dandelon.com
© 2008 AGI-Information Management Consultants
May be used for personal purposes only or by
libraries associated to dandelon.com network.

Learning Bayesian Networks

Richard E. Neapolitan

Northeastern Illinois University
Chicago, Illinois

PEARSON
Prentice Hall

Upper Saddle River, NJ 07458

Contents

Preface **vn**

I Basics

1 Introduction to Bayesian Networks **1**
1.1 Basics of Probability Theory . . . 3
1.1.1 Probability Functions and Spaces . . . 3
1.1.2 Conditional Probability and Independence . . . 6
1.1.3 Bayes' Theorem . . . 8
1.1.4 Random Variables and Joint Probability Distributions . . . 10
1.2 Philosophical Foundations of Probability . . . 17
1.2.1 The Relative Frequency Approach to Probability . . . 17
1.2.2 The Subjective/Bayesian Approach to Probability . . . 21
1.3 Random Variables in Statistics . . . 27
1.3.1 Meaning of Random Variables . . . 28
1.3.2 A Definition of Random Variables for Statistics . . . 31
1.4 Large Instances / Bayesian Networks . . . 35
1.4.1 The Difficulties Inherent in Large Instances . . . 35
1.4.2 The Markov Condition . . . 36
1.4.3 Bayesian Networks . . . 45
1.4.4 A Large Bayesian Network . . . 47
1.5 Creating Bayesian Networks Using Causal Edges . . . 48
1.5.1 Ascertaining Causal Influences Using Manipulation . . . 49
1.5.2 Causation and the Markov Condition . . . 56

2 More DAG/Probability Relationships **69**
2.1 Entailed Conditional Independencies . . . 70
2.1.1 Examples of Entailed Conditional Independencies . . . 70
2.1.2 d-Separation . . . 74
2.1.3 Finding d-Separations . . . 80
2.2 Markov Equivalence . . . 88
2.3 Entailing Dependencies with a DAG . . . 96
2.3.1 Faithfulness . . . 97

2.3.2 Embedded Faithfulness . . . 102
2.4 Minimality . . . 106
2.5 Markov Blankets and Boundaries . . . 109
2.6 More on Causal DAGs . . . 111
2.6.1 The Causal Minimality Assumption . . . 112
2.6.2 The Causal Faithfulness Assumption . . . 112
2.6.3 The Causal Embedded Faithfulness Assumption . . . 113

II Inference

3 Inference: Discrete Variables 125
3.1 Examples of Inference . . . 126
3.2 Pearl's Message-Passing Algorithm . . . 128
3.2.1 Inference in Trees . . . 129
3.2.2 Inference in Singly Connected Networks . . . 144
3.2.3 Inference in Multiply Connected Networks . . . 155
3.2.4 Complexity of the Algorithm . . . 157
3.3 The Noisy OR-Gate Model . . . 158
3.3.1 The Model . . . 158
3.3.2 Doing Inference with the Model . . . 162
3.3.3 Further Models . . . 163
3.4 Other Algorithms that Employ the DAG . . . 163
3.5 The SPI Algorithm . . . 163
3.5.1 The Optimal Factoring Problem . . . 165
3.5.2 Application to Probabilistic Inference . . . 170
3.6 Complexity of Inference . . . 172
3.7 Relationship to Human Reasoning . . . 173
3.7.1 The Causal Network Model . . . 173
3.7.2 Studies Testing the Causal Network Model . . . 174

4 More Inference Algorithms 183
4.1 Continuous Variable Inference . . . 183
4.1.1 The Normal Distribution . . . 184
4.1.2 An Example Concerning Continuous Variables . . . 185
4.1.3 An Algorithm for Continuous Variables . . . 187
4.2 Approximate Inference . . . 206
4.2.1 Logic Sampling . . . 206
4.2.2 Likelihood Weighting . . . 212
4.3 Abductive Inference . . . 216
4.3.1 Abductive Inference in Bayesian Networks . . . 216
4.3.2 A Best-First Search Algorithm for Abductive Inference . . . 219

5 Influence Diagrams **233**
5.1 Decision Trees . . . 233
5.1.1 Simple Examples . . . 233
5.1.2 Probabilities, Time, and Risk Attitudes . . . 237
5.1.3 Solving Decision Trees . . . 239
5.1.4 More Examples . . . 239
5.2 Influence Diagrams . . . 252
5.2.1 Representing with Influence Diagrams . . . 252
5.2.2 Solving Influence Diagrams . . . 260
5.3 Dynamic Networks . . . 265
5.3.1 Dynamic Bayesian Networks . . . 265
5.3.2 Dynamic Influence Diagrams . . . 272

III Learning

6 Parameter Learning: Binary Variables **285**
6.1 Learning a Single Parameter . . . 286
6.1.1 Probability Distributions of Relative Frequencies . . . 286
6.1.2 Learning a Relative Frequency . . . 295
6.2 More on the Beta Density Function . . . 302
6.2.1 Nonintegral Values of *a* and *b* . . . 302
6.2.2 Assessing the Values of *a* and *b* . . . 305
6.2.3 Why the Beta Density Function? . . . 306
6.3 Computing a Probability Interval . . . 310
6.4 Learning Parameters in a Bayesian Network . . . 315
6.4.1 Urn Examples . . . 315
6.4.2 Augmented Bayesian Networks . . . 323
6.4.3 Learning Using an Augmented Bayesian Network . . . 327
6.4.4 A Problem with Updating; Using an Equivalent Sample Size . . . 339
6.5 Learning with Missing Data Items . . . 348
6.5.1 Data Items Missing at Random . . . 349
6.5.2 Data Items Missing Not at Random . . . 354
6.6 Variances in Computed Relative Frequencies . . . 355
6.6.1 A Simple Variance Determination . . . 355
6.6.2 The Variance and Equivalent Sample Size . . . 357
6.6.3 Computing Variances in Larger Networks . . . 363
6.6.4 When Do Variances Become Large? . . . 364

7 More Parameter Learning **373**
7.1 Multinomial Variables . . . 373
7.1.1 Learning a Single Parameter . . . 373
7.1.2 More on the Dirichlet Density Function . . . 380
7.1.3 Computing Probability Intervals and Regions . . . 381

7.1.4 Learning Parameters in a Bayesian Network 384
7.1.5 Learning with Missing Data Items 390
7.1.6 Variances in Computed Relative Frequencies 390
7.2 Continuous Variables 391
7.2.1 Normally Distributed Variable 391
7.2.2 Multivariate Normally Distributed Variables 406
7.2.3 Gaussian Bayesian Networks 418

8 Bayesian Structure Learning 433
8.1 Learning Structure: Discrete Variables 433
8.1.1 Schema for Learning Structure 434
8.1.2 Procedure for Learning Structure 437
8.1.3 Learning From a Mixture of Observational and Experimental Data 441
8.1.4 Complexity of Structure Learning 443
8.2 Model Averaging 443
8.3 Learning Structure with Missing Data 444
8.3.1 Monte Carlo Methods 445
8.3.2 Large-Sample Approximations 454
8.4 Probabilistic Model Selection 460
8.4.1 Probabilistic Models 460
8.4.2 The Model Selection Problem 464
8.4.3 Using the Bayesian Scoring Criterion for Model Selection 465
8.5 Hidden Variable DAG Models 469
8.5.1 Models Containing More Conditional Independencies than DAG Models 469
8.5.2 Models Containing the Same Conditional Independencies as DAG Models 471
8.5.3 Dimension of Hidden Variable DAG Models 476
8.5.4 Number of Models and Hidden Variables 478
8.5.5 Efficient Model Scoring 479
8.6 Learning Structure: Continuous Variables 483
8.6.1 The Density Function of D 483
8.6.2 The Density function of D Given a DAG pattern 487
8.7 Learning Dynamic Bayesian Networks 498

9 Approximate Bayesian Structure Learning 503
9.1 Approximate Model Selection 503
9.1.1 Algorithms That Search over DAGs 505
9.1.2 Algorithms That Search over DAG Patterns 510
9.1.3 An Algorithm Assuming Missing Data or Hidden Variables 521
9.2 Approximate Model Averaging 523
9.2.1 A Model Averaging Example 524
9.2.2 Approximate Model Averaging Using MCMC 525

10 Constraint-Based Learning **533**
10.1 Algorithms Assuming Faithfulness . . . 534
10.1.1 Simple Examples . . . 534
10.1.2 Algorithms for Finding Faithful DAG Patterns . . . 537
10.1.3 Algorithm for Determining Admission of Faithfulness . . . 544
10.1.4 Application to Probability . . . 552
10.2 Assuming Only Embedded Faithfulness . . . 553
10.2.1 Inducing Chains . . . 554
10.2.2 A Basic Algorithm . . . 560
10.2.3 An Improved Algorithm . . . 570
10.2.4 Application to Probability . . . 581
10.2.5 Application to Learning Causal Influences . . . 582
10.3 Obtaining the d-separations . . . 590
10.3.1 Discrete Bayesian Networks . . . 590
10.3.2 Gaussian Bayesian Networks . . . 594
10.4 Relationship to Human Reasoning . . . 595
10.4.1 Background Theory . . . 595
10.4.2 A Statistical Notion of Causality . . . 597

11 More Structure Learning **609**
11.1 Comparing the Methods . . . 609
11.1.1 A Simple Example . . . 610
11.1.2 Learning College Attendance Influences . . . 612
11.1.3 Conclusions . . . 615
11.2 Data Compression Scoring Criteria . . . 616
11.3 Parallel Learning of Bayesian Networks . . . 616
11.4 Examples . . . 616
11.4.1 Structure Learning . . . 616
11.4.2 Inferring Causal Relationships . . . 626

IV Applications

12 Applications **639**
12.1 Applications Based on Bayesian Networks . . . 639
12.2 Beyond Bayesian Networks . . . 645

Bibliography **647**

Index **667**