

Modelling stochastic processes on evolutionary trees

Introduction

Process models are required to test hypotheses of how things have evolved through through the history of life on Earth. For example, the value of a continuously evolving character (such as body mass) through time is commonly modelled as the location of a particle moving under one-dimensional Brownian motion with constant rate [1]. The Brownian motion model is best suited to characters evolving under neutral drift, since the trajectory followed by the evolving character has a constant rate of change and no directional tendency. Unfortunately, characters of interest to biologists are not expected to evolve neutrally at all, but are subject to natural selection and other processes. In this project we will explore the use of non-Brownian stochastic processes to model the evolutionary process.

Methodology

Suppose we have modelled evolution as a Brownian process with some fixed rate of change, σ^2 . Branches of a phylogenetic tree where natural selection has caused the diversification of biological characteristics should exhibit a rate of change higher than σ^2 , while branches where natural selection has prevented the diversification of biological characteristics should exhibit a rate of change lower than σ^2 . A model of non-neutral evolution should thus permit the rate of evolution (the rate of change of the Brownian particle) to vary across branches of a phylogeny. We have recently implemented a non-Brownian model in which increments to evolving characters are drawn from a heavy tailed distribution [2]. An alternative is to assume that the rate of Brownian evolutionary change is itself a variable drawn from some probability distribution, and that each branch (or, alternatively, each generation of evolving creatures) draws its rate of change from that distribution. The problem is, how to we implement such a model and how do we infer its parameters from observed trait values for living species of animals?

A number of ways of implementing variable rates models of evolution suggest themselves.

1. The rate of change on each branch, σ^2 takes its value from a set of possible rates σ_i^2 , each of which occurs with probability p_i . If $B(x, y, \sigma_i^2)$ is the

probability of the value of a character changing from x to y along a branch with rate σ_i^2 , then the probability under our model is $\sum_i p_i B(x, y, \sigma_i^2)$. Given some observed values of the character in living species, along with a tree, we could use any global optimization method [3] or MCMC [4] to estimate the probabilities and values of rate categories that maximize the likelihood of the observed data.

2. As above, but instead of weighting the probabilities of change on each branch (as in $\sum_i p_i B(x, y, \sigma_i^2)$) we could assume that each branch possesses some specific rate of change with probability one, and that these rates vary across branches of a tree according to a Markov process with parameters to be estimated from the data.

3. We can even consider more complex models for trait evolution using kernel density methods [5], in which the distribution of site rate variation is nonparametric and multimodal.

We will implement Brownian motion models and alternatives using the R programming environment (or any alternative programming language that individuals are happy to use).

Expected outcomes

We will introduce participants to the BM model, its heavy-tailed generalization, and also to empirical likelihood methods for modelling even more complex patterns of evolution. Together we will learn how such models might be fit to real datasets using various types of global optimization and Markov chain Monte Carlo. There is lots of room for inventiveness and creativity in how to proceed in this task, so we will need to do some brain-storming. At the end of the workshop, participants will be comfortable manipulating phylogenetic trees and datasets, and will understand the costs and benefits of different model-fitting and optimization strategies. A minimum outcome is the ability to simulate the evolutionary process under complex models. The best outcome is that, together, we will have developed methods to fit a completely novel model of continuous character evolution, with variable rates of change, to real data.

References

General reading: Martins, ed. (1996) *Phylogenies and the comparative method in animal behavior* (OUP)

1. See chapter 24 of Felsenstein (2004) "Inferring Phylogenies" (Sinauer)
2. www.sfu.ca/~mooers/papers/elliott_mooers_m.pdf
3. http://en.wikipedia.org/wiki/Global_optimization
4. <http://mpdc.mae.cornell.edu/Courses/UQ/2988428.pdf>
5. http://en.wikipedia.org/wiki/Kernel_density_estimation