

# Graphical Models in Applied Multivariate Statistics

Joe Whittaker  
*Department of Mathematics*  
*University of Lancaster*  
*UK*

JOHN WILEY & SONS

Chichester • New York • Brisbane • Toronto • Singapore

# Contents

Preface	ix
Glossary	xiii
1 Introduction	1
1.1 Mathematics Marks	1
1.2 Infant Survival	6
1.3 Graphical Models and Modelling	8
1.4 Notational Preliminaries	14
1.5 Overview	18
2 Independence and Interaction	23
2.1 Independent Events	24
2.2 Independent Random Vectors	30
2.3 Mixed Derivative Measures of Interaction	35
2.4 The Bernoulli Distribution	38
2.5 Three Dimensional Bernoulli Distribution	42
2.6 The Normal Distribution	47
2.7 Exercises	51
3 Independence Graphs	56
3.1 Graph Theory	58
3.2 Independence Graphs	60
3.3 Separation	63
3.4 Markov Properties	68
3.5 Directed Acyclic Independence Graphs	71
3.6 Chain Independence Graphs	77
3.7 Exercises	82
4 Information Divergence	86
4.1 Kullback-Leibler Information Divergence	87
4.2 Divergence: a Heuristic Derivation	91
4.3 Properties of Information Divergence	94

4.4	Independence and Information Proper	99
4.5	Information in an Independence Graph	105
4.6	Divergence and Collapsibility	108
4.7	Iterative Proportional Fitting	112
4.8	Exercises	117
5	The Inverse Variance	120
5.1	Random Vectors, Expectation and Covariance	121
5.2	Linear Least Squares Prediction	125
5.3	Properties of the Predictor	128
5.4	Predicting the Mathematics Marks	132
5.5	The Partial Covariance	134
5.6	Invariance, Additivity and Recurrence	137
5.7	The Inverse Variance	141
5.8	Inverse Variance Lemma: Corollaries	143
5.9	Variance: Trace and Determinant	148
5.10	Exercises	150
6	Graphical Gaussian Models	155
6.1	Graphical Models and Modelling	156
6.2	The Multivariate Normal Distribution	158
6.3	Marginal and Conditional Distributions	162
6.4	Divergence between Normal Distributions	166
6.5	The Likelihood Function	170
6.6	Maximum Likelihood Estimates	173
6.7	Direct and Indirect Estimates	177
6.8	The Analysis of Deviance	185
6.9	Wishart, Bartlett and Mahalanobis	190
6.10	Exercises	194
7	Graphical Log-linear Models	198
7.1	The Cross-classified Multinomial Distribution	200
7.2	Log-linear Expansions and $u$ -terms	204
7.3	Graphical Log-linear Models	207
7.4	The Likelihood Function	213
7.5	Simple Examples of MLE	217
7.6	Estimates for Conditional Independence Models	222
7.7	Partitioning the Deviance	226
7.8	Diagnostics for Log-linear Models	232
7.9	Exercises	235

8	Model Selection	241
8.1	Issues in Model Selection	242
8.2	Log-linear Model Selection	246
8.3	Graphical Model Search Strategy	251
8.4	Model Selection: a Continuous Example	254
8.5	Model Selection: a Discrete Example	261
8.6	Exercises	265
9	Methods for Sparse Tables	269
9.1	The Partial Cross-product Ratio	270
9.2	The All Two-way Interaction Model	276
9.3	A Case Study: Rochdale	279
9.4	Exact Conditional Tests	285
9.5	Exact Tests for Graphical Models	292
9.6	Exercises	296
10	Regression and Graphical Chain Models	300
10.1	Conditional Probability Models	304
10.2	Fitting Regression Models in the Joint Distribution	310
10.3	A Case Study: Noctuid Moth Trappings	313
10.4	Several blocks: A Case Study	319
10.5	Regression Models for Continuous Variables	322
10.6	Regression: Some Sampling Results	329
10.7	Logistic Regression	333
10.8	Two Case Studies with Categorical Variables	336
10.9	Exercises	342
11	Models for Mixed Variables	345
11.1	The CG Distribution	347
11.2	Interaction Model Formulae	352
11.3	The Likelihood Function	355
11.4	Issues in Modelling	357
11.5	Case Studies	359
11.6	Exercises	370
12	Decompositions and Decomposability	374
12.1	Factorisation	375
12.2	Partial Factorisations: Decompositions	381
12.3	Irreducible Components	384
12.4	Decomposability	388

12.5	Collapsibility	394
12.6	Exercises	401
Appendices		403
A.1	Computing Packages	403
A.2	Outline Answers	408
References		426
Author Index		436
Subject Index		439