

Michael I. Jordan

Department of Electrical Engineering and Computer Science
Department of Statistics
UC Berkeley

Combinatorial Stochastic Processes and Nonparametric Bayesian Modeling

Computer Science has historically been strong on data structures and weak on inference from data, whereas Statistics has historically been weak on data structures and strong on inference from data. One way to draw on the strengths of both disciplines is to pursue the study of "inferential methods for data structures"; i.e., methods that update probability distributions on recursively-defined objects such as trees, graphs, grammars and function calls. This is accommodated in the world of "nonparametric Bayes," where prior and posterior distributions are allowed to be general stochastic processes. Both statistical and computational considerations lead one to certain classes of stochastic processes, and these tend to have interesting connections to combinatorics. I will focus on nonparametric Bayesian modeling based on Dirichlet processes and completely random processes, giving examples of how recursions based on these processes lead to useful models in several applied problem domains, including natural language parsing, computational vision, statistical genetics and protein structural modeling.