

DISSERTATION PROPOSAL PRESENTATION

Empirical Bayes smoothing methods in high-throughput sequencing data

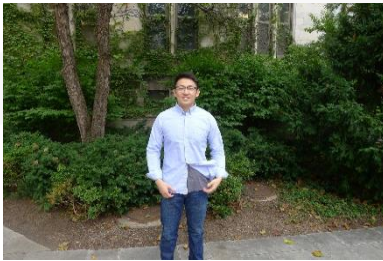
WHEN

November 9, 2021
12:30 PM, CST

WHERE

Via ZOOM

ZOOM information will be provided in the email announcement for this seminar.

**Dongyue Xie, PhD candidate**

High-throughput sequencing data provide biological information at high resolution. Sequencing techniques such as RNA-seq produce counts at base level genome wide that measure the intensity of the underlying biological processes. The biological processes are expected to be spatially-structured along the genome. In the first part of the talk, we outline an empirical Bayes smoothing method for count data. Specifically, methods assuming Poisson distribution in general fail to smooth the sequencing data. To deal with the issue, for the univariate smoothing problem, we introduce a nugget effect to handle the extra variation in the count data. We show in real RNA-seq examples the method produces better visualization of the intensity along the genome. Matrix factorization techniques are often used as soft clustering methods to explore structures of genetic data. By looking at the low-rank representation of multiple sequences, we could find meaningful latent biological processes. Motivated by this, we introduce a framework of smoothed matrix factorization based on multiscale decomposition to analyze sequencing data at the base-pair resolution. We show the potential of the method to jointly decompose different biological assays - RNA-seq, ATAC-seq and histone modification mark data.

stat.uchicago.edu



THE UNIVERSITY OF
CHICAGO

DEPARTMENT OF STATISTICS