

MATH 660
INFORMATION ON THE FINAL EXAM

The final in MATH 660 is Friday, **May 4 at 8am**. You should bring a (dumb) calculator so that you can do some elementary arithmetic quickly if need be. The exam will *heavily* emphasize ideas since the midterm. That is, the emphasis will be on chapters 6-13 in the notes. You do, however, need to know things like what problem NJ is trying to address (closest taxa metrically may not be sister), MC, parsimony informative site, etc. You will not, for example, be asked to do a weighted parsimony analysis. Computational questions will focus on things like estimating model parameters or performing an ML maximization problem or computing the expected time to coalescence under Kingman's coalescent model of n gene lineages.

You may bring a cheat sheet with you to the exam where you write all the formulas for the pairwise distances we learned. If you think you need to have something else written down, let me know in advance and I will make a decision and let other students know if needed.

The sort of computations/numerical questions you should expect will be modeled on those in the HW or examples from class. For instance, given a joint distribution of patterns for two sequences, argue which model might be a good fit for the data and compute a distance from the table. Something like finding a maximizer of a log-Likelihood is fair too. Being about to compute a simple posterior distribution (with either a discrete or continuous prior) may also be asked. Understanding NNI, SPR, and TBR moves in tree space will be tested. There are sure to be questions on the coalescent model.

The other information you need to know are definitions and statements of important results. This includes what it means for an estimator $\hat{\theta}_n$ to be consistent, and which of our methods for tree reconstruction are consistent. It is also important that you can discuss clear and concisely the difference between a 'statistically good' property (theoretical property) and practical questions (okay, so $\hat{T}_{MLE,n}$ is consistent, but how big should n be?) There will most certainly be questions asking you to compare and contrast various methods of tree construction.

5	Distance Methods	61
5.1	Dissimilarity Measures	61
5.2	An Algorithmic Construction: UPGMA	64
5.3	Unequal Branch Lengths	66
5.4	The Four-point Condition	70
5.5	The Neighbor Joining Algorithm	74
5.6	Additional Comments	76
5.7	Exercises	77
6	Probabilistic Models of DNA Mutation	85
6.1	A first example	85
6.2	Markov Models on Trees	91
6.3	Jukes-Cantor and Kimura Models	97
6.4	Time-reversible Models	101
6.5	Exercises	103
7	Model-based Distances	109
7.1	Jukes-Cantor Distance	109
7.2	Kimura and GTR Distances	114
7.3	Log-det Distance	115
7.4	Exercises	117
8	Maximum Likelihood	121
8.1	Probabilities and Likelihoods	121
8.2	ML Estimators for One-edge Trees	127
8.3	Inferring Trees by ML	128
8.4	Efficient ML Computation	130
8.5	Exercises	134
9	Tree Space	137
9.1	What is Tree Space?	137
9.2	Moves in Tree Space	139
9.3	Searching Tree space	144
9.4	Metrics on Tree Space	146
9.5	Metrics on Metric Tree Space	148
9.6	Additional Remarks	149
9.7	Exercises	149
10	Rate-variation and Mixture Models	153
10.1	Invariable Sites Models	153
10.2	Rates-Across-Sites Models	156
10.3	The Covarion Model	158
10.4	General Mixture Models	161
10.5	Exercises	162

11 Consistency and Long Branch Attraction	165
11.1 Statistical Consistency	166
11.2 Parsimony and Consistency	167
11.3 Consistency of Distance Methods	170
11.4 Consistency of Maximum Likelihood	171
11.5 Performance with Misspecified models	173
11.6 Performance on Finite-length Sequences	173
11.7 Exercises	174
12 Bayesian Inference	177
12.1 Bayes' theorem and Bayesian inference	177
12.2 Prior Distributions	182
12.3 MCMC Methods	183
12.4 Summarizing Posterior Distributions	185
12.5 Exercises	187
13 Gene trees and species trees	189
13.1 Gene Lineages in Populations	190
13.2 The Coalescent Model	193
13.3 Coalescent Gene Tree Probabilities	198
13.4 The Multispecies Coalescent Model	200
13.5 Inferring Species Trees	205
13.6 Exercises	212
14 Notation	215
15 Selected Solutions	217
Bibliography	229