

Preface

For the past 10 to 15 years, I have been investigating molecular genetics, and the application of Bayesian networks in this area. I decided to summarize this material into an interdisciplinary course on *probabilistic bioinformatics* at Northeastern Illinois University (NEIU). When searching for a textbook for it, I learned that there is not really anything appropriate. The closest thing was the edited volume *Probabilistic Modelling in Bioinformatics and Medical Informatics*, by Dirk Husmeier, Richard Dybowski, and Stephen Roberts. However, that book is not sufficiently accessible to be useful for a college-level course, at least at a mainstream university such as NEIU. So I decided to write this book, which serves not only as a text for an upper-level undergraduate or graduate course in probabilistic bioinformatics, but also should be a useful reference for a practitioner in the field.

This is not a book on biology, but rather one on the application of probability and statistics, in particular Bayesian networks, to genetics; so, there is no effort to discuss chemical structure or behavior. For example, a DNA sequence is simply treated as a string composed of the characters A, G, C, and T. I do, however, present some material on high-level features of biological processes so as to provide a context for the applications.

As in most books I write, I have taken the following measures to make the content accessible. First, ordinarily I do not introduce a concept or prove a theorem concerning the most general case. Rather, I discuss a specific, concrete case and then mention that the result can be generalized. Second, I provide ample figures and examples. A picture may be worth a 1000 words, but an example is worth 10,000 words.

The book has three parts. Part I provides background material on probability, statistics, and genetics, which the student or reader may already know. Many readers should be able to skip this part. Part II discusses Bayesian networks, inference in Bayesian networks, and learning Bayesian networks from data. Part III covers applications to bioinformatics. Specifically, Chapter 9 concerns non-molecular evolutionary genetics, Chapter 10 presents molecular evolutionary genetics, Chapter 11 shows an application of Bayesian networks to molecular phylogenetics, Chapter 12 introduces analyzing gene expression data, and Chapter 13 concerns gene linkage analysis.

For the purpose of classification, it seemed best to group the material into these three parts. However, I do not cover the material in this order in the

course. If the material were covered in the same order as here, one would not get to bioinformatics until the course was half-over. By that time, rather than being motivated by the interesting applications of probability and statistics to genetics, the student might be motivated to drop the course. So I cover the material in an order that gets to applications to genetics as quickly as possible. That is, I do not discuss any particular background material until that material is necessary to the understanding of a chapter on bioinformatics. The order is as follows:

- Chapter 1, 2 all
- Chapter 3 Sections 3.1, 3.3
- Chapter 4 all
- Chapter 9 Sections 9.1 – 9.2
- Chapter 3, Section 3.2
- Chapter 9, Sections 9.3 – 9.5
- Chapter 10, all
- Chapters 5, 6, 7, 8 all
- Chapters 11, 12, 13 all

Reader too may prefer to read the book in this order.

Rich Neapolitan
RE-Neapolitan@neiu.edu