

Contents

Table of Notation	xi
1 Bioinformatics and Gene Expression Experiments	1
1.1 Introduction	1
1.2 About This Book	3
2 Gene Expression Data: Basic Biology and Experiments	5
2.1 Background Biology	5
2.1.1 DNA Structures and Transcription	6
2.2 Gene Expression Microarray Experiments	9
2.2.1 Microarray Designs	11
2.2.2 Work Flow	11
2.2.3 Data Cleaning	15
3 Bayesian Linear Models for Gene Expression	21
3.1 Introduction	21
3.2 Bayesian Analysis of a Linear Model	22
3.2.1 Analysis via Conjugate Priors	23
3.2.2 Bayesian Variable Selection	25
3.2.3 Model Selection Priors	26
3.2.4 Priors on Regression Coefficients	27
3.2.5 Sparsity Priors	29
3.3 Bayesian Linear Models for Differential Expression	30
3.3.1 Relevant Work	32
3.4 Bayesian ANOVA for Gene Selection	34
3.4.1 The Basic Bayesian ANOVA Model	35
3.4.2 Differential Expression via Model Selection	36
3.5 Robust ANOVA model with Mixtures of Singular Distributions	38
3.6 Case Study	40
3.7 Accounting for Nuisance Effects	43
3.8 Summary and Further Reading	49
4 Bayesian Multiple Testing and False Discovery Rate Analysis	51
4.1 Introduction to Multiple Testing	51

4.2	False Discovery Rate Analysis	53
4.2.1	Theoretical Developments	53
4.2.2	FDR Analysis with Gene Expression Arrays	57
4.3	Bayesian False Discovery Rate Analysis	60
4.3.1	Theoretical Developments	60
4.4	Bayesian Estimation of FDR	61
4.5	FDR and Decision Theory	65
4.6	FDR and bFDR Summary	65
5	Bayesian Classification for Microarray Data	69
5.1	Introduction	69
5.2	Classification and Discriminant Rules	71
5.3	Bayesian Discriminant Analysis	72
5.4	Bayesian Regression Based Approaches to Classification	74
5.4.1	Bayesian Analysis of Generalized Linear Models	75
5.4.2	Link Functions	75
5.4.3	GLM using Latent Processes	76
5.4.4	Priors and Computation	76
5.4.5	Bayesian Probit Regression using Auxiliary Variables	77
5.5	Bayesian Nonlinear Classification	79
5.5.1	Classification using Interactions	79
5.5.2	Classification using Kernel Methods	82
5.6	Prediction and Model Choice	84
5.7	Examples	85
5.8	Discussion	87
6	Bayesian Hypothesis Inference for Gene Classes	89
6.1	Interpreting Microarray Results	89
6.2	Gene Classes	90
6.2.1	Enrichment Analysis	92
6.3	Bayesian Enrichment Analysis	94
6.4	Multivariate Gene Class Detection	95
6.4.1	Extending the Bayesian ANOVA Model	98
6.4.2	Bayesian Decomposition	105
6.5	Summary	107
7	Unsupervised Classification and Bayesian Clustering	109
7.1	Introduction to Bayesian Clustering for Gene Expression Data	109
7.2	Hierarchical Clustering	111
7.3	K-Means Clustering	112
7.4	Model-Based Clustering	114
7.5	Model-Based Agglomerative Hierarchical Clustering	115
7.6	Bayesian Clustering	116
7.7	Principal Components	117
7.8	Mixture Modeling	119

7.8.1	Label Switching	124
7.9	Clustering Using Dirichlet Process Prior	126
7.9.1	Infinite Mixture of Gaussian Distributions	131
8	Bayesian Graphical Models	137
8.1	Introduction	137
8.2	Probabilistic Graphical Models	138
8.3	Bayesian Networks	138
8.4	Inference for Network Models	141
8.4.1	Multinomial-Dirichlet Model	143
8.4.2	Gaussian Model	145
8.4.3	Model Search	146
8.4.4	Example	148
9	Advanced Topics	151
9.1	Introduction	151
9.2	Analysis of Time Course Gene Expression Data	151
9.2.1	Gene Selection	152
9.2.2	Functional Clustering	153
9.2.3	Dynamic Bayesian Networks	155
9.3	Survival Prediction Using Gene Expression Data	155
9.3.1	Gene Selection for Time-to-Event Outcomes	156
9.3.2	Weibull Regression Model	156
9.3.3	Proportional Hazards Model	157
9.3.4	Accelerated Failure Time Model	158
Appendix A:	Basics of Bayesian Modeling	159
A.1	Basics	159
A.1.1	The General Representation Theorem	160
A.1.2	Bayes' Theorem	161
A.1.3	Models Based on Partial Exchangeability	162
A.1.4	Modeling with Predictors	162
A.1.5	Prior Distributions	163
A.1.6	Decision Theory and Posterior and Predictive Inferences	165
A.1.7	Predictive Distributions	168
A.1.8	Examples	168
A.2	Bayesian Model Choice	172
A.3	Hierarchical Modeling	175
A.4	Bayesian Mixture Modeling	181
A.5	Bayesian Model Averaging	183
Appendix B:	Bayesian Computation Tools	185
B.1	Overview	185
B.2	Large-Sample Posterior Approximations	186
B.2.1	The Bayesian Central Limit Theorem	186

B.2.2	Laplace's Method	188
B.3	Monte Carlo Integration	191
B.4	Importance Sampling	192
B.5	Rejection Sampling	193
B.6	Gibbs Sampling	195
B.7	The Metropolis Algorithm and Metropolis-Hastings	198
B.8	Advanced Computational Methods	202
B.8.1	Block MCMC	203
B.8.2	Truncated Posterior Spaces	204
B.8.3	Latent Variables and the Auto-Probit Model	204
B.8.4	Bayesian Simultaneous Credible Envelopes	205
B.8.5	Proposal Updating	206
B.9	Posterior Convergence Diagnostics	207
B.10	MCMC Convergence and the Proposal	208
B.10.1	Graphical Checks for MCMC Methods	211
B.10.2	Convergence Statistics	212
B.10.3	MCMC in High-throughput Analysis	213
B.11	Summary	214
	References	217
	Index	237
	Appendix A: Basics of Bayesian Modeling	
A.1	Basics	
A.1.1	The General Representation Theorem	
A.1.2	Bayes' Theorem	
A.1.3	Models Based on Partial Independence	
A.1.4	Modeling with Plate Notation	
A.1.5	Prior Distributions	
A.1.6	Decision Theory and Posterior and Credible Intervals	
A.1.7	Factorial Distributions	
A.1.8	Examples	
A.2	Bayesian Model Comparison	
A.3	Statistical Modeling in Genetics	
A.4	Bayesian Mixture Modeling	
A.5	Bayesian Model Averaging	
	Appendix B: Bayesian Computation	
B.1	Overview	
B.2	Large-Sample Posterior Approximation	
B.3	The Bayesian Central Limit Theorem	