

Synopsis

Probabilistic models are becoming increasingly important in analyzing the huge amount of data being produced by large-scale DNA-sequencing efforts such as the Human Genome Project. For example, hidden Markov models are used for analyzing biological sequences, linguistic-grammar-based probabilistic models for identifying RNA secondary structure, and probabilistic evolutionary models for inferring phylogenies of sequences from different organisms. This book gives a unified, up-to-date and self-contained account, with a Bayesian slant, of such methods, and more generally to probabilistic methods of sequence analysis. Written by an interdisciplinary team of authors, it is accessible to molecular biologists, computer scientists, and mathematicians with no formal knowledge of the other fields, and at the same time presents the state of the art in this new and important field.

Book Information

Paperback: 356 pages

Publisher: Cambridge University Press; 1 edition (May 13, 1998)

Language: English

ISBN-10: 0521629713

ISBN-13: 978-0521629713

Product Dimensions: 6.8 x 0.8 x 9.7 inches

Shipping Weight: 1.7 pounds (View shipping rates and policies)

Average Customer Review: 4.5 out of 5 stars [See all reviews](#) (24 customer reviews)

Best Sellers Rank: #220,006 in Books (See Top 100 in Books) #43 in [Books > Computers & Technology > Computer Science > Bioinformatics](#) #132 in [Books > Science & Math > Biological Sciences > Biology > Molecular Biology](#) #213 in [Books > Engineering & Transportation > Engineering > Bioengineering > Biochemistry](#)

Customer Reviews

This book is a very well written overview to hidden Markov models and context-free grammar methods in computational biology. The authors have written a book that is useful to both biologists and mathematicians. Biologists with a background in probability theory equivalent to a senior-level course should be able to follow along without any trouble. The approach the author's take in the book is very intuitive and they motivate the concepts with elementary examples before moving on to the more abstract definitions. Exercises also abound in the book, and they are straightforward enough to work out, and should be if one desires an in-depth understanding of the main text. In

addition, there is a software package called HMMER, developed by one of the authors (Eddy) that is in the public domain and can be downloaded from the Internet. The package specifically uses hidden Markov models to perform sequence analysis using the methods outlined in the book. Probabilistic modeling has been applied to many different areas, including speech recognition, network performance analysis, and computational radiology. An overview of probabilistic modeling is given in the first chapter, and the authors effectively introduce the concepts without heavy abstract formalism, which for completeness they delegate to the last chapter of the book. Bayesian parameter estimation is introduced as well as maximum likelihood estimation. The authors take a pragmatic attitude in the utility of these different approaches, with both being developed in the book. This is followed by a treatment of pairwise alignment in Chapter Two, which begins with substitution matrices. They point out, via some exercises, the role of physics in influencing particular alignments (hydrophobicity for example).

I picked up this book at the recommendation of a number of colleagues in computational linguistics and speech processing as a way to find out what's going on in biological sequence analysis. I was hoping to learn about applications of the kinds of algorithms I know for handling speech and language, such as HMM decoding and context-free grammar parsing, to biological sequences. This book delivered, as recommended. As the title implies, "Biological Sequence Analysis" focuses almost exclusively on sequence analysis. After a brief overview of statistics (more a reminder than an introduction), the first half of the book is devoted to alignment algorithms. These algorithms take pairs of sequences of bases making up DNA or sequences of amino acids making up proteins and provide optimal alignments of the sequences or of subsequences according to various statistical models of match likelihoods. Methods analyzed include edit distances with various substitution and gapping penalties (penalties for sections that don't match), Hidden Markov Models (HMMs) for alignment and also for classification against families, and finally, multiple sequence alignment, where alignment is generalized from pairs to sets of sequences. I found the section on building phylogenetic trees by means of hierarchical clustering to be the most fascinating section of the book (especially given its practical application to classifying wine varietals!). The remainder of the book is devoted to higher-order grammars such as context-free grammars, and their stochastic generalization. Stochastic context-free grammars are applied to the analysis of RNA secondary structure (folding).

[Download to continue reading...](#)

Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids Genome-Scale

Algorithm Design: Biological Sequence Analysis in the Era of High-Throughput Sequencing
Probabilistic Graphical Models: Principles and Techniques (Adaptive Computation and Machine Learning series)
Biological Modeling and Simulation: A Survey of Practical Models, Algorithms, and Numerical Methods (Computational Molecular Biology)
Unsupervised Machine Learning in Python: Master Data Science and Machine Learning with Cluster Analysis, Gaussian Mixture Models, and Principal Components Analysis
Machine Learning: A Probabilistic Perspective (Adaptive Computation and Machine Learning series)
Probabilistic Reasoning in Expert Systems: Theory and Algorithms
Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference (Morgan Kaufmann Series in Representation and Reasoning)
Numerical Methods of Statistics (Cambridge Series in Statistical and Probabilistic Mathematics)
Practical Probabilistic Programming
The Imago Sequence and Other Stories
Sequence Knitting: Simple Methods for Creating Complex Reversible Fabrics
Blood of Innocents: The Sorcery Ascendant Sequence, Book 2
Partials (Partials Sequence)
A Crucible of Souls: The Sorcery Ascendant Sequence, Book 1
Skill Building Sequence for Choral Ensembles: Teacher's Guide for Children's Choir: Volume 1
Over Sea, Under Stone (The Dark is Rising Sequence)
The Analysis of Biological Data
Analytics: Data Science, Data Analysis and Predictive Analytics for Business (Algorithms, Business Intelligence, Statistical Analysis, Decision Analysis, Business Analytics, Data Mining, Big Data)
Art Models 6: The Female Figure in Shadow and Light (Art Models series)

[Dmca](#)