



THE UNIVERSITY OF CHICAGO

Department of Statistics

STATISTICS COLLOQUIUM

YUN SONG

Electrical Engineering and Computer Sciences
University of California, Berkeley

Information-Theoretic Limits on Inferring Population History

MONDAY, May 11, 2015, at 4:00 PM

Eckhart 133, 5734 S. University Avenue

Refreshments following the seminar in Eckhart 110.

ABSTRACT

Numerous empirical studies in evolutionary biology have employed a summary statistic called the sample frequency spectrum (SFS), which summarizes the information in a sample of DNA sequences. In particular, recently there has been growing interest in analyzing the SFS data from multiple populations to infer complex population histories. Despite their popularity, the accuracy of inference methods based on the SFS is difficult to characterize theoretically, and it is currently unknown how the estimation accuracy improves as more sites in the genome are used. In this talk, I will present information-theoretic limits on the accuracy of all estimators that use the SFS to infer population histories. I will consider the rate of convergence to the truth as the amount of data increases, and describe the surprising result that it is exponentially worse than known convergence rates for many classical estimation problems in statistics. This is joint work with Jonathan Terhorst.