

Contents

<i>Preface</i>	<i>page</i> ix
<i>Guide to the chapters</i>	xi
<i>Acknowledgment of support</i>	xii
Part I Introduction to the four themes	1
1 Statistics <i>L. Pachter and B. Sturmfels</i>	3
1.1 Statistical models for discrete data	4
1.2 Linear models and toric models	9
1.3 Expectation Maximization	17
1.4 Markov models	24
1.5 Graphical models	33
2 Computation <i>L. Pachter and B. Sturmfels</i>	43
2.1 Tropical arithmetic and dynamic programming	44
2.2 Sequence alignment	49
2.3 Polytopes	59
2.4 Trees and metrics	67
2.5 Software	75
3 Algebra <i>L. Pachter and B. Sturmfels</i>	85
3.1 Varieties and Gröbner bases	86
3.2 Implicitization	94
3.3 Maximum likelihood estimation	102
3.4 Tropical geometry	109
3.5 The tree of life and other tropical varieties	117
4 Biology <i>L. Pachter and B. Sturmfels</i>	125
4.1 Genomes	126
4.2 The data	132
4.3 The problems	137
4.4 Statistical models for a biological sequence	141
4.5 Statistical models of mutation	147

Part II Studies on the four themes	161
5 Parametric Inference <i>R. Mihaescu</i>	165
5.1 Tropical sum-product decompositions	166
5.2 The polytope propagation algorithm	169
5.3 Algorithm complexity	173
5.4 Specialization of parameters	177
6 Polytope Propagation on Graphs <i>M. Joswig</i>	181
6.1 Polytopes from directed acyclic graphs	181
6.2 Specialization to hidden Markov models	185
6.3 An implementation in polymake	186
6.4 Returning to our example	191
7 Parametric Sequence Alignment	
<i>C. Dewey and K. Woods</i>	193
7.1 Few alignments are optimal	193
7.2 Polytope propagation for alignments	195
7.3 Retrieving alignments from polytope vertices	199
7.4 Biologically correct alignments	202
8 Bounds for Optimal Sequence Alignment	
<i>S. Elizalde and F. Lam</i>	206
8.1 Alignments and optimality	206
8.2 Geometric interpretation	208
8.3 Known bounds	211
8.4 The square root conjecture	212
9 Inference Functions <i>S. Elizalde</i>	215
9.1 What is an inference function?	215
9.2 The few inference functions theorem	217
9.3 Inference functions for sequence alignment	220
10 Geometry of Markov Chains <i>E. Kuo</i>	226
10.1 Viterbi sequences	226
10.2 Two- and three-state Markov chains	229
10.3 Markov chains with many states	231
10.4 Fully observed Markov models	233
11 Equations Defining Hidden Markov Models	
<i>N. Bray and J. Morton</i>	237
11.1 The hidden Markov model	237
11.2 Gröbner bases	238
11.3 Linear algebra	240
11.4 Combinatorially described invariants	247

12 The EM Algorithm for Hidden Markov Models	
<i>I. B. Hallgrímsdóttir, R. A. Milowski and J. Yu</i>	250
12.1 The EM algorithm for hidden Markov models	250
12.2 An implementation of the Baum–Welch algorithm	254
12.3 Plots of the likelihood surface	257
12.4 The EM algorithm and the gradient of the likelihood	261
13 Homology Mapping with Markov Random Fields	<i>A. Caspi</i> 264
13.1 Genome mapping	264
13.2 Markov random fields	267
13.3 MRFs in homology assignment	270
13.4 Tractable MAP inference in a subclass of MRFs	273
13.5 The Cystic Fibrosis Transmembrane Regulator	276
14 Mutagenetic Tree Models	<i>N. Beerenwinkel and M. Drton</i> 278
14.1 Accumulative evolutionary processes	278
14.2 Mutagenetic trees	279
14.3 Algebraic invariants	282
14.4 Mixture models	287
15 Catalog of Small Trees	
<i>M. Casanellas, L. D. Garcia, and S. Sullivan</i>	291
15.1 Notation and conventions	291
15.2 Fourier coordinates	295
15.3 Description of website features	297
15.4 Example	298
15.5 Using the invariants	303
16 The Strand Symmetric Model	
<i>M. Casanellas and S. Sullivan</i>	305
16.1 Matrix-valued Fourier transform	306
16.2 Invariants for the 3-taxa tree	310
16.3 G -tensors	314
16.4 Extending invariants	318
16.5 Reduction to $K_{1,3}$	319
17 Extending Tree Models to Splits Networks	<i>D. Bryant</i> 322
17.1 Trees, splits and splits networks	322
17.2 Distance-based models for trees and splits graphs	325
17.3 A graphical model on a splits network	326
17.4 Group-based mutation models	327
17.5 Group-based models for trees and splits	330
17.6 A Fourier calculus for splits networks	332

viii	<i>Contents</i>	
18	Small Trees and Generalized Neighbor-Joining	
	<i>M. Contois and D. Levy</i>	335
	18.1 From alignments to dissimilarity	335
	18.2 From dissimilarity to trees	337
	18.3 The need for exact solutions	342
	18.4 Jukes–Cantor triples	344
19	Tree Construction using Singular Value Decomposition	
	<i>N. Eriksson</i>	347
	19.1 The general Markov model	347
	19.2 Flattenings and rank conditions	348
	19.3 Singular Value Decomposition	351
	19.4 Tree-construction algorithm	352
	19.5 Performance analysis	355
20	Applications of Interval Methods to Phylogenetics	
	<i>R. Sainudiin and R. Yoshida</i>	359
	20.1 Brief introduction to interval analysis	360
	20.2 Enclosing the likelihood of a compact set of trees	366
	20.3 Global optimization	366
	20.4 Applications to phylogenetics	371
21	Analysis of Point Mutations in Vertebrate Genomes	
	<i>J. Al-Aidroos and S. Snir</i>	375
	21.1 Estimating mutation rates	375
	21.2 The ENCODE data	378
	21.3 Synonymous substitutions	379
	21.4 The rodent problem	381
22	Ultra-Conserved Elements in Vertebrate and Fly Genomes	
	<i>M. Drton, N. Eriksson and G. Leung</i>	387
	22.1 The data	387
	22.2 Ultra-conserved elements	390
	22.3 Biology of ultra-conserved elements	392
	22.4 Statistical significance of ultra-conservation	400
	<i>References</i>	403
	<i>Index</i>	418