Some Statistical Problems in Cancer Genomics

Abstract:

The starting point for this talk comes from population genetics: how should we estimate evolutionarily relevant parameters from DNA sequence data taken from samples of individuals? I will give a brief overview of what we learned, starting from the Ewens Sampling Formula and touching on Approximate Bayesian Computation as an inference method when likelihoods are intractable. To illustrate ABC, I will give an example concerning inference of the number of distinct DNA sequences in a sample, given only information about the frequency of point mutations in the samples. This example provides an introduction to inference from typical cancer sequencing data, in which individuals are replaced by cells and in which typically we do not know which mutations occur in which cells. I will give a brief overview of what cancer evolution is about, the sort of statistical and computational problems it poses, and where we are in addressing some of them. Time permitting, I will describe some novel experimental methods we are developing to understand the 3D structure of tumors, paving the way for some challenging inferential problems that will require engagement from data scientists and others.

Special Note:
Light refreshments will be served at 4.30-5 p.m., preceding the talk.

Organizers: UCSD Mathematics Department & Physical Sciences

Thursday, February 7, 2019
5:00 PM
Natural Sciences Building Auditorium