

Invariants of Mixtures of Mutagenetic Trees With at Most Four Nodes

Elizabeth Dong, Seattle Pacific University
Guanming Lang, New College of Florida

Abstract:

Mutagenetic trees are used to model accumulative evolutionary processes resulting from permanent genetic changes. A mutagenetic tree induces a probability distribution on the set of all mutational patterns among a discrete number of genetic events. For example, mixture models of mutagenetic trees have been used to model the evolution of drug resistance in the HIV genome. The problem of model selection in this setting consists of finding the mixture of mutagenetic trees model that most accurately explains a given data set of biological mutations. Recently, it has been shown that the dimension of a model, i.e., the minimum number of parameters needed to specify the model is crucial in popular model selection methods such as the Bayesian Information Criterion (BIC). Using methods from algebraic statistics, we compute the dimension of all mixtures of two mutagenetic trees with at most 4 nodes. We also computed the polynomial invariants for all these models which are used in Frequentist methods for model selection. This information will be compiled in a valuable database that will be important for biologists and statisticians in the model selection process.