Group final Projects
STA 624-001 Spring 2010
Due: May 4th, Tue, 2010

During the final each group has to present a 50 minutes presentation. Each person in a group has to present at least 5 minutes each.

Each group has to make an appointment with me for 50 minutes interview to present the project.

Here we list possible final projects but not limited to. You may pick your own topic.

Problem 1 *Read the book by Durbin et al.*

1. *Explain the probabilistic theory behind computing a scoring matrix.*
2. *Explain the statistical model used for computing the PAM1 matrix.*
3. *Find 100 complete protein sequences for cystathionine gamma-synthase in fungi (Dr. Schardl can give instructions how to do this), and use these to recompute a PAM matrix. Then find 500 complete protein sequences for phosphoenolpyruvate carboxykinase in bacteria, and use these to compute a PAM matrix. List and explain 3 biological reasons why you might expect these PAM matrices to differ from one another. Compare them; do they differ?*
4. *Compute the limiting distribution for the Markov processes with the score matrix you computed and compared the limiting distribution you computed with the limiting distribution for the PAM matrix. Note that the limiting distribution is the distribution for a long run over the Markov chain. Are they different? If so show how they are different by taking the log odd ratio for each pair of the letters.*

Problem 2 *Read "A novel test for significant codivergence between cool-season grasses and their symbiotic fungal endophytes" by C. L. Schardl, K. D. Craven, A. Lindstrom, S. Speakman, A. Stromberg, R. Yoshida. Systematic Biology. Volume 57, Issue 3, (2008), p483 - 498. They developed a statistical hypothesis test for coevolution between hosts and parasites.*

1. *What are the hypotheses? What is the test statistic? They developed the MRCALink method. What is a novel idea behind this method? Why do we care?*
2. *Compare the statistical method they used in the paper with the MRCALink algorithm with partition homogeneity and likelihood ratio tests (the Shimodaira-Hasegawa test: Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Molecular Biology and Evolution 16: 1114-1116) using the same data sets they used in the paper. Also what is the biological hypothesis being tested in each case.*

Problem 3 *Read "A Bayesian Framework for the Analysis of Cospeciation" by John P. Huelsenbeck; Bruce Rannala; Bret Larget Evolution, Vol. 54, No. 2. (Apr., 2000), pp. 352-364.* `http://links.jstor.org/sici?sici=0014-3820%28200004%2954%3A2%3C352%3AABFFTA%3E2.0.CO%3B2-Z`

1. *Describe their models for analysis on cospeciation between hosts and their parasites.*

2. *Implement the software for the Baysian framework for analysis of cospeciation between hosts and their parasites.*

3. *Apply the method to detect possible host switchings of endophytes between plants described in the paper "A novel test for significant codivergence between cool-season grasses and their symbiotic fungal endophytes" by C. L. Schardl, K. D. Craven, A. Lindstrom, S. Speakman, A. Stromberg, R. Yoshida. Systematic Biology. Volume 57, Issue 3, (2008), p483 - 498. Comment on how similar or different are the results they got compared with the paper by Schardl et al, and speculate as to biological reasons why.*

Problem 4 *Read "Phylogenetic inference under recombination using Bayesian stochastic topology selection" by Alex Webb, John M. Hancock, and Chris C. Holmes Bioinformatics 2009 25(2):197-203; doi:10.1093/bioinformatics/btn607* `http://bioinformatics.oxfordjournals.org/cgi/content/full/25/2/197`

1. *State the HMM used in this paper, namely the observed state space, the hidden state space, the transition probabilities, and emission probabilities.*

2. *Download their data sets and software from* `http://www.stats.ox.ac.uk/~webb`*.*

3. *Using the software you downloaded and you implemented above experiment with the data sets available at* `http://www.stats.ox.ac.uk/~webb/` *and* `http://www.bioss.ac.uk/staff/dirk/Supplements/Glasgow/Data/`*.*