point hypotheses 16
2.2.6 Posterior predictive checks 17
2.3 Inference on normal parameters, mean and variance unknown 19
2.3.1 Normal distribution: mean and variance unknown and interdependent 22
2.3.2 Comparison of variances 23
2.3.3 Comparison of several means 24
2.4 The t density as a heavy-tailed alternative to the normal distribution 25
2.5 Categorical data models: binomial, Poisson and multinomial 27

2.5.1 Binomial outcomes 28
2.5.2 Using prior knowledge to estimate the beta prior 30
2.6 Poisson distribution for event counts 34
2.7 Polytomous outcomes via the multinomial density 37
2.8 Multivariate continuous data: MVN and multivariate t densities 41
2.8.1 Multivariate normal with unknown mean and variance 42
2.8.2 Hypotheses on multivariate means 52
2.9 Power and sample size calculations 53
2.9.1 Sample size specification to achieve a desired accuracy 53
2.9.2 Sample size specifications for testing hypotheses 55
Notes 61

Chapter 3 Models for Association and Classification 63
3.1 Introduction: association and classification 63
3.2 Analysing association in observational studies: follow-up and case control studies 65
3.2.1 Simulating controls through historic exposure 69
3.2.2 Accounting for stratification in study design 71
3.3.3 Attributable risk: effects of eliminating postulated causes 73
3.4 Controlled experiments and trials 75
3.5 Classification, screening and diagnostic accuracy 80
3.5.1 Multivariate discrimination 86

Chapter 4 Normal Linear Regression, General Linear Models and Log-Linear Models 91
4.1 The context for Bayesian regression methods 91
4.2 The univariate linear regression model 94
4.2.1 Posterior for β, precision known 94
4.2.2 Posterior for β and τ, precision unknown 95
4.3 Multicollinearity and robust regression methods 99
4.4 Nonparametric regression via splines 105
4.5 General linear models: logit and Poisson regression 109
4.5.1 Poisson regression 118
4.5.2 General nonlinear models 121
4.6 General linear models for survey data 127
4.7 Log-linear models for crossed categorical outcomes 129
4.7.1 The two-way contingency table 130
4.7.2 Modelling choices and strategies 132

4.7.3 Fitting log-linear models for a two-way model 133
4.7.4 The quasi-symmetry model 136
4.8 Variable and model selection in regression and log-linear models 139
4.8.1 Selection among regression models 140
4.8.2 Selection of log-linear models 143
4.9 Log-linear models for cumulated event data 145
4.10 Markov chain models 149
4.11 Multivariate categorical outcomes 151
4.12 Ordered outcomes 154
4.12.1 Scores for ordered outcomes 157

Chapter 5 Ensemble Estimates: Hierarchical Priors for Pooling Strength 163
5.1 Hierarchical priors: improved precision through borrowing strength 163
5.1.1 Pooling strength: Poisson outcomes 165
5.1.2 Pooling strength: binomial outcomes 168
5.2 Overdispersed categorical data and random effects models 174
5.3 Models for polytomies and histograms: extra-variation in multinomial models 178
5.3.1 Histogram smoothing 180
5.4 League tables and meta-analyses 181
5.5 Meta-analysis 185
5.5.1 Combining over design types and selection bias 188
5.6 Estimating population parameters from survey data 190
5.6.1 Modelling missing cells in multi-way designs 194
5.6.2 Two-stage sampling for proportions 196

Chapter 6 Latent Variables, Mixture Analysis and Models for NonResponse 199
6.1 Introduction 199
6.1.1 Missing data models 200
6.1.2 Latent traits vs latent classes 200
6.1.3 Regression and transition mixtures 202
6.1.4 Latent factor models, latent class analysis and item analysis 202
6.1.5 Item analysis 205
6.1.6 Polytomous outcomes 207
6.1.7 Latent class analysis (LCA) 209
6.1.8 Multiple categorical outcomes 213
6.2 Mixtures of distributions, regression mixtures and transition mixtures 214
6.2.1 BUGS implementation 216

6.2.2 Normal mixtures 220
6.3 Latent transition matrices and mixtures of transition matrices 223
6.4 Models for nonresponse and incomplete data 230
6.5 Missingness patterns: factored likelihoods for multivariate missing data 233
6.5.1 Multiple imputation 236
6.6 Ignorable and non-ignorable missingness: estimating missing counts in contingency tables 239
6.6.1 Hierarchical models for non-ignorable nonresponse 241
6.6.2 Parameterisation of differential nonresponse 243
6.6.3 Polytomous and multivariate nonresponse patterns 251
6.6.4 Mixtures of continuous and categorical data 257
6.7 Nonparametric mixture modelling via Dirichlet process priors 260
6.7.1 Data smoothing 266
6.7.2 Other nonparametric priors 267

Chapter 7 Correlated Data Models 275
7.1 Introduction 275
7.1.1 Time series analysis 275
7.1.2 Spatial modelling 277
7.1.3 Contemporaneous correlation in systems of equations and in measurement error models 279
7.2 Temporal correlation described by autoregressive processes 281
7.2.1 Priors on autoregressive coefficients 282
7.2.2 Stationarity priors in the $AR(p)$ model 283
7.2.3 Initial conditions as latent data 284
7.2.4 Predictions 285
7.3 Autoregressive moving average models 287
7.3.1 Initial conditions in the ARMA model 289
7.3.2 Normality assumptions 290
7.4 Outlier and intervention models 295
7.5 Regression with autocorrelated errors in time 297
7.5.1 Cointegration and unit root testing 303
7.6 Linear dynamic models and time varying coefficients 305
7.6.1 The general DLM framework 307
7.6.2 On-line vs retrospective analysis 308
7.6.3 Random walk and other autoregressive priors 309
7.6.4 Discounted priors for time-specific variances 310
7.6.5 Priors under interventions 311

7.7 Spectral time series analysis 322
7.8 Time series formulations of nonparametric regression 327
7.8.1 Continuous time prior 327
7.8.2 Fourier series prior 329
7.8.3 Discrete time priors 329
7.8.4 Implementation of time series priors 330
7.9 Spatial correlation 335
7.9.1 Hypothesis tests 336
7.9.2 Regression models for area outcomes 336
7.10 Continuous spatial processes 340
7.11 Measurement error, seemingly unrelated Regressions, and simultaneous equations 346
7.11.1 Double regressions 350
7.11.2 Epidemiological applications 353
7.12 Sets of relationships 356
7.13 Simultaneous equations 358
7.13.1 Using instruments: two-and three-stage least squares 360
7.13.2 The consumption model 361
Notes 365

Chapter 8 Multilevel Models, Multivariate Analysis and Longitudinal Models 367
8.1 Introduction 367
8.1.1 Multivariate robustness and identification 369
8.1.2 Ecological and atomistic fallacies 370
8.2 General linear multilevel models for hierarchical data sets 371
8.3 Panel data models and growth curves 375
8.3.1 Marginal and transition models 375
8.3.2 Random effects models 376
8.3.3 Autocorrelation in errors 378
8.4 Dynamic models for longitudinal and growth curve data 384
8.4.1 Growth curve models as DLMs 387
8.5 Multivariate regression, factor analysis and structural equation models 390
8.5.1 Multivariate regression for imputation of missing data 393
8.6 Factor analysis and structural equation models 394
8.6.1 Stochastic volatility by factor analysis 401
8.6.2 Structural equation modelling (SEM) 403
8.6.3 Different group means and covariances on latent variables 407
8.7 Multivariate hierarchical models 411

8.8 Small area and survey domain estimation 415
8.8.1 Modelling small area rates 416

Chapter 9 Life Table and Survival Analysis 423
9.1 Introduction 423
9.2 Survival analysis in continuous time: parametric modelling of duration and covariates 425
9.2.1 Modelling covariate impacts and time dependence in the hazard rate 427
9.2.2 Censored time observations 428
9.2.3 Survival analysis in BUGS 428
9.3 Non-monotonic hazards 432
9.4 Analysing competing risks in continuous time 433
9.5 Variations in proneness: models for frailty 438
9.6 Counting process models 443
9.7 Discrete time survival models and life tables in discrete time 449
9.7.1 Discrete time event history models 449
9.8 Life tables: demographic and medical applications 451
9.8.1 Clinical and epidemiological applications 451
9.9 Population life tables 456
9.10 Competing risks population life tables 459

Chapter 10 Bayesian Estimation and Model Assessment 465
10.1 Introduction 465
10.2 Model comparison and fit 469
10.3 Separate model estimation 472
10.3.1 Overall fit measures 472
10.3.2 Predictive fit measures 477
10.3.3 Posterior likelihood comparison methods 479
10.4 Simultaneous model assessment 480
Appendix 493

References 495
Index 515