

Contents

Preface	ix
I Basics	1
1 Introduction to Bayesian Networks	3
1.1 Basics of Probability Theory	5
1.1.1 Probability Functions and Spaces	6
1.1.2 Conditional Probability and Independence	9
1.1.3 Bayes' Theorem	12
1.1.4 Random Variables and Joint Probability Distributions . .	13
1.2 Bayesian Inference	20
1.2.1 Random Variables and Probabilities in Bayesian Applica- tions	20
1.2.2 A Definition of Random Variables and Joint Probability Distributions for Bayesian Inference	24
1.2.3 A Classical Example of Bayesian Inference	27
1.3 Large Instances / Bayesian Networks	29
1.3.1 The Difficulties Inherent in Large Instances	29
1.3.2 The Markov Condition	31
1.3.3 Bayesian Networks	40
1.3.4 A Large Bayesian Network	43
1.4 Creating Bayesian Networks Using Causal Edges	43
1.4.1 Ascertaining Causal Influences Using Manipulation	44
1.4.2 Causation and the Markov Condition	51
2 More DAG/Probability Relationships	65
2.1 Entailed Conditional Independencies	66
2.1.1 Examples of Entailed Conditional Independencies	66
2.1.2 d-Separation	70
2.1.3 Finding d-Separations	76
2.2 Markov Equivalence	84
2.3 Entailing Dependencies with a DAG	92
2.3.1 Faithfulness	95

2.3.2	Embedded Faithfulness	99
2.4	Minimality	104
2.5	Markov Blankets and Boundaries	108
2.6	More on Causal DAGs	110
2.6.1	The Causal Minimality Assumption	110
2.6.2	The Causal Faithfulness Assumption	111
2.6.3	The Causal Embedded Faithfulness Assumption	112
II	Inference	121
3	Inference: Discrete Variables	123
3.1	Examples of Inference	124
3.2	Pearl's Message-Passing Algorithm	126
3.2.1	Inference in Trees	127
3.2.2	Inference in Singly-Connected Networks	142
3.2.3	Inference in Multiply-Connected Networks	153
3.2.4	Complexity of the Algorithm	155
3.3	The Noisy OR-Gate Model	156
3.3.1	The Model	156
3.3.2	Doing Inference With the Model	160
3.3.3	Further Models	161
3.4	Other Algorithms that Employ the DAG	161
3.5	The SPI Algorithm	162
3.5.1	The Optimal Factoring Problem	163
3.5.2	Application to Probabilistic Inference	168
3.6	Complexity of Inference	170
3.7	Relationship to Human Reasoning	171
3.7.1	The Causal Network Model	171
3.7.2	Studies Testing the Causal Network Model	173
4	More Inference Algorithms	181
4.1	Continuous Variable Inference	181
4.1.1	The Normal Distribution	182
4.1.2	An Example Concerning Continuous Variables	183
4.1.3	An Algorithm for Continuous Variables	185
4.2	Approximate Inference	205
4.2.1	A Brief Review of Sampling	205
4.2.2	Logic Sampling	211
4.2.3	Likelihood Weighting	217
4.3	Abductive Inference	221
4.3.1	Abductive Inference in Bayesian Networks	221
4.3.2	A Best-First Search Algorithm for Abductive Inference	224

5	Influence Diagrams	239
5.1	Decision Trees	239
5.1.1	Simple Examples	239
5.1.2	Probabilities, Time, and Risk Attitudes	242
5.1.3	Solving Decision Trees	245
5.1.4	More Examples	245
5.2	Influence Diagrams	259
5.2.1	Representing with Influence Diagrams	259
5.2.2	Solving Influence Diagrams	266
5.3	Dynamic Networks	272
5.3.1	Dynamic Bayesian Networks	272
5.3.2	Dynamic Influence Diagrams	279
III	Learning	291
6	Parameter Learning: Binary Variables	293
6.1	Learning a Single Parameter	294
6.1.1	Probability Distributions of Relative Frequencies	294
6.1.2	Learning a Relative Frequency	303
6.2	More on the Beta Density Function	310
6.2.1	Non-integral Values of a and b	311
6.2.2	Assessing the Values of a and b	313
6.2.3	Why the Beta Density Function?	315
6.3	Computing a Probability Interval	319
6.4	Learning Parameters in a Bayesian Network	323
6.4.1	Urn Examples	323
6.4.2	Augmented Bayesian Networks	331
6.4.3	Learning Using an Augmented Bayesian Network	336
6.4.4	A Problem with Updating; Using an Equivalent Sample Size	348
6.5	Learning with Missing Data Items	357
6.5.1	Data Items Missing at Random	358
6.5.2	Data Items Missing Not at Random	363
6.6	Variances in Computed Relative Frequencies	364
6.6.1	A Simple Variance Determination	364
6.6.2	The Variance and Equivalent Sample Size	366
6.6.3	Computing Variances in Larger Networks	372
6.6.4	When Do Variances Become Large?	373
7	More Parameter Learning	381
7.1	Multinomial Variables	381
7.1.1	Learning a Single Parameter	381
7.1.2	More on the Dirichlet Density Function	388
7.1.3	Computing Probability Intervals and Regions	389
7.1.4	Learning Parameters in a Bayesian Network	392

7.1.5	Learning with Missing Data Items	398
7.1.6	Variances in Computed Relative Frequencies	398
7.2	Continuous Variables	398
7.2.1	Normally Distributed Variable	399
7.2.2	Multivariate Normally Distributed Variables	413
7.2.3	Gaussian Bayesian Networks	425
8	Bayesian Structure Learning	441
8.1	Learning Structure: Discrete Variables	441
8.1.1	Schema for Learning Structure	442
8.1.2	Procedure for Learning Structure	445
8.1.3	Learning From a Mixture of Observational and Experi- mental Data.	449
8.1.4	Complexity of Structure Learning	450
8.2	Model Averaging	451
8.3	Learning Structure with Missing Data	452
8.3.1	Monte Carlo Methods	453
8.3.2	Large-Sample Approximations	462
8.4	Probabilistic Model Selection	468
8.4.1	Probabilistic Models	468
8.4.2	The Model Selection Problem	472
8.4.3	Using the Bayesian Scoring Criterion for Model Selection	473
8.5	Hidden Variable DAG Models	476
8.5.1	Models Containing More Conditional Independencies than DAG Models	477
8.5.2	Models Containing the Same Conditional Independencies as DAG Models	479
8.5.3	Dimension of Hidden Variable DAG Models	484
8.5.4	Number of Models and Hidden Variables	486
8.5.5	Efficient Model Scoring	487
8.6	Learning Structure: Continuous Variables	491
8.6.1	The Density Function of \mathbf{D}	491
8.6.2	The Density function of \mathbf{D} Given a DAG pattern	495
8.7	Learning Dynamic Bayesian Networks	505
9	Approximate Bayesian Structure Learning	511
9.1	Approximate Model Selection	511
9.1.1	Algorithms that Search over DAGs	513
9.1.2	Algorithms that Search over DAG Patterns	518
9.1.3	An Algorithm Assuming Missing Data or Hidden Variables	529
9.2	Approximate Model Averaging	531
9.2.1	A Model Averaging Example	532
9.2.2	Approximate Model Averaging Using MCMC	533

10 Constraint-Based Learning	541
10.1 Algorithms Assuming Faithfulness	542
10.1.1 Simple Examples	542
10.1.2 Algorithms for Determining DAG patterns	545
10.1.3 Determining if a Set Admits a Faithful DAG Representation	552
10.1.4 Application to Probability	560
10.2 Assuming Only Embedded Faithfulness	561
10.2.1 Inducing Chains	562
10.2.2 A Basic Algorithm	568
10.2.3 Application to Probability	590
10.2.4 Application to Learning Causal Influences ¹	591
10.3 Obtaining the d-separations	599
10.3.1 Discrete Bayesian Networks	600
10.3.2 Gaussian Bayesian Networks	603
10.4 Relationship to Human Reasoning	604
10.4.1 Background Theory	604
10.4.2 A Statistical Notion of Causality	606
11 More Structure Learning	617
11.1 Comparing the Methods	617
11.1.1 A Simple Example	618
11.1.2 Learning College Attendance Influences	620
11.1.3 Conclusions	623
11.2 Data Compression Scoring Criteria	624
11.3 Parallel Learning of Bayesian Networks	624
11.4 Examples	624
11.4.1 Structure Learning	625
11.4.2 Inferring Causal Relationships	633
IV Applications	647
12 Applications	649
12.1 Applications Based on Bayesian Networks	649
12.2 Beyond Bayesian networks	655
Bibliography	657
Index	686

¹The relationships in the examples in this section are largely fictitious.