



THE UNIVERSITY OF CHICAGO

Department of Statistics

STATISTICS COLLOQUIUM

SIMON TAVARÉ

Director, Irving Institute for Cancer Dynamics &
Department of Statistics & Systems Biology, Columbia University

Some Statistical Problems in Cancer Genomics

MONDAY, November 4, 2019 at 4:30 PM

Eckhart 133, 5734 S. University Avenue

Refreshments before the seminar at 4:00PM in Jones 111

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 relative 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.

For further information and inquiries about building access for persons with disabilities, please contact Jonathan Rodriguez at 773.702.8333 or send him an email at jgrodriguez@galton.uchicago.edu. If you wish to subscribe to our email list, please visit the following website:
<https://lists.uchicago.edu/web/subscribe/statseminars>.