Bocconi
Università Commerciale Luigi Bocconi

Department of Decision Sciences

Statistics Seminar

Fragmentation-Coagulation Processes: a Model for Aligned Sequences and an Application in Genetics

Yee Whye Teh
University College London

Thursday, 12 May 2011
12:30pm Room 3-E4-SR03 Via Rontgen 1 Milano

Abstract

In this talk I will describe recent work on fragmentation-coagulation processes, an alternative to the popular hidden Markov models that are suitable for modelling aligned sequences often arising in genetics. Fragmentation-coagulation processes are Markov processes over partitions that evolve by either splitting a cluster into multiple clusters or merging multiple clusters into one.
Fragmentation-coagulation processes are suitable for modelling genetic sequences whereby the clustering structure at a point of the genome roughly corresponds to the genealogy at that point. Due to the genetic processes of recombination and gene conversion, the genealogies at different points of the genome can (and often do) differ, with the genealogies of far apart points being (on average) more different than close by points, and it is this effect that makes
the analysis of genetic data interesting. Using fragmentation-coagulation processes, this changing genealogy is captured by the partition structure which changes by splits and merges as it moves along the genome.
I will describe some interesting theoretical properties of fragmentation-coagulation processes, relate them to a number of Bayesian nonparametric models of partitions and hierarchical clusterings, and present some preliminary results on the problems of genotype imputation and recombination rate estimation.
Joint work with Charles Blundell, Vinayak Rao, Lloyd Elliott, and Andriy Mnih.

Department of Decision Sciences

Via Röntgen 1 - 20136Milano

Tel. 02 5836.5632
Fax 02 5836.5630