SpringerBriefs in Statistics

Monia Lupparelli · Giovanni Maria Marchetti · Claudia Tarantola

Regression Graph Models for Categorical Data

Parameterization and Inference

Springer

SpringerBriefs in Statistics

SpringerBriefs present concise summaries of cutting-edge research and practical applications across a wide spectrum of fields. Featuring compact volumes of 50 to 125 pages, the series covers a range of content from professional to academic. Typical topics might include:

- A timely report of state-of-the art analytical techniques
- A bridge between new research results, as published in journal articles, and a contextual literature review
- A snapshot of a hot or emerging topic
- An in-depth case study or clinical example
- A presentation of core concepts that students must understand in order to make independent contributions

SpringerBriefs in Statistics showcase emerging theory, empirical research, and practical application in Statistics from a global author community.

SpringerBriefs are characterized by fast, global electronic dissemination, standard publishing contracts, standardized manuscript preparation and formatting guidelines, and expedited production schedules.

Monia Lupparelli · Giovanni Maria Marchetti · Claudia Tarantola

Regression Graph Models for Categorical Data

Parameterization and Inference

Springer

Monia Lupparelli
Department of Statistics, Computer Science, Applications
University of Florence
Florence, Italy

Giovanni Maria Marchetti
Department of Statistics, Computer Science, Applications
University of Florence
Florence, Italy

Claudia Tarantola
Department of Economics, Management and Quantitative Methods
University of Milan
Milan, Italy

ISSN 2191-544X ISSN 2191-5458 (electronic)
SpringerBriefs in Statistics
ISBN 978-3-031-99796-9 ISBN 978-3-031-99797-6 (eBook)
https://doi.org/10.1007/978-3-031-99797-6

Mathematics Subject Classification: 62H17, 62F15

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To Hélèn Massam

Preface

In this work we have attempted to summarize our activity in the field of chain graph models, with a particular emphasis on the parameterizations and inference of regression chain graph models, often referred as regression graph models, for the analysis of categorical data. Unlike other families of graphical models, such as the Bayesian networks (that is directed acyclic graphs) or the undirected graph models, which are by now extensively known and used, the regression graphs do not enjoy yet the same popularity. Our motivation for writing this book stems from the conviction that this class of models can be highly beneficial for statistical data analysis. This family of graphs defines a framework of recursive multivariate regressions that find application in important fields such as structural equation models and marginal models for qualitative data. In contrast to undirected graphical models, the regression graph models allow the variables considered on equal standing to be grouped into chain components that can serve different roles in the data-generating mechanism. Regression graphs also bridge some limitations of Bayesian networks used to define univariate recursive regression models that do not account for additional forms of dependencies for instance induced by unobserved variables.

This work has a twofold intention. The first is to consolidate knowledge on regression graphs, their interpretation in terms of sequences of multivariate regressions, interpretable parameterizations for categorical data, and inference and model selection within the frequentist and Bayesian approaches. The second objective of this work is encouraging both the applications and the research concerning this class of graphical models.

The book is intended primarily for graduate and Ph.D. students in Statistics and Data Science who are familiar with the basics of graphical Markov models and of categorical data analysis, and for motivated researchers in specific applied fields. Data and R code used in the book are made available at the website https://github.com/StaThin/RGM.

The writing of this book has been stimulated and enriched by collaborations and discussions over the time with colleagues and friends from the community of graphical models and categorical data analysis. We are especially grateful to Nanny Wermuth, David Cox, Wicher Bergsma, Tamás Rudas, Ioannis Ntzoufras,

Anna Gottard, Kayvan Sadeghi, Alberto Roverato, Luca La Rocca, Mathias Drton, Robin Evans, Francesco Bartolucci, Antonio Forcina, Guido Consonni, Søren Højsgaard. We would also thank the Editors of Springer Publications and the Anonymous Reviewers for their careful reading, constructive comments and suggestions that have thoroughly improved this manuscript.

Florence, Italy
June 2025

Monia Lupparelli
Giovanni Maria Marchetti
Claudia Tarantola

Competing Interests The authors have no competing interests to declare that are relevant to the content of this manuscript.

Contents

1 Regression Graph Models 1
1.1 Graphical Markov Models 1
1.1.1 Introduction 1
1.1.2 Marginal Independence 2
1.1.3 Conditional Independence 3
1.1.4 Role of the Variables 3
1.2 Marginal Independence Models 4
1.2.1 The Connected Set Markov Property 6
1.2.2 Conditional Covariance Graphs 8
1.3 Multivariate Regression Graphs 9
1.4 Sequences of Regressions 13
1.5 The Regression Chain Graph Markov Property 16
1.6 Some Applications to Categorical Data 18
1.7 Interpretation via Latent Variables 22
1.8 Appendix: Recursive Multivariate Regressions for a Gaussian Distribution 24
1.9 Bibliographic Notes 25

2 Multivariate Logistic Regression Models 29
2.1 An Introduction to Marginal Models for Categorical Data 29
2.2 The Multivariate Logistic Model 31
2.2.1 The Parameterization 31
2.2.2 Graphical Models of Marginal Independence 34
2.3 Multivariate Regression Models for Categorical Responses 36
2.4 Conditional Independence Constraints 39
2.5 Log-Linear Marginal Models 43
2.6 Variation Independence 46
2.7 Marginal Log-Linear Parameterization for Bi-directed Graphs 47

2.8 Marginal Log-Linear Parameterizations for Regression Graphs . . . 48
2.9 Appendix: Link Functions for Marginal Log-Linear Parameterizations . . . 49
2.10 Bibliographic Notes . . . 50

3 Maximum Likelihood Inference . . . 53
3.1 Likelihood Function for Discrete Regression Graph Models . . . 53
3.2 MLEs for a Generalized Class of Non-linear Regression Models . . . 54
3.2.1 The Algorithm . . . 55
3.2.2 Asymptotic Properties and Model Comparison . . . 57
3.3 Fitting a Regression Graph Model . . . 58
3.4 Strategies for Structure Learning of Regression Chain Graphs . . . 63
3.5 Appendix: Hessian Matrix Approximations . . . 69
3.6 Bibliographic Notes . . . 70

4 Bayesian Inference . . . 73
4.1 Basic Background . . . 73
4.1.1 Parametric Inference . . . 73
4.1.2 Model Comparison . . . 74
4.2 Bayesian Analysis of Regression Graphs . . . 76
4.3 Markov Equivalence . . . 77
4.4 Cell Probability Parameterization . . . 80
4.4.1 Prior Distribution . . . 81
4.4.2 Model Selection . . . 84
4.4.3 Bayesian Inference for Coppen's Data . . . 85
4.5 Marginal Log-Linear Parameterization . . . 90
4.5.1 Prior Distribution . . . 91
4.5.2 MCMC Algorithm . . . 92
4.6 Concluding Remark . . . 96
4.7 Appendix: Jacobian and Link Function . . . 97
4.7.1 Jacobian Matrix . . . 97
4.7.2 Contrast Matrix $\boldsymbol{C}_{\mathcal{M}}$ with Sum to Zero Constraints . . . 98
4.8 Bibliographic Notes . . . 99

References . . . 103

Index . . . 107