DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE
THE UNIVERSITY OF HONG KONG

Departmental Seminar

Dr. Yan ZHOU

Date Changed

College of Mathematics and Statistics, Institute of Statistical Sciences
Shenzhen University, China

will give a talk
entitled

CLASSIFYING NEXT-GENERATION SEQUENCING DATA USING A ZERO-INFLATED POISSON MODEL

Abstract

Motivation: With the development of high-throughput techniques, RNA-sequencing (RNA-seq) is becoming increasingly popular as an alternative for gene expression analysis, such as RNAs profiling and classification. Identifying which type of diseases a new patient belongs to with RNA-seq data has been recognized as a vital problem in medical research. As RNA-seq data are discrete, statistical methods developed for classifying microarray data cannot be readily applied for RNA-seq data classification. Witten proposed a Poisson linear discriminant analysis (PLDA) to classify the RNA-seq data in 2011. Note, however, that the count datasets are frequently characterized by excess zeros in real RNA-seq or microRNA sequence data (i.e. when the sequence depth is not enough or small RNAs with the length of 18–30 nucleotides). Therefore, it is desired to develop a new model to analyze RNA-seq data with an excess of zeros.

Results: In this paper, we propose a Zero-Inflated Poisson Logistic Discriminant Analysis (ZIPLDA) for RNA-seq data with an excess of zeros. The new method assumes that the data are from a mixture of two distributions: one is a point mass at zero, and the other follows a Poisson distribution. We then consider a logistic relation between the probability of observing zeros and the mean of the genes and the sequencing depth in the model. Simulation studies show that the proposed method performs better than, or at least as well as, the existing methods in a wide range of settings. Two real datasets including a breast cancer RNA-seq dataset and a microRNA-seq dataset are also analyzed, and they coincide with the simulation results that our proposed method outperforms the existing competitors.

Availability and implementation: The software is available at http://www.math.hkbu.edu.hk/.

on **Thursday, August 22, 2019**

~~Wednesday, August 21, 2019~~

(Refreshments will be served from 10:45 a.m. <u>outside Room 301 Run Run Shaw Building</u>)

11:00 a.m. – 12:00 noon

at

Room 301, Run Run Shaw Building

<u>**Visitors Please Note**</u> **that the University has limited parking space. If you are driving please call the Department at 3917 2466 for parking arrangement.**

All interested are welcome