

Contents

1	Gene Genealogies and Genetic Data	1
1.1	Genealogies and Genealogical Thinking	4
1.2	Mutation and Mutation Models	7
1.3	Measures of DNA Sequence Polymorphism	9
1.4	Variation at the PDHA1 Locus in Humans	12
1.5	Exercises	16
2	Probability Theory for the Coalescent	17
2.1	Fundamentals of Probability Theory	17
2.1.1	Events, Probabilities, and Random Variables	17
2.1.2	Four Famous Probability Distributions	26
2.2	Poisson Processes	33
2.2.1	Poisson Process Results for the Coalescent	35
2.2.2	Convolutions of Exponential Distributions	37
2.3	Exercises	39
3	The Coalescent	41
3.1	Population Genetic Models	41
3.1.1	The Wright-Fisher Model	42
3.1.2	The Moran Model	43
3.2	The Standard Coalescent Model	45
3.2.1	Wright-Fisher Model Derivation	47
3.2.2	Moran Model Derivation	51
3.2.3	Breeding Structure and Exchangeability	53
3.3	Some Properties of Coalescent Genealogies	57
3.3.1	Two Measures of the Size of a Genealogy	57
3.3.2	The Branching Structure of Genealogies	61
3.4	Human-Neanderthal Couples?	63
3.5	Exercises	69
4	Neutral Mutations and Genetic Polymorphisms	71
4.1	The Infinite Sites Model and DNA Sequence Polymorphism	72
4.1.1	The Number Segregating Sites	73
4.1.2	Pairwise Differences	75
4.1.3	Site Frequencies	79
4.2	The Infinite Alleles Model and the Ewens Sampling Formula	83
4.3	Deviations from the Standard Model: Testing “Neutrality”	86
4.3.1	Test Statistics Based on Site Frequencies	87
4.3.2	Demographic History and Patterns of Polymorphism	90
4.4	A Footprint of Positive Selection <i>Drosophila simulans</i>	94

4.5	Exercises	99
5	Population Structure	101
5.1	Markov Processes	101
5.2	The Structured Coalescent	101
5.3	Geographic Structure	101
5.3.1	Subdivision with Migration	101
5.3.2	Divergence without Gene Flow	101
5.3.3	More Complicated Histories	101
5.4	Strong Natural Selection	101
5.4.1	Selective Sweeps	101
5.4.2	Balancing Selection	101
5.5	Gene Trees versus Species Trees in the Great Apes	101
5.6	Exercises	102
6	Splitting Lineages: Recombination and Weak Selection	103
6.1	Branching Processes and Branching-Coalescing Graphs	103
6.2	The Ancestral Recombination Graph	103
6.2.1	Breaking up the Sequence	103
6.2.2	Expectations and Variances	103
6.2.3	Comparing Patterns within and between Loci	103
6.3	The Ancestral Selection Graph	103
6.4	Correlation of Polymorphism Levels across the Human Genome	103
6.5	Exercises	104
7	Separation of Time Scales and Robustness Results	105
7.1	Markov Processes with Two Time Scales	105
7.2	Robustness Results for the Coalescent	105
7.2.1	Two Sexes	105
7.2.2	Partial Selfing	105
7.2.3	Age Structure	105
7.2.4	Metapopulation Structure	105
7.3	Effective Population Size in <i>Drosophila</i>	105
7.4	Exercises	106
8	Simulations and Simulation-based Methods	107
8.1	Simulating Genealogies: Monte-Carlo Integration	107
8.2	Computing Likelihoods of Full Data Patterns	107
8.2.1	Importance Sampling Methods	107
8.2.2	Markov Chain Monte Carlo Methods	107
8.3	The Bayesian Approach to Inference	107
8.4	Human Migrations into the Americas	107
8.5	Exercises	108
	Solutions to Odd-Numbered Exercises	109
	Bibliography	111
	Index	119