Stockholm University

Mathematical Statistics
Stockholm University
Master Thesis **2022:4**
http://www.math.su.se

Statistical modelling and inference of single-cell gene expression profile

Lai Mei Yip Lundström*

June 2022

Abstract

Single-cell ribonucleic acids (scRNA) sequencing technologies have made it possible to measure genetic information at the cellular level, thereby facilitating the characterisation of a cell by its gene expression levels. This thesis sets out to model the messenger RNA (mRNA) transcriptional and degeneration process of a given gene in a cell by means of a bivariate Markov Chain as well as to derive its stationary distribution. The steady-state stationary distribution of the number of mRNA molecules that are synthesized by a given gene in a cell is approximated using perturbation techniques and the parameters are inferred using maximum likelihood. The stationary distributions of the different genes with the inferred parameters then form the gene expression profile of the cell. The result is that the negative binomial distribution is shown to be the exact solution to the simpler problem in the perturbative solution. Furthermore, it is shown that biologically relevant quantities such as a gene's mRNA transcriptional frequency and transcriptional size are related to the parameters of the negative binomial distribution. In addition, a comprehensive study of the probability currents in the Markov Chain has also found them to be closely connected to the mean of the distribution. The model is then applied to scRNA sequencing data and the results are presented and discussed.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: laimei.yip@gmail.com. Supervisor: Chun-Biu Li.