
Contents

Part I Building Blocks for Bayesian Networks

| | | |
|----------|--|----|
| 1 | Introduction | 3 |
| 1.1 | An Example Bayes Network | 4 |
| 1.2 | Cognitively Diagnostic Assessment | 7 |
| 1.3 | Cognitive and Psychometric Science | 11 |
| 1.4 | Ten Reasons for Considering Bayesian Networks | 14 |
| 1.5 | What Is in This Book | 16 |
| 2 | An Introduction to Evidence-Centered Design | 19 |
| 2.1 | Overview | 20 |
| 2.2 | Assessment as Evidentiary Argument | 21 |
| 2.3 | The Process of Design | 23 |
| 2.4 | Basic ECD Structures | 26 |
| 2.4.1 | The Conceptual Assessment Framework | 27 |
| 2.4.2 | Four-Process Architecture for Assessment Delivery | 34 |
| 2.4.3 | Pretesting and Calibration | 38 |
| 2.5 | Conclusion | 39 |
| 3 | Bayesian Probability and Statistics: a Review | 41 |
| 3.1 | Probability: Objective and Subjective | 41 |
| 3.1.1 | Objective Notions of Probability | 42 |
| 3.1.2 | Subjective Notions of Probability | 43 |
| 3.1.3 | Subjective–Objective Probability | 45 |
| 3.2 | Conditional Probability | 46 |
| 3.3 | Independence and Conditional Independence | 51 |
| 3.3.1 | Conditional Independence | 53 |
| 3.3.2 | Common Variable Dependence | 54 |
| 3.3.3 | Competing Explanations | 55 |
| 3.4 | Random Variables | 57 |
| 3.4.1 | The Probability Mass and Density Functions | 57 |
| 3.4.2 | Expectation and Variance | 60 |

| | | |
|----------|--|------------|
| 3.5 | Bayesian Inference | 62 |
| 3.5.1 | Re-expressing Bayes Theorem | 63 |
| 3.5.2 | Bayesian Paradigm | 63 |
| 3.5.3 | Conjugacy | 67 |
| 3.5.4 | Sources for Priors | 72 |
| 3.5.5 | Noninformative Priors | 74 |
| 3.5.6 | Evidence-Centered Design and the Bayesian Paradigm | 76 |
| 4 | Basic Graph Theory and Graphical Models | 81 |
| 4.1 | Basic Graph Theory | 82 |
| 4.1.1 | Simple Undirected Graphs | 83 |
| 4.1.2 | Directed Graphs | 83 |
| 4.1.3 | Paths and Cycles | 84 |
| 4.2 | Factorization of the Joint Distribution | 86 |
| 4.2.1 | Directed Graph Representation | 86 |
| 4.2.2 | Factorization Hypergraphs | 88 |
| 4.2.3 | Undirected Graphical Representation | 90 |
| 4.3 | Separation and Conditional Independence | 91 |
| 4.3.1 | Separation and D-Separation | 91 |
| 4.3.2 | Reading Dependence and Independence from Graphs | 93 |
| 4.3.3 | Gibbs–Markov Equivalence Theorem | 94 |
| 4.4 | Edge Directions and Causality | 95 |
| 4.5 | Other Representations | 97 |
| 4.5.1 | Influence Diagrams | 97 |
| 4.5.2 | Structural Equation Models | 99 |
| 4.5.3 | Other Graphical Models | 100 |
| 5 | Efficient Calculations | 105 |
| 5.1 | Belief Updating with Two Variables | 106 |
| 5.2 | More Efficient Procedures for Chains and Trees | 111 |
| 5.2.1 | Propagation in Chains | 112 |
| 5.2.2 | Propagation in Trees | 116 |
| 5.2.3 | Virtual Evidence | 119 |
| 5.3 | Belief Updating in Multiply Connected Graphs | 122 |
| 5.3.1 | Updating in the Presence of Loops | 122 |
| 5.3.2 | Constructing a Junction Tree | 123 |
| 5.3.3 | Propagating Evidence Through a Junction Tree | 134 |
| 5.4 | Application to Assessment | 135 |
| 5.4.1 | Proficiency and Evidence Model Bayes Net Fragments | 137 |
| 5.4.2 | Junction Trees for Fragments | 139 |
| 5.4.3 | Calculation with Fragments | 143 |
| 5.5 | The Structure of a Test | 145 |
| 5.5.1 | The Q -Matrix for Assessments Using Only Discrete Items | 146 |
| 5.5.2 | The Q -Matrix for a Test Using Multi-observable Tasks | 147 |
| 5.6 | Alternative Computing Algorithms | 149 |

| | | |
|----------|---|------------|
| 5.6.1 | Variants of the Propagation Algorithm | 150 |
| 5.6.2 | Dealing with Unfavorable Topologies | 150 |
| 6 | Some Example Networks | 157 |
| 6.1 | A Discrete IRT Model | 158 |
| 6.1.1 | General Features of the IRT Bayes Net | 161 |
| 6.1.2 | Inferences in the IRT Bayes Net | 162 |
| 6.2 | The “Context” Effect | 166 |
| 6.3 | Compensatory, Conjunctive, and Disjunctive Models | 172 |
| 6.4 | A Binary-Skills Measurement Model | 178 |
| 6.4.1 | The Domain of Mixed Number Subtraction | 178 |
| 6.4.2 | A Bayes Net Model for Mixed-Number Subtraction . . . | 180 |
| 6.4.3 | Inferences from the Mixed-Number Subtraction Bayes Net | 184 |
| 6.5 | Discussion | 190 |
| 7 | Explanation and Test Construction | 197 |
| 7.1 | Simple Explanation Techniques | 198 |
| 7.1.1 | Node Coloring | 198 |
| 7.1.2 | Most Likely Scenario | 200 |
| 7.2 | Weight of Evidence | 201 |
| 7.2.1 | Evidence Balance Sheet | 202 |
| 7.2.2 | Evidence Flow Through the Graph | 205 |
| 7.3 | Activity Selection | 209 |
| 7.3.1 | Value of Information | 209 |
| 7.3.2 | Expected Weight of Evidence | 213 |
| 7.3.3 | Mutual Information | 215 |
| 7.4 | Test Construction | 215 |
| 7.4.1 | Computer Adaptive Testing | 216 |
| 7.4.2 | Critiquing | 217 |
| 7.4.3 | Fixed-Form Tests | 220 |
| 7.5 | Reliability and Assessment Information | 224 |
| 7.5.1 | Accuracy Matrix | 225 |
| 7.5.2 | Consistency Matrix | 230 |
| 7.5.3 | Expected Value Matrix | 230 |
| 7.5.4 | Weight of Evidence as Information | 232 |

Part II Learning and Revising Models from Data

| | | |
|----------|--|------------|
| 8 | Parameters for Bayesian Network Models | 241 |
| 8.1 | Parameterizing a Graphical Model | 241 |
| 8.2 | Hyper-Markov Laws | 244 |
| 8.3 | The Conditional Multinomial—Hyper-Dirichlet Family | 246 |
| 8.3.1 | Beta-Binomial Family | 247 |

| | | |
|-----------|--|------------|
| 8.3.2 | Dirichlet-Multinomial Family | 248 |
| 8.3.3 | The Hyper-Dirichlet Law | 248 |
| 8.4 | Noisy-OR and Noisy-AND Models | 250 |
| 8.4.1 | Separable Influence | 254 |
| 8.5 | DiBello's Effective Theta Distributions | 254 |
| 8.5.1 | Mapping Parent Skills to θ Space | 256 |
| 8.5.2 | Combining Input Skills | 257 |
| 8.5.3 | Samejima's Graded Response Model | 260 |
| 8.5.4 | Normal Link Function | 263 |
| 8.6 | Eliciting Parameters and Laws | 267 |
| 8.6.1 | Eliciting Conditional Multinomial and Noisy-AND | 269 |
| 8.6.2 | Priors for DiBello's Effective Theta Distributions | 272 |
| 8.6.3 | Linguistic Priors | 273 |
| 9 | Learning in Models with Fixed Structure | 279 |
| 9.1 | Data, Models, and Plate Notation | 279 |
| 9.1.1 | Plate Notation | 280 |
| 9.1.2 | A Bayesian Framework for a Generic Measurement Model | 282 |
| 9.1.3 | Extension to Covariates | 284 |
| 9.2 | Techniques for Learning with Fixed Structure | 287 |
| 9.2.1 | Bayesian Inference for the General Measurement Model | 288 |
| 9.2.2 | Complete Data Tables | 289 |
| 9.3 | Latent Variables as Missing Data | 297 |
| 9.4 | The EM Algorithm | 298 |
| 9.5 | Markov Chain Monte Carlo Estimation | 305 |
| 9.5.1 | Gibbs Sampling | 308 |
| 9.5.2 | Properties of MCMC Estimation | 309 |
| 9.5.3 | The Metropolis-Hastings Algorithm | 312 |
| 9.6 | MCMC Estimation in Bayes Nets in Assessment | 315 |
| 9.6.1 | Initial Calibration | 316 |
| 9.6.2 | Online Calibration | 321 |
| 9.7 | Caution: MCMC and EM are Dangerous! | 324 |
| 10 | Critiquing and Learning Model Structure | 331 |
| 10.1 | Fit Indices Based on Prediction Accuracy | 332 |
| 10.2 | Posterior Predictive Checks | 335 |
| 10.3 | Graphical Methods | 342 |
| 10.4 | Differential Task Functioning | 347 |
| 10.5 | Model Comparison | 350 |
| 10.5.1 | The DIC Criterion | 350 |
| 10.5.2 | Prediction Criteria | 353 |
| 10.6 | Model Selection | 354 |
| 10.6.1 | Simple Search Strategies | 355 |
| 10.6.2 | Stochastic Search | 356 |

| | |
|---|------------|
| 10.6.3 Multiple Models | 357 |
| 10.6.4 Priors Over Models | 357 |
| 10.7 Equivalent Models and Causality | 358 |
| 10.7.1 Edge Orientation | 358 |
| 10.7.2 Unobserved Variables | 358 |
| 10.7.3 Why Unsupervised Learning cannot Prove Causality | 360 |
| 10.8 The “True” Model | 362 |
| 11 An Illustrative Example | 371 |
| 11.1 Representing the Cognitive Model | 372 |
| 11.1.1 Representing the Cognitive Model as a Bayesian Network | 372 |
| 11.1.2 Representing the Cognitive Model as a Bayesian Network | 377 |
| 11.1.3 Higher-Level Structure of the Proficiency Model; i.e., $p(\theta \lambda)$ and $p(\lambda)$ | 379 |
| 11.1.4 High Level Structure of the Evidence Models; i.e., $p(\pi)$ | 381 |
| 11.1.5 Putting the Pieces Together | 382 |
| 11.2 Calibrating the Model with Field Data | 382 |
| 11.2.1 MCMC Estimation | 383 |
| 11.2.2 Scoring | 389 |
| 11.2.3 Online Calibration | 392 |
| 11.3 Model Checking | 397 |
| 11.3.1 Observable Characteristic Plots | 398 |
| 11.3.2 Posterior Predictive Checks | 401 |
| 11.4 Closing Comments | 405 |

Part III Evidence-Centered Assessment Design

| | |
|--|------------|
| 12 The Conceptual Assessment Framework | 411 |
| 12.1 Phases of the Design Process and Evidentiary Arguments | 414 |
| 12.1.1 Domain Analysis and Domain Modeling | 414 |
| 12.1.2 Arguments and Claims | 418 |
| 12.2 The Student Proficiency Model | 424 |
| 12.2.1 Proficiency Variables | 424 |
| 12.2.2 Relationships Among Proficiency Variables | 428 |
| 12.2.3 Reporting Rules | 433 |
| 12.3 Task Models | 438 |
| 12.4 Evidence Models | 443 |
| 12.4.1 Rules of Evidence (for Evidence Identification) | 444 |
| 12.4.2 Statistical Models of Evidence (for Evidence Accumulation) | 448 |
| 12.5 The Assembly Model | 453 |
| 12.6 The Presentation Model | 458 |

| | |
|---|------------|
| 12.7 The Delivery Model | 460 |
| 12.8 Putting It All Together | 461 |
| 13 The Evidence Accumulation Process | 467 |
| 13.1 The Four-Process Architecture | 468 |
| 13.1.1 A Simple Example of the Four-Process Framework | 471 |
| 13.2 Producing an Assessment | 474 |
| 13.2.1 Tasks and Task Model Variables | 474 |
| 13.2.2 Evidence Rules | 478 |
| 13.2.3 Evidence Models, Links, and Calibration | 486 |
| 13.3 Scoring | 488 |
| 13.3.1 Basic Scoring Protocols | 489 |
| 13.3.2 Adaptive Testing | 493 |
| 13.3.3 Technical Considerations | 497 |
| 13.3.4 Score Reports | 500 |
| 14 Biomass: An Assessment of Science Standards | 507 |
| 14.1 Design Goals | 507 |
| 14.2 Designing Biomass | 510 |
| 14.2.1 Reconceiving Standards | 510 |
| 14.2.2 Defining Claims | 513 |
| 14.2.3 Defining Evidence | 514 |
| 14.3 The Biomass Conceptual Assessment Framework | 515 |
| 14.3.1 The Proficiency Model | 515 |
| 14.3.2 The Assembly Model | 519 |
| 14.3.3 Task Models | 523 |
| 14.3.4 Evidence Models | 529 |
| 14.4 The Assessment Delivery Processes | 535 |
| 14.4.1 Biomass Architecture | 536 |
| 14.4.2 The Presentation Process | 538 |
| 14.4.3 Evidence Identification | 540 |
| 14.4.4 Evidence Accumulation | 541 |
| 14.4.5 Activity Selection | 543 |
| 14.4.6 The Task/Evidence Composite Library | 543 |
| 14.4.7 Controlling the Flow of Information Among the Processes | 544 |
| 14.5 Conclusion | 545 |
| 15 The Biomass Measurement Model | 549 |
| 15.1 Specifying Prior Distributions | 550 |
| 15.1.1 Specification of Proficiency Variable Priors | 552 |
| 15.1.2 Specification of Evidence Model Priors | 554 |
| 15.1.3 Summary Statistics | 560 |
| 15.2 Pilot Testing | 561 |
| 15.2.1 A Convenience Sample | 561 |

| | | |
|-----------|--|-----|
| 15.2.2 | Item and other Exploratory Analyses | 564 |
| 15.3 | Updating Based on Pilot Test Data | 566 |
| 15.3.1 | Posterior Distributions | 566 |
| 15.3.2 | Some Observations on Model Fit | 575 |
| 15.3.3 | A Quick Validity Check | 577 |
| 15.4 | Conclusion | 579 |
| 16 | The Future of Bayesian Networks in Educational | |
| | Assessment | 583 |
| 16.1 | Applications of Bayesian Networks | 583 |
| 16.2 | Extensions to the Basic Bayesian Network Model | 586 |
| 16.2.1 | Object-Oriented Bayes Nets | 586 |
| 16.2.2 | Dynamic Bayesian Networks | 588 |
| 16.2.3 | Assessment-Design Support | 592 |
| 16.3 | Connections with Instruction | 593 |
| 16.3.1 | Ubiquitous Assessment | 594 |
| 16.4 | Evidence-Centered Assessment Design and Validity | 596 |
| 16.5 | What We Still Do Not Know | 597 |
| A | Bayesian Network Resources | 601 |
| A.1 | Software | 601 |
| A.1.1 | Bayesian Network Manipulation | 602 |
| A.1.2 | Manual Construction of Bayesian Networks | 603 |
| A.1.3 | Markov Chain Monte Carlo | 603 |
| A.2 | Sample Bayesian Networks | 604 |
| | References | 607 |
| | Author Index | 639 |

<http://www.springer.com/978-1-4939-2124-9>

Bayesian Networks in Educational Assessment

Almond, R.G.; Mislevy, R.J.; Steinberg, L.; Yan, D.;

Williamson, D.

2015, XXXIII, 662 p. 155 illus., 87 illus. in color.,

Hardcover

ISBN: 978-1-4939-2124-9