Ludwig Fahrmeir Gerhard Tutz

Multivariate Statistical Modelling Based on Generalized Linear Models

Second Edition

With contributions from Wolfgang Hennevogl

With 51 Figures



Contents

~

Prei Prei List List List	face t face t of E of Fi of Ta	o the S o the F xamples gures . ables .	econd Edition	v vii vii xxi xxi xxv
1.	Intr	oducti	on	1
	1.1	Outlin	e and Examples	2
	1.2	Remar	ks on Notation	13
	1.3	Notes	and Further Reading	14
2.	Mo	delling	and Analysis of Cross-Sectional Data: A Review	
	of U	Jnivari	ate Generalized Linear Models	15
	2.1	Univar	riate Generalized Linear Models	16
		2.1.1	Data	16
			Coding of Covariates	16
			Grouped and Ungrouped Data	17
		2.1.2	Definition of Univariate Generalized Linear Models .	18
		2.1.3	Models for Continuous Responses	22
			Normal Distribution	22
			Gamma Distribution	23
			Inverse Gaussian Distribution	24
		2.1.4	Models for Binary and Binomial Responses	24
			Linear Probability Model	25
			Probit Model	26
			Logit Model	26
			Complementary Log-Log Model	26
			Complementary Log-Model	26
			Binary Models as Threshold Models of Latent	
			Linear Models	29
			Parameter Interpretation	29
			Overdispersion	35
		2.1.5	Models for Count Data	36
			Log-linear Poisson Model	36
			~	

			Linear Poisson Model	36
	2.2	Likelił	nood Inference	38
		2.2.1	Maximum Likelihood Estimation	38
			Log-likelihood, Score Function and Information Matrix	39
			Numerical Computation of the MLE by	
			Iterative Methods	41
			Uniqueness and Existence of MLEs*	43
			Asymptotic Properties	44
			Discussion of Regularity Assumptions*	46
			Additional Scale or Overdispersion Parameter	47
		2.2.2	Hypothesis Testing and Goodness-of-Fit Statistics	47
			Goodness-of-Fit Statistics	50
	2.3	Some	Extensions	55
		2.3.1	Quasi-likelihood Models	55
			Basic Models	55
			Variance Functions with Unknown Parameters	58
			Nonconstant Dispersion Parameter	59
		2.3.2	Bayesian Models	6 0
		2.3.3	Nonlinear and Nonexponential Family Regression	
			Models*	65
-	2.4	Notes	and Further Reading	67
_				
3.	Mo	dels fo	r Multicategorical Responses: Multivariate	20
	Ext	ension	is of Generalized Linear Models	69
	3.1		Categorical Response Models	10
		3.1.1	Multinomial Distribution	70
		3.1.Z		11
		3.1.3	I ne Multivariate Model	(2
	10	3.1.4 M. J.	Multivariate Generalized Linear Models	75 77
	3.2	Mode	Is for Nominal Responses	77
		3.2.1	The Principle of Maximum Random Utility	((
		3.2.2	Modelling of Explanatory Variables: Choice of	=0
		N. 1.	Design Matrix	79
	3.3	Mode	C latin M l l TP The shall A	81
		3.3.1	Cumulative Models: The Inreshold Approach	83
			Cumulative Logistic Model or Proportional	0.0
			Odds Model	83
			Grouped Cox Model or Proportional Hazards Model .	80
			Extreme Maximal-value Distribution Model	80
		ა.ა.2 ე ე ე ე	Extended versions of Cumulative Models	87
		J.J.J	Link runctions and Design Matrices for Cumulative	00
		2 9 4	Wodels	88
		3.3.4	Sequential Models	92
			Generalized Sequential Models	95
			Link Functions of Sequential Models	-98

х

			•	
		3.3.5	Strict Stochastic Ordering*	99
		3.3.6	Two-Step Models	100
			Link Function and Design Matrix for	
			Two-Step Models	102
		3.3.7	Alternative Approaches	103
	3.4	Statist	tical Inference	105
		3.4.1	Maximum Likelihood Estimation	105
			Numerical Computation	107
		3.4.2	Testing and Goodness-of-Fit	107
			Testing of Linear Hypotheses	107
			Goodness-of-Fit Statistics	107
		3.4.3	Power-Divergence Family*	109
			Asymptotic Properties under Classical "Fixed Cells"	
			Assumptions	111
			Sparseness and "Increasing-Cells" Asymptotics	112
	3.5	Multi	variate Models for Correlated Responses	112
	4	3.5.1	Conditional Models	114
			Asymmetric Models	114
			Symmetric Models	116
		3.5.2	Marginal Models	119
			Marginal Models for Correlated Univariate Responses	120
			The Generalized Estimating Approach for Statistical	
			Inference	123
			Marginal Models for Correlated	
			Categorical Responses	129
			Likelihood-based Interence for Marginal Models	135
	3.6	Notes	and Further Reading	136
			Bayesian Inference	136
4.	Sel	ecting	and Checking Models	139
	4.1	Varia	ble Selection	139
•		4.1.1	Selection Criteria	140
		4.1.2	Selection Procedures	142
			All-Subsets Selection	142
			Stepwise Backward and Forward Selection	143
	4.2	Diagn	lostics	145
		4.2.1	Diagnostic Tools for the Classical Linear Model	146
		4.2.2	Generalized Hat Matrix	147
		4.2.3	Residuals and Goodness-of-Fit Statistics	151
		4.2.4	Case Deletion	156
	4.3	Gener	ral Tests for Misspecification*	161
		4.3.1	Estimation under Model Misspecification	162
		4.3.2	Hausman-type Tests	165
			Hausman Tests	165
			Information Matrix Test	166

4.3.3	Tests for Nonnested Hypotheses 167
	Tests Based on Artificial Nesting 168
	Generalized Wald and Score Tests 168
4.4 Notes	and Further Reading
	Bayesian Model Determination
	Robust Estimates
	Model Tests Against Smooth Alternatives 172
	-

5. Semi- and Nonparametric Approaches to

Reg	ressio	n Analysis
5.1	Smoot	thing Techniques for Continuous Responses
	5.1.1	Regression Splines and Other Basis Functions 174
		Regression Splines 176
		Other Basis Functions
		Regularization
	5.1.2	Smoothing Splines 181
	5.1.3	Local Estimators 183
		Simple Neighborhood Smoothers 183
		Local Regression
		Bias-Variance Trade-off
		Relation to Other Smoothers 189
	5.1.4	Selection of Smoothing Parameters 190
5.2	Smoot	thing for Non-Gaussian Data 193
	5.2.1	Basis Function Approach 193
		Fisher Scoring for Penalized Likelihood* 194
	5.2.2	Penalization and Spline Smoothing 195
		Fisher Scoring for Generalized Spline Smoothing* 196
		Choice of Smoothing Parameter 197
	5.2.3	Localizing Generalized Linear Models 198
		Local Fitting by Weighted Scoring 201
5.3	Mode	lling with Multiple Covariates
	5.3.1	Modelling Approaches
		Generalized Additive Models 207
		Partially Linear Models 208
		Varying-Coefficient Models 208
		Projection Pursuit Regression 209
		Basis Function Approach 210
	5.3.2	Estimation Concepts 213
		Backfitting Algorithm for Generalized
		Additive Models 213
		Backfitting with Spline Functions
		Choice of Smoothing Parameter
		Partial Linear Models 220
5.4	Semip	parametric Bayesian Inference for
	Gener	alized Regression

		5.4.1	Gaussian Responses	. 221
			Smoothness Priors Approaches	. 221
			Basis Function Approaches	. 227
			Models with Multiple Covariates	. 228
		5.4.2	Non-Gaussian Responses	. 231
			Latent Variable Models for Categorical Responses	. 234
	5.5	Notes	and Further Reading	. 239
6	Fix	ad Par	rameter Models for Time Series and	
0.	Lon	gitudi	nal Data	241
	6.1	Time	Series	242
	0.1	611	Conditional Models	242
		0.1.1	Generalized Autoregressive Models	242
			Quasi-Likelihood Models and Generalized Autore-	. 212
			gression Moving Average Models	246
		612	Statistical Inference for Conditional Models	249
		613	Marginal Models	255
		0.1.0	Estimation of Marginal Models	258
	62	Longi	tudinal Data	260
	0.2	621	Conditional Models	261
		0.2.1	Generalized Autoregressive Models Quasi-Likelihood	. 201
			Models	261
			Statistical Inference	262
			Transition Models	264
			Subject-specific Approaches and	. 201
			Conditional Likelihood	. 264
		6.2.2	Marginal Models	267
		0.2.12	Statistical Inference	268
		6.2.3	Generalized Additive Models for Longitudinal Data	274
	6.3	Notes	and Further Reading	278
	0.0	1.0005		. 210
7.	Rai	ndom I	Effects Models	. 283
	7.1	Linea	r Random Effects Models for Normal Data	. 285
		7.1.1	Two-stage Random Effects Models	. 285
			Random Intercepts	. 286
			Random Slopes	. 287
			Multilevel Models	. 288
		7.1.2	Statistical Inference	. 289
			Known Variance-Covariance Components	. 289
			Unknown Variance-Covariance Components	. 289
			Derivation of the EM algorithm [*]	. 291
	7.2	Rand	om Effects in Generalized Linear Models	. 292
			Generalized Linear Models with Random Effects	. 293
			Examples	. 294
	7.3	Estim	ation Based on Posterior Modes	. 298

		7.3.1	Known Variance-Covariance Components	298
		7.3.2	Unknown Variance-Covariance Components	299
		7.3.3	Algorithmic Details*	300
			Fisher Scoring for Given Variance-Covariance	
			Components	3 00
			EM Type Algorithm	302
	7.4	\mathbf{Estim}	ation by Integration Techniques	303
		7.4.1	Maximum Likelihood Estimation of Fixed Parameters	303
			Direct Maximization Using Fitting Techniques for	
			GLMs	305
			Nonparametric Maximum Likelihood for Finite Mix-	
			tures	308
		7.4.2	Posterior Mean Estimation of Random Effects	310
		7.4.3	Indirect Maximization Based on the EM Algorithm*.	311
		7.4.4	Algorithmic Details for Posterior Mean Estimation* .	315
~	7.5	Exam	ples	318
	7.6	Bayes	ian Mixed Models	321
			Bayesian Generalized Mixed Models	321
			Generalized Additive Mixed Models	322
	7.7	Margi	nal Estimation Approach to Random Effects Models	325
	7.8	Notes	and Further Reading	328
8.	Stat	te Sna	ce and Hidden Markov Models	331
0.	81	Linear	r State Space Models and the Kalman Filter	332
	0.1	811	Linear State Space Models	332
		8.1.2	Statistical Inference	337
		0.1.2	Linear Kalman Filtering and Smoothing	338
			Kalman Filtering and Smoothing as Posterior Mode	
			Estimation*	340
			Unknown Hyperparameters	342
-			EM Algorithm for Estimating Hyperparameters*	343
	8.2	Non-N	Normal and Nonlinear State Space Models	345
		8.2.1	Dynamic Generalized Linear Models	345
			Categorical Time Series	347
		8.2.2	Nonlinear and Nonexponential Family Models*	349
	8.3	Non-N	Normal Filtering and Smoothing	350
		8.3.1	Posterior Mode Estimation	351
			Generalized Extended Kalman Filter and Smoother*.	352
			Gauss-Newton and Fisher-Scoring Filtering and	
			Smoothing*	354
			Estimation of Hyperparameters [*]	356
			Some Applications	356
		8.3.2	Markov Chain Monte Carlo and Integration-based	
~			Approaches	361
			MCMC Inference	362

1

....

		Posterior Mode Smoothing
		Fully Bayesian Inference via MCMC
	9.5	Remarks and Further Reading
А.		
	A.1	Exponential Families and Generalized Linear Models 433
	A.2	Basic Ideas for Asymptotics
	A.3	EM Algorithm
	A.4	Numerical Integration
	A.5	Monte Carlo Methods
B.	Soft	ware for Fitting Generalized Linear Models
	and •	Extensions
Bib	liogra	uphy
Aut	hor I	ndex
Sub	ject	Index

.