

Statistical Modelling by Exponential Families

This book is a readable, digestible introduction to exponential families, encompassing statistical models based on the most useful distributions in statistical theory, such as the normal, gamma, binomial, Poisson, and negative binomial. Strongly motivated by applications, it presents the essential theory and then demonstrates the theory's practical potential by connecting it with developments in areas such as item response analysis, social network models, conditional independence and latent variable structures, and point process models. Extensions to incomplete data models and generalized linear models are also included. In addition, the author gives a concise account of the philosophy of Per Martin-Löf in order to connect statistical modelling with ideas in statistical physics, such as Boltzmann's law. Written for graduate students and researchers with a background in basic statistical inference, the book includes a vast set of examples demonstrating models for applications and numerous exercises embedded within the text as well as at the ends of chapters.

ROLF SUNDBERG is Professor Emeritus of Statistical Science at Stockholm University. His work embraces both theoretical and applied statistics, including principles of statistics, exponential families, regression, chemometrics, stereology, survey sampling inference, molecular biology, and paleoclimatology. In 2003, he won (with M. Linder) the award for best theoretical paper in the *Journal of Chemometrics* for their work on multivariate calibration, and in 2017 he was named Statistician of the Year by the Swedish Statistical Society.

INSTITUTE OF MATHEMATICAL STATISTICS
TEXTBOOKS

Editorial Board

Nancy Reid (University of Toronto)
Ramon van Handel (Princeton University)
Xuming He (University of Michigan)
Susan Holmes (Stanford University)

ISBA Editorial Representative

Peter Müller (University of Texas at Austin)

IMS Textbooks give introductory accounts of topics of current concern suitable for advanced courses at master's level, for doctoral students and for individual study. They are typically shorter than a fully developed textbook, often arising from material created for a topical course. Lengths of 100–290 pages are envisaged. The books typically contain exercises.

In collaboration with the International Society for Bayesian Analysis (ISBA), selected volumes in the IMS Textbooks series carry the “with ISBA” designation at the recommendation of the ISBA editorial representative.

Other Books in the Series (*with ISBA)

1. *Probability on Graphs*, by Geoffrey Grimmett
2. *Stochastic Networks*, by Frank Kelly and Elena Yudovina
3. *Bayesian Filtering and Smoothing*, by Simo Särkkä
4. *The Surprising Mathematics of Longest Increasing Subsequences*, by Dan Romik
5. *Noise Sensitivity of Boolean Functions and Percolation*, by Christophe Garban and Jeffrey E. Steif
6. *Core Statistics*, by Simon N. Wood
7. *Lectures on the Poisson Process*, by Günter Last and Mathew Penrose
8. *Probability on Graphs (Second Edition)*, by Geoffrey Grimmett
9. *Introduction to Malliavin Calculus*, by David Nualart and Eulàlia Nualart
10. *Applied Stochastic Differential Equations*, by Simo Särkkä and Arno Solin
11. **Computational Bayesian Statistics*, by M. Antónia Amaral Turkman, Carlos Daniel Paulino, and Peter Müller
12. *Statistical Modelling by Exponential Families*, by Rolf Sundberg

Statistical Modelling by Exponential Families

ROLF SUNDBERG
Stockholm University



Cambridge University Press
978-1-108-47659-1 — Statistical Modelling by Exponential Families
Rolf Sundberg
Frontmatter
[More Information](#)

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom
One Liberty Plaza, 20th Floor, New York, NY 10006, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre,
New Delhi – 110025, India
79 Anson Road, #06–04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org

Information on this title: www.cambridge.org/9781108476591

DOI: 10.1017/9781108604574

© Rolf Sundberg 2019

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 2019

Printed and bound in Great Britain by Clays Ltd, Elcograf S.p.A.

A catalogue record for this publication is available from the British Library.

Library of Congress Cataloguing-in-Publication data

Names: Sundberg, Rolf, 1942– author.

Title: Statistical modelling by exponential families / Rolf Sundberg (Stockholm University).

Description: Cambridge ; New York, NY : Cambridge University Press, 2019. |

Series: Institute of Mathematical Statistics textbooks ; 12 |

Includes bibliographical references and index.

Identifiers: LCCN 2019009281 | ISBN 9781108476591 (hardback : alk. paper) |

ISBN 9781108701112 (pbk. : alk. paper)

Subjects: LCSH: Exponential families (Statistics)—Problems, exercises, etc. |

Distribution (Probability theory)—Problems, exercises, etc.

Classification: LCC QA276.7 .S86 2019 | DDC 519.5—dc23

LC record available at <https://lcn.loc.gov/2019009281>

ISBN 978-1-108-47659-1 Hardback

ISBN 978-1-108-70111-2 Paperback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

*To Margareta,
and to Per;
celebrating 50 years*

Contents

<i>Examples</i>	ix
<i>Preface</i>	xii
1 What Is an Exponential Family?	1
2 Examples of Exponential Families	6
2.1 Examples Important for the Sequel	6
2.2 Examples Less Important for the Sequel	15
2.3 Exercises	21
3 Regularity Conditions and Basic Properties	24
3.1 Regularity and Analytical Properties	24
3.2 Likelihood and Maximum Likelihood	31
3.3 Alternative Parameterizations	36
3.4 Solving Likelihood Equations Numerically	45
3.5 Conditional Inference for Canonical Parameter	46
3.6 Common Models as Examples	50
3.7 Completeness and Basu's Theorem	59
3.8 Mean Value Parameter and Cramér–Rao (In)equality	61
4 Asymptotic Properties of the MLE	64
4.1 Large Sample Asymptotics	64
4.2 Small Sample Refinement: Saddlepoint Approximations	70
5 Testing Model-Reducing Hypotheses	75
5.1 Exact Tests	76
5.2 Fisher's Exact Test for Independence, Homogeneity, Etc.	80
5.3 Further Remarks on Statistical Tests	84
5.4 Large Sample Approximation of the Exact Test	86
5.5 Asymptotically Equivalent Large Sample Tests	90
5.6 A Poisson Trick for Deriving Test Statistics	94

Contents

vii

6	Boltzmann's Law in Statistics	100
6.1	Microcanonical Distributions	100
6.2	Boltzmann's Law	102
6.3	Hypothesis Tests in a Microcanonical Setting	108
6.4	Statistical Redundancy	109
6.5	A Modelling Exercise in the Light of Boltzmann's Law	114
7	Curved Exponential Families	118
7.1	Introductory Examples	118
7.2	Basic Theory for ML Estimation and Hypothesis Testing	124
7.3	Statistical Curvature	129
7.4	More on Multiple Roots	131
7.5	Conditional Inference in Curved Families	136
8	Extension to Incomplete Data	143
8.1	Examples	143
8.2	Basic Properties	147
8.3	The EM Algorithm	150
8.4	Large-Sample Tests	155
8.5	Incomplete Data from Curved Families	155
8.6	Blood Groups under Hardy–Weinberg Equilibrium	156
8.7	Hidden Markov Models	159
8.8	Gaussian Factor Analysis Models	161
9	Generalized Linear Models	164
9.1	Basic Examples and Basic Definition	164
9.2	Models without Dispersion Parameter	169
9.3	Models with Dispersion Parameter	175
9.4	Exponential Dispersion Models	181
9.5	Quasi-Likelihoods	183
9.6	GLMs versus Box–Cox Methodology	184
9.7	More Application Areas	186
10	Graphical Models for Conditional Independence Structures	191
10.1	Graphs for Conditional Independence	192
10.2	Graphical Gaussian Models	195
10.3	Graphical Models for Contingency Tables	201
10.4	Models for Mixed Discrete and Continuous Variates	205
11	Exponential Family Models for Social Networks	210
11.1	Social Networks	210
11.2	The First Model Stage: Bernoulli Graphs	211
11.3	Markov Random Graphs	212

viii	<i>Contents</i>	
11.4	Illustrative Toy Example, $n = 5$	218
11.5	Beyond Markov Models: General ERGM Type	225
12	Rasch Models for Item Response and Related Models	228
12.1	The Joint Model	229
12.2	The Conditional Model	231
12.3	Testing the Conditional Rasch Model Fit	234
12.4	Rasch Model Conditional Analysis by Log-Linear Models	239
12.5	Rasch Models for Polytomous Response	240
12.6	Factor Analysis Models for Binary Data	241
12.7	Models for Rank Data	243
13	Models for Processes in Space or Time	246
13.1	Models for Spatial Point Processes	246
13.2	Time Series Models	254
14	More Modelling Exercises	258
14.1	Genotypes under Hardy–Weinberg Equilibrium	258
14.2	Model for Controlled Multivariate Calibration	259
14.3	Refindings of Ringed Birds	259
14.4	Statistical Basis for Positron Emission Tomography	262
<i>Appendix A</i>	Statistical Concepts and Principles	265
<i>Appendix B</i>	Useful Mathematics	268
B.1	Some Useful Matrix Results	268
B.2	Some Useful Calculus Results	269
	<i>Bibliography</i>	271
	<i>Index</i>	278

Examples

1.1	The structure function for repeated Bernoulli trials	4
2.1	Bernoulli trials and the binomial	6
2.2	Logistic regression models	7
2.3	Poisson distribution family for counts	7
2.4	The multinomial distribution	8
2.5	Multiplicative Poisson and other log-linear models	9
2.6	‘Time to first success’ (geometric distribution)	10
2.7	Exponential and gamma distribution models	10
2.8	A sample from the normal distribution	11
2.9	Gaussian linear models	11
2.10	The multivariate normal distribution	12
2.11	Covariance selection models	13
2.12	Exponential tilting families	14
2.13	Finite Markov chains	15
2.14	Von Mises and Fisher distributions for directional data	16
2.15	Maxwell–Boltzmann model in statistical physics	17
2.16	The Ising model	19
3.1	Marginality and conditionality for Gaussian sample	41
3.2	Two Poisson variates	47
3.3	Conditional inference for Gaussian sample, cont’d	48
3.4	Bernoulli trials, continued from Example 2.1	50
3.5	Logistic regression, continued from Example 2.2	50
3.6	Conditional inference in logistic regression	51
3.7	Poisson sample, continued from Example 2.3	52
3.8	Multiplicative Poisson model, continued from Example 2.5	52
3.9	Exponential and gamma, continued from Example 2.7	53
3.10	A Gaussian sample, continued from Example 2.8	54
3.11	The multivariate normal, continued from Example 2.10	55
3.12	Covariance selection models, continued	55

x	<i>Examples</i>	
4.1	Bernoulli trials, continued from Example 3.4	68
4.2	Time to n successes, continued from Example 2.6	68
4.3	A sample from the normal, continued from Example 3.10	69
4.4	Time to n successes, continued from Example 4.2	73
5.1	Variance test for a normal distribution	78
5.2	Mean value test for a normal distribution	79
5.3	The score test for homogeneity or independence or multiplicity in 2×2 -tables, continuation from Section 5.2	88
5.4	Test for independence or homogeneity	96
6.1	Bernoulli trials (again)	100
6.2	Cross-classification of n items by two binary criteria	101
6.3	Ideal gas model	101
6.4	Bernoulli trials, continued from Example 6.1	102
6.5	Binary cross-classifications, continued from Example 6.2	102
6.6	Ideal gas, continued from Example 6.3	102
6.7	Bernoulli trials, continued from Example 6.4	103
6.8	Cross-classifications, continued from Example 6.5	103
6.9	Ideal gas, continued from Example 6.6	103
6.10	Bernoulli trials, continued from Example 6.7	103
6.11	Sex of newborns as Bernoulli trials	111
6.12	Sex of newborns, continued from Example 6.11	112
6.13	Independence between income and number of children	113
7.1	Normal sample with known coefficient of variation	118
7.2	Two normal samples with $\mu_1 = \mu_2$ ('Behrens–Fisher')	119
7.3	Probit regression	119
7.4	SUR models in multivariate linear regression	119
7.5	Covariance structures	121
7.6	Exponential life-times, with censored data	122
7.7	Binary fission, or Yule process	123
7.8	'Shaved dice' inference	123
7.9	Exponential life-times with censored data, continued	126
7.10	Behrens–Fisher model, continued from Example 7.2	131
7.11	Two parametrically related Poisson variates	136
7.12	Normal sample with known coefficient of variation	138
7.13	Multinomial model with nonunique ancillary statistic	138
7.14	Two Poisson variates, continued	140
8.1	Grouped or censored continuous type data	143
8.2	Aggregated multinomial data	144
8.3	The folded normal distribution	144

	<i>Examples</i>	xi
8.4	Mixture of two normal distributions	145
8.5	Random data with measurement errors	145
8.6	Missing data in multivariate analysis	146
8.7	The negative binomial distribution	146
8.8	The Cauchy distribution family	147
8.9	The folded normal distribution, continued	150
9.1	Gaussian linear models, continued from Example 2.9	164
9.2	Multiplicative Poisson and log-linear models for contingency tables, continued from Example 2.5	165
9.3	Probit and logit regression models for binary data	166
9.4	Normal distribution with dispersion parameter	175
9.5	Exponential and gamma (with dispersion parameter)	177
9.6	Negative binomial to allow Poisson overdispersion	179
10.1	A conditional independence graph with five nodes	192
12.1	Two model test examples	237
12.2	The influenza example again	240
13.1	A spatially inhomogeneous Poisson process	247
13.2	Strauss model for Swedish pines data	250
13.3	A dynamic logit model in practice	256

Preface

The theoretical importance of exponential families of distributions has long been evident. Every book on general statistical theory includes a few sections introducing exponential families. However, there are three gaps in the literature that I want to fill with the present text:

- It is the first book devoted wholly to exponential families and written at the Master's/PhD level. As a course book, it can stand by itself or supplement other literature on parametric statistical inference.
- It aims to demonstrate not only the essentials of the elegant and powerful general theory but also the extensive development of exponential and related models in a diverse range of applied disciplines.
- It gives an account of the half-century-old but still influential ideas of Per Martin-Löf about such models. Part of this innovative material has previously been available only in my handwritten lecture notes in Swedish from 1969–1970.

Two older books are devoted to the general theory of exponential families: Ole Barndorff-Nielsen's *Information and Exponential Families in Statistical Theory* (Barndorff-Nielsen, 1978) and Larry Brown's *Fundamentals of Statistical Exponential Families* (Brown, 1988). The present text is quite different in intention. The general presentation is less focused on mathematical rigour, and instead the text is much more directed toward applications (even though it does not include much real data). My ambition is not to present a self-contained stringent theory, which would have required more space and distracted the reader from my main messages. Thus I refer to other works for some proofs or details of proofs not short enough or instructive enough to be included.

Instead, in order to show the theory's potential, I make room to describe developments in numerous areas of applied modelling. The topics treated include incomplete data, generalized linear models, conditional independence and latent variable structures, social network models, models for

item analysis and spatial point process models. The methodology of generalized linear models will take us a bit beyond exponential family theory. We will also be outside the scope of the standard exponential families when the parameter space or the data are incomplete, but not so far outside that the theory is not beneficial.

The inferential philosophy underlying this text is that of a frequentist statistician. In particular, conditional inference plays a substantial role in the text. This should not preclude potential interest from a Bayesian point of view, however. For both frequentists and Bayesians the likelihood is central, and it plays a dominant role in the present text.

This textbook contains more material than typically covered in a single course. My own proposed selection for a basic course on the topic would cover Chapters/Sections 1, 2.1, 3.1–6, 4.1, 5, 7.1–2, 8.1–3, and 9.1–3. The choice of additional material should be driven by the interests of students and/or the instructor. Instruction on exponential families can also be combined with other aspects of statistical theory. At Stockholm University, a theoretical graduate level course consists of exponential families jointly with more general theory of parametric inference.

As mentioned, the text is written with graduate students primarily in mind. It has a large number of examples and exercises, many of them appearing recurrently. They range from small instructional illustrations to application areas of independent interest and of great methodological or applied utility in their own right, represented by separate sections or chapters. This simplifies the choice of topics by taste. Chapter 2 introduces a large number of common and less common distributions and models. Most of these examples reappear in later chapters. Chapters 10–13 can be regarded as elaborated examples per se.

Exercises are typically found at the ends of sections, but there are also sections entirely devoted to exercises. Solutions to exercises are not included here, but can be found at www.cambridge.org/9781108701112.

As prerequisites, readers should have a general knowledge of the theory of statistics and experience with the use of statistical methods (the more, the better). Appendix A defines some statistical concepts and lists the principles assumed to be in the mind of the reader. Some particularly useful calculus and matrix algebra is found in the other short appendices.

Application of an exponential family model in practice typically requires a statistical computer package. I give some references to the huge, free set of **R** packages, but these references could be regarded as examples, because alternatives often exist in other environments, or even in **R** itself.

My original inspiration to study the theory and methodology of exponential families goes 50 years back in time, in particular to unpublished lectures given by Per Martin-Löf in 1969–1970 (Martin-Löf, 1970). While skipping some technicalities, I try to do justice to many of his original thoughts and deep results, in particular in Chapter 6. Per Martin-Löf's lecture notes were written in Swedish, which of course severely restricted their accessibility and influence outside the Scandinavian languages. Nevertheless, these notes have got about one hundred Google Scholar citations. The Swedish original is available on my department's website: www.math.su.se/PML1970.

Implicit above is my deep gratitude to Per Martin-Löf for introducing me to the area, for suggesting a fruitful thesis topic and for his sharp mind challenging me to achieve a clear understanding myself.

As mentioned, earlier versions of this text have been used as course texts in my department during their development. I'm particularly grateful to Michael Höhle, who took over the course after me, and whose many questions and comments have led to important corrections and other improvements. I am also grateful to students commenting on previous versions, and to colleagues inside and outside Stockholm University for comments and stimulation.

I appreciate greatly the support of David Cox and Diana Gillooly, who have both been eager to see these notes published in the IMS Textbook series with Cambridge University Press. After many years the goal is now achieved, and I hope they like the result.

Finally, my deep thanks go to my wife Margareta, my wife since 50 years ago, without whose patience and support this project would not have come to completion.