

STATISTICS COLLOQUIUM

MONDAY, NOVEMBER 21, 2011

TALK: 4:00 PM — SCIENCE CENTER RM. 309

RECEPTION: 5:15 PM — SCIENCE CENTER, 7TH FLOOR

“Tree-Structured Stick Breaking Processes for Hierarchical Data”

**Ryan Adams
Computer Science
SEAS
Harvard University**

ABSTRACT

The Dirichlet process and related distributions over infinite-dimensional random measures have had a large impact on Bayesian nonparametric modeling, but they can be limited by the lack of structure they place over their random partitions. In this work, we use a recursive stick breaking process to construct a Bayesian nonparametric prior for inferring hierarchies in data, yielding an exchangeable urn scheme on trees of unbounded width and depth. We use a novel slice sampling approach based on an auxiliary variable model to update the structure of the tree, as well as Gibbs moves that take advantage of invariance under size-biased permutation. We use this MCMC sampling scheme to perform hierarchical clustering on a set of 50,000 color images, and also to discover structure from natural language in a corpus of scientific documents.

This is joint work with Zoubin Ghahramani (Cambridge) and Michael Jordan (Berkeley).