

For favour of posting



DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE  
THE UNIVERSITY OF HONG KONG

50<sup>th</sup> Anniversary Seminar Series

**Professor Lei SUN**

Department of Statistical Sciences, Faculty of Arts and Science  
Division of Biostatistics, Dalla Lana School of Public Health  
University of Toronto, Canada

will give a talk  
entitled

**THE POWER OF SIMPLE STATISTICAL TECHNIQUES IN  
THE ERA OF BIG AND COMPLEX DATA: SOME RECENT  
EXAMPLES FROM GENETIC ASSOCIATION STUDIES**

Abstract

Genetic association studies aim to identify genetic markers associated with a heritable trait/outcome of interest. Data used are typically big and challenging. Whole-genome studies scan through millions of variables, missing data and measurement errors are often present, individuals from the same family are correlated, and complex genetic etiologies imply complex models. Development favoring machine learning type of approaches is rising rapidly, across different scientific studies. As a complementary approach, in this talk I present some recent examples where we efficiently and reliably extract information from large-scale genetic association studies, by reconsidering some of the classical statistical techniques in newer settings. We first revisit the well-known Fisher's method, commonly used in meta-analyses to combine p-values from the *same* test applied to *K independent* samples. Here we propose to use it to combine p-values from *different* tests applied to the *same* sample, when analyzing multiple genetic variants simultaneously (Derkach, Lawless and Sun 2014, *Statistical Science*; 2015, *Genetic Epidemiology*), or when jointly capturing main and interaction effects (Soave et al. 2015, *American Journal of Human Genetics*). In both settings, we show that there are two classes of complementary tests that are asymptotically independent of each other under a global null hypothesis; this is a desirable feature for analyzing big data. We then revisit the simple linear regression and its celebrated extensions in novel context. We first show that Levene's scale-test for variance heterogeneity can be derived from a two-stage regression framework, and this allows us to generalize the test, with ease, for more complex data (Soave and Sun 2017, *Biometrics*). If time permits, I will discuss on-going work, with graduate student Lin Zhang, on how to use a regression model to test Hardy-Weinberg equilibrium; this leads to a new allele-based association test with theoretical insights on its robustness. We also provide supporting evidence from applications including genetic association studies of complications related to type 1 diabetes and cystic fibrosis.

on

**Wednesday, May 24, 2017**

*(Refreshments will be served from 2:15 p.m. outside Room 301 Run Run Shaw Building)*

**2:30 p.m. – 3:30 p.m.**

at

**Room 301, Run Run Shaw Building**

**Visitors Please Note 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