

Contents

List of Tables	xv
List of Figures	xvii
Preface	xxiii
1 Estimation and Testing in Time-Course Microarray Experiments	1
<i>C. Angelini, D. De Canditiis, and M. Pensky</i>	
1.1 Abstract	1
1.2 Introduction	2
1.3 Data Structure	3
1.3.1 Data Structure in the One-Sample Case	4
1.3.2 Data Structure in the Multi-Sample Case	5
1.4 Statistical Methods for the One-Sample Microarray Experiments	6
1.4.1 Multivariate Bayes Methodology	6
1.4.2 Functional Data Approach	7
1.4.3 Empirical Bayes Functional Data Approach	8
1.5 Statistical Methods for the Multi-Sample Microarray Experiments	11
1.5.1 Multivariate Bayes Methodology in the Two-Sample Case	11
1.5.2 Estimation and Testing in Empirical Bayes Functional Data Model	11
1.5.3 Bayesian Hierarchical Model with Subject-Specific Effect	15
1.5.4 Hidden Markov Models	16
1.6 Software	16
1.7 Comparisons between Techniques	19
1.8 Discussion	20
References	23
2 Classification for Differential Gene Expression Using Bayesian Hierarchical Models	27
<i>Natalia Bochkina and Alex Lewin</i>	
2.1 Introduction	27

2.2	Bayesian Hierarchical Model for Differential Expression . . .	30
2.2.1	Normalization	30
2.2.2	Shrinkage of Gene Variances	31
2.3	Differential Expression between Two Conditions	33
2.3.1	Classification Framework	33
2.3.2	Mixture Model	36
2.3.3	Unstructured Prior, Fixed Interval Null	40
2.3.4	Unstructured Prior, Adaptive Interval Null	42
2.3.5	Unstructured Prior, Bayesian Decision Rule, Point Null	42
2.3.6	Unstructured Prior, Point Null and Marginal P-Values	43
2.3.7	Discussion	44
2.4	Testing Hypotheses for Multiclass Data	47
2.5	Predictive Model Checking	50
2.5.1	Checking Gene Variances	51
2.5.2	Checking Mixture Prior	53
	References	59
3	Applications of MOSS for Discrete Multi-Way Data	63
	<i>Adrian Dobra, Laurent Briollais, Hamdi Jarjanazi, Hilmi Ozelik, and Hélène Massam</i>	
3.1	Introduction	64
3.2	MOSS	66
3.3	Conjugate Priors for Hierarchical Log-Linear Models	69
3.3.1	Model Parameterization	69
3.3.2	The Multinomial for Hierarchical Log-Linear Models	70
3.3.3	The Diaconis-Ylvisaker Conjugate Prior	71
3.3.4	Computing the Marginal Likelihood of a Regression	72
3.4	Discretizing Ordered Variables	73
3.5	Bayesian Model Averaging	74
3.6	Simulated Examples	76
3.6.1	First Simulated Example	76
3.6.2	Second Simulated Example	77
3.7	Real Examples: Gene Expression	78
3.7.1	Breast Cancer Data	78
3.7.2	Leukemia Data	79
3.7.3	Lymph Node Data	80
3.8	Real Examples: Genome-Wide Analysis of Estrogen Response with Dense SNP Array Data	82
3.9	Discussion	87
	References	89

4 Nonparametric Bayesian Bioinformatics	95
<i>David B. Dunson</i>	
4.1 Introduction	95
4.2 Dirichlet Process Mixture Models	96
4.3 Multiple Testing and High-Dimensional Regression	104
4.4 Clustering and Functional Data Analysis	108
4.5 Additional Topics	111
4.6 Discussion	114
References	117
5 Measurement Error and Survival Model for cDNA Microarrays	123
<i>Jonathan A. L. Gelfond and Joseph G. Ibrahim</i>	
5.1 Introduction	123
5.2 The Data Structure	126
5.3 The General Model	127
5.4 Priors	132
5.5 Model Fit	133
5.6 Case Study in Breast Cancer	134
5.6.1 Estimating the Measurement Error Parameters	135
5.6.2 Data Preprocessing	135
5.6.3 Results: Genes Identified by the Gene Only Model	137
5.7 Robustness Analysis and Operating Characteristics	138
5.7.1 Deviation from Normality in the Data	138
5.7.2 Simulations Demonstrating Robustness to Nonnormality	138
5.8 Discussion	142
References	143
Appendix	147
6 Bayesian Robust Inference for Differential Gene Expression	149
<i>Raphael Gottardo</i>	
6.1 Introduction	149
6.2 Model and Prior Distributions	150
6.3 Parameter Estimation	152
6.4 Application to Experimental Data	153
6.4.1 Data Description	153
6.4.2 Results	154
6.5 Discussion	158
References	161
7 Bayesian Hidden Markov Modeling of Array CGH Data	165
<i>Subharup Guha</i>	
7.1 Introduction	165
7.2 A Bayesian Model	169

7.3	Characterizing Array CGH Profiles	171
7.3.1	Classification Scheme	171
7.3.2	Posterior Inference	172
7.4	Illustrations	173
7.4.1	Pancreatic Adenocarcinoma Data	173
7.4.2	Comparisons with Some Existing Methods	176
7.5	Simulation Studies	178
7.5.1	Comparison with Non-Bayesian HMM and CBS Algorithms	178
7.5.2	Prior Sensitivity	181
7.6	Conclusion	183
7.7	Appendix: An MCMC Algorithm	184
	References	187
8	Bayesian Approaches to Phylogenetic Analysis	193
	<i>Mark T. Holder, Jeet Sukumaran, and Rafe M. Brown</i>	
8.1	Background	193
8.1.1	Calculating the Likelihood for a Phylogenetic Model	194
8.1.2	Maximum Likelihood Approaches	196
8.2	Bayesian Version of "Standard" Model-Based Phylogenetics	197
8.3	Current Directions in Model-Based Phylogenetics	199
8.3.1	Modeling Variability in the Rate of Sequence Substitution	200
8.3.2	Modeling Variability in Rates of Different Types of Substitution	201
8.3.3	Allowing Process of Molecular Evolution to Change over the Tree	202
8.3.4	Context-Dependent Models	204
8.4	Integration with Structured Coalescent	206
8.5	Divergence Time Estimation	208
8.6	Simultaneous Alignment and Inference	212
8.7	Example Analysis	214
8.8	Conclusions	218
	References	221
9	Gene Selection for the Identification of Biomarkers in High-Throughput Data	233
	<i>Jaesik Jeong, Marina Vannucci, Kim-Anh Do, Bradley Broom, Sinae Kim, Naijun Sha, Mahlet Tadesse, Kai Yan, and Lajos Pusztai</i>	
9.1	Introduction	233
9.2	Variable Selection in Linear Settings	235
9.2.1	Continuous Response	235
9.2.2	Categorical Response	236
9.2.3	Survival Time	237
9.2.4	Markov Chain Monte Carlo Algorithm	237

9.2.5	Posterior Inference	238
9.3	Model-Based Clustering	240
9.3.1	Finite Mixture Models	240
9.3.2	Dirichlet Process Mixture Models	241
9.3.3	Prior Setting	242
9.3.4	MCMC Implementation	242
9.3.5	Posterior Inference	243
9.4	Breast Cancer DNA Microarrays	244
9.4.1	Experimental Design	244
9.4.2	Pre-Processing	245
9.4.3	Classification via Bayesian Variable Selection	246
9.5	Conclusion	249
	References	251
10	Sparsity Priors for Protein-Protein Interaction Predictions	255
	<i>Inyoung Kim, Yin Liu, and Hongyu Zhao</i>	
10.1	Introduction	256
10.2	Model	258
10.2.1	Example	259
10.3	A Bayesian Approach	260
10.4	Real Data Analysis	261
10.5	Conclusion and Discussion	264
	References	267
11	Learning Bayesian Networks for Gene Expression Data	271
	<i>Faming Liang</i>	
11.1	Introduction	271
11.2	Bayesian Networks	274
11.3	Learning Bayesian Networks Using SAMC	275
11.3.1	A Review of the SAMC Algorithm	275
11.3.2	Learning Bayesian Networks Using SAMC	278
11.4	Numerical Examples	280
11.4.1	A Simulated Example	280
11.4.2	The Yeast Cell Cycle Data	283
11.5	Discussion	286
	References	287
12	In-Vitro to In-Vivo Factor Profiling in Expression Genomics	293
	<i>Joseph E. Lucas, Carlos M. Carvalho, Daniel Merl, and Mike West</i>	
12.1	Introduction	293
12.2	Modeling Gene Expression	295
12.2.1	Sparsity Priors	296
12.2.2	Signature Scores	297
12.2.3	Sparse, Semi-Parametric Latent Factor Models	298
12.3	Signature Factor Profiling Analysis	300

12.4 The E2F3 Signature in Ovarian, Lung, and Breast Cancer . . .	301
12.4.1 Indications of E2F3 Co-Regulation with NF-Y	302
12.4.2 Adenocarcinoma versus Squamous Cell Carcinoma . . .	304
12.4.3 Survival and the E2F3 Pathway	305
12.5 Closing Comments	309
References	313
13 Proportional Hazards Regression Using Bayesian Kernel Machines	317
<i>Arnab Maity and Bani K. Mallick</i>	
13.1 Introduction	317
13.2 Background and Framework	321
13.2.1 Reproducing Kernel Hilbert Space	321
13.2.2 Support Vector Machine	321
13.2.3 Relevance Vector Machine	323
13.3 Proportional Hazards Model Based on Kernel Regression . . .	323
13.3.1 Modeling of $p(t W)$ and the Baseline Hazard Function . . .	324
13.3.2 Modeling of $p(W f)$ and f Using RKHS	325
13.4 Relevance Vector Machine-Based PH Regression	326
13.4.1 Conditional Distribution and Posterior Sampling . . .	327
13.5 Support Vector Machine-Based PH Regression	328
13.6 Prediction	330
13.7 Application	331
13.7.1 Breast Carcinomas Data Set	331
13.7.2 Diffuse Large B-cell Lymphoma (DLBCL) Data Set . . .	334
13.7.3 Model Comparison	336
13.8 Discussion	336
References	339
14 A Bayesian Mixture Model for Protein Biomarker Discovery 343	
<i>Peter Müller, Keith Baggerly, Kim Anh Do, and Raj Bandyopadhyay</i>	
14.1 Introduction	344
14.1.1 Background	345
14.1.2 Statistical Methods for Classification of Mass Spectrometry Data	347
14.2 The Data	347
14.3 Likelihood Based Inference for Proteomic Spectra	348
14.4 A Hierarchical Beta Mixture Model	350
14.5 Posterior Inference	352
14.6 Results	354
14.7 Discussion	360
References	363

15 Bayesian Methods for Detecting Differentially Expressed Genes	365
<i>Fang Yu, Ming-Hui Chen, and Lynn Kuo</i>	
15.1 Introduction	365
15.2 Models for Microarray Gene Expression Data	367
15.2.1 Normal Model	368
15.2.2 Linear Model	368
15.2.3 Gamma Model	369
15.3 Prior Elicitation	369
15.3.1 Traditional Conjugate Prior	369
15.3.2 Power Prior	370
15.3.3 Mixture Non-Parametric Prior	372
15.4 Gene Selection Algorithms	373
15.4.1 Ordered Bayes Factor	373
15.4.2 Calibrated Bayes Factor	373
15.4.3 Two-Criterion	375
15.4.4 BH Method	377
15.4.5 Bayesian FDR	377
15.4.6 Posterior Odds	377
15.5 A Simulation Study	378
15.6 Real Data Example	381
15.7 Discussion	383
15.7.1 Bimodal Data	383
15.7.2 Time Course Data	385
References	387
16 Bayes and Empirical Bayes Methods for Spotted Microarray Data Analysis	393
<i>Dabao Zhang</i>	
16.1 Introduction	393
16.2 Multiplicative Background Correction	395
16.3 Bayesian Normalization	397
16.3.1 Measurement-Error Models	397
16.3.2 Bayesian Framework	399
16.3.3 Identifying Differentially Expressed Genes	401
16.4 Generalized Empirical Bayes Method for Multiple Arrays	402
16.4.1 Generalized Empirical Bayes Estimators	403
16.4.2 Identifying Differentially Expressed Genes	405
16.4.3 Reducing ChIP-Chip Data	405
16.5 Software Packages	406
16.5.1 MicroBayes	406
16.5.2 GEBCauchy	407
References	409

17 Bayesian Classification Method for QTL Mapping	413
<i>Min Zhang</i>	
17.1 Introduction	413
17.1.1 Non-Bayesian Methods	413
17.1.2 Bayesian Methods	414
17.2 Bayesian Classification to Map QTL	415
17.2.1 Model and Prior Specification	416
17.2.2 Bayesian Inference and the Two-Stage Procedure	417
17.3 QTLBayes: Software for QTL Mapping with Bayesian Classification	419
17.3.1 Availability	419
17.3.2 Model, Bayesian Inference, and Algorithm	419
17.3.3 Input Data and Format	420
17.3.4 Specifications and Default Setting	420
17.3.5 Output and Results	422
17.3.6 Summary	423
17.4 An Example of Data Analysis Using QTLBayes	423
17.4.1 The Data	423
17.4.2 QTL Analysis	423
17.5 Conclusion	425
17.6 Acknowledgment	425
References	427
Index	431