Title:
Random Partition Distribution
Indexed by Pairwise Information

Speaker:
Dr. David Dahl
Associate Professor
Department of Statistics
Brigham Young University

Time & Place:
Wednesday, March 5, 2014, 4:00-5:00pm
Room 140 Bardeen
(Cookies & Coffee @ 3:30 in Rm 1210 MSC)

Abstract:
While exchangeable probability models for partitions have attractive properties, they make strong assumptions. For example, the Dirichlet process induces a partition distribution in which the probability that an item is clustered with another is uniform across all items. We propose a random partition distribution indexed by pairwise information such that items which are "close" to each other are more likely to cluster together. Our approach is explicitly nonexchangeable. The proposed distribution is a natural extension of the Chinese restaurant process. We can explicitly write the associated probability mass function as well as the distribution of the number of clusters and its first two moments. We show that a typical Bayesian nonparametric model can be readily modified to use our random partition distribution based on pairwise information. Applying our distribution as a prior clustering distribution in a model for protein structure prediction, we find that the method incorporating pairwise information substantially improve predictive accuracy.