

Computational Biology HS 19 - Exercise 5

Answers to Theory Questions

Toby Law

5 December 2019

Question 1

Why can't Fisher's exact test be directly used to check for correlation on the values of discrete traits at the tips of a phylogenetic tree?

Fisher's exact test assumes independent observations of discrete traits, therefore neglects their phylogenetic dependence and the fact that the species/individuals share an evolutionary history. Performing the Fisher's exact test directly on the values of the discrete traits violates the assumption of independent observations with normally distributed errors, and might lead to false conclusions about correlations between the traits of interest.

Question 2

In the independent contrast method, the contrasts Z_k are mutually independent and have identical variance. Explain why this is done.

To investigate if there are any true correlations between traits on a phylogenetic tree, we need to first mathematically eliminate the inter-dependencies of the trait observations, before performing linear regression. Independent contrasts are the recomputed original data points that are corrected to be independent and have identically distributed errors, so they can then be used for linear regression analyses.

Question 3

Explain which steps of the algorithm focus on defining a set of independent variables, and which steps ensure that these have identical variance.

To define the set of independent variables, the algorithm recursively divides the entire tree into independent parts (i.e. recursively defining cherries into an

independent tip), and then calculates normalized contrasts Z_k for pairs of tips to give the independent variables at each internal node/cherry.

These independent variables are derived from differences between trait values at the tips/child nodes. As in the Brownian motion model, variance is proportional to evolutionary time, to make sure that these independent variables have identical variance, (1) the tip trait values are weighed according to their evolutionary distances from the parent node. When merging a cherry into a tip as we move up the tree, we also (2) calculate a corrected branch length that represents the evolutionary time associated with that whole cherry, so that the contrast value associated with that parent node can be normalized with the evolutionary times/variance associated with its tip nodes.

Question 4

You are studying the height of canines and the length of claws among apes. What could make you think that they evolved independently, or on the contrary, that they evolved in a correlated way? In which case would the two traits show a correlation? In which case would they show correlated normalized contrasts?

Knowing the values of both traits and the phylogenetic relationship amongst the apes species in question, we could use the independent contrast method and determine the normalized contrasts for both traits between the ape species, then perform linear regression on them (height of canines vs. length of claws). The closer the *coefficient of determination* R^2 (with a p-value of <0.05) is to 1, the higher the chance the two traits evolved in a correlated way, with moderate to strong correlation between the normalized contrasts, and on the contrary, the closer R^2 is to zero, the weaker the correlation between the normalized contrasts of the two traits, hence the more probable the two traits have evolved independently.

Question 5

A strategy for comparing discrete characters while accounting for relatedness between individuals was presented in lecture 8. Imagine you want to perform a discrete character comparison using this strategy on two discrete traits that you are able to accurately observe on individuals at the tips of a known tree. What problem might you encounter when trying to fill in the contingency table of character changes?

In lecture 8, to account for the relatedness between individuals, the Fisher's exact test was done on the frequencies of phylogenetic tree branches with character changes instead of frequencies of individuals with the traits of interest.

However, this requires us to know exactly where in evolutionary time or in which common ancestor did the character changes arise, and this piece of information is often not accurately known or and not observable since it happened sometime in the history.