UNIVERSITÄT
SALZBURG

Mathematisches Kolloquium

Mittwoch, 28. November 2012
16 Uhr c.t.
Hörsaal 423

Dr. Ruriko Yoshida
University of Kentucky

Nonparametric Estimation of Phylogenetic Tree Distributions

Abstract

As population-based models of gene trees such as coalescent models have been developed to more accurately model distributions of gene trees across genomes, meanwhile detection of horizontal gene transfers and discordances among gene trees have become important problems in phylogenetics. Here we focus on the problem of discordance among gene trees, and the distribution on gene trees as a whole. We view "typical" gene trees as samples from some distribution f that generates gene trees as independent samples. We also suppose there may be rare outlier gene trees which are not "typical", and are samples from some other distribution f' very different from f.

Given the tremendous amount of ongoing effort to develop better parametric models for gene tree distributions, here we take a nonparametric approach. One advantage of nonparametric estimation is that modeling decisions and assumptions are avoided. In contrast to parametric models such as coalescents, using a nonparametric approach avoids issues such as model mis-specification which might potentially confound detection of outlier trees. While most of methods to detect outliers apply statistical methods over an Euclidean (vector) space, using a Markov Chain Monte Carlo technique, we propose to develop statistical methods to estimate the distribution of trees over the space of trees which is not a Euclidean space.