

For Favour of posting



The Hong Kong University of Science and Technology

Department of Mathematics

Mathematics Colloquium

Statistical methods for integrative analysis of big genomic data

By

Dr. Can YANG

Hong Kong Baptist University

Abstract

We are at a point in history that Big Data is transforming the way of doing science. To answer a scientific question, we design experiments, collect data and analyze the data while it is often the case that we have limited power to claim significant findings solely based on the data in our hands. The rising of Big Data is revolutionizing the way of doing research -- we may be able to answer the scientific question by combining our data with other existing information that hides in the Big Data Mountain.

In the field of human genetics and genomics, it is often difficult to get approval for the access to individual-level data due to many practical issues, such as privacy protection and logistics of data transportation and storage. In contrast, many research groups are willing to share their summary statistics with each other and release summary statistics to the public. A natural question arises: Instead of getting approval of access to individual-level data or making efforts to collect new data, is it possible to achieve comparable performance using existing summary statistics?

In this talk, we introduce summary-statistics-based approaches to integrative analysis of big genomic data. The goals of our integrative analysis include estimation of overall genetic contribution to a complex disease, identification of genetic risk variants and disease risk prediction. These are achieved by the development of statistically rigorous models and computationally efficient algorithms. We have also developed parallel computing software packages, supporting real data analysis in the genome-wide scale. Our real data analysis results suggest that summary-statistics-based approaches can often be nearly as efficient as the approaches that require individual-level data. We have also observed that summary-statistics-based analysis can have advantages over individual-level-data-based analysis because a much larger sample size is often available in summary-statistics-based analysis. We believe our recent development can greatly help biomedical researchers make the most efficient use of the vast amounts of valuable data and gain new scientific insights on the genetic architecture of complex diseases.

Date: Tuesday, 14 February 2017

Time: 2:00 p.m. – 3:00 p.m.

***Venue: Room 5506, Academic Building
(near Lifts 25&26), HKUST***

All are welcome!

The speaker is a candidate for a faculty position