



THE UNIVERSITY OF
CHICAGO

DEPARTMENT OF STATISTICS

MASTER'S THESIS PRESENTATION

MICHELLE LI

Department of Statistics
The University of Chicago

Trajectory Exploration and Inference for scRNA-seq Data on Whole versus Endpoints Only
Trajectories

THURSDAY, October 20, 2022, at 11:00 AM

Zoom Meeting, 5747 S. Ellis Avenue

ABSTRACT

With the development of single cell RNA-sequencing (scRNA-seq), researchers are able to measure mRNA molecule counts, or gene expression levels, for single cells. These cells are often positioned along a developmental trajectory. But the positions of these cells and, in some cases, the trajectory are unknown. Many methods have been developed to infer the position of the cells and the trajectory. In this paper, we introduced a simulation method for scRNA-seq count data using Gaussian Process gene expression models and multinomial measurement models (Sarkar and Stephens 2021) along arbitrary rooted, directed tree trajectories. We also simulated data for cells that exist only at the endpoints of the trajectories, reflecting experiments where cells are only observed at maturation, or the end of development. Four methods were applied to the simulated count data in order to explore and infer the trajectories underlying the data: PCA reduced dimension visualizations, hierarchical clustering, binary splitting in PCA reduced dimensions (Li et al 2020), and the trajectory inference method Slingshot (Street et al 2018). The results for each of the methods can be summarized into three main effects: proper normalization, whole tree trajectory versus the leaf ends only, and the complexity of the tree trajectory structure.