
OLIVIER FRANÇOIS, Universités de Grenoble

Probabilistic analysis of tree shape statistics

The explosion of phylogenetic studies not only provides a clear snapshot of biodiversity, but also makes it possible to infer how the diversity has arisen. To this aim, variation in speciation and extinction rates have been investigated through their signatures in the shapes of phylogenetic trees. This issue is of great importance, as fitting stochastic models to tree data would help to understand underlying macroevolutionary processes. Although the prevailing view is that it does not represent phylogenies so well, the most popular model of phylogenetic trees is a branching process introduced by Yule, in which lineages split at random. For two decades, the Colless index has been the most frequently used statistic for assessing the balance of phylogenetic trees. In this talk, this statistic is studied under two models of phylogenetic trees: the Yule and uniform model. The main tool of analysis is a coupling argument with another well-known index called the Sackin statistic. Asymptotics for the mean, variance and covariance of these two statistics are obtained, as well as their limiting joint distribution for large phylogenies. Under the Yule model, the limiting distribution arises as a solution of a functional fixed point equation. Under the uniform model, the limiting distribution is the Airy distribution. The cornerstone of this study is the fact that the probabilistic models for phylogenetic trees are strongly related to the random permutation and the Catalan models for binary search trees. Application to a large computer database of phylogenetic trees will also be shown.