

# Springer Series in Statistics

*Advisors:*

P. Bickel, P. Diggle, S.E. Fienberg, U. Gather,  
I. Olkin, S. Zeger

More information about this series at <http://www.springer.com/series/692>



Thomas W. Yee

# Vector Generalized Linear and Additive Models

With an Implementation in R

 Springer

Thomas W. Yee  
Department of Statistics  
University of Auckland  
Auckland, New Zealand

ISSN 0172-7397                      ISSN 2197-568X (electronic)  
Springer Series in Statistics  
ISBN 978-1-4939-2817-0            ISBN 978-1-4939-2818-7 (eBook)  
DOI 10.1007/978-1-4939-2818-7

Library of Congress Control Number: 2015945942

Springer New York Heidelberg Dordrecht London

© Thomas Yee 2015

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

Springer Science+Business Media LLC New York is part of Springer Science+Business Media ([www.springer.com](http://www.springer.com))

*To my parents and Selina and Annie*



# Preface

*Beauty will result from the form and correspondence of the whole, with respect to the several parts, of the parts with regard to each other, and of these again to the whole; that the structure may appear an entire and compleat body, wherein each member agrees with the other, and all necessary to compose what you intend to form.*

—The First Book of Andrea Palladio’s Architecture, 1570

In the early 1970s, the *generalized linear model* (GLM) class of statistical models was proposed by Nelder and Wedderburn (1972), providing a unified framework for several important regression models. They showed that the linear model, Poisson regression, logistic regression and probit analysis, and others could be treated as special cases of GLMs, and that one algorithm could be used to estimate them all. The unified GLM framework also provides an elegant overriding theoretical structure resulting in inference, diagnostics, software interface, etc. that applies to all of them. Prior to GLMs, these methods were largely treated as unrelated. Since then, GLMs have gained universal acceptance in the statistical world, and as Senn (2004, p. 7) writes, “Nelder and Wedderburn was a paper that changed the statistical landscape for ever and it is simply impossible now to envisage the modelling world without it.”

Although GLMs were a great advance, they are largely confined to one-parameter distributions from an exponential family. There are many situations where practical data analysis and regression modelling demands much greater flexibility than this. To this end, this book describes a much larger and more flexible statistical framework for fixed-effects regression modelling that greatly extends GLMs. It comprises about half-a-dozen major classes of statistical models, and having at its heart two classes, called *vector generalized linear models* (VGLMs) and *vector generalized additive models* (VGAMs). (Other classes are listed below.) Each class is related to each other in a natural way within this framework, e.g., VGAMs are a smooth or data-driven version of VGLMs.

The purpose of this book is to introduce the framework and each major subclass of models. VGLMs might be thought of loosely as multivariate GLMs, which

are not confined to the exponential family. VGAMs replace the linear functions of VGLMs by smooths. The remaining classes can be thought of as extensions of VGLMs and VGAMs. As a software implementation of our approach, the R package `VGAM` (and its companion `VGAMdata`) are used. Following the grand tradition of GLMs, we demonstrate its usefulness for data analysis over a wide range of problems and application areas. It is hoped that, in a way similar to what was pioneered by GLMs, the overall framework described in this book will provide a natural vehicle for thinking about regression modelling, and performing much more of applied statistics as a coherent whole.

What advantages does the VGLM/VGAM framework confer? It provides the same benefits that GLMs gave, but on a much larger scale. For example, the theory and methodology of this book unifies areas such as univariate distributions, categorical data analysis, aspects of quantile regression, and extremes. And the advantages which generalized additive models add to GLM-type analyses in terms of smoothing, VGAMs do for VGLMs. It has long been my belief that what should be ‘ordinary’ and ‘routine’ regression modelling in applied statistics has been hampered by a lack of common framework. For VGLMs/VGAMs, the user has the freedom to easily vary model elements within a large flexible framework, e.g., currently `VGAM` implements over 150 family functions.

The underlying conception is to treat almost all distributions and classical models as generalized regression models. By ‘classical’, we mean models and distributions which are amenable to estimation by first and second derivative methods, particularly Fisher scoring. For years there has been a need to broaden the scope of GLMs and GAMs, and to fortify it with some necessary infrastructure to make them more fully operable. The end is that the framework instils structure in many classical regression models.

## Audience

The book was written with three types of readers in mind. The first are existing users of the `VGAM` R package needing a primary source for the methodology, and having sufficient examples and details to be useful. The second are people interested in a new and very general modelling framework. The third are students/teachers in courses on general regression following a basic regression course on GLMs. For these, some notes for instructors are given below.

Some assumptions have been made about the background of the reader. Firstly is a basic working knowledge of R. There are now many books that provide this, e.g., Venables and Ripley (2002), Dalgaard (2008), Maindonald and Braun (2010), Cohen and Cohen (2008), Zuur et al. (2009), de Vries and Meys (2012). Further references are listed at the end of Chap. 8. Second is a mid-undergraduate level knowledge of statistical theory and practice. Thirdly, some chapters assume a basic familiarity with linear algebra and calculus up to mid-undergraduate level.

Some of the applied chapters (in Part II) would benefit from some prior exposure to those subject areas, e.g., econometrics, plant ecology, etc. because only a minimal attempt can be made to establish the background, motivation, and notation there.

## How This Book is Organized

There are two parts. **Part I** describes the general theory and computational details behind each major class of models. These classes are displayed in the flowchart of Fig. 1.2. Readers familiar with LMs, GLMs, and GAMs might jump directly into



the VGLM and VGAM chapters (Chap. 2 is a summary of these topics), otherwise they could go through sequentially from the beginning. However, Chap. 1 (especially Sect. 1.3) is crucial, because it sets the scene and gives a brief overview of the entire framework. The other major classes of models are various generalizations of reduced-rank methods. The main motivation for these is dimension reduction, and they consequently operate on latent variables. They are:

- RR-VGLMs: these *reduced-rank VGLMs* (Chap. 5) are based on linear combinations of the explanatory variables, and should be useful for many readers.
- QRR-VGLMs: these *quadratic RR-VGLMs* (Chap. 6) being relevant to mainly ecologists could be generally skipped on first reading.
- RR-VGAMs: these allow for *constrained additive ordination* (CAO; Chap. 7), which is a smooth version of RR-VGLMs.

Chapter 8 on the VGAM package should be read by all software users.

Part II explores some major application areas. Practitioners of certain topics in Part II will usually not need to concern themselves with the other parts, because they are mainly separate. However, readers will be able to at least browse through the range of other applications there. The breadth of coverage should demonstrate the versatility of the framework developed in Part I; I feel this is a major attractive feature of the book. Chapter 18 is a specialist chapter addressed more to R programmers wishing to extend the capabilities of the current software, e.g., by writing new VGAM family functions.

Most acronyms are listed on pp. [xxiii–xxiv](#), and the appendices, notation, and glossary are located at the back of the book. R packages are denoted in sans-serif fonts, e.g., `stats`, `splines`; and R commands in typewriter font and functions ending in parentheses, e.g., `coef()`. All logarithms are to base  $e$  unless otherwise specified.

The scope of this book is broad. To keep things to a manageable size, many topics which I would like to have included have had to be omitted. Rather than citing hundreds of journal articles, I have largely cited books and review papers. Other references and sources of data, etc. can be found in the packages' online help files. I apologize beforehand for many peoples' work that should probably have been cited but are not, due to space limitations or ignorance. A slightly unfortunate consequence of the book's breadth is that some of the notation had to be recycled, however, despite some symbols having multiple meanings, this should not raise excessive confusion because the topics where they are used are far enough apart.

Topics considered too specialized or technical for most readers have been marked with a dagger (†); these may be safely skipped without losing the overall discourse.

## Book Website and Software

Resources are available at the book's webpage, starting at

<http://www.stat.auckland.ac.nz/~yee>

Amongst other things, these should include R scripts, complements, errata, and any latest information. Readers wishing to run the software examples will need access to R (see <http://www.R-project.org>), and the VGAM and VGAMdata packages installed along side. The latter comprises some data used here. Version 1.0-0 or later for both packages are needed for compatibility with this book.

## Note for Instructors

Most of the contents are pitched at the level of a senior undergraduate or first-year postgraduate in statistics. It might be considered a follow-on course after a first course in GLMs. Students are assumed to have a reasonable working R background already, because only a few important R topics that are crucial are covered (at a cursory level too) in the first chapter. The first few chapters of Fox and Weisberg (2011) would be useful preparation, as well as those listed above.

For a course based on this book, it is recommended that about half of the time be devoted to Part I, so that the overall framework can be seen. The remaining time might be used on a selection of topics drawn from Part II depending on interest and need. Each Part II chapter is at an introductory level, and the references at the end of each chapter can be pursued for a deeper coverage.

Some of the exercises might be suitable for homework or tutorial purposes. Overall, they are a blend of short analyses involving real and simulated data, mathematical statistics problems, and statistical computing (R programming) tasks. While most problems are short, some are less straightforward and require more advanced R programming to obtain an elegant solution. The most demanding or time-consuming problems are marked with a dagger (†) and should be considered optional or for a sizeable project.

The book is primarily focused on estimation. Theoretical rigour that was unnecessary to the main messages has been omitted, and the presentation is chiefly driven by a pragmatic approach. Most of the material is self-contained, and I have tried to strike a balance between theory, computational details and practice for an applied statistician.

## Acknowledgements

This is a synthesis of work that started while I was a PhD student at the University of Auckland, and it has spanned the two decades since. Not surprisingly, there are many people to thank, both directly and indirectly. I am indebted to my PhD supervisors Chris Wild and Neil Mitchell who had a very positive effect during the formative years of my academic life. Ross Ihaka and Trevor Hastie pointed me in the right direction soon after in those very early days.

Many graduate students have contributed to this body of work in one way or another over the years, especially James Gray, Kai Huang, Víctor Miranda, and Alvin Sou. Successive heads of departments have helped arrange my teaching to allow for a faster progression of this work. Additionally, I wish to acknowledge the hospitality of many (mostly statistics) departments all across this terrestrial ball, which I have had the good fortune to visit over the years.

Many thanks are extended to Garry Tee for proofreading the book in its entirety. Also to many others who have given valuable feedback on portions thereof; those are Michel Dedual, Stéphane Guindon, Lian Heng, Richard Huggins, Murray Jorgensen, Ioannis Kosmidis, Louise McMillan, Alastair Scott, Alec Stephenson, Jakub Stoklosa, Berwin Turlach, Yong Wang, Alan Welsh, and Chris Wild. All errors, however, are my own. Thanks to John Young for his translation of several quotes from the ancients used at the beginning and ending of several chapters. For help with the typesetting (especially L<sup>A</sup>T<sub>E</sub>X and knitr), my thanks to Trevor Hastie, Alan Izenman, Andrew Robinson, and Yihui Xie. Also, the R-core members and the R community at large have been of great direct and indirect help; these include

the countless users who have sent in bug reports and suggestions over the years. Feedback from workshop attendees has been helpful too. I would also like to thank Hannah Bracken at Springer.

Last but not least, thanks to my wife Selina who has been very understanding during the intensive preparations.

## Future Directions

This book is a snapshot from a project that is a work in progress, and it will probably never be finished. Large tracts of theory and methodology presently lie undeveloped, and many existing parts require refinement. It is hoped that, over time, bugs will be attended to and further additions made to the software. Consequently, a warning to the reader: low-level internal changes to the software are ongoing, however high-level usage will change less. (Good coding practice such as the use of the generic functions listed in Chap. 8, as well as the suggestions of Sect. 8.5.1, should help minimize such problems). The NEWS file in the packages lists successive changes between versions.

Over time, I hope to further develop the theory and continue to add to the functionality of VGAM. Due to the volume of emails received (which exceeds the number I can reply to), please send bug reports only. And my apologies if I am unable to reply at all!

Auckland, New Zealand  
May 2015

Thomas W. Yee

*It is a good thing for an uneducated man to read books of quotations.*  
—Winston Churchill, *My Early Life*



# Contents

## Part I General Theory

<b>1</b>	<b>Introduction</b>	3
1.1	Introduction	3
1.1.1	Outline of This Book and a Quick Start	4
1.2	Six Illustrative Models	4
1.2.1	The Linear Model	5
1.2.2	Poisson and Negative Binomial Regression	6
1.2.3	Bivariate Odds Ratio Model	8
1.2.4	Proportional Odds and Multinomial Logit Models	11
1.3	General Framework	13
1.3.1	Vector Generalized Linear Models	13
1.3.2	Vector Generalized Additive Models	14
1.3.3	RR-VGLMs	15
1.3.4	QRR-VGLMs and Constrained Ordination	16
1.3.5	RR-VGAMs and Constrained Ordination	17
1.3.6	RCIMs	17
1.4	An Overview of VGAM	17
1.5	Some Background Topics	18
1.5.1	The Penalty Function Approach	19
1.5.2	Snippets of the S Language	20
1.5.3	More on the Number of Parameters and Terms	29
1.6	Summary	29
	Exercises	30
<b>2</b>	<b>LMs, GLMs and GAMs</b>	33
2.1	Introduction	33
2.2	LMs	33
2.2.1	The Hat Matrix	36
2.2.2	WLS and GLS	38
2.2.3	Fitting LMs in R	39
2.3	GLM Basics	39
2.3.1	Inference	42
2.3.2	GLM Residuals and Diagnostics	43

2.3.3	Estimation of $\phi$ . . . . .	44
2.3.4	Fitting GLMs in R . . . . .	45
2.3.5	Quasi-Likelihood Models . . . . .	45
2.3.6	Binary Responses . . . . .	46
2.4	Univariate Smoothing Methods . . . . .	49
2.4.1	The Classical Smoothing Problem . . . . .	49
2.4.2	Polynomial Regression . . . . .	51
2.4.3	Regression Splines . . . . .	52
2.4.4	Smoothing Splines . . . . .	59
2.4.5	P-Splines . . . . .	64
2.4.6	Local Regression . . . . .	66
2.4.7	Some General Theory . . . . .	74
2.5	Generalized Additive Models . . . . .	81
2.5.1	Why Additive Models? . . . . .	82
2.5.2	Binomial ‘example’ . . . . .	83
	Exercises . . . . .	85
<b>3</b>	<b>VGLMs</b> . . . . .	<b>91</b>
3.1	Introduction . . . . .	91
3.2	Iteratively Reweighted Least Squares . . . . .	92
3.2.1	Computation † . . . . .	94
3.2.2	Advantages and Disadvantages of IRLS . . . . .	96
3.3	Constraints on the Component Functions . . . . .	96
3.3.1	Fitting Constrained Models in VGAM . . . . .	99
3.4	The $\text{xij}$ Argument . . . . .	103
3.4.1	The Central Formula . . . . .	105
3.4.2	Using the $\text{xij}$ Argument in VGAM . . . . .	106
3.4.3	More Complicated Examples . . . . .	107
3.4.4	Smoothing . . . . .	110
3.4.5	Last Word . . . . .	114
3.5	Other Topics . . . . .	115
3.5.1	Multiple Responses . . . . .	115
3.5.2	Deviance . . . . .	116
3.5.3	Convergence . . . . .	117
3.5.4	Half-Stepping . . . . .	118
3.6	Inference . . . . .	120
3.6.1	Regularity Conditions . . . . .	120
3.6.2	Parameter Link Functions . . . . .	120
3.6.3	Hypothesis Testing . . . . .	120
3.6.4	Residual Degrees of Freedom . . . . .	120
3.7	Residuals and Diagnostics . . . . .	121
3.7.1	Working Residuals . . . . .	121
3.7.2	Pearson Residuals . . . . .	122
3.7.3	Response Residuals . . . . .	122
3.7.4	Deviance Residuals . . . . .	123
3.7.5	Hat Matrix . . . . .	123
	Exercises . . . . .	125

<b>4</b>	<b>VGAMs</b> .....	127
4.1	Introduction .....	127
4.1.1	Smoothing for VGLMs and VGAMs .....	127
4.1.2	The Vector Smoothing Problem .....	129
4.2	Vector Smoothing Methods .....	129
4.2.1	Vector Splines .....	130
4.2.2	Local Regression for Vector Responses † .....	141
4.2.3	On Linear Vector Smoothers .....	152
4.3	The Vector Additive Model and VGAM Estimation .....	153
4.3.1	Penalized Likelihood .....	153
4.3.2	Vector Backfitting .....	154
4.3.3	Degrees of Freedom and Standard Errors .....	156
4.3.4	Score Tests for Linearity † .....	157
4.4	On More Practical Aspects .....	159
4.4.1	Using the Software .....	159
4.4.2	<code>vsmooth.spline()</code> .....	159
4.4.3	Example: Cats and Dogs .....	160
	Exercises .....	164
<b>5</b>	<b>Reduced-Rank VGLMs</b> .....	167
5.1	Introduction .....	167
5.2	What Are RR-VGLMs? .....	168
5.2.1	Why Reduced-Rank Regression? .....	169
5.2.2	Normalizations .....	170
5.2.3	The Stereotype Model .....	171
5.3	A Few Details .....	171
5.3.1	Alternating Algorithm .....	171
5.3.2	SEs of RR-VGLMs .....	172
5.4	Other RR-VGLM Topics .....	172
5.4.1	Summary of RR-VGLM Software .....	172
5.4.2	Convergence .....	173
5.4.3	Latent Variable Plots and Biplots .....	173
5.4.4	Miscellaneous Notes .....	174
5.5	RR-VGLMs with Two Linear Predictors .....	174
5.5.1	Two-Parameter Rank-1 RR-VGLMs .....	174
5.5.2	Some Examples .....	176
5.6	RR-VGLM Examples .....	179
5.6.1	RR-Multiple Binomial Model .....	179
5.6.2	Reduced Rank Regression for Time Series .....	181
5.7	Row-Column Interaction Models .....	183
5.7.1	<code>rcim()</code> .....	184
5.7.2	Examples .....	187
5.7.3	Quasi-Variations .....	190
	Exercises .....	195
<b>6</b>	<b>Constrained Quadratic Ordination</b> .....	201
6.1	Introduction .....	201
6.1.1	Ordination .....	201
6.1.2	Prediction and Calibration .....	204

6.2	Quadratic RR-VGLMs for CQO	205
6.2.1	An Example: Hunting Spiders Data	207
6.2.2	Normalizations for QRR-VGLMs	211
6.3	Fitting QRR-VGLMs	213
6.3.1	Arguments <code>I.tolerances</code> and <code>eq.tolerances</code>	215
6.3.2	Initial Values and the <code>isd.latvar</code> Argument	217
6.3.3	Estimation	218
6.4	Post-Fitting Analyses	219
6.4.1	Arguments <code>varI.latvar</code> and <code>refResponse</code>	219
6.4.2	Ordination Diagrams	221
6.4.3	Perspective Plots	223
6.4.4	Trajectory Plots	224
6.4.5	Calibration	225
6.5	Some Practical Considerations	227
6.6	A Further Example: Trout Data	230
6.7	Unconstrained Quadratic Ordination	232
6.7.1	RCIMs and UQO	232
	Exercises	236
<b>7</b>	<b>Constrained Additive Ordination</b>	<b>239</b>
7.1	Introduction	239
7.2	Constrained Additive Ordination	239
7.2.1	Controlling Function Flexibility	240
7.2.2	Estimation	241
7.2.3	Practical Advice	241
7.3	Examples	242
7.3.1	Hunting Spiders	242
7.3.2	Trout Data	242
7.3.3	Diseases in a Cross-Sectional Study	245
7.4	Some Afterthoughts	246
	Exercises	247
<b>8</b>	<b>Using the VGAM Package</b>	<b>249</b>
8.1	Introduction	249
8.1.1	Naming Conventions of VGAM Family Functions	249
8.1.2	Naming Conventions of Arguments	249
8.2	Basic Usage of VGAM	252
8.2.1	Some Miscellaneous Arguments	252
8.2.2	Constraints	253
8.2.3	Control Functions	254
8.2.4	Convergence Criteria	255
8.2.5	Smart Prediction	256
8.3	More Advanced Usage of VGAM	258
8.3.1	Initial Values	258
8.3.2	Speeding Up the Computations	259
8.4	Some Details on Selected Methods Functions	260
8.4.1	The <code>fitted()</code> Generic	260
8.4.2	The <code>summary()</code> Generic	263
8.4.3	The <code>simulate()</code> Generic	264



8.4.4	The <code>plot()</code> Generic	265
8.4.5	The <code>predict()</code> Generic	267
8.5	Some Suggestions for Fitting VGLMs and VGAMs	268
8.5.1	Doing Things in Style	269
8.5.2	Some Useful Miscellanea	270
8.6	Slots in <code>vgam()</code> Objects	271
8.7	Solutions to Some Specific Problems	272
8.7.1	Obtaining the LM-Type Model Matrix for $\eta_j$	272
8.7.2	Predicting at $\bar{\mathbf{x}}$	273
	Exercises	274
<b>9</b>	<b>Other Topics</b>	277
9.1	Introduction	277
9.2	Computing the Working Weights †	277
9.2.1	Weight Matrices Not Positive-Definite	278
9.2.2	Simulated Fisher Scoring	278
9.2.3	The BHHH Method	279
9.2.4	Quasi-Newton Updates	280
9.2.5	Numerical Derivatives	280
9.3	Model Selection by AIC and BIC	281
9.4	Bias-Reduction	282
9.4.1	Binary Case	285
9.4.2	Software Implementation	285
	Exercises	286
<b>Part II Some Applications</b>		
<b>10</b>	<b>Some LM and GLM Variants</b>	291
10.1	Introduction	291
10.2	LM Variants	291
10.2.1	Varying-Coefficient Models	291
10.2.2	The Tobit Model	294
10.2.3	Seemingly Unrelated Regressions	297
10.2.4	The AR(1) Time Series Model	303
10.3	Binomial Variants	305
10.3.1	Two-Stage Sequential Binomial	305
10.3.2	The Bradley-Terry Model	306
10.3.3	Bivariate Responses: The Bivariate Probit Model	309
10.3.4	Binary Responses: Loglinear Models	310
10.3.5	Double Exponential Models	312
	Exercises	313
<b>11</b>	<b>Univariate Discrete Distributions</b>	317
11.1	Introduction	317
11.1.1	<code>dpqr</code> -Type Functions	318
11.2	Dispersion Models	321
11.3	Negative Binomial Regression	322
11.3.1	Computational Details	323
11.3.2	A Second Parameterization— <code>polyaR()</code>	324
11.3.3	Canonical Link	325

11.3.4	Fitting Other NB Variants	325
11.3.5	Some Practical Suggestions	327
11.4	The Beta-Binomial Model	329
11.5	Lagrangian Probability Distributions	330
	Exercises	332
<b>12</b>	<b>Univariate Continuous Distributions</b>	<b>343</b>
12.1	Introduction	343
12.2	Some Basics	345
12.2.1	Location, Scale and Shape Parameters	345
12.2.2	Initial Values	348
12.2.3	About the Tables	350
12.2.4	Q-Q Plots	351
12.2.5	Scale and Shape Parameters	353
12.3	A Few Groups of Distributions	353
12.3.1	Size Distributions	353
12.3.2	Pearson System	354
12.3.3	Poisson Points in the Plane and Volume	355
	Exercises	355
<b>13</b>	<b>Bivariate Continuous Distributions</b>	<b>371</b>
13.1	Introduction	371
13.1.1	Bivariate Distribution Theory—A Short Summary	371
13.1.2	dpr-Type Functions	373
13.2	Two Bivariate Distributions	373
13.2.1	Bivariate Normal Distribution	373
13.2.2	Plackett's Bivariate Distribution	375
13.3	Copulas	376
13.3.1	Dependency Measures	378
13.3.2	Archimedean Copulas	379
13.3.3	Initial Values	380
	Exercises	380
<b>14</b>	<b>Categorical Data Analysis</b>	<b>385</b>
14.1	Introduction	385
14.1.1	The Multinomial Distribution	385
14.2	Nominal Responses—The Multinomial Logit Model	386
14.2.1	The $x_{ij}$ Argument	388
14.2.2	Marginal Effects and Deviance	390
14.2.3	Stereotype Model	390
14.2.4	Summation Constraints	393
14.3	Using the Software	393
14.3.1	The Poisson Trick	394
14.4	Ordinal Responses	396
14.4.1	Models Involving Cumulative Probabilities	396
14.4.2	Coalminers Example I	401
14.4.3	Coalminers Example II	402
14.4.4	Models Involving Stopping- and Continuation-Ratios	404
14.4.5	Models Involving Adjacent Categories	406
14.4.6	Convergence	406

14.5	Genetic Models	407
14.5.1	The Dirichlet Distribution	408
14.5.2	The Dirichlet-Multinomial Distribution	410
	Exercises	411
<b>15</b>	<b>Quantile and Expectile Regression</b>	<b>415</b>
15.1	Introduction	415
15.1.1	Some Notation and Background	416
15.2	LMS-Type Methods	416
15.2.1	The Box-Cox-Normal Distribution Version	417
15.2.2	Other Versions	419
15.2.3	Example	420
15.2.4	More Advanced Usage of Some Methods Functions	422
15.3	Classical Quantile Regression by Scoring	424
15.3.1	Classical Quantile Regression	425
15.3.2	An Asymmetric Laplace Distribution	427
15.3.3	Example	429
15.3.4	The Onion Method	431
15.4	Expectile Regression	433
15.4.1	Introduction	433
15.4.2	Expectiles for the Linear Model	437
15.4.3	ALS Example	439
15.4.4	Poisson Regression	440
15.4.5	Melbourne Temperatures and the Onion Method	442
15.5	Discussion	442
	Exercises	443
<b>16</b>	<b>Extremes</b>	<b>447</b>
16.1	Introduction	447
16.1.1	Classical EV Theory	448
16.2	GEV	450
16.2.1	The $r$ -Largest Order Statistics	452
16.2.2	The Gumbel and Block-Gumbel Models	452
16.3	GPD	454
16.4	Diagnostics	456
16.4.1	Probability and Quantile Plots	456
16.4.2	Gumbel Plots	457
16.4.3	Mean Excess Plots	457
16.5	Some Software Details	457
16.6	Examples	458
16.6.1	Port Pirie Sea Levels: GEV Model	458
16.6.2	Venice Sea Levels: The Block-Gumbel Model	460
16.6.3	Daily Rainfall Data: The GPD Model	463
	Exercises	466
<b>17</b>	<b>Zero-Inflated, Zero-Altered and Positive Discrete Distributions</b>	<b>469</b>
17.1	Introduction	469
17.1.1	Software Details	472
17.1.2	A Zero-Inflated Poisson Example	472

17.2	The Positive-Bernoulli Distribution	474
17.2.1	Further Software Details	480
17.2.2	Deermice Example	480
17.2.3	Prinia Example	482
17.2.4	A $\mathcal{M}_{tbh}$ Example	484
17.2.5	Using <code>Select()</code>	486
17.2.6	Ephemeral and Enduring Memory Example	487
17.3	The Zero-Inflated Binomial Distribution	489
17.4	RR-ZIPs and RR-ZA Models	489
17.4.1	RR-ZAPs and RR-ZABs and Other Variants	490
	Exercises	491
<b>18</b>	<b>On VGAM Family Functions</b>	<b>499</b>
18.1	Introduction	499
18.2	Link Functions	501
18.2.1	Chain Rule Formulas	502
18.3	Family Function Basics	504
18.3.1	A Simple VGAM Family Function	504
18.3.2	Initial Values	509
18.3.3	Arguments in VGAM Family Functions	509
18.3.4	Extending the Exponential Distribution Family	510
18.3.5	The <code>wz</code> Data Structure	514
18.3.6	Implementing Constraints Within Family Functions	517
18.4	Some Other Topics	519
18.4.1	Writing R Packages and Documentation	519
18.4.2	Some S4 Issues	519
18.5	Examples	520
18.5.1	The Kumaraswamy Distribution Family	520
18.5.2	Simulated Fisher Scoring	524
18.6	Writing Smart Functions †	526
	Exercises	530
	<b>Erratum</b>	<b>E1</b>
<b>A</b>	<b>Background Material</b>	<b>3</b>
A.1	Some Classical Likelihood Theory	3
A.1.1	Likelihood Functions	3
A.1.2	Maximum Likelihood Estimation	4
A.1.3	Properties of Maximum Likelihood Estimators	12
A.1.4	Inference	14
A.2	Some Useful Formulas	19
A.2.1	Change of Variable Technique	19
A.2.2	Series Expansions	19
A.2.3	Order Notation	19
A.2.4	Conditional Expectations	21
A.2.5	Random Vectors	21
A.3	Some Linear Algebra	21
A.3.1	Cholesky Decomposition	22
A.3.2	Sherman-Morrison Formulas	23
A.3.3	QR Method	23
A.3.4	Singular Value Decomposition	24

A.4	Some Special Functions	24
A.4.1	Gamma, Digamma and Trigamma Functions	25
A.4.2	Beta Function	26
A.4.3	The Riemann Zeta Function	26
A.4.4	Erf and Erfc	26
A.4.5	The Marcum Q-Function	27
A.4.6	Exponential Integral, Debye Function	27
A.4.7	Bessel Functions	27
	Exercises	29
	<b>Glossary</b>	31
	<b>References</b>	37
	<b>Index</b>	55



# Acronyms

ALD	Asymmetric Laplace distribution (Sect. 15.3.2)
ALS	Asymmetric least squares (Sect. 15.4.1)
AML	Asymmetric maximum likelihood (Sect. 15.4.2)
BPM	Bivariate probit model (Sect. 10.3.3)
CAO	Constrained additive ordination (Chap. 7, Fig. 1.2)
CDF	Cumulative distribution function, $P(Y \leq y)$ or $P(Y_1 \leq y_1, Y_2 \leq y_2, \dots)$ (Sect. 11.1.1)
CQO	Constrained quadratic ordination, cf. QRR-VGLM (Chap. 6, Fig. 1.2)
CV	Cross-validation (Sect. 2.4.7.6)
EIM	Expected (Fisher) information matrix (3.11)
ENDF	Effective nonlinear degrees of freedom (Sects. 2.4.7.4, 7.2.1)
EVT	Extreme value theory (Chap. 16)
flop	Floating point operation
GAM	Generalized additive model (Sect. 2.5)
GEV	Generalized extreme value (Sect. 16.2)
GLM	Generalized linear model (Sect. 2.3)
GLS	Generalized least squares (Sect. 2.2.2)
GPD	Generalized Pareto distribution (Sect. 16.3)
GRC	Goodman's $RC(R)$ model (Sect. 5.7)
HWE	Hardy-Weinberg equilibrium (Sect. 14.5)
iff	If and only if, $\iff$
i.i.d.	Independent and identically distributed
IRLS	Iteratively reweighted least squares (Sect. 3.2)
LHS	Left hand side
LM	Linear model (Sect. 2.2)
LRT	Likelihood ratio test (Sect. A.1.4.1)
LS	Least squares (Sect. 2.2)
LSD	Least significant difference (Sect. 5.7.3.4)
MGF	Moment generating function
MLE	Maximum likelihood estimation or estimator
MLM	Multinomial logit model (Sects. 1.2.4, 14.2)
MPV	Median predicted value (Sect. 16.2.2)
MVUE	Minimum variance unbiased estimator (Sect. A.1.3.1)
NBD	Negative binomial distribution (Sects. 1.2.2, 11.3)

NCS	Natural cubic spline ((2.47), Sects. 2.4.3.2, 4.2.1)
NEF	Natural exponential family (Sect. 11.2)
NR	Newton-Raphson (Sects. 3.2, A.1.2.4)
OIM	Observed information matrix ((3.15), Sect. A.1.2.4)
OLS	Ordinary least squares (Sect. 2.2.2)
PDF	Probability density function, $f(y)$ or $f(\mathbf{y})$
PGF	Probability generating function (Sect. 11.5)
PMF	Probability mass function, $P(Y = y)$ or $P(Y_1 = y_1, Y_2 = y_2, \dots)$ (Sect. 11.1)
QLP	Quasi-likelihood Poisson (Sect. 11.3.4.1)
QRR-VGLM	Quadratic reduced-rank vector generalized linear model (Chap. 6, Fig. 1.2)
QSE	Quasi-standard error (Sect. 5.7.3)
RCIM	Row-column interaction model (Sect. 5.7, Fig. 1.2)
ResSS	Residual sum of squares (2.4)
RHS	Right hand side
RKHS	Reproducing kernel Hilbert space (Sect. 4.2.1.7)
RR	Reduced-rank (Sect. 1.3.3, Chap. 5)
RRR	Reduced-rank regression (Sect. 1.3.3, Chap. 5)
SE	Standard error (an estimated standard deviation)
SFS	Simulated Fisher scoring (Sect. 9.2.2)
SUR	Seemingly unrelated regressions (Sects. 4.2.2, 10.2.3)
SVD	Singular value decomposition (Sect. A.3.4)
UQO	Unconstrained quadratic ordination (Sect. 6.7)
VCM	Varying coefficient model (Sect. 10.2.1)
VGAM	Vector generalized additive model (Chap. 4, Fig. 1.2)
VGLM	Vector generalized linear model (Chap. 3, Fig. 1.2)
VLM	Vector linear model (3.22), Fig. 1.2
VMM	Vector measurement model (4.2)
WLS	Weighted least squares (Sect. 2.2.2)
ZA	Zero-altered (Sect. 17.1)
ZI	Zero-inflated (Sect. 17.1)