THE UNIVERSITY OF CHICAGO | DEPARTMENT OF STATISTICS

PhD Dissertation Presentation

Dongyue Xie
Department of Statistics
The University of Chicago

"Empirical Bayes methods for analyzing sequencing count data"

Thursday, June 29, 2023, at 10:00 AM
Jones 303, 5747 S. Ellis Avenue

Abstract

High-throughput sequencing (HTS) techniques such as RNA-seq, ChIP-seq and ATAC-seq have enabled researchers to investigate complex biological processes in unprecedented detail. One common feature of HTS data is that they often consist of counts. For example, in RNA-seq, the counts typically represent the number of times a RNA molecule has been sequenced and are a proxy for the expression level. Recently, the advent of single-cell sequencing techniques such as scRNA-seq has unveiled the transcriptome at cell-level resolution. However, the single-cell count data are sparse and come with high levels of technical noise. With the emergence of large, sparse and noisy sequencing data, there is a need for rigorous statistical methods that can accurately model these counts.

On the other hand, due to the complex structure of the sequencing data exhibited, the statistical methods developed for the data should be flexible enough to incorporate different assumptions and structural information. For instance, matrix factorization has been extensively employed to uncover the latent structure of gene expression across a variety of cell types. The incorporation of sparsity assumptions into these latent structures has been shown to yield a more parsimonious representation and enhance the interpretability of results. Consequently, it would be beneficial to integrate sparsity assumptions when modeling the structure of sequencing data.

In this dissertation, we introduce a novel variational empirical Bayes (VEB) framework designed for analyzing count sequencing data. In particular, the method enables us to apply well-developed Gaussian empirical Bayes methods for inference on non-Gaussian models. The resulting variational inference algorithm is modular, alternating between a step that handles the count data and a step that fits the Gaussian model. We demonstrate the framework by applying it to an empirical Bayes Poisson matrix factorization model. This results in a general empirical Bayes method for Poisson factor analysis, allowing for various prior families on both loadings and

For information about building access for persons with disabilities, please contact Keisha Prowoznik at 773.702-0541 or send an email to kprowoznik@statistics.uchicago.edu. If you wish to subscribe to our email list, please visit the following website: https://lists.uchicago.edu/web/info/statseminars.

factors. Numerical studies are performed to show the benefits of directly modeling the count using Poisson distribution, and how the new VEB method gives flexible and accurate latent structure recovery.

For information about building access for persons with disabilities, please contact Keisha Prowoznik at 773.702-0541 or send an email to kprowoznik@statistics.uchicago.edu. If you wish to subscribe to our email list, please visit the following website: https://lists.uchicago.edu/web/info/statseminars.