

# Contents

|                                                          |           |
|----------------------------------------------------------|-----------|
| <i>Preface</i>                                           | vii       |
| <i>About the Author</i>                                  | ix        |
| <b>1. Introduction and Review</b>                        | <b>1</b>  |
| 1.1 Statistical Inference . . . . .                      | 1         |
| 1.2 Bayesian Inference . . . . .                         | 2         |
| 1.2.1 Main steps . . . . .                               | 3         |
| 1.2.2 Some basic models . . . . .                        | 4         |
| 1.3 Conditional Independence . . . . .                   | 10        |
| 1.3.1 Definitions . . . . .                              | 10        |
| 1.3.2 CI in statistical inference . . . . .              | 11        |
| <b>Part 1 Latent Structure Models</b>                    | <b>13</b> |
| <b>2. Incomplete Data and the EM Algorithm</b>           | <b>15</b> |
| 2.1 Assumptions . . . . .                                | 15        |
| 2.1.1 Ignorability . . . . .                             | 16        |
| 2.1.2 Observed-data likelihood and posterior . . . . .   | 17        |
| 2.2 The EM Algorithm and Its Properties . . . . .        | 18        |
| 2.2.1 The algorithm . . . . .                            | 18        |
| 2.2.2 EM as an MM algorithm . . . . .                    | 21        |
| 2.2.3 Properties of the EM . . . . .                     | 22        |
| 2.2.4 Missing information and convergence rate . . . . . | 24        |
| 2.2.5 Another example . . . . .                          | 25        |

|           |                                                            |           |
|-----------|------------------------------------------------------------|-----------|
| 2.3       | EM for Exponential Families . . . . .                      | 26        |
| 2.3.1     | Exponential families . . . . .                             | 26        |
| 2.3.2     | MLE for complete data . . . . .                            | 27        |
| 2.3.3     | EM for incomplete data . . . . .                           | 28        |
| 2.4       | Incomplete Normal Data . . . . .                           | 30        |
| 2.4.1     | The complete-data MLE . . . . .                            | 31        |
| 2.4.2     | EM algorithm for incomplete normal data . . . . .          | 32        |
| 2.5       | Bayesian Inference with Missing Data . . . . .             | 34        |
| 2.5.1     | Data augmentation . . . . .                                | 35        |
| 2.5.2     | Examples . . . . .                                         | 36        |
| 2.6       | Problems . . . . .                                         | 39        |
| <b>3.</b> | <b>Mixture Modeling</b>                                    | <b>41</b> |
| 3.1       | Mixture Models . . . . .                                   | 41        |
| 3.1.1     | Definition . . . . .                                       | 41        |
| 3.1.2     | Maximum likelihood estimate using EM . . . . .             | 43        |
| 3.2       | Model-Based Clustering . . . . .                           | 46        |
| 3.3       | Motif Discovery . . . . .                                  | 49        |
| 3.3.1     | Problem formulation . . . . .                              | 49        |
| 3.3.2     | Maximum likelihood via EM . . . . .                        | 50        |
| 3.3.3     | Bayesian inference via Gibbs sampler . . . . .             | 53        |
| 3.4       | Problems . . . . .                                         | 54        |
| <b>4.</b> | <b>Hidden Markov Models</b>                                | <b>57</b> |
| 4.1       | Elements of an HMM . . . . .                               | 57        |
| 4.1.1     | Example and definition . . . . .                           | 57        |
| 4.1.2     | Graphical representation . . . . .                         | 58        |
| 4.2       | Maximum Likelihood Estimate via the EM Algorithm . . . . . | 59        |
| 4.2.1     | Problem setup . . . . .                                    | 60        |
| 4.2.2     | Forward and backward summations . . . . .                  | 61        |
| 4.3       | The Viterbi Algorithm . . . . .                            | 65        |
| 4.4       | Extensions . . . . .                                       | 66        |
| 4.4.1     | Continuous observation . . . . .                           | 66        |
| 4.4.2     | Kalman filtering . . . . .                                 | 67        |
| 4.5       | Problems . . . . .                                         | 69        |
| <b>5.</b> | <b>Random Graphs for Modeling Network Data</b>             | <b>71</b> |
| 5.1       | Network Data . . . . .                                     | 71        |
| 5.2       | Latent Space Models . . . . .                              | 72        |
| 5.3       | Stochastic Block Models . . . . .                          | 73        |

|       |                                             |    |
|-------|---------------------------------------------|----|
| 5.3.1 | Model structure . . . . .                   | 73 |
| 5.3.2 | Variational EM algorithm . . . . .          | 75 |
| 5.3.3 | Variational inference for the SBM . . . . . | 77 |
| 5.3.4 | Community detection . . . . .               | 81 |
| 5.3.5 | Extensions and discussion . . . . .         | 83 |
| 5.4   | Graphon . . . . .                           | 84 |
| 5.5   | Problems . . . . .                          | 87 |

## Part 2 Causal Graphical Models 89

### 6. Undirected Graphical Models 91

|       |                                           |     |
|-------|-------------------------------------------|-----|
| 6.1   | Graphoid . . . . .                        | 91  |
| 6.2   | Conditional Independence Tests . . . . .  | 93  |
| 6.3   | Undirected Graphs . . . . .               | 95  |
| 6.4   | Markov Properties . . . . .               | 96  |
| 6.4.1 | Definitions . . . . .                     | 96  |
| 6.4.2 | Examples . . . . .                        | 97  |
| 6.4.3 | Equivalence . . . . .                     | 99  |
| 6.5   | Conditional Independence Graphs . . . . . | 99  |
| 6.5.1 | Gaussian graphical models . . . . .       | 100 |
| 6.5.2 | Discrete graphical models . . . . .       | 103 |
| 6.6   | Faithfulness . . . . .                    | 105 |
| 6.7   | Markov Blanket . . . . .                  | 106 |
| 6.8   | Neighborhood Lattice . . . . .            | 108 |
| 6.9   | Problems . . . . .                        | 111 |

### 7. Directed Acyclic Graphs 113

|       |                                      |     |
|-------|--------------------------------------|-----|
| 7.1   | Terminology for DAGs . . . . .       | 113 |
| 7.2   | $d$ -Separation . . . . .            | 114 |
| 7.3   | Markov Properties . . . . .          | 116 |
| 7.3.1 | Definitions and relations . . . . .  | 116 |
| 7.3.2 | Moral graphs . . . . .               | 118 |
| 7.3.3 | Markov equivalence . . . . .         | 120 |
| 7.4   | Parameterizations . . . . .          | 122 |
| 7.4.1 | Gaussian Bayesian networks . . . . . | 122 |
| 7.4.2 | Discrete Bayesian networks . . . . . | 124 |

|            |                                                               |            |
|------------|---------------------------------------------------------------|------------|
| 7.5        | Chain Graphs . . . . .                                        | 125        |
| 7.5.1      | Definition and characterization . . . . .                     | 125        |
| 7.5.2      | Markov properties . . . . .                                   | 127        |
| 7.6        | Faithfulness . . . . .                                        | 128        |
| 7.7        | Overview of Topics . . . . .                                  | 129        |
| 7.8        | Problems . . . . .                                            | 131        |
| <b>8.</b>  | <b>Causal Inference Based on Directed Acyclic Graphs</b>      | <b>133</b> |
| 8.1        | Causal DAGs and Intervention . . . . .                        | 133        |
| 8.2        | Truncated Factorization . . . . .                             | 137        |
| 8.3        | Linear Structural Equation Models . . . . .                   | 139        |
| 8.4        | Estimation of Causal Effect . . . . .                         | 142        |
| 8.4.1      | Back-door adjustment . . . . .                                | 142        |
| 8.4.2      | Front-door adjustment . . . . .                               | 145        |
| 8.4.3      | Instrumental variables . . . . .                              | 147        |
| 8.5        | Potential Outcome Approach . . . . .                          | 149        |
| 8.6        | Problems . . . . .                                            | 152        |
| <b>9.</b>  | <b>Structure Learning of Directed Acyclic Graphs</b>          | <b>153</b> |
| 9.1        | Overview and Assumptions . . . . .                            | 153        |
| 9.2        | Equivalence Class and Completed Partially DAGs . . . . .      | 155        |
| 9.3        | Constraint-Based Learning . . . . .                           | 158        |
| 9.3.1      | The PC algorithm . . . . .                                    | 159        |
| 9.3.2      | $p$ -value adjacency thresholding . . . . .                   | 162        |
| 9.4        | Score-Based Learning . . . . .                                | 166        |
| 9.4.1      | Scoring functions . . . . .                                   | 166        |
| 9.4.2      | Consistency . . . . .                                         | 169        |
| 9.4.3      | Search strategies . . . . .                                   | 170        |
| 9.5        | Experimental Data . . . . .                                   | 172        |
| 9.6        | Continuous Optimization . . . . .                             | 175        |
| 9.6.1      | Maximum penalized likelihood . . . . .                        | 176        |
| 9.6.2      | Differentiable acyclicity constraint . . . . .                | 180        |
| 9.7        | Problems . . . . .                                            | 182        |
| <b>10.</b> | <b>Learning Generalized Directed Acyclic Graphical Models</b> | <b>185</b> |
| 10.1       | DAG Identifiability . . . . .                                 | 185        |
| 10.1.1     | Equal error variance . . . . .                                | 186        |
| 10.1.2     | Generalized linear DAGs . . . . .                             | 187        |

|            |                                                    |            |
|------------|----------------------------------------------------|------------|
| 10.2       | Linear Non-Gaussian DAGs . . . . .                 | 189        |
| 10.2.1     | LiNGAM discovery algorithm . . . . .               | 189        |
| 10.2.2     | Sequential order learning . . . . .                | 190        |
| 10.3       | Nonlinear DAG Models . . . . .                     | 194        |
| 10.3.1     | Additive noise models . . . . .                    | 194        |
| 10.3.2     | Learning with neural networks . . . . .            | 196        |
| 10.4       | Network Data . . . . .                             | 198        |
| 10.4.1     | Gaussian DAG for dependent data . . . . .          | 198        |
| 10.4.2     | Parameter estimation . . . . .                     | 200        |
| 10.4.3     | Structure learning via de-correlation . . . . .    | 202        |
| <b>11.</b> | <b>Directed Mixed Graphs for Latent Variables</b>  | <b>203</b> |
| 11.1       | Semi-Markov Causal Models . . . . .                | 203        |
| 11.2       | Acyclic Directed Mixed Graphs . . . . .            | 206        |
| 11.3       | Factorizations on ADMGs . . . . .                  | 207        |
| 11.3.1     | District factorization . . . . .                   | 207        |
| 11.3.2     | Tian factorization . . . . .                       | 209        |
| 11.3.3     | Nested factorization . . . . .                     | 211        |
| 11.3.4     | Finding GCI constraints . . . . .                  | 213        |
| 11.4       | Identification of Causal Effects . . . . .         | 214        |
| 11.5       | Linear SEM Associated with ADMG . . . . .          | 217        |
| 11.5.1     | Parameterization and identifiability . . . . .     | 217        |
| 11.5.2     | Structure learning . . . . .                       | 219        |
| 11.6       | Ancestral Graphs . . . . .                         | 220        |
| 11.6.1     | Maximal ancestral graphs . . . . .                 | 221        |
| 11.6.2     | Partial ancestral graphs . . . . .                 | 222        |
| 11.6.3     | The FCI algorithm . . . . .                        | 225        |
| 11.7       | Latent Factor Analysis . . . . .                   | 226        |
| 11.7.1     | Graphical representation . . . . .                 | 227        |
| 11.7.2     | Structure recovery by clique search . . . . .      | 228        |
| <b>12.</b> | <b>Partitioned, Federated, and Active Learning</b> | <b>231</b> |
| 12.1       | Divide and Conquer . . . . .                       | 231        |
| 12.1.1     | Partition and estimation . . . . .                 | 232        |
| 12.1.2     | Fusion . . . . .                                   | 234        |
| 12.2       | Federated Learning . . . . .                       | 235        |
| 12.2.1     | Global objective function . . . . .                | 236        |
| 12.2.2     | Distributed optimization . . . . .                 | 237        |
| 12.2.3     | Theoretical guarantees . . . . .                   | 239        |

|        |                                                      |     |
|--------|------------------------------------------------------|-----|
| 12.3   | Causal Bandit via Sequential Interventions . . . . . | 240 |
| 12.3.1 | Multi-armed bandit . . . . .                         | 240 |
| 12.3.2 | Causal bandit . . . . .                              | 242 |
| 12.3.3 | Bayesian backdoor bandit . . . . .                   | 244 |
|        | <i>Bibliography</i>                                  | 249 |
|        | <i>Subject Index</i>                                 | 261 |
|        | <i>Author Index</i>                                  | 267 |