
Contents

Preface	xvii
1 Linear Models	1
1.1 A simple linear model	2
Simple least squares estimation	2
1.1.1 Sampling properties of $\hat{\beta}$	3
1.1.2 So how old is the universe?	4
1.1.3 Adding a distributional assumption	7
Testing hypotheses about β	7
Confidence intervals	8
1.2 Linear models in general	9
1.3 The theory of linear models	11
1.3.1 Least squares estimation of β	12
1.3.2 The distribution of $\hat{\beta}$	13
1.3.3 $(\hat{\beta}_i - \beta_i)/\hat{\sigma}_{\hat{\beta}_i} \sim t_{n-p}$	13
1.3.4 F-ratio results I	14
1.3.5 F-ratio results II	14
1.3.6 The influence matrix	16
1.3.7 The residuals, $\hat{\epsilon}$, and fitted values, $\hat{\mu}$	16
1.3.8 Results in terms of $\mathbf{X}$	17
1.3.9 The Gauss Markov Theorem: What's special about least squares?	18
1.4 The geometry of linear modelling	19
1.4.1 Least squares	19
1.4.2 Fitting by orthogonal decompositions	20
1.4.3 Comparison of nested models	21
1.5 Practical linear modelling	22
1.5.1 Model fitting and model checking	23
1.5.2 Model summary	28
1.5.3 Model selection	30
1.5.4 Another model selection example	31
A follow-up	34
1.5.5 Confidence intervals	35
1.5.6 Prediction	36
1.5.7 Co-linearity, confounding and causation	36

1.6	Practical modelling with factors	38
1.6.1	Identifiability	39
1.6.2	Multiple factors	40
1.6.3	'Interactions' of factors	41
1.6.4	Using factor variables in R	43
1.7	General linear model specification in R	46
1.8	Further linear modelling theory	47
1.8.1	Constraints I: General linear constraints	47
1.8.2	Constraints II: 'Contrasts' and factor variables	48
1.8.3	Likelihood	49
1.8.4	Non-independent data with variable variance	50
1.8.5	Simple AR correlation models	52
1.8.6	AIC and Mallows' statistic	52
1.8.7	The wrong model	54
1.8.8	Non-linear least squares	54
1.8.9	Further reading	56
1.9	Exercises	56
2	Linear Mixed Models	61
2.1	Mixed models for balanced data	61
2.1.1	A motivating example	61
	The wrong approach: A fixed effects linear model	62
	The right approach: A mixed effects model	64
2.1.2	General principles	65
2.1.3	A single random factor	66
2.1.4	A model with two factors	69
2.1.5	Discussion	74
2.2	Maximum likelihood estimation	74
2.2.1	Numerical likelihood maximization	76
2.3	Linear mixed models in general	77
2.4	Linear mixed model maximum likelihood estimation	78
2.4.1	The distribution of $\mathbf{b} \mathbf{y}, \hat{\boldsymbol{\beta}}$ given $\boldsymbol{\theta}$	79
2.4.2	The distribution of $\hat{\boldsymbol{\beta}}$ given $\boldsymbol{\theta}$	80
2.4.3	The distribution of $\hat{\boldsymbol{\theta}}$	81
2.4.4	Maximizing the profile likelihood	81
2.4.5	REML	83
2.4.6	Effective degrees of freedom	83
2.4.7	The EM algorithm	84
2.4.8	Model selection	85
2.5	Linear mixed models in R	86
2.5.1	Package nlme	86
2.5.2	Tree growth: An example using lme	87
2.5.3	Several levels of nesting	91
2.5.4	Package lme4	93
2.5.5	Package mgcv	94

2.6	Exercises	95
3	Generalized Linear Models	101
3.1	GLM theory	102
3.1.1	The exponential family of distributions	103
3.1.2	Fitting generalized linear models	105
3.1.3	Large sample distribution of $\hat{\beta}$	107
3.1.4	Comparing models	108
	Deviance	108
	Model comparison with unknown ϕ	109
	AIC	110
3.1.5	Estimating ϕ , Pearson's statistic and Fletcher's estimator	110
3.1.6	Canonical link functions	111
3.1.7	Residuals	112
	Pearson residuals	112
	Deviance residuals	113
3.1.8	Quasi-likelihood	113
3.1.9	Tweedie and negative binomial distributions	115
3.1.10	The Cox proportional hazards model for survival data	116
	Cumulative hazard and survival functions	118
3.2	Geometry of GLMs	119
3.2.1	The geometry of IRLS	121
3.2.2	Geometry and IRLS convergence	122
3.3	GLMs with R	124
3.3.1	Binomial models and heart disease	125
3.3.2	A Poisson regression epidemic model	131
3.3.3	Cox proportional hazards modelling of survival data	136
3.3.4	Log-linear models for categorical data	138
3.3.5	Sole eggs in the Bristol channel	142
3.4	Generalized linear mixed models	147
3.4.1	Penalized IRLS	148
3.4.2	The PQL method	149
3.4.3	Distributional results	151
3.5	GLMMs with R	151
3.5.1	glmmPQL	151
3.5.2	gam	154
3.5.3	glmer	155
3.6	Exercises	156
4	Introducing GAMs	161
4.1	Introduction	161
4.2	Univariate smoothing	162
4.2.1	Representing a function with basis expansions	162
	A very simple basis: Polynomials	162
	The problem with polynomials	162

	The piecewise linear basis	164
	Using the piecewise linear basis	165
4.2.2	Controlling smoothness by penalizing wiggleness	166
4.2.3	Choosing the smoothing parameter, λ , by cross validation	169
4.2.4	The Bayesian/mixed model alternative	172
4.3	Additive models	174
4.3.1	Penalized piecewise regression representation of an additive model	175
4.3.2	Fitting additive models by penalized least squares	177
4.4	Generalized additive models	180
4.5	Summary	182
4.6	Introducing package <code>mgcv</code>	182
4.6.1	Finer control of <code>gam</code>	184
4.6.2	Smooths of several variables	187
4.6.3	Parametric model terms	189
4.6.4	The <code>mgcv</code> help pages	191
4.7	Exercises	191
5	Smoothers	195
5.1	Smoothing splines	195
5.1.1	Natural cubic splines are smoothest interpolators	196
5.1.2	Cubic smoothing splines	198
5.2	Penalized regression splines	199
5.3	Some one-dimensional smoothers	201
5.3.1	Cubic regression splines	201
5.3.2	A cyclic cubic regression spline	202
5.3.3	P-splines	204
5.3.4	P-splines with derivative based penalties	206
5.3.5	Adaptive smoothing	207
5.3.6	SCOP-splines	208
5.4	Some useful smoother theory	210
5.4.1	Identifiability constraints	211
5.4.2	'Natural' parameterization, effective degrees of freedom and smoothing bias	211
5.4.3	Null space penalties	214
5.5	Isotropic smoothing	214
5.5.1	Thin plate regression splines	215
	Thin plate splines	215
	Thin plate regression splines	217
	Properties of thin plate regression splines	218
	Knot-based approximation	219
5.5.2	Duchon splines	221
5.5.3	Splines on the sphere	222
5.5.4	Soap film smoothing over finite domains	223
5.6	Tensor product smooth interactions	227

5.6.1	Tensor product bases	227
5.6.2	Tensor product penalties	229
5.6.3	ANOVA decompositions of smooths	232
	Numerical identifiability constraints for nested terms	233
5.6.4	Tensor product smooths under shape constraints	234
5.6.5	An alternative tensor product construction	235
	What is being penalized?	236
5.7	Isotropy versus scale invariance	237
5.8	Smooths, random fields and random effects	239
5.8.1	Gaussian Markov random fields	240
5.8.2	Gaussian process regression smoothers	241
5.9	Choosing the basis dimension	242
5.10	Generalized smoothing splines	243
5.11	Exercises	245
6	GAM theory	249
6.1	Setting up the model	249
6.1.1	Estimating β given λ	251
6.1.2	Degrees of freedom and scale parameter estimation	251
6.1.3	Stable least squares with negative weights	252
6.2	Smoothness selection criteria	255
6.2.1	Known scale parameter: UBRE	255
6.2.2	Unknown scale parameter: Cross validation	256
	Leave-several-out cross validation	257
	Problems with ordinary cross validation	257
6.2.3	Generalized cross validation	258
6.2.4	Double cross validation	260
6.2.5	Prediction error criteria for the generalized case	261
6.2.6	Marginal likelihood and REML	262
6.2.7	The problem with $\log \mathbf{S}_\lambda _+$	264
6.2.8	Prediction error criteria versus marginal likelihood	266
	Unpenalized coefficient bias	267
6.2.9	The 'one standard error rule' and smoother models	268
6.3	Computing the smoothing parameter estimates	269
6.4	The generalized Fellner-Schall method	269
6.4.1	General regular likelihoods	271
6.5	Direct Gaussian case and performance iteration (PQL)	272
6.5.1	Newton optimization of the GCV score	273
6.5.2	REML	276
	$\log \mathbf{S}_\lambda _+$ and its derivatives	277
	The remaining derivative components	278
6.5.3	Some Newton method details	280
6.6	Direct nested iteration methods	280
6.6.1	Prediction error criteria	282
6.6.2	Example: Cox proportional hazards model	283

	Derivatives with respect to smoothing parameters	285
	Prediction and the baseline hazard	286
6.7	Initial smoothing parameter guesses	287
6.8	GAMM methods	288
6.8.1	GAMM inference with mixed model estimation	289
6.9	Bigger data methods	290
6.9.1	Bigger still	291
6.10	Posterior distribution and confidence intervals	293
6.10.1	Nychka's coverage probability argument	294
	Interval limitations and simulations	296
6.10.2	Whole function intervals	300
6.10.3	Posterior simulation in general	300
6.11	AIC and smoothing parameter uncertainty	301
6.11.1	Smoothing parameter uncertainty	302
6.11.2	A corrected AIC	303
6.12	Hypothesis testing and p-values	304
6.12.1	Approximate p-values for smooth terms	305
	Computing T_r	307
	Simulation performance	308
6.12.2	Approximate p-values for random effect terms	309
6.12.3	Testing a parametric term against a smooth alternative	312
6.12.4	Approximate generalized likelihood ratio tests	313
6.13	Other model selection approaches	315
6.14	Further GAM theory	316
6.14.1	The geometry of penalized regression	316
6.14.2	Backfitting GAMs	318
6.15	Exercises	320
7	GAMs in Practice: <code>mgcv</code>	325
7.1	Specifying smooths	325
7.1.1	How smooth specification works	327
7.2	Brain imaging example	328
7.2.1	Preliminary modelling	329
7.2.2	Would an additive structure be better?	333
7.2.3	Isotropic or tensor product smooths?	334
7.2.4	Detecting symmetry (with <code>by</code> variables)	336
7.2.5	Comparing two surfaces	337
7.2.6	Prediction with <code>predict.gam</code>	339
	Prediction with <code>lpmatrix</code>	341
7.2.7	Variances of non-linear functions of the fitted model	342
7.3	A smooth ANOVA model for diabetic retinopathy	343
7.4	Air pollution in Chicago	346
7.4.1	A single index model for pollution related deaths	349
7.4.2	A distributed lag model for pollution related deaths	351
7.5	Mackerel egg survey example	353

7.5.1	Model development	354
7.5.2	Model predictions	357
7.5.3	Alternative spatial smooths and geographic regression	359
7.6	Spatial smoothing of Portuguese larks data	361
7.7	Generalized additive mixed models with R	365
7.7.1	A space-time GAMM for sole eggs	365
	Soap film improvement of boundary behaviour	368
7.7.2	The temperature in Cairo	371
7.7.3	Fully Bayesian stochastic simulation: <code>jagam</code>	374
7.7.4	Random wiggly curves	376
7.8	Primary biliary cirrhosis survival analysis	377
7.8.1	Time dependent covariates	380
7.9	Location-scale modelling	383
7.9.1	Extreme rainfall in Switzerland	384
7.10	Fuel efficiency of cars: Multivariate additive models	387
7.11	Functional data analysis	390
7.11.1	Scalar on function regression	390
	Prostate cancer screening	391
	A multinomial prostate screening model	393
7.11.2	Function on scalar regression: Canadian weather	395
7.12	Other packages	397
7.13	Exercises	398
A	Maximum Likelihood Estimation	405
A.1	Invariance	405
A.2	Properties of the expected log-likelihood	406
A.3	Consistency	409
A.4	Large sample distribution of $\hat{\theta}$	410
A.5	The generalized likelihood ratio test (GLRT)	411
A.6	Derivation of $2\lambda \sim \chi_r^2$ under H_0	411
A.7	AIC in general	414
A.8	Quasi-likelihood results	416
B	Some Matrix Algebra	419
B.1	Basic computational efficiency	419
B.2	Covariance matrices	420
B.3	Differentiating a matrix inverse	420
B.4	Kronecker product	421
B.5	Orthogonal matrices and Householder matrices	421
B.6	QR decomposition	422
B.7	Cholesky decomposition	422
B.8	Pivoting	423
B.9	Eigen-decomposition	424
B.10	Singular value decomposition	425
B.11	Lanczos iteration	426

C Solutions to Exercises	429
C.1 Chapter 1	429
C.2 Chapter 2	432
C.3 Chapter 3	438
C.4 Chapter 4	440
C.5 Chapter 5	441
C.6 Chapter 6	443
C.7 Chapter 7	447
 Bibliography	 455
 Index	 467