
Preface

Recent advances in genome sequencing (genomics) and protein identification (proteomics) have given rise to challenging research problems that require combined expertise from statistics, biology, computer science, and other fields. Similar challenges arise dealing with modeling genetic data in the form of sequences of various types, e.g., SNIP, RAPD, and microsatellite data. The interdisciplinary nature of bioinformatics presents many research challenges related to integrating concepts, methods, software, and multi-platform data. In addition to new tools for investigating biological systems via high-throughput genomic and proteomic measurements, statisticians face many novel methodological research questions generated by such data. The work in this volume is dedicated to the development and application of Bayesian statistical methods in the analysis of high-throughput bioinformatics data that arise from problems in medical research, in particular cancer and other disease related research, molecular and structural biology.

This volume does not aim to be comprehensive in all areas of bioinformatics. Rather, it presents a broad overview of statistical inference, clustering, and classification problems related to two main high-throughput platforms: microarray gene expression and phylogenetic analysis. The volume's main focus is on the design, statistical inference, and data analysis, from a Bayesian perspective, of data sets arising from such high-throughput experiments. All the chapters included in the volume focus on Bayesian methodologies, parametric as well as nonparametric, but each is self contained and independent. There are topics closely related to each other, but we feel there is no need to maintain any kind of sequencing for reading the chapters. The chapters have been arranged purely in alphabetical order according to the last name of the first author and, in no way, should it be taken to mean any preferential order. Chapter 1 provides estimation and testing in time-course microarray experiments. This chapter discusses a variety of recently developed Bayesian techniques for detection of differentially expressed genes on the basis of time series of microarray data. Chapter 2 considers classification for differential gene expression using hierarchical Bayesian models. Chapter 3 deals with applications of mode-oriented stochastic search algorithms for discrete multi-way data for genome wide studies. Chapter 4 develops novel Bayesian nonparametric approaches to bioinformatics problems. Dirichlet process mixture models are introduced for multiple testing, high-dimensional regression, clustering, and functional data analysis. Chapter 5 develops measurement error and survival models for cDNA microarrays. Chapter 6 focuses on robust Bayesian

inference for differential gene expression. Chapter 7 develops Bayesian hidden Markov modeling approach for CGH array data. Bayesian approaches to phylogenetic analysis are developed in Chapter 8. Chapter 9 is concerned with gene selection for identification of biomarkers in high-throughput data. Chapter 10 develops sparsity priors for protein-protein interaction predictions. Chapter 11 describes Bayesian networks learning for gene expression data. Chapter 12 describes in-vitro to in-vivo factor profiling in genomic expressions. Proportional hazards regression using Bayesian kernel machines is considered in Chapter 13. Chapter 14 focuses on a Bayesian mixture model for protein biomarker discovery using matrix-assisted laser desorption and ionization mass spectrometry. Chapter 15 focuses on general Bayesian methodology for detecting differentially expressed genes. Bayes and empirical Bayes methods for spotted microarray data analysis are considered in Chapter 16. Finally Chapter 17 focuses on the Bayesian classification method for QTL mapping.

We thank all the collaborators for contributing their ideas and insights toward this volume in a timely manner. We fail in our duties if we do not express our sincere indebtedness to our referees, who were quite critical and unbiased in giving their opinions. We do realize that despite of their busy schedules the referees offered us every support in commenting on various chapters. Undoubtedly, it is the joint endeavor of the contributors and the referees that emerged in the form of such an important and significant volume for the Bayesian world.

We thankfully acknowledge David Grubbs from Chapman & Hall, CRC, for his continuous encouragement and support of this project. Our special thanks go to our spouses, Rita, Amrita, and Mou, for their continuous support and encouragement.

We are excited by the continuing opportunities for statistical challenges in the area of bioinformatics data analysis and modeling from a Bayesian point of view. We hope our readers will join us in being engaged with changing technologies and statistical development in this area of bioinformatics.

Dipak K. Dey
Samiran Ghosh
Bani Mallick

MATLAB® is a registered trademark of The Math Works, Inc. For product information, please contact:

The Math Works
3 Apple Hill Drive
Natick, MA
Tel: 508-647-7000
Fax: 508-647-7001
Email: info@mathworks.com
Web: <http://www.mathworks.com>