

# BTRY 4840: Project Proposal

Garrett Tetrault, grt43

December 2, 2019

---

For the final project, I would like to focus on the following paper by Felsenstein and Churchill that presents a model to compute the likelihood of a phylogeny, allowing for unequal evolutionary rates at different sites in the molecular sequences. This paper was noted in Durbin's text in chapter eight when the discussion turns to creating a more realistic model of evolution.

J Felsenstein, G A Churchill, A Hidden Markov Model  
approach to variation among sites in rate of evolution.,  
Molecular Biology and Evolution, Volume 13, Issue 1,  
Jan 1996, Pages 93–104,

<https://doi.org/10.1093/oxfordjournals.molbev.a025575>

The underpinning of the model is a Hidden Markov Model whose states represent different rates of genetic mutation for a site. This model requires a prior distribution of rates and transition probabilities between rates. The project would be divided into two main sections.

1. The first would be to implement the algorithm to compute a phylogeny's likelihood and compute a maximum likelihood phylogeny from this. The paper describes an algorithm to do this using the above model to compute likelihoods and estimating branch length by the Newton-Raphson method.
2. The next portion of the project would be to consider different prior distributions of rates. For example, the paper points out that a "discrete distribution with four well-chosen classes" performed well. Alternatively, I would like to try some variation of the truncated geometric distribution we encountered in our homework. Additionally, different amounts of hidden states could shed light on the effects of over-fitting.

Possible data sets for the project would be that which was assigned in Problem Set 4, or that which the paper detailed.