

## Index

- 4TI2, 75
- algebraic  
 invariant, 7, 237, 238, 285, 302, 306  
 number, 88  
 statistical model, 7  
 statistics, 3  
 variety, 87
- alignment, 136, 220  
 graph, 51  
 polygon, 209
- assignment problem, 48
- autosome, 126
- Bézout's theorem, 88
- backward probability, 252
- Baum–Welch algorithm, 30, 250–254, *see also*  
 Expectation Maximization
- Bayesian network, *see* directed graphical model
- BLAST, 80
- bounded region, 10
- branch length, 150
- Broad Institute, 135
- Buchberger's algorithm, 90, 93
- cancer, 278
- CFTR, *see* Cystic Fibrosis Transmembrane  
 Regulator
- Chapman–Kolmogorov equations, 148
- character group, 328
- cherry, 72, 342  
 picking theorem, 72
- chromosome, 126
- clique, 37, 267
- codon, 128
- common refinement, 216
- comparative genomics, 130
- compatible  
 splits, 69, 322  
 state, 280
- concave function, 9, 10
- conditional independence, 34, 249, 283
- contigs, 134
- convex  
 hull, 60  
 polyhedron, 60  
 polytope, 60
- coverage, 133
- CpG island, 4, 128, 141
- Cramer's rule, 121
- Cystic Fibrosis Transmembrane Regulator, 82,  
 276, 378
- Delannoy numbers, 51
- DiaNA, 1, 4–11, 20, 21, 92, 96, 106, 114, 116,  
 141, 157
- diploid, 126
- directed graphical model, 38, 281
- dissimilarity map, 67, 74, 336
- DNA, 126, 138  
 intergenic, 138
- Drosophila*, 388
- dynamic programming, 44
- EM algorithm, *see* Expectation Maximization
- empirical distribution, 6, 8
- ENCODE regions, 141, 356, 378, 388
- entropy, 14
- error, types of, 359
- estimation, 4
- euchromatin, 135
- Euclidean algorithm, 94
- eukaryote, 126
- evolutionary model, 148–155, 375
- exon, 139
- Expectation Maximization, 17, 287
- explanation, 142, 165, 215
- exponential family, 11
- expression, 129
- extended alignment graph, 54
- f-vector, 61
- FASTA format, 82
- fastDNAm1, 344, 373
- Felsenstein  
 algorithm, 57  
 hierarchy, 153, 305, 375
- finished genome, 135
- flattening, 348
- Floyd–Warshall algorithm, 46
- Fort Lauderdale agreement, 132
- forward  
 algorithm, 166  
 probability, 252
- forward algorithm, *see also* sum-product  
 decomposition
- four-point condition, 70, 118
- Fourier transform, 152, 295, 308, 332
- Frobenius norm, 352
- fully observed tree model, 31
- G-tensor, 314

- Galois group, 88  
 Gaussian elimination, 94  
 GENBANK, 136  
 gene, 128  
   annotation, 380  
   finding, 129, 145–147  
   ontology, 394  
 genetic code, 128, 379  
 genome, 3  
   signature, 129  
 global  
   alignment, 53  
   Markov property, 283  
 globin, 203  
 gradient, 13  
 graphical model, 33–42, 165, 215, 278, 326  
 Gröbner basis, 5, 89, 237, 238, 282  
 group-based model, 307, 327
- H-polytope, 60  
 Hadamard conjugation, *see* Fourier transform  
 Hardy–Weinberg curve, 86  
 Hessian matrix, 9  
 heterochromatin, 134  
 hidden  
   Markov model, 29–31, 42, 98, 167, 185, 237, 250  
   tree model, 31, 148, 330  
 Hilbert  
   basis theorem, 89, 242  
   Nullstellensatz, 90  
 HMM, *see* hidden Markov model  
 homogeneous polynomial, 87  
 homology, 131  
   mapping, 264–277  
 human genome, 125, 135, 378, 388  
 human immunodeficiency virus, 278  
 hyperplane  
   arrangement, 10, 116  
   tropical, 120  
 hypersurface, 87
- ideal, 89  
 implicitization, 96, 238–240  
 independence model, 15  
 inference  
   discriminative, 41  
   function, 115, 117, 215–225  
   generative, 41  
   parametric, 165–180  
 initial  
   form, 65  
   ideal, 89  
   monomial, 86  
 integer linear program, 47, 275  
 interval arithmetic, 360–366  
 intron, 139  
 Ising model, 274  
 isometric  
   coloring, 323  
   embedding, 323
- Joint Genome Institute, 135  
 Jukes–Cantor  
   correction, 155  
   model, 11, 149, 153, 293, 401
- Kimura  
   2-parameter model, 154, 293, 329  
   3-parameter model, 154, 293, 329
- Kullback–Leibler distance, 20
- Lagrange multiplier, 13, 108  
 Lander–Waterman model, 134  
 lattice, 282  
   polytope, 174  
 lexicographic monomial order, 86  
 likelihood, 5  
   equations, 10  
   function, 5, 8  
 linear model, 9  
 log-likelihood function, 5, 8, 9  
 log-linear model, 12  
 log-odds ratio, 56  
 long branch attraction, 385
- MACAULAY2, 76  
 MAGMA, 77  
 MAP inference, *see* maximum a posteriori inference  
 MAPLE, 77  
 marginalization, 237  
   of a product function, 42  
 map, 17, 31  
 Markov  
   chain, 27  
   random field, 36, 267–270  
 MATHEMATICA, 78  
 MATLAB, 78  
 matrix exponential, 149  
 MAVID, 82  
 maximum a posteriori inference, 25, 40, 142, 169, 273  
 maximum likelihood estimation, 5, 8, 375–386  
 metric  
   cone, 68  
   network, 326  
   space, 68  
 microbial genomes, 125  
 Minkowski sum, 64, 183, 216  
 mixture model, 7, 287  
 model  
   invariant, 26  
   selection, 33  
 molecular clock, 293  
 moment map, 14  
 monomial, 86  
   order, 86, 282  
 Moore’s law, 125  
 mRNA, 139, 203, 396  
 multigraded ring, 243  
 mutagenetic tree, 279  
   mixture model, 287  
   model, 280
- naive Bayes model, 102, 288  
 natural selection, 130, 138  
 Needleman–Wunsch algorithm, 54, 168, *see also* sum-product decomposition  
 neighbor-joining  
   algorithm, 74  
   generalized, 342, 374  
 Newick tree notation, 376  
 Newton polytope, 65, 115, 170, 196, 216  
 non-identifiability, 21  
 normal  
   cone, 61  
   fan, 61  
 nucleotide, 126

420

occasionally dishonest casino, 30, 245, 248  
 oncogenetic trees, 278  
 open reading frame, 145

pair hidden Markov model, 55, 157, 209  
 pairwise compatible splits, 69, 322  
 PAML, 83, 375  
 parameter specialization, 177  
 parametric  
   alignment, 55, 193–206  
   inference, *see* inference, parametric  
 partition function, 12, 268  
 Petersen graph, 119  
 PHMM, *see* pair hidden Markov Model  
 PHYLIP, 83  
 phylogenetic  
   diversity, 336  
   hidden Markov model, 157  
   invariant, 152, 302, 306, *see also* algebraic  
     invariant  
   shadowing, 156  
   surface, 122  
 Poisson model, 134  
 polyA tail, 138  
 polyhedron, 10  
 polymake, 79  
 polynomial  
   map, 7  
   ring, 86  
 polytomy, 119  
 polytope, 14, 60  
   algebra, 64  
   lattice, 218  
   propagation, 66, 170, 181  
 position specific score matrix, 143  
 Potts model, 274  
 pre-mRNA, 138  
 probability simplex, 6  
 prokaryote, 126  
 promoter, 129, 139  
 protein, 128  
 PSSM, *see* position specific score matrix  
 purine, 126  
 pyrimidine, 126

quality score, 133  
 quartet, 70

random variable, 6  
 rate matrix, 148, 376  
 rational normal curve, 289  
 read, 132  
 reduced Gröbner basis, 89, 282  
 RefSeq genes, 139, 357, 380  
 regulation, 147  
 reversal, 131  
 rice genome, 134  
 rodent problem, 375

saddle point, 23, 258  
 sample size, 8  
 Sanger Institute, 136  
 scoring scheme, 52  
 secant variety, 288  
 Segre variety, 286, 288  
 sex chromosome, 126  
 signed permutation, 131  
 SINGULAR, 79  
 singular value decomposition, 351–352  
 space of trees, 68, 118

Index

splice site, 400  
   acceptor, 139  
   donor, 139, 144  
   model, 144  
 split, 69, 322  
   splits graph, 323  
   splits network, 323  
   SPLITSTREE, 84  
 square root conjecture, 212–214  
 state space, 6  
   diagram, 144  
 statistical  
   genetics, 126  
   inference, 3  
   model, 4, 6  
 substitution matrix, 148  
 sufficient statistic, 8  
 sum-product decomposition, 166  
 super-contigs, 134  
 surjectively positive map, 114  
 SVD, *see* singular value decomposition  
 synonymous substitution, 379, 399

term ordering, 239  
 tetrahedron, 6, 7  
 toric  
   Markov chain, 25, 228  
   model, 11  
   variety, 25  
 trace, 133  
 transcription, 129, 138  
 transition, 155  
 translation, 129, 139  
 transversion, 154, 155  
 tree metric, 68  
 tropical  
   arithmetic, 44–49  
   determinant, 48  
   geometry, 109–117  
   Grassmannian, 118  
   hypersurface, 111  
   product, 44  
   semiring, 44  
   sum, 44  
 tropicalization, 47, 112  
 tumor development, 278

ultra-conserved element, 390  
 undirected graphical model, *see* Markov random  
   field  
 uniform star model, 289  
 unrooted tree, 69  
 UTR, 139

V-polytope, 60  
 Varchenko's formula, 10  
 variety, *see* algebraic variety  
 Veronese map, 289  
 Viterbi  
   algorithm, 57, 168, *see also* sum-product  
     decomposition  
   sequence, 56, 57, 145, 227  
   pseudo, 228

weakly compatible splits, 326  
 weight matrix, 143, *see also* position specific  
   score matrix  
 whole genome shotgun, 132

Zariski closure, 238  
 zonotope, 116, 183