Contents

Preface VII

1 Introduction and Examples 1
1.1 Correlated Data 1
1.2 Longitudinal Data Analysis 2
1.3 Data Examples 6
1.3.1 Indonesian Children's Health Study 6
1.3.2 Epileptic Seizures Data 7
1.3.3 Retinal Surgery Data 9
1.3.4 Orientation of Sandhoppers 10
1.3.5 Schizophrenia Clinical Trial 11
1.3.6 Multiple Sclerosis Trial 13
1.3.7 Tretinoin Emollient Cream Trial 13
1.3.8 Polio Incidences in USA 14
1.3.9 Tokyo Rainfall Data 15
1.3.10 Prince George Air Pollution Study 16
1.4 Remarks 19
1.5 Outline of Subsequent Chapters 20

2 Dispersion Models 23
2.1 Introduction 23
2.2 Dispersion Models 25
2.2.1 Definitions 26
2.2.2 Properties 28
2.3 Exponential Dispersion Models 30
2.4 Residuals 35
2.5 Tweedie Class 36
2.6 Maximum Likelihood Estimation 37
2.6.1 General Theory 38
2.6.2 MLE in the ED Models 41
2.6.3 MLE in the Simplex GLM 42

2.6.4 MLE in the von Mises GLM ... 49

3 Inference Functions ... 55
3.1 Introduction ... 55
3.2 Quasi-Likelihood Inference in GLMs ... 56
3.3 Preliminaries ... 58
3.4 Optimal Inference Functions ... 61
3.5 Multi-Dimensional Inference Functions ... 65
3.6 Generalized Method of Moments ... 68

4 Modeling Correlated Data ... 73
4.1 Introduction ... 73
4.2 Quasi-Likelihood Approach ... 76
4.3 Conditional Modeling Approaches ... 80
4.3.1 Latent Variable Based Approach ... 80
4.3.2 Transitional Model Based Approach ... 82
4.4 Joint Modeling Approach ... 84

5 Marginal Generalized Linear Models ... 87
5.1 Model Formulation ... 88
5.2 GEE: Generalized Estimating Equations ... 89
5.2.1 General Theory ... 90
5.2.2 Some Special Cases ... 93
5.2.3 Wald Test for Nested Models ... 95
5.3 GEE2 ... 95
5.3.1 Constant Dispersion Parameter ... 96
5.3.2 Varying Dispersion Parameter ... 100
5.4 Residual Analysis ... 101
5.4.1 Checking Distributional Assumption ... 102
5.4.2 Checking Constant Dispersion Assumption ... 102
5.4.3 Checking Link Functions ... 102
5.4.4 Checking Working Correlation ... 102
5.5 Quadratic Inference Functions ... 103
5.6 Implementation and Softwares ... 106
5.6.1 Newton-Scoring Algorithm ... 106
5.6.2 SAS PROC GENMOD ... 107
5.6.3 SAS MACRO QIF ... 108
5.7 Examples ... 109
5.7.1 Longitudinal Binary Data ... 110
5.7.2 Longitudinal Count Data ... 112
5.7.3 Longitudinal Proportional Data ... 116

6 Vector Generalized Linear Models 121
6.1 Introduction 121
6.2 Log-Linear Model for Correlated Binary Data 122
6.3 Multivariate ED Family Distributions 125
6.3.1 Copulas 126
6.3.2 Construction 127
6.3.3 Interpretation of Association Parameter 129
6.4 Simultaneous Maximum Likelihood Inference 136
6.4.1 General Theory 136
6.4.2 VGLMs for Correlated Continuous Outcomes 137
6.4.3 VGLMs for Correlated Discrete Outcomes 138
6.4.4 Scores for Association Parameters 139
6.5 Algorithms 141
6.5.1 Algorithm I: Maximization by Parts 142
6.5.2 Algorithm II: Gauss-Newton Type 146
6.6 An Illustration: VGLMs for Trivariate Discrete Data 146
6.6.1 Trivariate VGLMs 147
6.6.2 Comparison of Asymptotic Efficiency 148
6.7 Data Examples 150
6.7.1 Analysis of Two-Period Cross-Over Trial Data 150
6.7.2 Analysis of Hospital Visit Data 152
6.7.3 Analysis of Burn Injury Data 153

7 Mixed-Effects Models: Likelihood-Based Inference 157
7.1 Introduction 157
7.2 Model Specification 161
7.3 Estimation 165
7.4 MLE Based on Numerical Integration 167
7.5 Simulated MLE 175
7.6 Conditional Likelihood Estimation 176
7.7 MLE Based on EM Algorithm 178
7.8 Approximate Inference: PQL and REML 182
7.9 SAS Software 192
7.9.1 PROC MIXED 192
7.9.2 PROC NLMIXED 193
7.9.3 PROC GLIMMIX 194

8 Mixed-Effects Models: Bayesian Inference 195
8.1 Bayesian Inference Using MCMC Algorithm 195
8.1.1 Gibbs Sampling: A Practical View 195
8.1.2 Diagnostics 198
8.1.3 Enhancing Burn-in 201
8.1.4 Model Selection 202
8.2 An Illustration: Multiple Sclerosis Trial Data 203
8.3 Multi-Level Correlated Data 206

8.4 WinBUGS Software . . . 212
8.4.1 WinBUGS Code in Multiple Sclerosis Trial Data Analysis . . . 213
8.4.2 WinBUGS Code for the TEC Drug Analysis . . . 214

9 Linear Predictors . . . 217
9.1 General Results . . . 217
9.2 Estimation of Random Effects in GLMMs . . . 221
9.2.1 Estimation in LMMs . . . 221
9.2.2 Estimation in GLMMs . . . 221
9.3 Kalman Filter and Smoother . . . 222
9.3.1 General Forms . . . 222

10 Generalized State Space Models . . . 227
10.1 Introduction . . . 227
10.2 Linear State Space Models . . . 231
10.3 Shift-Mean Model . . . 232
10.4 Monte Carlo Maximum Likelihood Estimation . . . 235

11 Generalized State Space Models for Longitudinal Binomial Data . . . 239
11.1 Introduction . . . 239
11.2 Monte Carlo Kalman Filter and Smoother . . . 240
11.3 Bayesian Inference Based on MCMC . . . 246

12 Generalized State Space Models for Longitudinal Count Data . . . 261
12.1 Introduction . . . 261
12.2 Generalized Estimating Equation . . . 264
12.3 Monte Carlo EM Algorithm . . . 265
12.4 KEE in Stationary State Processes . . . 267
12.4.1 Setup . . . 267
12.4.2 Kalman Filter and Smoother . . . 269
12.4.3 Godambe Information Matrix . . . 271
12.4.4 Analysis of Polio Incidences Data . . . 272
12.5 KEE in Non-Stationary State Processes . . . 275
12.5.1 Model Formulation . . . 275
12.5.2 Kalman Filter and Smoother . . . 278
12.5.3 Parameter Estimation . . . 280
12.5.4 Model Diagnosis . . . 281
12.5.5 Analysis of Prince George Data . . . 283

13 Missing Data in Longitudinal Studies 291
13.1 Introduction 291
13.2 Missing Data Patterns 293
13.2.1 Patterns of Missingness 293
13.2.2 Types of Missingness and Effects 297
13.3 Diagnosis of Missing Data Types 300
13.3.1 Graphic Approach 301
13.3.2 Testing for MCAR 302
13.4 Handling MAR Mechanism 306
13.4.1 Simple Solutions and Limitations 307
13.4.2 Multiple Imputation 307
13.4.3 EM Algorithm 311
13.4.4 Inverse Probability Weighting 317
13.5 Handling NMAR Mechanism 320
13.5.1 Parametric Modeling 320
13.5.2 A Semiparametric Pattern Mixture Model 322

References 329

Index 343