

Computational Molecular Evolution

Ziheng Yang (z.yang@ucl.ac.uk)
University College London

Preface

Part I: Modeling Molecular Evolution

CHAPTER 1 Models of Nucleotide Substitution

- 1.1 Introduction
- 1.2 Markov Models of Nucleotide Substitution and Distance Estimation
 - 1.2.1 The JC69 Model
 - 1.2.2 The K80 Model
 - 1.2.3 HKY85, F84, TN93 etc.
 - 1.2.4 The Transition/Transversion Rate Ratio
- 1.3 Variable Substitution Rates Across Sites
- 1.4 Maximum Likelihood Estimation
 - 1.4.1 The JC69 Model
 - 1.4.2 The K80 Model
 - *1.4.3 Profile and Integrated Likelihood Methods
- 1.5 Markov Chains and Distance Estimation under General Models
 - 1.5.1 General Theory
 - 1.5.2 The General Time-Reversible (GTR) Model
- 1.6 Discussions
- 1.7 Exercises

CHAPTER 2 Models of Amino Acid and Codon Substitution

- 2.1 Introduction
- 2.2 Models of Amino Acid Replacement
 - 2.2.1 Empirical Models
 - 2.2.2 Mechanistic Models
 - 2.2.3 Among-Site Heterogeneity
- 2.3 Estimation of Distance Between Two Protein Sequences
 - 2.3.1 The Poisson model
 - 2.3.2 Empirical models
 - 2.3.2 Gamma Distances
 - 2.3.3 Example: Distance between Cat and Rabbit p53 Genes
- 2.4 Models of Codon Substitution
- 2.5 Estimation of Synonymous and Nonsynonymous Substitution Rates
 - 2.5.1 Counting Methods
 - 2.5.2 Maximum Likelihood Method
 - 2.5.3 Comparison of Methods
 - *2.5.4 Interpretation and a Plethora of Distances
- *2.6 Numerical Calculation of the Transition-Probability Matrix
- 2.7 Exercises

Part II: Phylogeny Reconstruction

CHAPTER 3 PHYLOGENY RECONSTRUCTION: Overview

- 3.1 Tree Concepts
 - 3.1.1 Terminology
 - 3.1.2 Topological Distance Between Trees
 - 3.1.3 Consensus Trees
 - 3.1.4 Gene Trees and Species Trees
 - 3.1.5 Classification of Tree-Reconstruction Methods
- 3.2 Exhaustive and Heuristic Tree Search
 - 3.2.1 Exhaustive Tree Search
 - 3.2.2 Heuristic Tree Search
 - 3.2.3 Branch Swapping

- 3.2.4 Local Peaks in the Tree Space
 - 3.2.5 Stochastic Tree Search
- 3.3 Distance Methods
 - 3.3.1 Least Squares Method
 - 3.3.2 Neighbor-Joining Method
- 3.4 Maximum Parsimony
 - 3.4.1 Brief History
 - 3.4.2 Counting the Minimum Number of Changes Given the Tree
 - 3.4.3 Weighted Parsimony and Transversion Parsimony
 - 3.4.4 Long-Branch Attraction
 - 3.4.5 Assumptions of Parsimony

CHAPTER 4 Maximum Likelihood Methods

- 4.1 Introduction
- 4.2 Likelihood Calculation on Tree
 - 4.2.1 Data, Model, Tree, and Likelihood
 - 4.2.2 The Pruning Algorithm
 - 4.2.3 Time Reversibility, the Root of Tree and the Molecular Clock
 - 4.2.4 Missing Data and Alignment Gaps
 - 4.2.5 An Numerical Example: Phylogeny of Apes
- 4.3 Likelihood Calculation under More-Complex Models
 - 4.3.1 Models of Variable Rates Among Sites
 - 4.3.2 Models for Combined Analysis of Multiple Data Sets
 - 4.3.3 Nonhomogeneous and Nonstationary Models
 - 4.3.4 Amino Acid and Codon Models
- 4.4 Reconstruction of Ancestral States
 - 4.4.1 Overview
 - 4.4.2 Empirical and Hierarchical Bayes Reconstruction
 - 4.4.3 Discrete Morphological Characters
 - 4.4.4 Systematic Biases in Ancestral Reconstruction
- *4.5 Numerical Algorithms for Maximum Likelihood Estimation
 - 4.5.1 Univariate Optimization
 - 4.5.2 Multivariate Optimization
 - 4.5.3 Optimization on a Fixed Tree
 - 4.5.4 Multiple Local Peaks on the Likelihood Surface for a Fixed Tree
 - 4.5.5 Search for the Maximum Likelihood Tree
- 4.6 Approximations to Likelihood
- 4.7 Model Selection and Robustness
 - 4.7.1 LRT, AIC, and BIC
 - 4.7.2 Model Adequacy and Robustness
- 4.8 Exercises

CHAPTER 5 Bayesian Methods

- 5.1 The Bayesian Paradigm
 - 5.1.1 Overview
 - 5.1.2 Bayes Theorem
 - 5.1.3 Classical versus Bayesian Statistics
- 5.2 Prior
- 5.3 Markov Chain Monte Carlo
 - 5.3.1 Monte Carlo Integration
 - 5.3.3 Metropolis-Hastings Algorithm
 - 5.3.4 Single-Component Metropolis-Hastings Algorithm
 - 5.3.5 Gibbs Sampler
 - 5.3.5 Metropolis-Coupled MCMC (MCMCMC)
- 5.4 Simple Moves and Their Proposal Ratios
 - 5.4.1 Sliding Window Using Uniform Proposal
 - 5.4.2 Sliding Window Using Normal Proposal
 - 5.4.3 Sliding Window Using Multivariate Normal Proposal
 - 5.4.4 Proportional Shrinking and Expanding
- 5.5 Monitoring Markov Chains and Processing Output
 - 5.5.1 Validating and Diagnosing MCMC Algorithms
 - 5.5.2 Potential Scale Reduction Statistic

- 5.5.3 Processing Output
- 5.6 Bayesian Phylogenetics
 - 5.6.1 Brief History
 - 5.6.2 General Framework
 - 5.6.3 Summarizing MCMC Output
 - 5.6.4 Bayesian versus Likelihood
 - 5.6.5 A Numerical Example: Phylogeny of Apes
- 5.7 MCMC Algorithms under the Coalescent Model
 - 5.7.1 Overview
 - 5.7.2 Estimation of θ
- 5.8 Exercises

CHAPTER 6 Comparison of Methods and Tests on Trees

- 6.1 Statistical Performance of Tree-Reconstruction Methods
 - 6.1.1 Criteria
 - 6.1.2 Performance
- 6.2 Likelihood
 - 6.2.1 Contrast with Conventional Parameter Estimation
 - 6.2.2 Consistency
 - 6.2.3 Efficiency
 - 6.2.4 Robustness
- 6.3 Parsimony
 - 6.3.1 Equivalence with Misbehaved Likelihood Models
 - 6.3.2 Equivalence with Well-Behaved Likelihood Models
 - 6.3.3 Assumptions and Justifications
- 6.4 Testing Hypotheses Concerning Trees
 - 6.4.1 Bootstrap
 - 6.4.2 Interior Branch Test
 - 6.4.3 Kishino-Hasegawa Test and Modifications
 - 6.4.4 Indexes Used in Parsimony Analysis
 - 6.4.5 Example: Phylogeny of Apes
- *6.5 Appendix: Tuffley and Steel's Likelihood Analysis of One Character

Part III: Advanced Topics

CHAPTER 7 Molecular Clock and Estimation of Species Divergence Times

- 7.1 Overview
- 7.2 Tests of the Molecular Clock
 - 7.2.1 Relative Rate Tests
 - 7.2.2 Likelihood Ratio Test
 - 7.2.3 Limitations of the clock test
 - 7.2.4 Index of Dispersion
- 7.3 Likelihood Estimation of Divergence Times
 - 7.3.1 Global Clock Model
 - 7.3.2 Local-Clock Models
 - 7.3.3 Heuristic Rate-Smoothing Methods
 - 7.3.4 Dating Primate Divergences
 - *7.3.5 Uncertainties in Fossils
- 7.4 Bayesian Estimation of Divergence Times
 - 7.4.1 General Framework
 - 7.4.2 Calculation of the Likelihood
 - 7.4.3 Prior on Rates
 - 7.4.4 Uncertainties in Fossils and Prior on Divergence Times
 - 7.4.5 Application to Primate and Mammalian Divergences
- 7.5 Perspectives

CHAPTER 8 Neutral and Adaptive Protein Evolution

- 8.1 Introduction
- 8.2 The Neutral Theory and Tests of Neutrality
 - 8.2.1 The Neutral and Nearly Neutral Theory

- 8.2.2 Tajima's D statistic
- 8.2.3 Fu and Li's D and Fay and Wu's H Statistics
- 8.2.4 McDonald-Kreitman Test and Estimation of Selective Strength
- 8.2.5 Hudson-Kreitman-Aquade Test
- 8.3 Lineages Undergoing Adaptive Evolution
 - 8.3.1 Heuristic Methods
 - 8.3.2 Likelihood Method
- 8.4 Amino Acid Sites Undergoing Adaptive Evolution
 - 8.4.1 Three Strategies
 - 8.4.2 Likelihood Ratio Test of Positive Selection under Random-Sites Models
 - 8.4.3 Identification of Sites under Positive Selection
 - 8.4.4 Positive Selection in the Human Major Histocompatibility (MHC) Locus
- 8.5 Adaptive Evolution Affecting Particular Sites and Lineages
 - 8.5.1 Branch-Site Test of Positive Selection
 - 8.5.2 Other Similar Models
 - 8.5.3 Adaptive Evolution in Angiosperm Phytochromes
- 8.6 Assumptions, Limitations and Comparisons
 - 8.6.1 Limitations of Current Methods
 - 8.6.2 Comparison Between Tests of Neutrality and Tests Based on dN and dS
- 8.7 Adaptively Evolving Genes

CHAPTER 9 Simulating Molecular Evolution

- 9.1 Introduction
- 9.2 Random Number Generator
- 9.3 Generation of Continuous Random Variables
- 9.4 Generation of Discrete Random Variables
 - 9.4.1 Discrete Uniform Distribution
 - 9.4.2 Binomial Distribution
 - 9.4.3 General Discrete Distribution
 - 9.4.4 Multinomial Distribution
 - 9.4.5 The Composition Method for Mixture Distributions
 - 9.4.6* The Alias Method for Sampling from a Discrete Distribution
- 9.5 Simulating Molecular Evolution
 - 9.5.1 Simulating Sequences on a Fixed Tree
 - 9.5.2 Generating Random Trees
- 9.6 Exercises

CHAPTER 10 Perspectives

Appendixes

- A. Functions of Random Variables
- B. The Delta Technique
- C. Phylogenetics Software

Reference

Index

Computational Molecular Evolution. 374 Pages · 2006 · 6.39 MB · 7,816 Downloads · English. by Ziheng Yang. Preview. Download. · illustrations and jar How Computers Work: The Evolution of Technology Ron White|Tim Downs How to Hack Computers: how to hack computers, hacking for beginners, penetration testing, hacking for dummies, computer security, computer hacking, hacking techniques, network scanning. 138 Pages · 2015 · 679 KB · 92,593 Downloads · New! (Course programs, previous years). Online lectures. Common descent (11:52). Natural selection (14:57). Evidence for evolution (part 1) (9:34). Evidence for evolution (part 2) (20:54). Population growth and selection (18:13). Course material. Lecture notes. Slides, week 1. Computer exercise. Population Growth, Fitness, and Selection. Online lectures. Neutral Theory of Molecular Evolution (11:28). Genetic Drift (9:35). Trees: Terminology and Representation (9:41). Homology and Homoplasy (9:07). Molecular evolution refers to the changes in living systems that have occurred thereafter, eventually leading to the formation of unicellular and multicellular organisms. From: Encyclopedia of Bioinformatics and Computational Biology, 2019. Related terms: Peptide Sequence.